

UNIVERSIDAD NACIONAL DEL LITORAL

Facultad de Bioquímica y Ciencias Biológicas

Facultad de Ingeniería Química

Instituto de Física del Litoral



Tesis para la obtención del Grado Académico de Magíster en Física

**Caracterización óptica de películas delgadas
mediante interferometría holográfica**

Nicolás Balducci

Director de Tesis: Dr. Nicolás Budini

Co-director de Tesis: Dr. Raúl Urteaga

Laboratorio de Física de semiconductores – Instituto de Física del Litoral

– Universidad Nacional del Litoral

-2020-

Y dijo Dios: Sea la luz; y fue la luz.

Y vio Dios que la luz era buena;

Génesis 1:3-4.

Porque Jehová da la sabiduría,
y de su boca viene el conocimiento y la inteligencia.

Bienaventurado el hombre que halla la sabiduría,
y que obtiene la inteligencia;
porque su ganancia es mejor que la ganancia de la plata,
y sus frutos más que el oro fino.

Proverbios 2:6 – 3:13-14.

***Este trabajo está dedicado con amor a mi esposa Noelia,
y a mis hijos Uziel y Sarah.***

Agradecimientos:

Cada persona posee una cosmovisión con la cual intenta interpretar el universo que nos rodea. En mi cosmovisión, quiero agradecer primeramente a Dios por la oportunidad de poder realizar esta maestría, ya que si fuese solo por esfuerzo propio y antecedentes académicos, no hubiera alcanzado un título de grado, mucho menos pensar en realizar un posgrado. Conocer el mensaje de Jesús cambió lo que para mí eran fracasos en bendiciones, me da un propósito y sentido de significado a mi vida. Además, como si fuera poco, luego me regaló el amor de mi esposa, y posteriormente el de mis hijos, ellos son lo mejor que puedo dejar en este mundo.

Quiero agradecer a mi esposa Noelia por su apoyo incondicional, su paciencia al soportar tantas horas de ausencia de mi parte, y sobre todo por sus palabras de ánimo cuando las cosas se ponían difíciles, siempre alentándome a no bajar los brazos. A mis hijos Uziel y Sarah, que siempre me regalaron un abrazo y una sonrisa cuando regresaba de mis horas de trabajo y estudio.

También quiero agradecer a mi Director, el Dr. Nicolás Budini, por su guía y ayuda en el trabajo de laboratorio y por las correcciones del borrador de este manuscrito. A mi Codirector, el Dr. Raúl Urteaga, por su ayuda, aportes y colaboración. A los dos les agradezco por dedicar de su tiempo y conocimiento para que yo pueda realizar este modesto peregrinaje dentro de la óptica física e intentar plasmarlo en este trabajo.

A los docentes de la Maestría: Dr. Roberto Arce, Dr. Fernando Bonetto, Dr. Claudio Bonin, Lic. Jorge Caram, Dr. Julio Ferrón, Dra. Edith Goldberg, Dr. Roberto Koropecski, Dr. Gustavo Ruano, Dr. Javier Schmidt y Dr. Silvano Sferco, por su dedicación, disponibilidad y apoyo en cada uno de los cursos que realicé.

Por último quiero agradecer a mis suegros Raúl y María Ester por todo su apoyo.

Resumen	6
Abstract	8
Capítulo I: Introducción	
Perfilómetros	10
1.1- Perfilómetros de contacto	11
1.1.1- Perfilómetro de aguja	11
1.1.2- Microscopio de punta de escaneo	14
1.1.2.1- Microscopio de efecto túnel	14
1.1.2.2- Microscopio de fuerza atómica	15
1.2- Perfilómetros ópticos	16
1.2.1- Perfilómetro con sensor óptico de focalización	17
1.2.2- Microscopio confocal	18
1.2.2.1- Procesamiento y apilamiento de las imágenes	19
1.3- Perfilómetros ópticos interferométricos	20
1.3.1- Configuraciones más usuales	21
1.3.2- Resolución	22
1.3.3- Microscopía holográfica	23
Objetivos	24
Referencias	25
Capítulo II: Óptica física y difracción	
2.1- Interferencia	26
2.1.1- Experimento de Young	28
2.1.2- Coherencia espacial y temporal	30
2.2- Difracción	35
2.2.1- Teoría escalar de la perturbación luminosa	35
2.2.2- Difracción por una pantalla plana con abertura	37
2.2.3- El principio de Huygens – Fresnel	40
2.2.4- La aproximación de Fresnel	41
2.2.5- Integral de convolución de Huygens – Fresnel	43
2.2.6- Aproximación de Fraunhofer	43
2.2.7- Propagación del espectro angular	44
2.2.8- Difracción de una lente delgada	47
2.2.9- Difracción por una abertura cuadrada	51

Referencias -----	55
Capítulo III: Holografía e interferometría holográfica	
3.1- Holografía -----	56
3.1.1- Registro del holograma -----	56
3.1.2- Reconstrucción de la onda objeto -----	58
3.1.3- Holograma de Gabor -----	60
3.1.4- Holograma de Leith – Upatnieks -----	62
3.1.5- Otros tipos de hologramas	
3.1.5.1- Holograma de Fourier -----	66
3.1.5.2- Holograma de Fourier sin lentes -----	67
3.1.5.3- Hologramas imagen, de Fresnel y de Fraunhofer -----	69
3.1.6- Posiciones de las imágenes y aumentos -----	70
3.2- Interferometría holográfica -----	73
3.2.1- Interferograma holográfico de doble exposición -----	73
3.2.2- Variación de la fase de interferencia -----	76
3.2.3- Ambigüedad del signo de la fase de interferencia -----	78
3.3- Elementos utilizados en el esquema holográfico -----	82
3.3.1- El láser -----	82
3.3.2- Componentes ópticos -----	83
3.3.3- Cámara CCD -----	83
Referencias -----	86
Capítulo IV: Holografía digital	
4.1- Teorema del muestreo -----	87
4.2- Digitalización del holograma -----	90
4.3- Reconstrucción numérica de la imagen del objeto -----	96
4.3.1- Onda de referencia numérica -----	96
4.3.2- Reconstrucción numérica por la transformada de Fresnel -----	98
4.3.3- Eliminación del término de continua -----	103
4.3.4- Eliminación de la imagen gemela -----	107
4.3.5- Reconstrucción numérica por la convolución de Huygens – Fresnel -----	108
4.3.6- Reconstrucción numérica por la propagación del espectro angular -----	109
4.4- Interferometría holográfica digital -----	111
4.4.1- Demodulación de la fase de interferencia -----	114
4.5- Topografía superficial con holografía digital -----	115
4.5.1- Variación de la longitud de onda -----	115

4.5.2- Variación del índice de refracción -----	117
4.5.3- Variación de la dirección de iluminación -----	118
Referencias -----	120
Capítulo V: Microscopía holográfica digital, trabajo experimental, procesamientos y resultados	
5.1- Esquema óptico para la DHM -----	121
5.2- Microscopía de fase cuantitativa -----	122
5.3- DHM de una muestra de c-Si -----	124
5.4- DHM de una grilla de Ag depositada sobre c-Si -----	135
5.4.1- Determinación de la distancia focal de reconstrucción -----	135
5.4.2- Corrección digital de aberraciones en la imagen de fase -----	144
5.4.3- Determinación del espesor de película de Ag mediante la imagen de fase -----	149
5.5- DHM de una película de SiO ₂ sobre c-Si -----	154
5.6- Comprobación del procesamiento digital por métodos numéricos -----	160
Referencias -----	165
Capítulo VI: Conclusiones -----	166
Apéndice -----	170

Resumen

El presente trabajo fue planteado como un recorrido de aprendizaje sobre fenómenos de la óptica física, desde la interferencia hasta las técnicas holográficas digitales, con el fin de elaborar un sistema óptico experimental para realizar microscopía holográfica digital de fase cuantitativa.

A partir de un esquema de interferómetro de *Michelson*, el objetivo es medir espesores de películas delgadas del orden de fracciones de la longitud de onda del láser utilizado, y lograr la reconstrucción tridimensional de las superficies como una imagen del contraste de fase a través del desarrollo de algoritmos de procesamiento de los hologramas digitales por medio de la transformada discreta de *Fourier*.

Para el desarrollo del trabajo experimental, en una primera etapa se utilizó una muestra que contiene una película de plata depositada sobre silicio cristalino, de espesor conocido por mediciones realizadas en un trabajo anterior. Se aplicó la técnica de microscopía holográfica digital a dicha muestra, para obtener imágenes del diagrama de fases, como forma de ajustar el montaje óptico y medir las distancias focales.

La segunda etapa consistió en determinar la distancia focal óptima, con tolerancias del orden de la décima de milímetro, situando la cámara (plano imagen) a diferentes distancias de la muestra (plano objeto) y analizando diferentes operadores de medición de foco mediante algoritmos propuestos por diversos autores, utilizando el software *Matlab*[®], para adoptar el método de mejor desempeño en la configuración del esquema óptico montado.

En la tercera etapa se desarrolló la corrección de las principales aberraciones ópticas presentes en el diagrama de fases por métodos de procesamiento numérico. Estas aberraciones, existentes en todo esquema óptico, se deben principalmente al desenfoque, a la inclinación del plano objeto respecto del plano imagen (*tilt*) y a las deformaciones de los frentes de onda del objeto y de referencia con respecto a una onda plana ideal.

Una vez cumplidos los tres pasos anteriores se procedió a obtener hologramas de la muestra de referencia. Procesando las imágenes de fases se obtuvieron mediciones del espesor del depósito de plata de aproximadamente 160 nanómetros, con una tolerancia de 10 nanómetros, que está en acuerdo con mediciones realizadas sobre la muestra en las experiencias precedentes mencionadas.

Además se graficaron perfiles de la superficie y se elaboraron gráficas tridimensionales de la topografía superficial de la muestra.

A continuación, para cumplir con el objetivo propuesto en este trabajo, se tomaron hologramas de una muestra que contiene una película delgada de óxido de silicio sobre un sustrato de silicio

cristalino. Las imágenes de fases obtenidas y procesadas arrojaron mediciones de los espesores del óxido de silicio de aproximadamente 90 nanómetros, con una tolerancia de 10 nanómetros, obteniéndose también gráficas de los perfiles de la superficie y gráficas tridimensionales de la topografía superficial, donde se aprecia claramente la diferencia de nivel provocada por el recubrimiento de óxido de silicio.

Por último, para comprobar la eficacia del método de procesamiento digital, se creó un holograma en forma numérica que contenía regiones donde los patrones de franjas de interferencia fueron desplazados arbitrariamente, para simular diferencias de nivel que se calcularon de manera teórica. Luego de procesar este holograma numérico y obtener las gráficas del perfil simulado por el mismo, los resultados obtenidos discreparon en aproximadamente 10 nanómetros de los valores teóricos, estando en acuerdo con las tolerancias de las mediciones realizadas a las muestras.

Con el desarrollo y aplicación de esta técnica de microscopía holográfica digital, se demuestra que este método es aplicable para la medición y caracterización de recubrimientos de películas delgadas de diversos materiales, ya sean opacos o transparentes, sobre sustratos de silicio cristalino.

Palabras claves: microscopía holográfica digital, recubrimiento de película delgada, silicio cristalino.

Abstract.

The present work was proposed as a learning journey on phenomena of physical optics, from interference to digital holographic techniques, in order to develop an experimental optical system to perform quantitative phase digital holographic microscopy.

Beginning with a Michelson interferometer scheme, the objective is to measure thin film thicknesses on the order of fractions of the wavelength of the laser used, and achieve three-dimensional reconstruction of surfaces as a phase contrast image through the development of digital hologram processing algorithms by means of the discrete Fourier transform.

For the development of experimental work, in a first stage a sample containing a silver film deposited on crystalline silicon was used, known thickness from measurements made in previous work. The digital holographic microscopy technique was applied to said sample, to get images of the phase diagram, as a way to adjust the optical assembly and measure focal lengths.

The second stage consisted of determining the optimal focal length, with tolerances on the order of a tenth of a millimeter, placing the camera (image plane) at different distances from the sample (object plane) and analyzing different focus measurement operators using algorithms proposed by various authors, using Matlab® software, to adopt the best performance method in the configuration of the mounted optical scheme.

In the third stage, the correction of the main optical aberrations present in the phase diagram was developed by numerical processing methods. These aberrations, existing in any optical scheme, are mainly due to blurring, the inclination of the object plane with respect to the image plane (tilt) and the deformations of the object and reference waves fronts with respect to an ideal plane wave.

Once the three previous steps had been completed, holograms of the reference sample were obtained. Processing the phase images obtained measurements of the thickness of the silver deposit of approximately 160 nanometers, with a tolerance of 10 nanometers, which is in agreement with measurements made on the sample in the preceding experiences mentioned. In addition, surface profiles were plotted and three-dimensional graphs of the surface topography of the sample were made.

Next, to fulfill the objective proposed in this work, holograms were taken of a sample containing a thin film of silicon oxide on a crystalline silicon substrate. The phase images obtained and processed gave measurements of the thickness of silicon oxide of approximately 90 nanometers, with a tolerance of 10 nanometers, also obtaining graphs of the surface profiles and three-dimensional graphs of the surface topography, where the difference in level caused by the silicon oxide coating can be clearly seen.

Finally, to check the effectiveness of the digital processing method, a hologram was created in numerical form containing regions where interference fringe patterns were arbitrarily displaced, to simulate level differences that were calculated theoretically. After processing this numerical hologram and obtaining the graphs of the profile simulated by it, the results obtained differed by approximately 10 nanometers from the theoretical values, being in agreement with the tolerances of the measurements made to the samples.

With the development and application of this digital holographic microscopy technique, it is shown that this method is applicable for the measurement and characterization of thin film coatings of various materials, whether opaque or transparent, on crystalline silicon substrates.

Keywords: digital holographic microscopy, thin film coating, crystalline silicon.

Capítulo I

Introducción

“Medir es conocer. Si no puedes medir algo, no lo puedes mejorar.” – Lord Kelvin.

Las modernas técnicas de deposición de películas delgadas poseen un rol importante en la fabricación de filtros ópticos, componentes opto-electrónicos, dispositivos semiconductores, celdas solares, etc., por citar sólo algunos pocos ejemplos [1].

El espesor de la película depositada es comúnmente controlado utilizando técnicas de alto vacío, entre las que podemos citar a las técnicas de haces moleculares, sputtering (bombardeo catódico), evaporación, y de deposición química desde la fase vapor [2]. Otras técnicas de deposición que no requieren vacío (aunque a veces requieren atmósfera controlada en cierta medida) son también utilizadas ampliamente, como por ejemplo la deposición capa a capa de nanopartículas y el recubrimiento por inmersión (*dip coating*) o centrífugo (*spin coating*) [3].

En todos los casos, sin excepción, medir de manera precisa el espesor de la capa delgada depositada es de enorme importancia para todas las aplicaciones industriales y de investigación científica y tecnológica. Estimar adecuadamente el espesor de la película permite ajustar los parámetros de un proceso de deposición determinado, para lograr producir películas con las características requeridas, ya que el espesor de las mismas influye directamente en sus propiedades eléctricas, ópticas, mecánicas, etc. En resumen, el espesor de la película depositada es una variable fundamental, y debe medirse y controlarse de la mejor manera posible para obtener películas con las propiedades deseadas [4].

Para medir el espesor de películas delgadas y la topografía superficial en general de una muestra de pequeñas dimensiones se utilizan habitualmente perfilómetros, que son instrumentos de uso básico en cualquier laboratorio de metrología, tanto industrial como en ciencia.

Perfilómetros:

Gracias al desarrollo de nuevas tecnologías, en los últimos años el campo de la metrología superficial se ha expandido tanto en sofisticación como en rango de aplicación, debido a la creciente demanda para caracterizar superficies de manera precisa, como se mencionó anteriormente.

En cuanto a la caracterización óptica de superficies, los avances en las fuentes de iluminación, como el láser, en detectores de estado sólido y en dispositivos opto-electrónicos, han permitido el desarrollo de diversos instrumentos que no solo mapean la topografía de la superficie, sino que también determinan parámetros como el espesor, el índice de refracción, etc. Gracias a estas tecnologías se han incrementado las posibilidades de caracterización óptica de materiales [5].

En todo proceso de medición deben considerarse dos aspectos fundamentales: la repetibilidad y la reproducibilidad. La repetibilidad se define como una desviación estándar de un parámetro en un objeto que es medido varias veces en un corto período de tiempo, sin cambios en el sistema de medición. La reproducibilidad se refiere a la distribución obtenida de múltiples mediciones del mismo parámetro en un largo período de tiempo y en diferentes condiciones de medición.

Algunas técnicas de medición superficial serán descritas a continuación, comenzando con los perfilómetros de contacto o mecánicos, siguiendo luego con los ópticos, para concluir con los ópticos interferométricos, que son el objeto de estudio del presente trabajo de tesis.

1.1- Perfilómetros de contacto:

Los dos principales perfilómetros de contacto son el de aguja y el microscopio de punta de escaneo; ambos utilizan una punta que está en contacto con la superficie para así obtener una gráfica del perfil de la misma.

1.1.1- Perfilómetro de aguja:

Este instrumento mueve una pequeña punta a través de la superficie, y censa las variaciones en el movimiento vertical de dicha punta para así determinar el perfil de alturas de la superficie de la muestra. Este perfilómetro puede medir diferencias de alturas de hasta 1 mm. La manera en la que opera es, normalmente, moviendo la muestra bajo la punta de escaneo, pero existen algunos modelos que mueven la punta dejando fija la muestra.

El movimiento vertical de la punta es detectado, en general, por un transductor diferencial variable que transforma las variaciones de altura en señales eléctricas.

La punta de escaneo del instrumento se construye de material de alta dureza, como el diamante. Posee un radio de curvatura en la punta de entre 0,05 a 50 μm , lo que determina, como veremos, la resolución lateral del instrumento.

Para asegurar que la superficie no sufra daños durante la medición, la carga de la punta sobre la superficie varía desde 0,1 hasta 50 miligramos. Estas cargas tan pequeñas permiten que la punta se mueva sobre la superficie sin deformarla. Podemos observar un esquema del perfilómetro de aguja en la Figura 1.1.

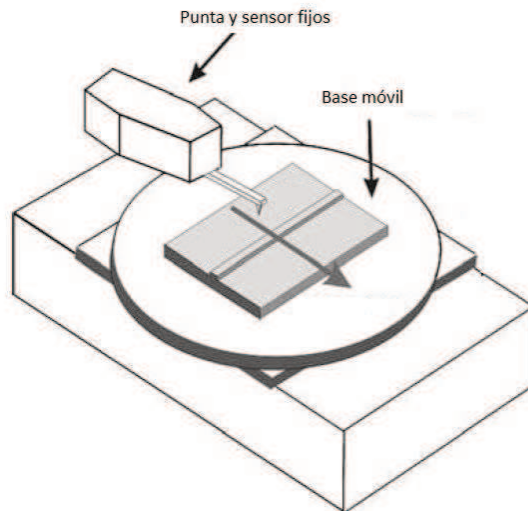


Figura 1.1. Esquema de perfilómetro de aguja (extraído de Veeco Instruments).

La forma y dimensión de la punta de la aguja determinan la profundidad de penetración de esta en la superficie bajo medición. La gráfica arrojada por el perfilómetro es el resultado de la convolución entre el tamaño y forma de la punta con el perfil real de la superficie. Elegir correctamente la configuración de la punta es muy importante para asegurarse de que esta penetre hasta el fondo de escalones estrechos y no redondee demasiado los picos agudos de la superficie. En la Figura 1.2 se muestra el efecto de una punta, de forma y dimensión determinada, en la medición de canales con anchos y altos de distintos tamaños.

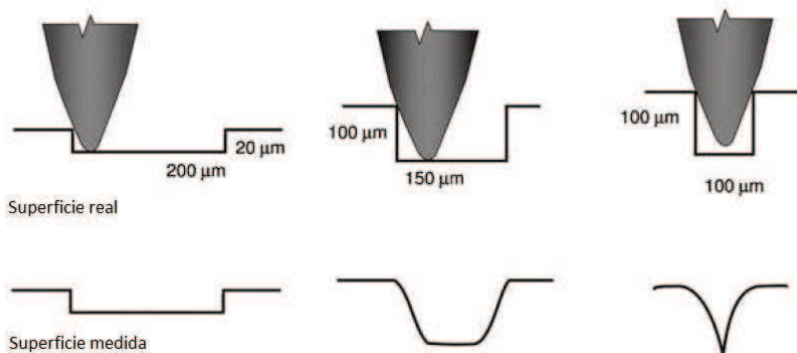


Figura 1.2. Gráfica resultante dada por el perfilómetro para tres canales distintos, utilizando una punta esférica de 50 μm de diámetro (extraído de Schmit, Creath y Wyant, 2007).

Podemos observar que si el radio de la punta es más pequeño, el perfilómetro copia con mayor exactitud la superficie del objeto. Pero si la punta es muy aguda, la presión aplicada sobre la superficie bajo el área de contacto puede resultar muy grande, llegando a deformar la muestra. Si la deformación es elástica, la pieza no se daña, pero el perfil obtenido es incorrecto. Obviamente que si la deformación es plástica la pieza queda dañada y se obtiene también un perfil incorrecto.

Por otro lado, cuando se utiliza una punta de pequeño radio, la velocidad de escaneo debe ser muy reducida, como lo debe ser también la carga, para asegurarse una medición precisa.

Los perfilómetros de aguja de mayor precisión poseen un radio de punta de décimas de micrones y cargas del orden del miligramo o menos. Estos instrumentos requieren un sistema de hermeticidad libre de polvo, como así también estar aislados de vibraciones, y pueden completar el escaneo de unos mil puntos en pocos minutos.

La resolución lateral del perfilómetro de aguja está determinada, como vimos, por el radio de la punta, la forma de la superficie y el intervalo de muestreo.

Por ejemplo, para una aguja de punta esférica que mide una superficie con un perfil sinusoidal, el período más pequeño de la senoide que puede llegar a medirse no solo depende del radio de la punta sino también de la amplitud de la senoide. Se puede calcular el menor período medible d , mediante la siguiente ecuación

$$d = 2\pi\sqrt{ar} \quad (1.1)$$

donde a es la amplitud del perfil sinusoidal y r el radio de la punta. Mínimamente se necesitan dos puntos por período para reconstruir la senoide (criterio de Nyquist), entonces la resolución lateral será $d/2$. Si se posee una punta de $10 \mu\text{m}$ de radio midiendo una superficie con una variación de altura de 1 nm , la resolución lateral es de unos $0,3 \mu\text{m}$, con dos mediciones por período. Cuantas más mediciones por período pueda realizar el perfilómetro la resolución lateral será de menor longitud, incrementándose así la precisión del instrumento [6].

Los perfilómetros de aguja son normalmente calibrados por patrones provistos por institutos de normalización. El patrón mas común es el escalón construido de acero inoxidable o de cromo sobre vidrio fundido. Debe tenerse en cuenta que la altura del escalón debe ser la apropiada para el rango de medición del instrumento.

Para calibrar la resolución lateral se utilizan piezas con una rugosidad superficial patrón. Este patrón de rugosidad posee un perfil de onda cuadrada, con una altura y período determinados, adecuadas al rango del perfilómetro.

Algunos perfilómetros de aguja son capaces de medir escalones de 100 nm de altura, con una repetibilidad de 6 \AA , y una resolución lateral de $0,1 \mu\text{m}$.

Para medir espesores de películas delgadas con la exactitud adecuada este instrumento necesita que el depósito forme un escalón lo más definido posible entre el sustrato y la película [7].

1.1.2 – Microscopio de punta de escaneo:

El microscopio de punta de escaneo o *SPM* (*scanning probe microscope*) es capaz de obtener resoluciones en escala atómica. Trabaja moviendo una fina punta ubicada muy próxima a la superficie bajo estudio, usualmente a pocos angstroms de dicha superficie. En algunos modelos la punta puede estar en contacto directo con la muestra, aplicando fuerzas muy pequeñas, del orden de nano newtons.

El primer microscopio utilizado de este tipo fue el microscopio de efecto túnel o *STM* (*scanning tunneling microscope*), luego se desarrolló el microscopio de fuerza atómica o *AFM* (*atomic force microscope*). Existen variadas formas de operación del *AFM*, de las que podemos mencionar el microscopio de fuerza magnética y el microscopio de fuerza eléctrica.

Los diseños de los *SPM* están evolucionando de manera rápida; constantemente son desarrolladas nuevas puntas de prueba para una gran variedad de aplicaciones, desde circuitos integrados hasta muestras biológicas [8].

1.1.2.1 – Microscopio de efecto túnel (*STM*):

En este instrumento una punta de metal es desplazada sobre la superficie de la muestra, sin estar en contacto con ella. La superficie de esta muestra debe ser de material conductor o semiconductor para que logre establecerse una corriente entre la punta y la superficie bajo estudio. Esta corriente, denominada corriente túnel, es creada aplicando un voltaje entre punta y superficie. Esta corriente túnel solo puede ser detectada si la punta está a menos de 1 nm de distancia sobre la superficie. Cuando la punta se mueve acercándose a dicha superficie la corriente túnel aumenta exponencialmente, una variación de 0,1 nm en la separación aumenta 10 veces el valor de la corriente túnel.

Normalmente la punta escanea a unos pocos angstroms sobre la superficie, y detecta las variaciones de altura de la misma utilizando un transductor piezoeléctrico. La necesidad de que la superficie de la muestra sea conductora limita la aplicación del *STM*, además la resolución que se obtiene depende mucho de la geometría de la punta. La Figura 1.3 es un esquema de un *STM*, que consiste en una sonda con una punta muy fina montada en un transductor piezoeléctrico.

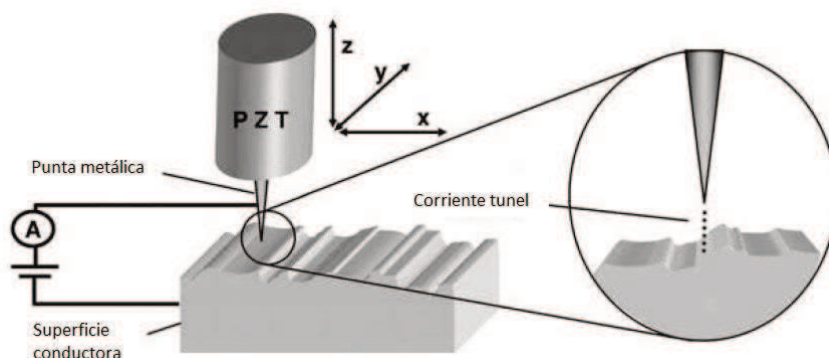


Figura 1.3. Esquema de un *STM* (extraído de Schmit, Creath y Wyant, 2007).

Los modelos de *STM* se diferencian constructivamente en dos formas de trabajo, o modos distintos, que son a corriente constante o a altura constante. El modo de corriente constante utiliza un sistema de retroalimentación para variar la altura de la punta, o de la muestra, durante el escaneo, para lograr mantener la corriente túnel en un valor constante. El modo de altura constante mide la corriente túnel en función de la posición, mientras que la punta se mantiene a una altura fija. El modo de corriente constante es el más utilizado, debido a que pueden medirse variaciones de altura mayores que con el otro modo. Sin embargo el modo de altura constante es más rápido, pero tiene el inconveniente de que la punta puede chocar con la superficie si esta es muy rugosa, y el choque no solo rompería la punta sino que también dañaría la superficie de la muestra.

La resolución del *STM* depende, como ya mencionamos, del tipo de punta y también de la configuración del transductor piezoeléctrico. Se pueden llegar a obtener resoluciones verticales de 0,01 nm y laterales de 0,1 nm [9].

1.1.2.2 – Microscopio de fuerza atómica:

El AFM posee la ventaja de que es apto para escanear todo tipo de superficies, no sólo aquellas que sean las conductoras eléctricas, dado que utiliza la fuerza de interacción entre la punta y la superficie, en vez de la corriente túnel.

Dos tipos de fuerza atómica son utilizadas en el AFM: la fuerza repulsiva y la atractiva. Las mismas se corresponden con los dos modos de operación denominados de contacto y de no contacto.

Utilizando la fuerza repulsiva, la sonda se mueve en contacto con la superficie. La punta de este instrumento tiene, normalmente, unos 10 μm de longitud, con un diámetro en el extremo de entre 4 a 20 nm. Esta punta está ubicada en el extremo libre de una ménsula, usualmente construida de un material muy flexible (como nitruro de silicio), con longitudes que varían entre 100 a 500 μm y espesores de 2 a 10 μm .

La fuerza de interacción repulsiva entre la punta y la superficie de la muestra provoca que la ménsula se flexione, y esta flexión es detectada para medir la topografía del objeto bajo estudio.

La detección de la flexión puede ser realizada por un láser que es reflejado en un espejo montado en la ménsula y luego dirigido hacia un fotodetector. La flexión es medida en base al principio de triangulación óptica. Otra técnica utiliza un diodo láser ubicado sobre la ménsula, cuyo haz es reflejado por la superficie de dicha ménsula y dirigida hacia un detector. Ambos esquemas constructivos se muestran en la Figura 1.4.

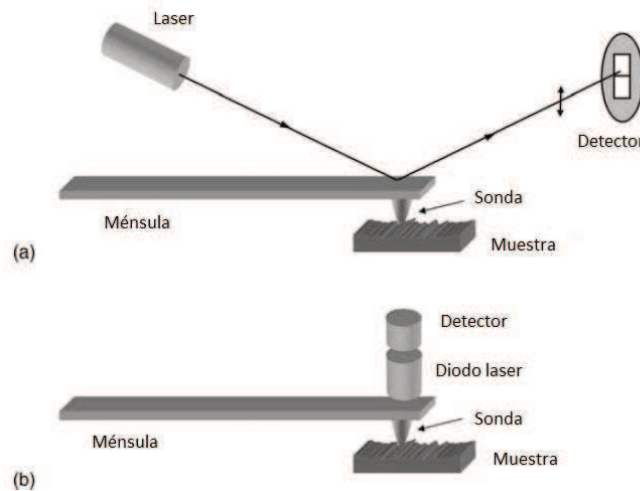


Figura 1.4. Esquema de un *AFM*. a) Detección de desviación óptica. b) Detección con diodo láser (extraído de Schmit, Creath y Wyant, 2007).

Utilizando la fuerza atractiva en el AFM, o modo de no contacto, pueden medirse muestras de material muy suave. Dado que la fuerza atractiva es demasiado pequeña, la punta llega a oscilar a alta frecuencia mientras escanea la superficie, y lo que se detecta, en realidad, son los cambios de amplitud y fase en la vibración de la ménsula, provocados por el aumento o disminución en la energía cinética durante la interacción entre punta y superficie.

El AFM no solo permite determinar la topografía de una muestra, sino que también es capaz de caracterizar las propiedades eléctricas y magnéticas de la misma.

En el AFM pueden llegar a obtenerse resoluciones verticales de 0,1 nm, y laterales desde 2 a 10 nm, dependiendo del modo de funcionamiento y del tipo de muestra [9].

1.2- Perfilómetros ópticos:

A diferencia de los perfilómetros mecánicos, los instrumentos ópticos no están en contacto con la superficie de las muestras. Los perfilómetros ópticos determinan la topografía de la superficie procesando la mejor distancia de focalización sobre el objeto analizado. Dichos instrumentos generan las mediciones a través del registro de la distancia focal de puntos individuales en la superficie, ajustando la altura de la lente hasta lograr el enfoque. La distancia que debe

desplazarse la lente indica la altura de la superficie en ese punto en particular. De esta manera, tomando la suficiente cantidad de puntos de muestreo, toda la superficie es escaneada para luego generar un perfil de alturas, en dos o en tres dimensiones.

Se utilizan diferentes métodos para obtener el perfil superficial, algunos emplean un microscopio convencional y otros un microscopio estereoscópico, ambos equipados con sistemas de enfoque especiales. Todos los métodos se basan en el procesamiento de las imágenes obtenidas en el escaneo, durante el cual se logra la mejor focalización de cada punto de muestreo.

1.2.1- Perfilómetro con sensor óptico de focalización:

En este tipo de instrumento, una fuente de iluminación láser es enfocada en la superficie de la muestra y el haz reflejado por dicha superficie es dividido en dos partes utilizando un prisma. Cada mitad del haz incide sobre dos fotodetectores, y la diferencia en la señal generada por cada detector es procesada. Cuando la lente está ubicada a una distancia mayor que la distancia focal, el haz reflejado se enfoca sobre cada detector, pero generando una señal mayor en el denominado detector interno. Cuando la lente está a una distancia menor que la distancia focal, la señal mayor es generada por el denominado detector externo. La diferencia en las señales es utilizada para enviar una señal de error que mueve el sistema focalizador de la lente hacia la posición correcta. Cuando la lente se sitúa a la distancia focal ambos detectores generan la misma señal y su diferencia es nula. Generalmente se utilizan dos pares de detectores, para corregir las posibles inclinaciones en la posición de la muestra. Un esquema de perfilómetro de este tipo es mostrado en la Figura 1.5.

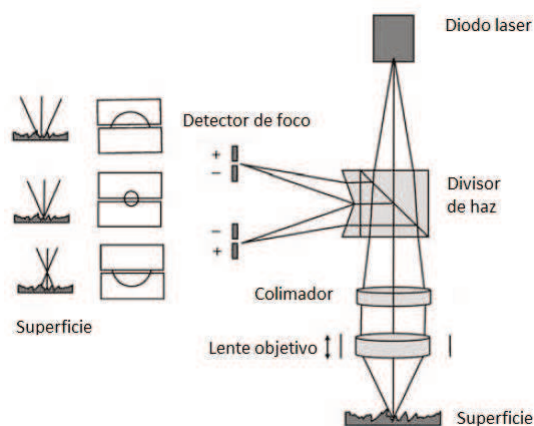


Figura 1.5. Esquema de un perfilómetro óptico con sensores de focalización (extraído de Brodman y Smilga, 1987).

Dado que el foco debe ser ajustado para anular la señal de error por cada punto muestreado de la superficie, a este tipo de perfilómetro le toma varios minutos generar un perfil tridimensional del objeto analizado.

La resolución lateral de este instrumento se encuentra limitada por el tamaño del punto focal del láser, que usualmente varía entre 1 y 1,5 μm de diámetro. En un punto dado, la altura de la superficie medida es un valor promedio de las alturas existentes en toda el área del punto focal.

Una limitación de este tipo de instrumento se debe al hecho de que el haz de luz reflejado por la superficie de la muestra siempre debe incidir sobre los sensores. Si existen pendientes pronunciadas en la superficie, el haz puede ser desviado fuera de dichos sensores, causando resultados erróneos, dado que solo se considera la diferencia de señales generadas por estos en el proceso de medición.

La resolución vertical queda definida por el rango de desplazamiento del sistema focalizador. Si la diferencia de altura a medir es grande se pueden realizar desplazamientos de mayor longitud de la lente y así reducir el tiempo de adquisición entre puntos de muestreo. Por el contrario, un desplazamiento pequeño de la lente, para lograr una focalización más fina en cada punto de muestreo, incrementa considerablemente los tiempos de adquisición de datos.

La calibración del sensor óptico de focalización es similar a la del perfilómetro de aguja y el SPM, utilizando un objeto patrón (escalón o perfil de onda cuadrada) adecuado al rango del instrumento en particular [10].

1.2.2- Microscopio confocal:

Este perfilómetro óptico es muy utilizado, y en su principio de funcionamiento se aplica el sistema de focalización.

El microscopio confocal se diferencia de los microscopios ópticos comunes por el uso de pequeñas aberturas circulares o *pinholes*, para asegurar que sólo el haz de luz del punto enfocado en la superficie de la muestra llegue al detector. La eliminación de los haces fuera de foco y la gran relación señal – ruido de las imágenes obtenidas, son las principales ventajas de este instrumento. La configuración más utilizada del confocal posee dos *pinholes* ubicados en planos conjugados del plano focal de la lente objetivo, uno frente a la fuente de iluminación y el otro frente al detector, como se muestra en la Figura 1.6.

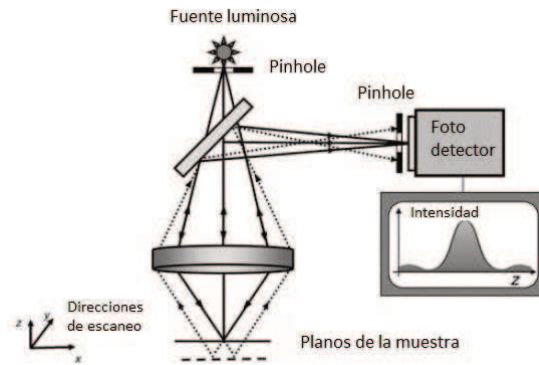


Figura 1.6. Esquema de un microscopio confocal de reflexión (extraído de Schmit, Creath y Wyant, 2007).

En algunos de estos instrumentos se utiliza una fuente de luz coherente (láser), eliminando la necesidad de utilizar el *pinhole* en la fuente. Sin embargo en este tipo de instrumento, denominado microscopio láser confocal, el *speckle* provocado por la iluminación láser puede ser un inconveniente. Dado que en el microscopio confocal se toma un punto de la superficie de la muestra por vez, para obtener la topografía del objeto se debe contar con un sistema sincronizado de escaneo lateral que obtenga la cantidad requerida de puntos de muestreo sobre todo el plano focal.

1.2.2.1- Procesamiento y apilamiento de las imágenes:

Escaneando punto a punto, de manera de formar una grilla de puntos en el plano x - y , y procesando la irradiancia que llega al fotodetector, obtenemos una imagen de la muestra a nivel del plano focal. Esta es una imagen en 2D de un plano que se encuentra a un nivel o altura dada de la superficie. Para obtener una imagen en 3D, la lente objetivo se mueve verticalmente, en la dirección z una distancia determinada, para luego escanear todo el plano horizontal, de manera de obtener la cantidad de imágenes suficientes, a diferentes niveles de la superficie de la muestra, procesarlas y reconstruir una imagen 3D de la misma. Esto es lo que se denomina “apilar las imágenes”, ya que da una idea de formar una pila de imágenes 2D para reconstruir la imagen 3D de la muestra.

Para cada punto de la muestra que se encuentra en las proximidades del plano focal del microscopio (sobre, arriba o debajo de este), la señal de irradiancia recolectada por el detector, también denominada función de dispersión de punto o *PSF* (*point spread function*), disminuye cuando la distancia entre dicho punto y el plano focal aumenta. El ancho de esta PSF para un punto dado de la muestra, utilizando un objetivo de microscopio sin aberraciones, con una apertura numérica o *NA* (*numeric aperture*) menor a 0,5, puede ser calculada por la siguiente expresión [11]

$$WPSF = \frac{0.9\lambda}{NA^2} \quad (1.2)$$

donde λ es la longitud de onda del láser utilizado.

Obviamente, la señal de irradiancia detectada de un punto de la muestra durante el escaneo, tiene su valor máximo si este punto corresponde al plano focal.

Se obtiene una mejor resolución si la señal de irradiancia detectada posee un pico máximo agudo, y esto depende tanto de la *NA* como del aumento de la lente objetivo.

La resolución vertical está determinada por el paso mínimo en el movimiento vertical de la lente. Para lentes objetivo de gran aumento puede ser de unos nanómetros, y para lentes de poco aumento entre 10 a 15 nm, llegando hasta los micrones.

El microscopio confocal es principalmente utilizado para medir el espesor de capas transparentes, donde se realiza un proceso llamado corte óptico. En este proceso, la luz penetra en el objeto bajo estudio y es reflejada por las interfaces de las capas. El procedimiento es comúnmente aplicado en biología para medir células y en la industria de semiconductores para medir depósitos transparentes.

La resolución vertical del corte óptico puede ser de 1,5 μm , para lentes objetivo con gran *NA*, mucho mayor que en el procedimiento descrito anteriormente [12].

1.3- Perfilómetros ópticos interferométricos:

La interferencia óptica se vale de la formación de un patrón o imagen de interferencia entre dos frentes de ondas de luz, una que proviene directamente de la fuente lumínica (denominada haz de referencia) y otra proveniente del reflejo (o la transmisión) sobre (o a través de) la superficie de la muestra (denominada haz objeto).

Existen numerosas configuraciones de interferómetros, que se describirán más adelante. Aquí se menciona brevemente que si la diferencia en el camino óptico entre los frentes de onda de referencia y objeto no es demasiado grande el patrón de interferencia formado consiste en franjas que pueden mostrar pequeñas desviaciones provocadas por la rugosidad de la superficie de la muestra [13].

Los perfilómetros interferométricos comerciales están basados en microscopios comunes donde las lentes objetivo son reemplazadas por un sistema de interferencia óptica. Estos instrumentos utilizan las imágenes de las franjas creadas por la interferencia para calcular la altura de diferentes puntos en la superficie de la muestra. Durante la medición la imagen de la interferencia puede ser modificada, variando la distancia óptica entre el haz que proviene de la muestra y el haz de referencia. Esta variación en el patrón de interferencia puede llevarse a cabo mediante técnicas de cambio de fase o también de escaneo vertical, donde se mueve la lente objetivo en relación a la superficie de la muestra mediante un transductor piezoeléctrico o un motor paso a paso.

1.3.1- Configuraciones más usuales:

Podemos considerar cuatro configuraciones básicas en la fabricación de perfilómetros interferométricos, que se muestran esquemáticamente en la Figura 1.7. Cada uno de ellos tiene un desempeño óptimo para una aplicación en particular.

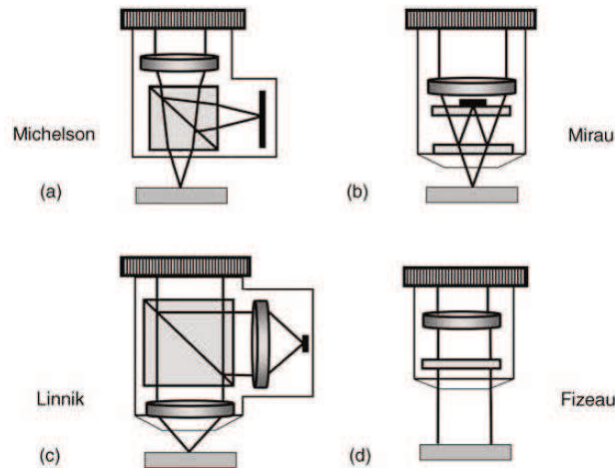


Figura 1.7. Esquemas de interferómetros. a) Michelson, b) Mirau, c) Linnik y d) Fizeau (extraído de Schmit, Creath y Wyant, 2007).

Observamos que las cuatro configuraciones de interferómetros se diferencian en la manera en cómo se divide el haz para formar el haz de referencia y el haz objeto. En todas las configuraciones, excepto en la de Fizeau, el espejo del haz de referencia está ubicado a la distancia focal de la lente objetivo. Para obtener la imagen con el mejor contraste de franjas, primero se debe poner en foco la muestra y luego mover la posición de dicho espejo hasta obtener el mejor contraste, que es la posición de máximo enfoque del instrumento.

El interferómetro de Michelson está compuesto por una lente objetivo, un divisor de haz y un espejo para el haz de referencia. En este esquema el conjunto objetivo posee una gran distancia entre la lente y la muestra, comparado con los otros esquemas, por lo que este interferómetro sólo puede ser utilizado con lentes de bajo aumento, con baja NA y gran distancia focal.

El interferómetro de Mireau posee dos vidrios planos entre la lente objetivo y la muestra. Un vidrio contiene un pequeño espejo para el haz de referencia, y el otro está recubierto de un solo lado, para funcionar como divisor de haz. Este instrumento es utilizado con lentes objetivo de aumento medio, dado que hay poca distancia entre la lente y la muestra. Aumentos menores a 10X no pueden ser utilizados, dado que el pequeño espejo de referencia oscurece demasiado el campo visual. Para aumentos mayores a 50X, la distancia focal de la lente resulta demasiado corta para este esquema que debe alojar a los dos vidrios planos.

El esquema del interferómetro de Linnik está compuesto por un cubo divisor de haz, dos lentes y un espejo para el haz de referencia. Las dos lentes deben ser iguales y estar alineadas con el

divisor de haz para obtener un frente de onda con la mínima aberración y el máximo contraste de franjas. Este instrumento se utiliza con lentes de gran aumento, dada la pequeña distancia entre la lente objeto y la muestra, aunque pueden utilizarse lentes de cualquier aumento. Sin embargo este interferómetro requiere un ajuste óptico difícil de realizar en la práctica, por lo que su uso comercial se vio algo limitado.

El interferómetro de Fizeau es un instrumento que posee distancias de camino óptico desiguales, a diferencia de los anteriores basados todos en el interferómetro de Michelson (con distancias ópticas iguales), por lo que este esquema requiere de una fuente de luz con gran longitud de coherencia, como el láser. La lente de este instrumento crea un haz colimado en la superficie de la muestra, y a su vez captura la imagen de la misma. Se pueden observar las franjas de interferencia sobre todo el campo visual, siempre y cuando la lente haga foco en la superficie de la muestra. El dispositivo puede utilizarse con lentes de una gran variedad de aumentos, pero debe considerarse que en su configuración el semiespejo del haz de referencia está situado entre la lente y la muestra, existiendo una distancia de separación determinada entre estos dos elementos.

Cabe mencionar que en todos estos interferómetros mencionados, excepto el de Fizeau, el haz que incide sobre la superficie bajo estudio forma un cono de luz, por lo tanto la posición de la lente objetivo debe ser ajustada cuidadosamente hasta lograr observar franjas de interferencia [14].

1.3.2- Resolución:

En estos instrumentos la determinación de la resolución lateral está basada en la de los microscopios ópticos, y depende fundamentalmente de la resolución de la lente objetivo utilizada. Esta es función de la longitud de onda de la fuente utilizada y de la *NA* de la lente. Utilizando el criterio de Sparrow, podemos calcularla por medio de la siguiente expresión [15]

$$LR = \frac{0.5\lambda}{NA} \quad (1.3)$$

Este criterio se aplica para luz incoherente, considerando dos fuentes de luz puntuales separadas una distancia dada, sin embargo es utilizada en este caso, ya que en la práctica da buenos resultados.

Por otro lado, en el tratamiento digital, la imagen registrada por una cámara *CCD* necesita de al menos tres píxeles para lograr formar una imagen de dos fuentes puntuales separadas.

La topografía de una muestra solo puede ser obtenida si las franjas de interferencia cubren toda la superficie. Si la muestra se mantiene en la posición focal, las franjas pueden ser observadas en todos los niveles de alturas de la superficie, pero es necesario que este rango de alturas se encuentre dentro de la longitud de la profundidad de campo de la lente objetivo.

La profundidad de campo de la lente objetivo depende de la longitud de onda de la fuente y de la NA, y se encuentra definida por [16]

$$DF = \frac{\lambda\sqrt{1-(NA)^2}}{(NA)^2} \quad (1.4)$$

La mínima diferencia en altura medible de una muestra está determinada por lo pequeña que sea la longitud de la profundidad de campo.

Los perfilómetros ópticos interferométricos comerciales más utilizados poseen una fuente de luz coherente monocromática, y utilizan la técnica de la interferometría de cambio de fase o *PSI* (*phase shifting interferometry*) para procesar las imágenes, cuyos detalles serán desarrollados en los próximos capítulos. Estos instrumentos pueden brindar imágenes con muy baja interferencia, por lo que superficies prácticamente planas llegan a ser medidas con precisión, en el orden de los angstroms. Sin embargo la técnica de *PSI* se encuentra limitada por el hecho de que sólo se pueden medir diferencias de altura menores a $\lambda/4$.

1.3.3- Microscopía Holográfica:

Es una técnica utilizada para realizar observaciones microscópicas mediante la obtención de imágenes holográficas. Con esta técnica pueden medirse espesores de películas delgadas, con la ventaja mencionada de que la muestra no está en contacto con el instrumento, puede estar compuesta de cualquier material y no requiere una preparación especial para ser observada.

Para realizar microscopía holográfica se emplean interferómetros tipo *Mach-Zehnder* y *Michelson*, como los descritos anteriormente.

En la realización de este trabajo adoptamos dicha técnica, con una configuración de interferómetro de *Michelson*, cuyas principales ventajas consisten en no necesitar lentes objetivo ni sistemas de escaneo para obtener los hologramas, como si ocurre con los dispositivos para microscopía holográfica construidos comercialmente, esto resultó en una configuración óptica menos sofisticada, de fácil ajuste y sobre todo más económica en comparación a los equipos existentes en el mercado.

Para medir los espesores de películas delgadas, que es el principal objetivo del trabajo, se deben procesar digitalmente las imágenes holográficas capturadas por una cámara *CMOS*, con dicho procesamiento calcular el espesor de capa delgada, y obtener gráficas de perfiles y reconstrucciones en 3D de la topografía superficial de las muestras. Todo este procedimiento es descrito detalladamente en el capítulo V donde se describe el trabajo experimental.

Objetivos:

Los objetivos planteados en el desarrollo de esta tesis se describen a continuación:

- Estudiar el fundamento teórico dado por la óptica física para aplicarlo a un sistema de obtención de imágenes holográficas.
- Realizar el montaje de los elementos ópticos. Se adopta un esquema de interferómetro de Michelson – Morley, el cual debe ajustarse para lograr la interferencia óptica (formación de franjas en la imagen).
- Obtener imágenes con la calidad necesaria, minimizando las vibraciones, las fuentes de luz externas, y las interferencias introducidas en el CMOS.
- Elaborar y analizar los algoritmos en Matlab®, utilizando la transformada discreta de Fourier para procesar la imagen holográfica, a partir de la cual se obtienen los perfiles y las imágenes tridimensionales de la superficies de la muestras.
- Estudiar e implementar diversos algoritmos de ajuste de la distancia focal, con el software Matlab®, adoptando el que mejor se desempeña para nuestro esquema óptico, determinando la distancia de reconstrucción óptima de la imagen.
- Analizar e implementar algoritmos, en Matlab®, para la corrección de aberraciones ópticas presentes en la imagen de fase, utilizando los polinomios de Zernike.
- Por medio de la imagen de las diferencias de fase obtenida del holograma, y luego de corregir la distancia de reconstrucción y las aberraciones, determinar los espesores de las películas delgadas, en segmentos previamente seleccionados de la superficie de la muestra, mediante algoritmos de Matlab®.
- A partir de la imagen de fase corregida, mencionada en el punto anterior, obtener una imagen tridimensional de un sector seleccionado de la superficie de la muestra, mediante algoritmos de Matlab®.

Por último, demostrar como comprobación de viabilidad del procesamiento adoptado, el reenfoque numérico que se tiene en holografía digital, y que es una característica única de esta técnica. Para ello se deben realizarse simulaciones numéricas, y sus resultados compararse con valores teóricos y con los resultados experimentales obtenidos, como forma de validar esta propuesta de trabajo.

Referencias:

- [1] Z Yuan, X Zhu, X Wang, X Cai, B Zhang, D Qiu, H Wu (2011). Thin Solid Films.
- [2] K Seshan (2002). Handbook of Thin Film Deposition Processes and Techniques: Principles, Methods, Equipment and Applications. Andrew ed.
- [3] S Sahu, S Majee, A Pal (2007). Applied Physics Letter.
- [4] B Singh (1999). Applied Surface Science.
- [5] F Blais (2004). Review of 20 Year of Range Sensor Development, Journal of Electronic Imaging.
- [6] J Bennett, J Dancy (1981). Stylus Profiling Instrument for Measuring Statistical Properties of Smooth Optical Surfaces, Applied Optics.
- [7] J Schmit, K Creath, J Wyant (2007). Surface Profilers, Multiple Wavelength and White Light Interferometry.
- [8] B Bhushan, H Fuchs, S Hosaka (2004). Applied Scanning Probe Methods, Springer ed.
- [9] S Cohen, K Lightbody (1999). Atomic Force Microscopy / Scanning Tunneling Microscopy, Springer ed.
- [10] U Breitmeier, R Ahler (1987). Dynamically Focusing Electro – Optical Sensor Systems for Microprofilometry, SPIE ed.
- [11] T Corle, G Kino (1996). Confocal Scanning Optical Microscopy and Related Imaging Systems, Academic Press, N.Y..
- [12] C Sheppard (2003). Scanning Confocal Microscopy, Encyclopedia of Optical Engineering, Driggers ed.
- [13] T Kreis (2005). Handbook of Holographic Interferometry, Wiley – VCH ed.
- [14] P Hariharan (2007). Optical Interferometry, Academic Press, San Diego.
- [15] M Born, E Wolf (1999). Principles of Optics: Electromagnetic Theory of Propagation, Interference and Diffraction of Light, University Press, Cambridge.
- [16] J Benford (1966). Microscope Objectives, Applied Optics and Optical Engineering, Kingslake ed.

Capítulo II

Óptica física: interferencia y difracción

Los fenómenos denominados interferencia y difracción de la luz juegan un rol muy importante en la propagación de ondas luminosas. En este capítulo desarrollamos algunos fundamentos teóricos sobre la interferencia y sobre la teoría escalar de la difracción de ondas, dentro del área denominada óptica física, dado que para poder llegar a comprender las propiedades de los sistemas que forman las imágenes en la interferometría holográfica es esencial conocer estos fenómenos y las limitaciones que ellos imponen [1].

2.1- Interferencia:

Definimos la interferencia óptica como la interacción de dos o más ondas de luz, que producen una irradiancia resultante diferente a solo la suma de las irradiancias individuales de cada onda.

Obtengamos una expresión para esta irradiancia resultante de la interferencia de dos ondas luminosas. Para ello, consideremos dos fuentes puntuales que emiten ondas monocromáticas de igual frecuencia en un medio homogéneo, consideremos que los frentes de onda son planos, y que las ondas están linealmente polarizadas. Estas ondas están definidas por las siguientes expresiones

$$\vec{A}_1(\vec{r}, t) = \vec{A}_{01} \cdot \cos(\vec{k}_1 \cdot \vec{r} - \omega t + \Phi_1) \quad , \quad \vec{A}_2(\vec{r}, t) = \vec{A}_{02} \cdot \cos(\vec{k}_2 \cdot \vec{r} - \omega t + \Phi_2) \quad (2.1)$$

Para estudiar la interferencia debemos recurrir a la irradiancia, por que es el parámetro que puede medirse directamente con ayuda de sensores (los ojos, fotocélulas, fotodiodos, etc.), la misma se define por

$$I = \langle \vec{A}^2 \rangle_T \quad (2.2)$$

Que es el promedio temporal de la magnitud del campo electromagnético al cuadrado.

Dado que

$$\vec{A}^2 = \vec{A} \cdot \vec{A} = (\vec{A}_1 + \vec{A}_2) \cdot (\vec{A}_1 + \vec{A}_2) = \vec{A}_1^2 + \vec{A}_2^2 + 2 \cdot \vec{A}_1 \cdot \vec{A}_2 \quad (2.3)$$

donde

$$\vec{A}_1 \cdot \vec{A}_2 = \vec{A}_{01} \cdot \cos(\vec{k}_1 \cdot \vec{r} - \omega t + \Phi_1) \cdot \vec{A}_{02} \cdot \cos(\vec{k}_2 \cdot \vec{r} - \omega t + \Phi_2) \quad (2.4)$$

o de manera equivalente

$$\vec{A}_1 \cdot \vec{A}_2 = \vec{A}_{01} \cdot \vec{A}_{02} [\cos(\vec{k}_1 \cdot \vec{r} + \Phi_1) \cdot \cos(\omega t) + \text{sen}(\vec{k}_1 \cdot \vec{r} + \Phi_1) \cdot \text{sen}(\omega t)] \cdot [\cos(\vec{k}_2 \cdot \vec{r} + \Phi_2) \cdot \cos(\omega t) + \text{sen}(\vec{k}_2 \cdot \vec{r} + \Phi_2) \cdot \text{sen}(\omega t)] \quad (2.5)$$

Si consideramos la definición dada en (2.2), obtenemos para la ecuación (2.3)

$$\langle \vec{A}_1^2 \rangle_T = I_1 \quad , \quad \langle \vec{A}_2^2 \rangle_T = I_2 \quad , \quad \langle \vec{A}_1 \cdot \vec{A}_2 \rangle_T = \frac{1}{T} \int_t^{t+T} (\vec{A}_1 \cdot \vec{A}_2) dt \quad (2.6)$$

El período de estas ondas es $\tau = 2\pi/\omega$, y en la integral $T \gg \tau$, por lo que es igual a

$$\langle \vec{A}_1 \cdot \vec{A}_2 \rangle_T = \frac{1}{2} \vec{A}_{01} \cdot \vec{A}_{02} \cdot \cos(\vec{k}_1 \cdot \vec{r} + \Phi_1 - \vec{k}_2 \cdot \vec{r} - \Phi_2) = \frac{1}{2} \vec{A}_{01} \cdot \vec{A}_{02} \cdot \cos \delta \quad (2.7)$$

Donde tenemos que la diferencia de fase es

$$\delta = (\vec{k}_1 - \vec{k}_2) \cdot \vec{r} + (\Phi_1 - \Phi_2) \quad (2.8)$$

El valor de las irradiancias individuales y la dada por la combinación de las dos ondas, resultan de resolver las integrales del promedio temporal, obteniéndose

$$I_1 = \frac{1}{2} \vec{A}_{01}^2 \quad , \quad I_2 = \frac{1}{2} \vec{A}_{02}^2 \quad , \quad I_{1-2} = \vec{A}_{01} \cdot \vec{A}_{02} \cdot \cos \delta \quad (2.9)$$

Por lo que la irradiancia total es igual a

$$I = I_1 + I_2 + 2\sqrt{I_1 \cdot I_2} \cos \delta \quad (2.10)$$

Un máximo de irradiancia se obtiene cuando $\delta = 0, \pm 2\pi, \pm 4\pi, \dots$ donde

$$I_{max} = I_1 + I_2 + 2\sqrt{I_1 \cdot I_2} \quad (2.11)$$

Este caso es denominado interferencia constructiva.

Un mínimo de irradiancia se obtiene si $\delta = \pm\pi, \pm 3\pi, \pm 5\pi \dots$ donde

$$I_{min} = I_1 + I_2 - 2\sqrt{I_1 \cdot I_2} \quad (2.12)$$

Este caso es denominado interferencia destructiva.

Las irradiancias máximas y mínimas son las responsables de producir el llamado patrón de franjas de interferencia.

Observamos que para el caso donde $I_1 = I_2$ obtenemos

$$I = 2I_1(1 + \cos\delta) = 4I_1 \cdot \cos^2\left(\frac{\delta}{2}\right) \quad (2.13)$$

Ahora podemos extender este análisis para el caso de ondas esféricas. Estas son emitidas por ambas fuentes, y las podemos definir por las siguientes expresiones

$$\vec{A}_1(\vec{r}_1, t) = \vec{A}_{01}(\vec{r}_1) \cdot \exp[j(\vec{k}\vec{r}_1 - \omega t + \Phi_1)] \quad \vec{A}_2(\vec{r}_2, t) = \vec{A}_{02}(\vec{r}_2) \cdot \exp[j(\vec{k}\vec{r}_2 - \omega t + \Phi_2)] \quad (2.14)$$

donde r_1 y r_2 son los radios de los frentes de onda esféricos que salen de cada fuente y se superponen en un punto P determinado. En este caso la diferencia de fase esta dado por

$$\delta = (\vec{r}_1 - \vec{r}_2) \cdot \vec{k} + (\Phi_1 - \Phi_2) \quad (2.15)$$

Si consideramos iguales a las irradiancias de cada fuente, entonces la irradiancia total es la misma que se obtuvo en la ecuación (2-13), donde tenemos que los máximos ocurren cuando $\delta = 2m \cdot \pi$, con $m = 0, \pm 1, \pm 2, \dots$, y los mínimos cuando $\delta = n \cdot \pi$, con $n = \pm 1, \pm 3, \pm 5, \dots$, o si se prefiere $n = 2m+1$ [2].

2.1.1 - Experimento de Young:

Sea una onda plana monocromática que ilumina una rendija larga y estrecha. Desde esa primera rendija la luz se difracta y emerge como una onda cilíndrica, esta onda a su vez incide sobre dos rendijas estrechas S_1 y S_2 que están ubicadas muy juntas y en forma paralela, tal y como vemos en la Figura 2.1

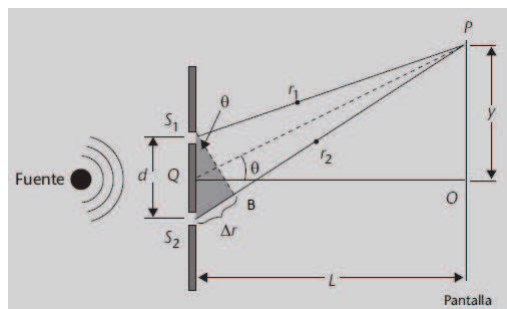


Figura 2.1. Geometría del experimento de Young (extraído de Marc Figueras Atienza, 2011).

Por simetría, el frente de onda primario llega a las dos rendijas con exactamente la misma fase, por lo que en las dos fuentes secundarias tenemos que $\Phi_1 = \Phi_2$.

La diferencia de camino óptico entre los rayos S_1P y S_2P puede calcularse a partir de una perpendicular a S_1P desde S_2 , y viene dada por

$$\overline{S_1B} = \overline{S_1P} - \overline{S_2P} = \vec{r}_1 - \vec{r}_2 \quad (2.16)$$

De la Figura 2.1 tenemos que

$$\vec{r}_1 - \vec{r}_2 = d \cdot \text{sen}\theta \approx d \cdot \theta \quad (2.17)$$

si el ángulo θ es muy pequeño. Además

$$\tan \theta \approx \theta \approx \frac{y}{L} \quad (2.18)$$

Entonces reemplazando en la ecuación (2.17), tenemos que

$$\vec{r}_1 - \vec{r}_2 = d \frac{y}{L} \quad (2.19)$$

Reescribiendo la ecuación (2.15) obtenemos para la diferencia de fase

$$\delta = \vec{k}(\vec{r}_1 - \vec{r}_2) = \frac{2\pi}{\lambda} \frac{dy}{L} \quad (2.20)$$

Quedándonos la ecuación (2.13) de la forma siguiente

$$I = 4I_1 \cdot \cos^2 \left(\frac{\pi d}{\lambda L} y \right) \quad (2.21)$$

Resultando en un patrón de franjas horizontales dispuestas a lo largo del eje vertical y . Las franjas brillantes producidas por la interferencia constructiva ocurren cuando

$$\frac{2\pi}{\lambda} (\vec{r}_1 - \vec{r}_2) = 2m\pi \quad \rightarrow \quad (\vec{r}_1 - \vec{r}_2) = m\lambda \quad (2.22)$$

Entonces la m -ésima franja brillante, tomando la ecuación (2.19), se encuentra a una distancia del eje de simetría igual a

$$y_m = m \frac{L}{d} \lambda \quad (2.23)$$

y la separación entre franjas podemos calcularla a través de

$$\Delta y = y_{m+1} - y_m = (m + 1) \frac{L}{d} \lambda - m \frac{L}{d} \lambda = \frac{L}{d} \lambda \quad (2.24)$$

En la Figura 2.2 podemos apreciar el patrón de franjas de interferencia del experimento de las dos rendijas de Young [3].



Figura 2.2. Patrón de interferencia de dos rendijas (extraído de Marc Figueras Atienza, 2011).

2.1.2 - Coherencia espacial y temporal:

Para poder observar una distribución de franjas de interferencia, las dos ondas tienen que marchar juntas, manteniendo una relación de fase bien definida, en esta condición se dice que ambas ondas son coherentes.

Debido a que las ondas monocromáticas son una idealización que no existe en la realidad física, ya que todas las ondas reales están formadas por trenes de onda finitos, debemos comenzar nuestro análisis suponiendo un pulso armónico determinado por la siguiente expresión en el campo espacial

$$A(x) = \begin{cases} A_0 \cos(k_p x) & , \text{ para } -L \leq x \leq L \\ 0 & , \text{ para } |x| > L \end{cases} \quad (2.25)$$

Aunque también el pulso puede ser función del tiempo.

A continuación tratemos de ver el efecto de la frecuencia espacial k_p , por lo que necesitamos obtener el espectro de frecuencias de la función (2.25) por medio de su transformada de Fourier

$$A(k) = \mathcal{F}[A(x)] = A_0 \left[\frac{\text{sen}[(k+k_p)L]}{(k+k_p)} + \frac{\text{sen}[(k-k_p)L]}{(k-k_p)} \right] \quad (2.26)$$

o si se prefiere

$$A(k) = A_0 \cdot L \left[\text{sinc}[(k+k_p)L] + \text{sinc}[(k-k_p)L] \right] \quad (2.27)$$

Cuando existen muchas oscilaciones en el tren de ondas, tenemos que $\lambda_p \ll L$ (como ocurre con las ondas electromagnéticas), por lo que $k_p \cdot L \gg 2\pi$, y la función $\text{sinc}[(k+k_p)L]$ tiene valores despreciables. Si $k=k_p$ la función $\text{sinc}[(k-k_p)L]$ es igual a 1, su máximo valor, y si como acabamos de mencionar $L \gg \lambda_p$, el pico de esta última función es muy estrecho, y lejos de él dicha función cae a valores muy pequeños. Para valores de frecuencia $k = k_p \pm \pi/L$ la función es cero, por lo que

el espectro se concentra en este intervalo. Así, de esta manera podemos considerar al espectro de frecuencia espacial simplificado en la siguiente expresión

$$A(k) = A_0 \cdot L \cdot \text{sinc}[(k - k_p)L] \quad (2.28)$$

En la Figura 2.3 podemos observar el pulso como función de la longitud (x) y el espectro de frecuencias espaciales

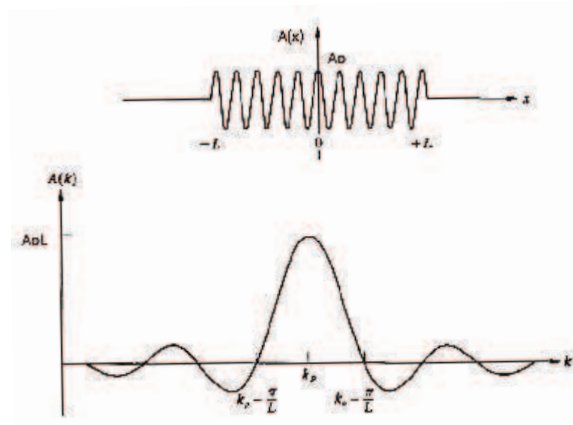


Figura 2.3. Pulso cosenoidal y su espectro de frecuencias (extraído de Hecht, 2015).

De la misma forma podemos trabajar en el campo temporal, si el pulso está definido en un período de tiempo $2T$, obtenemos la expresión

$$A(t) = \begin{cases} A_0 \cos(\omega_p t) & , \text{ para } -T \leq x \leq T \\ 0 & , \text{ para } |x| > T \end{cases} \quad (2.29)$$

Con su espectro de frecuencia temporal definido por

$$A(\omega) = A_0 \cdot T \cdot \text{sinc}[(\omega - \omega_p)T] \quad (2.30)$$

Donde ω y k están relacionadas por la velocidad de fase.

Ahora estamos en condiciones de definir la anchura de banda de frecuencia, si observamos la Figura 2.3, dada por

$$\Delta k = \frac{2\pi}{L} \quad , \quad \Delta \omega = \frac{2\pi}{T} \quad (2.31)$$

Si la onda se comporta como una senoide dentro de la longitud $\Delta x = 2L$, llamamos a esta longitud de coherencia Δl_c . Lo mismo sucede en el dominio temporal, siendo el intervalo de coherencia $\Delta t_c = 2T$.

Observamos que $\Delta l_c \cdot \Delta k = 4\pi$, lo mismo que $\Delta t_c \cdot \Delta \omega = 4\pi$, resulta en un valor constante.

Ambos intervalos de frecuencia están relacionados obviamente por

$$\Delta l_c = c \cdot \Delta t_c \quad (2.32)$$

La coherencia espacial describe la mutua correlación de diferentes partes del mismo frente de onda. La coherencia temporal describe la correlación de una onda consigo misma, observada en diferentes instantes de tiempo.

Para que dos haces puedan producir un patrón de interferencia, su frecuencia debe ser casi igual, dado que una diferencia de frecuencias significativa resulta en un desfase de rápida variación temporal, lo que haría que el promedio de irradiancia mutua I_{1-2} resultara nulo durante el intervalo de detección. Por otro lado, los haces no tienen que estar en fase entre si, siempre y cuando este desfase entre las fuentes permanezca constante, de esta manera dichas fuentes se denominan coherentes.

Podemos mencionar, por ejemplo, que una lámpara de mercurio de laboratorio posee una longitud de coherencia de unos tres milímetros, mientras que un laser de He-Ne tiene una longitud de coherencia de 400 metros [4].

Pasemos a ver ahora lo que se denomina el grado de coherencia. Para ello, consideremos dos fuentes puntuales que interfieren en un punto P (como en el experimento de Young), formando un campo complejo, en función del tiempo, igual a

$$\widetilde{A}_P(t) = \widetilde{K}_1 \cdot \widetilde{A}_1(t - t_1) + \widetilde{K}_2 \cdot \widetilde{A}_2(t - t_2) \quad (2.33)$$

donde el símbolo \sim denota una función compleja, $t_1 = r_1/c$ y $t_2 = r_2/c$. Las cantidades \widetilde{K}_1 y \widetilde{K}_2 se llaman propagadores, y dependen del tamaño de las aberturas y sus posiciones relativas. Ellos manifiestan una reducción del campo que podría darse por absorción, difracción, etc. En este desarrollo solo consideramos que los propagadores provocan un desfase de $\pi/2$ en cada campo, que puede introducirse multiplicando por $\exp(j\pi/2)$, por lo que \widetilde{K}_1 y \widetilde{K}_2 son imaginarios puros.

La irradiancia resultante medida en el punto P, en un intervalo de tiempo $T \gg \Delta t_c$, es

$$I = \langle \widetilde{A}_P(t) \widetilde{A}_P^*(t) \rangle_T = \widetilde{K}_1 \cdot \widetilde{K}_1^* \langle \widetilde{A}_1(t - t_1) \widetilde{A}_1^*(t - t_1) \rangle_T + \widetilde{K}_2 \cdot \widetilde{K}_2^* \langle \widetilde{A}_2(t - t_2) \widetilde{A}_2^*(t - t_2) \rangle_T + \widetilde{K}_1 \cdot \widetilde{K}_2^* \langle \widetilde{A}_1(t - t_1) \widetilde{A}_2^*(t - t_2) \rangle_T + \widetilde{K}_2 \cdot \widetilde{K}_1^* \langle \widetilde{A}_2(t - t_2) \widetilde{A}_1^*(t - t_1) \rangle_T \quad (2.34)$$

Si suponemos que los campos son estacionarios, de manera que el promedio temporal es independiente del origen de tiempo elegido, a los dos primeros promedios temporales de la ecuación (2.34) podemos escribirlos como

$$\langle \widetilde{A}_1(t) \widetilde{A}_1^*(t) \rangle_T = I_{S1} \quad \text{y} \quad \langle \widetilde{A}_2(t) \widetilde{A}_2^*(t) \rangle_T = I_{S2} \quad (2.35)$$

Definiendo $\tau = t_2 - t_1$, podemos desplazar el origen temporal una cantidad t_2 en los dos últimos términos de la ecuación (2.34), resultando

$$\widetilde{K}_1 \cdot \widetilde{K}_2^* \langle \widetilde{A}_1(t + \tau) \widetilde{A}_2^*(t) \rangle_T + \widetilde{K}_2 \cdot \widetilde{K}_1^* \langle \widetilde{A}_2(t) \widetilde{A}_1^*(t + \tau) \rangle_T = 2 \cdot \text{Re} [\widetilde{K}_1 \cdot \widetilde{K}_2^* \langle \widetilde{A}_1(t + \tau) \widetilde{A}_2^*(t) \rangle_T] \quad (2.36)$$

De la definición de los propagadores como imaginarios puros, tenemos que $\widetilde{K}_1 \cdot \widetilde{K}_2^* = \widetilde{K}_2 \cdot \widetilde{K}_1^* = |\widetilde{K}_1| \cdot |\widetilde{K}_2|$. Por otro lado, el promedio temporal de la ecuación (2.36) es una función de correlación cruzada, que designamos por

$$\widetilde{F}_{12}(\tau) = \langle \widetilde{A}_1(t + \tau) \widetilde{A}_2^*(t) \rangle_T \quad (2.37)$$

Esta es la llamada función de coherencia mutua de los campos luminosos.

Si reemplazamos lo anterior en la ecuación (2.34), obtenemos

$$I = |\widetilde{K}_1|^2 \cdot I_{S1} + |\widetilde{K}_2|^2 \cdot I_{S2} + 2 \cdot |\widetilde{K}_1| \cdot |\widetilde{K}_2| \text{Re} [\widetilde{F}_{12}(\tau)] = I_1 + I_2 + 2 \cdot |\widetilde{K}_1| \cdot |\widetilde{K}_2| \text{Re} [\widetilde{F}_{12}(\tau)] \quad (2.38)$$

De la definición dada en la ecuación (2.37) podemos también escribir

$$\widetilde{F}_{11}(\tau) = \langle \widetilde{A}_1(t + \tau) \widetilde{A}_1^*(\tau) \rangle_T \quad \text{y} \quad \widetilde{F}_{22}(\tau) = \langle \widetilde{A}_2(t + \tau) \widetilde{A}_2^*(\tau) \rangle_T \quad (2.39)$$

Obteniendo las denominadas funciones de auto-coherencia, donde para $\tau=0$ tenemos

$$\widetilde{F}_{11}(0) = I_{S1} \quad \text{y} \quad \widetilde{F}_{22}(0) = I_{S2} \quad (2.40)$$

Así pues, de la ecuación (2.38) resulta

$$I_1 = |\widetilde{K}_1|^2 \cdot \widetilde{F}_{11}(0) \quad , \quad I_2 = |\widetilde{K}_2|^2 \cdot \widetilde{F}_{22}(0) \quad (2.41)$$

De donde obtenemos que

$$|\widetilde{K}_1| \cdot |\widetilde{K}_2| = \frac{\sqrt{I_1 I_2}}{\sqrt{\widetilde{F}_{11}(0) \cdot \widetilde{F}_{22}(0)}} \quad (2.42)$$

En esta forma podemos escribir la expresión normalizada de la función de coherencia mutua, definida como

$$\widetilde{\gamma}_{12}(\tau) = \frac{\widetilde{I}_{12}(\tau)}{\sqrt{\widetilde{I}_{11}(0)\widetilde{I}_{22}(0)}} = \frac{\langle \widetilde{A}_1(t+\tau)\widetilde{A}_2^*(t) \rangle_T}{\sqrt{\langle |\widetilde{A}_1(t)|^2 \rangle_T \langle |\widetilde{A}_2(t)|^2 \rangle_T}} \quad (2.43)$$

Conocida como grado de coherencia complejo. Así entonces, la ecuación (2.38) adquiere la siguiente expresión

$$I = I_1 + I_2 + 2\sqrt{I_1 I_2} \text{Re}[\widetilde{\gamma}_{12}(\tau)] \quad (2.44)$$

Que es la ley general de interferencia para la luz parcialmente coherente.

De la ecuación (2.15) decíamos que

$$\delta = \vec{k}(\vec{r}_1 - \vec{r}_2) = \frac{2\pi}{\lambda}(\vec{r}_1 - \vec{r}_2) \quad (2.45)$$

Donde ahora $\bar{\lambda}$ es la longitud de onda promedio. Como $\widetilde{\gamma}_{12}(\tau)$ es una cantidad compleja, a la misma la podemos reescribir como

$$\widetilde{\gamma}_{12}(\tau) = |\widetilde{\gamma}_{12}(\tau)| \cdot \exp[j\Phi_{12}(\tau)] \quad (2.46)$$

El ángulo de fase es función de $\widetilde{I}_{12}(\tau)$ y de la fase δ entre los campos, por lo tanto definimos que

$$\Phi_{12}(\tau) = \alpha_{12}(\tau) - \delta \quad (2.47)$$

Resultando la irradiancia total de la ecuación (2.44) igual a

$$I = I_1 + I_2 + 2\sqrt{I_1 I_2} |\widetilde{\gamma}_{12}(\tau)| \cos[\alpha_{12}(\tau) - \delta] \quad (2.48)$$

observemos que esta es análoga a la ecuación (2.10).

Por último analicemos la ecuación (2.43), donde por medio de la desigualdad de Schwarz, se puede demostrar que $0 \leq |\widetilde{\gamma}_{12}(\tau)| \leq 1$. Para el caso en que $|\widetilde{\gamma}_{12}(\tau)| = 1$, tenemos la irradiancia generada por dos ondas coherentes desfasadas, como vimos en el punto 1-1. En el otro extremo donde $|\widetilde{\gamma}_{12}(\tau)| = 0$, no hay interferencia, resultando la irradiancia $I = I_1 + I_2$, dado que las dos ondas son incoherentes.

El $|\widetilde{\gamma}_{12}(\tau)|$ se denomina grado de coherencia, y dependiendo de su valor podemos definir que [5]

$$|\widetilde{\gamma}_{12}| = 1 \text{ es el límite coherente}$$

$$|\widetilde{\gamma}_{12}| = 0 \text{ es el límite incoherente}$$

$$0 < |\widetilde{\gamma}_{12}| < 1 \text{ es la coherencia parcial}$$

2.2- Difracción:

Si colocamos un cuerpo opaco entre una fuente luminosa puntual y una pantalla, este cuerpo proyecta una sombra compuesta de regiones claras y oscuras, muy diferente de la que podría esperarse utilizando los principios de la óptica geométrica. Este efecto denominado difracción es una característica general de los fenómenos ondulatorios, que ocurre siempre que una parte de un frente de onda se encuentra con algún tipo de obstrucción [2].

2.2.1- Teoría escalar de la perturbación luminosa:

Representemos a una onda de luz monocromática, situada en un punto P y en un instante de tiempo t, por la función escalar $u(P,t)$, definida por

$$u(P, t) = A(P) \cdot \cos[2\pi\nu t + \Phi(P)] \quad (2.49)$$

donde $A(P)$ es la amplitud y $\Phi(P)$ la fase de la onda en el punto P, siendo ν la frecuencia óptica. Podemos escribir la ecuación (2.49) en notación compleja, obteniendo

$$u(P, t) = \text{Re}[U(P) \cdot \exp(-j2\pi\nu t)] \quad (2.50)$$

aquí $U(P)$ es la función compleja espacial dada por

$$U(P) = A(P) \cdot \exp[-j\Phi(P)] \quad (2.51)$$

Dado que la perturbación real $u(P,t)$ representa una onda óptica, esta debe satisfacer en todo el espacio, libre de fuentes, la ecuación de onda escalar

$$\nabla^2[u(P, t)] - \frac{n^2}{c^2} \frac{\partial^2[u(P, t)]}{\partial t^2} = 0 \quad (2.52)$$

donde n es el índice de refracción del medio donde se propaga la onda.

Si reemplazamos la ecuación (2.50) en la (2.52) obtenemos la siguiente relación

$$\nabla^2 U(P) + k^2 U(P) = 0 \quad (2.53)$$

Esta expresión es llamada ecuación de Helmholtz, donde k es el número de onda.

$$k = 2\pi\nu \frac{n}{c} = \frac{2\pi}{\lambda} \quad (2.54)$$

A continuación podemos calcular la perturbación compleja $U(P)$ con ayuda del teorema de Green, enunciado como sigue:

Sean $U(P)$ y $G(P)$ dos funciones complejas, y sea S una superficie cerrada que delimita un volumen V . Si U , G y sus derivadas parciales primera y segunda están definidas y son continuas en el interior de S y sobre su superficie, entonces debe cumplirse que

$$\iiint (U \cdot \nabla^2 G - G \cdot \nabla^2 U) dV = \iint \left(U \frac{\partial G}{\partial \hat{n}} - G \frac{\partial U}{\partial \hat{n}} \right) dS \quad (2.55)$$

donde $\frac{\partial}{\partial \hat{n}}$ es la derivada parcial en dirección a la normal exterior de la superficie S .

Entonces, sea P_0 el punto de observación y sea S una superficie cerrada que lo rodea. Si escogemos una función G que represente una onda esférica de amplitud unitaria propagándose a partir de P_0 de manera que su valor en un punto P_1 venga dado por

$$G(P_1) = \frac{\exp(jk\vec{r}_{01})}{|\vec{r}_{01}|} \quad (2.56)$$

aquí \vec{r}_{01} es el vector desde P_0 hasta P_1 .

Para excluir la discontinuidad de G en el punto P_0 debemos adoptar alrededor de este punto una pequeña superficie esférica S' de radio ϵ , por lo que la superficie de integración será ahora $S'' = S + S'$, y el volumen $V'' = V - V'$, como observamos en la Figura 2.4. La perturbación G satisface la ecuación de Helmholtz en el interior de V'' , por lo tanto

$$(\nabla^2 + k^2)G = 0 \quad (2.57)$$

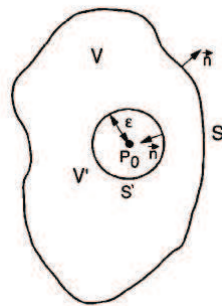


Figura 2.4. Superficie de integración (extraído de Goodman, 2008).

Sustituyendo las ecuaciones (2.53) y (2.57) en la parte izquierda de la ecuación (2.55), resulta que

$$\iiint (U \cdot \nabla^2 G - G \cdot \nabla^2 U) dV'' = \iiint (-U \cdot k^2 G + G \cdot k^2 U) dV'' = 0 \quad (2.58)$$

Entonces nos queda la parte derecha de la ecuación (2.55) igual a

$$\iint \left(U \frac{\partial G}{\partial \hat{n}} - G \frac{\partial U}{\partial \hat{n}} \right) dS'' = 0 \quad (2.59)$$

Separando ambas superficies obtenemos la siguiente igualdad

$$- \iint \left(U \frac{\partial G}{\partial \hat{n}} - G \frac{\partial U}{\partial \hat{n}} \right) dS' = \iint \left(U \frac{\partial G}{\partial \hat{n}} - G \frac{\partial U}{\partial \hat{n}} \right) dS \quad (2.60)$$

De la ecuación (2.56) obtenemos la derivada de G respecto a la normal, siendo igual a

$$\frac{\partial G(P_1)}{\partial \hat{n}} = \left(jk - \frac{1}{|\vec{r}_{01}|} \right) \frac{\exp(jk\vec{r}_{01})}{|\vec{r}_{01}|} \cos(\hat{n}, \vec{r}_{01}) \quad (2.61)$$

donde (\hat{n}, \vec{r}_{01}) es el ángulo formado por la normal exterior y el vector r_{01} .

Si el punto P_1 esta situado sobre la superficie S' , entonces $\cos(\hat{n}, \vec{r}_{01}) = -1$, y las ecuaciones (2.56) y (2.61) se transforman en

$$G(P_1) = \frac{\exp(jk\varepsilon)}{|\varepsilon|} \quad \text{y} \quad \frac{\partial G(P_1)}{\partial \hat{n}} = \left(\frac{1}{|\varepsilon|} - jk \right) \frac{\exp(jk\varepsilon)}{|\varepsilon|} \quad (2.62)$$

Eligiendo ε tan pequeño como se quiera, de la ecuación (2.60) podemos escribir

$$\lim_{\varepsilon \rightarrow 0} \iint \left(U \frac{\partial G}{\partial \hat{n}} - G \frac{\partial U}{\partial \hat{n}} \right) dS' = \lim_{\varepsilon \rightarrow 0} 4\pi\varepsilon^2 \left[U(P_0) \left(\frac{1}{|\varepsilon|} - jk \right) \frac{\exp(jk\varepsilon)}{|\varepsilon|} - \frac{\partial U(P_0)}{\partial \hat{n}} \frac{\exp(jk\varepsilon)}{|\varepsilon|} \right] = 4\pi \cdot U(P_0) \quad (2.63)$$

Sustituyendo esta ecuación (2.63) en la ecuación (2.60), finalmente obtenemos

$$U(P_0) = \frac{1}{4\pi} \iint \left[\frac{\partial U \exp(jk\vec{r}_{01})}{\partial \hat{n} |\vec{r}_{01}|} - U \frac{\partial}{\partial \hat{n}} \left(\frac{\exp(jk\vec{r}_{01})}{|\vec{r}_{01}|} \right) \right] dS \quad (2.64)$$

Esta ecuación se denomina teorema integral de Helmholtz – Kirchhoff, y permite expresar el campo en un punto cualquiera P_0 en función de los valores en los límites de la onda sobre una superficie cerrada que rodea a dicho punto [1].

2.2.2- Difracción por una pantalla plana con abertura:

Escojamos una superficie cerrada formada por dos partes, la primera es una superficie plana S_1 situada justo detrás de una pantalla difractante, y unida a esta superficie tenemos una segunda superficie S_2 que es un casquete esférico de radio R , centrada en el punto de observación P_0 , como vemos en la Figura 2.5.

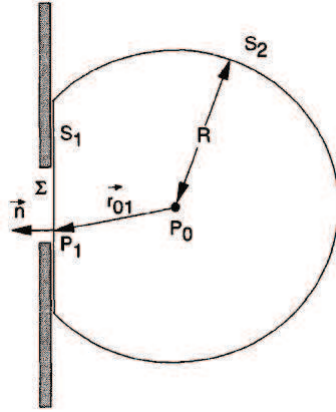


Figura 2.5. Difracción por una pantalla plana (extraído de Goodman, 2008).

Utilizando la ecuación (2.56), si la aplicamos a la superficie S_2 tenemos

$$G = \frac{\exp(jkR)}{|R|} \quad \text{y} \quad \frac{\partial G}{\partial \hat{n}} = \left(jk - \frac{1}{|R|} \right) \frac{\exp(jkR)}{|R|} \quad (2.65)$$

Si R es suficientemente grande entonces la derivada en la ecuación anterior resulta

$$\frac{\partial G}{\partial \hat{n}} \approx jk \frac{\exp(jkR)}{|R|} \approx jk \cdot G \quad (2.66)$$

Por lo que la integral sobre la superficie S_2 ahora se reduce a

$$\iint \left[G \frac{\partial U}{\partial \hat{n}} - U \cdot jkG \right] dS_2 = \int G \left(\frac{\partial U}{\partial \hat{n}} - jkU \right) R^2 d\omega_\Omega \quad (2.67)$$

donde Ω es el ángulo sólido bajo el cual observamos a la superficie S_2 desde el punto P_0 . Como el valor de $R \cdot G$ está uniformemente limitado sobre S_2 , la integral se anula cuando R tiende a infinito solo si la perturbación U cumple la siguiente propiedad

$$\lim_{R \rightarrow \infty} R \left(\frac{\partial U}{\partial \hat{n}} - jkU \right) = 0 \quad (2.68)$$

Esta expresión es conocida como la condición de radiación de Sommerfeld, y se cumple si U tiende a cero al menos tan rápido como G (siendo esta una onda esférica divergente), y dado que solo debemos tener ondas salientes de S_2 , la contribución de la integral de la ecuación (2.67) es nula. Entonces nos queda que la perturbación en P_0 la calculamos solo sobre el plano S_1 inmediatamente detrás de la pantalla, o sea

$$U(P_0) = \frac{1}{4\pi} \iint \left[\frac{\partial U}{\partial \hat{n}} G - U \frac{\partial G}{\partial \hat{n}} \right] dS_1 \quad (2.69)$$

La pantalla es opaca, excepto sobre una abertura de área Σ . A continuación podemos adoptar las denominadas condiciones de contorno de Kirchhoff para el plano S_1 , a saber:

- 1- En todo punto de la superficie Σ , el campo U y su derivada $\frac{\partial U}{\partial \hat{n}}$ son los mismos que en ausencia de la pantalla.
- 2- Sobre la parte del plano S_1 situada en la sombra geométrica de la pantalla, el campo U y su derivada $\frac{\partial U}{\partial \hat{n}}$ son igualmente nulos.

En consecuencia, dadas estas condiciones, podemos expresar la ecuación (2.69) de la siguiente forma

$$U(P_0) = \frac{1}{4\pi} \iint \left[\frac{\partial U}{\partial \hat{n}} G - U \frac{\partial G}{\partial \hat{n}} \right] d\Sigma \quad (2.70)$$

Para calcular la perturbación en P_0 , veamos la Figura 2.6, y consideremos que la distancia del punto de observación a la pantalla es mucho mayor que la longitud de onda, por lo que $k \gg 1/r_{01}$, entonces de la ecuación (2.61) podemos escribir

$$\frac{\partial G(P_1)}{\partial \hat{n}} = \left(jk - \frac{1}{|r_{01}|} \right) \frac{\exp(jk\vec{r}_{01})}{|r_{01}|} \cos(\hat{n}, \vec{r}_{01}) \approx jk \frac{\exp(jk\vec{r}_{01})}{|r_{01}|} \cos(\hat{n}, \vec{r}_{01}) \quad (2.71)$$

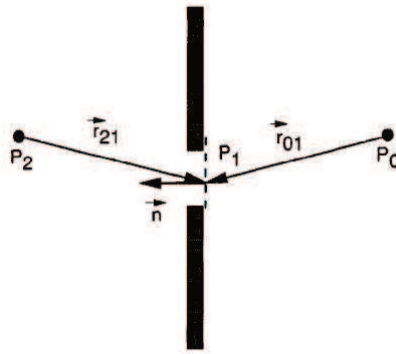


Figura 2.6. Pantalla plana con abertura iluminada (extraído de Goodman, 2008).

Supongamos ahora que iluminamos la abertura con una fuente puntual situada en el punto P_2 , que produce una onda esférica cuya expresión en el punto P_1 de la pantalla está dado por

$$U(P_1) = A \frac{\exp(jk\vec{r}_{21})}{|r_{21}|} \quad (2.72)$$

Considerando también que la distancia r_{21} es mucho mayor que la longitud de onda, tenemos que la derivada de la expresión (2.72) con respecto a la normal de la abertura es aproximadamente igual a

$$\frac{\partial U(P_1)}{\partial \hat{n}} \approx -jk \cdot A \frac{\exp(jk\vec{r}_{21})}{|\vec{r}_{21}|} \cos(\hat{n}, \vec{r}_{21}) \quad (2.73)$$

Así, si reemplazamos las ecuaciones (2.56), (2.71), (2.72) y (2.73) en la expresión (2.64), finalmente obtenemos

$$U(P_0) = \frac{A}{j\lambda} \iint \left[\frac{\exp[jk(\vec{r}_{21} + \vec{r}_{01})]}{|\vec{r}_{21}| |\vec{r}_{01}|} \left[\frac{\cos(\hat{n}, \vec{r}_{01}) - \cos(\hat{n}, \vec{r}_{21})}{2} \right] \right] d\Sigma \quad (2.74)$$

Esta expresión se aplica solo al caso de la iluminación producida por una fuente puntual, y es llamada fórmula de la difracción de Fresnel – Kirchhoff [1].

2.2.3- El principio de Huygens - Fresnel:

Si ubicamos a los puntos P_0 y P_2 en una posición simétrica con respecto a la pantalla, observamos que

$$|\vec{r}_{01}| = |\vec{r}_{21}| \quad \text{y} \quad \cos(\hat{n}, \vec{r}_{01}) = -\cos(\hat{n}, \vec{r}_{21}) = \cos\theta \quad (2.75)$$

De esta manera a la expresión (2.74) la podemos reescribir como

$$U(P_0) = \frac{1}{j\lambda} \iint U(P_1) \frac{\exp(jk\vec{r}_{01})}{|\vec{r}_{01}|} \cos\theta d\Sigma \quad (2.76)$$

Llamado el principio de difracción de Huygens – Fresnel. Esta ecuación expresa el campo $U(P_0)$ observado como una superposición de ondas esféricas divergentes, originadas a partir de fuentes secundarias localizadas en cada uno de los puntos P_1 de la abertura Σ . Cada fuente secundaria posee las siguientes propiedades:

- 1- Tiene una amplitud compleja proporcional a la amplitud de excitación $U(P_1)$
- 2- Tiene una amplitud inversamente proporcional a la longitud de onda λ
- 3- Tiene una fase que adelanta $\pi/2$ a la onda incidente, dado por el factor $1/j$
- 4- Cada una de las fuentes secundarias tiene un factor de oblicuidad $\cos\theta$.

Establezcamos la ecuación (2.76) en coordenadas rectangulares, con ayuda de la Figura 2.7. Suponemos que la abertura difractante está contenida en el plano (ξ, η) , y es iluminada en la dirección positiva del eje z .

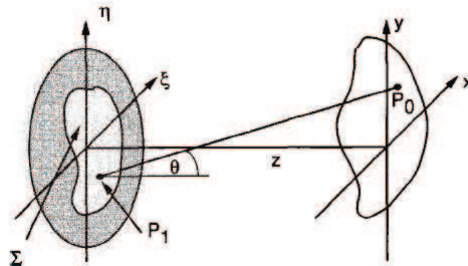


Figura 2.7. Configuración geométrica de la difracción (extraído de Goodman, 2008).

Necesitamos calcular el campo sobre el plano (x, y) paralelo al (ξ, η) , y separado de él una distancia z . El eje z atraviesa los orígenes de ambos planos.

En la Figura 2.7 vemos que θ , siendo el ángulo entre la normal hacia afuera (z negativo) y al vector r_{01} (apuntando desde P_0 a P_1), resulta

$$\cos \theta = \frac{z}{|\vec{r}_{01}|} \quad (2.77)$$

Donde el módulo del vector viene dado por

$$|\vec{r}_{01}| = \sqrt{(x - \xi)^2 + (y - \eta)^2 + z^2} \quad (2.78)$$

Entonces podemos reescribir la ecuación (2.76) como sigue

$$U(x, y) = \frac{z}{j\lambda} \iint U(\xi, \eta) \frac{\exp(jk\vec{r}_{01})}{|\vec{r}_{01}|^2} d\xi d\eta \quad (2.79)$$

2.2.4- La aproximación de Fresnel:

Comencemos desarrollando la expresión (2.78) mediante la serie de Taylor para la raíz cuadrada siguiente

$$\sqrt{1+b} = 1 + \frac{1}{2}b - \frac{1}{8}b^2 + \dots \quad (2.80)$$

siendo b un número menor que la unidad, y donde la exactitud de esta serie depende de la magnitud de b . A continuación apliquemos la ecuación (2.80) en la (2.78), quedándonos solo con los dos primeros términos de la serie

$$|\vec{r}_{01}| = z \sqrt{1 + \left(\frac{x-\xi}{z}\right)^2 + \left(\frac{y-\eta}{z}\right)^2} \approx z \left[1 + \frac{1}{2} \left(\frac{x-\xi}{z}\right)^2 + \frac{1}{2} \left(\frac{y-\eta}{z}\right)^2\right] \quad (2.81)$$

Observando la ecuación (2.79) vemos que en el caso del $|r_{01}|^2$ del denominador, el error introducido por considerar solo el termino z en la ecuación (2.81) y despreciar los otros dos, es generalmente pequeño. Por otro lado, para el r_{01} del exponente el error por despreciar los dos últimos términos es mas grande, dado que este r_{01} está multiplicando a número muy grande k (mayor a $1 \times 10^7 \text{ m}^{-1}$ en el visible), y también por que cambios de fase muy pequeños modifican significativamente el valor del exponencial. Aclarado esto, la expresión para el campo resultante en (x,y) es

$$U(x, y) = \frac{\exp(jkz)}{j\lambda z} \iint U(\xi, \eta) \cdot \exp \left[j \frac{k}{2z} [(x - \xi)^2 + (y - \eta)^2] \right] d\xi d\eta \quad (2.82)$$

Si desarrollamos los términos dentro de la exponencial en la integral, y tomamos límites de integración infinitos, considerando que la función $U(\xi, \eta)$ solo tiene valores no nulos dentro de la superficie de abertura Σ , podemos obtener la siguiente expresión

$$U(x, y) = \frac{\exp(jkz)}{j\lambda z} \exp \left[j \frac{k}{2z} (x^2 + y^2) \right] \iint_{-\infty}^{\infty} \left[U(\xi, \eta) \cdot \exp \left[j \frac{k}{2z} (\xi^2 + \eta^2) \right] \right] \exp \left[-j \frac{k}{z} (x\xi + y\eta) \right] d\xi d\eta \quad (2.83)$$

Esta expresión es conocida como la integral de difracción de Fresnel.

Si tomamos en el plano de la abertura (ξ, η) la expresión entre corchetes dentro de la integral de la ecuación (2.83), y la llamamos función abertura, de modo que

$$\mathcal{A}(\xi, \eta) = U(\xi, \eta) \cdot \exp \left[j \frac{k}{2z} (\xi^2 + \eta^2) \right] \quad (2.84)$$

y consideramos los términos fuera de la integral como un factor complejo multiplicativo de fase, obtenemos

$$U(x, y) = \tilde{C} \iint_{-\infty}^{\infty} \mathcal{A}(\xi, \eta) \cdot \exp \left[-j \frac{2\pi}{\lambda z} [x\xi + y\eta] \right] d\xi d\eta = \tilde{C} \cdot \mathcal{F}[\mathcal{A}(\xi, \eta)] = \tilde{C} \cdot A(f_x, f_y) \quad (2.85)$$

donde tenemos $f_x = x/(\lambda z)$ y $f_y = y/(\lambda z)$ igual a las frecuencias espaciales.

De modo que el campo de la distorsión en el plano (x,y) es la transformada de Fourier de la función abertura, siendo esta ultima el producto del campo justo a la derecha del plano (ξ, η) por una fase exponencial cuadrática.

Si esta aproximación es válida, decimos que estamos en la región de difracción de Fresnel, o en la región de campo cercano a la abertura [6].

2.2.5- Integral de convolución de Huygens - Fresnel:

Si tomamos la ecuación (2.82) y la reescribimos de la siguiente manera

$$U(x, y) = \iint U(\xi, \eta) \cdot h[(x - \xi), (y - \eta)] d\xi d\eta \quad (2.86)$$

Donde vemos claramente que queda expresada una convolución entre $U(\xi, \eta)$ y $h(x, y)$, siendo esta última función igual a

$$h(x, y) = \frac{\exp(jkz)}{j\lambda z} \exp\left[j\frac{k}{2z}(x^2 + y^2)\right] \quad (2.87)$$

denominada respuesta al impulso de la difracción de Fresnel.

Ahora, por las propiedades de la transformada de Fourier, podemos obtener fácilmente la transformada del campo en el plano (x, y) a partir de la transformada del campo en (ξ, η) , a saber

$$\mathcal{F}[U(x, y)] = \mathcal{F}[U(\xi, \eta)] \cdot \mathcal{F}[h(x, y)] \quad (2.88)$$

Luego haciendo la transformada inversa del resultado dado por la ecuación (2.88), obtenemos la expresión del campo en el plano (x, y) [6].

2.2.6- Aproximación de Fraunhofer:

Si además de las condiciones de la aproximación de Fresnel, se satisface la siguiente condición extra

$$z \gg \frac{k}{2} (\xi^2 + \eta^2)_{max} \quad (2.89)$$

Entonces el factor de fase cuadrática del plano (ξ, η) en la ecuación (2.83) es igual a la unidad sobre toda superficie de la abertura, y el campo en el plano (x, y) puede hallarse por

$$U(x, y) = \frac{\exp(jkz)}{j\lambda z} \exp\left[j\frac{k}{2z}(x^2 + y^2)\right] \iint_{-\infty}^{\infty} U(\xi, \eta) \exp\left[-j\frac{2\pi}{\lambda z}(x\xi + y\eta)\right] d\xi d\eta = \bar{C} \cdot \mathcal{F}[U(\xi, \eta)] \quad (2.90)$$

O sea, directamente igual a la transformada de Fourier de la distribución en el plano de la abertura multiplicada por el factor complejo de fase en (x, y) .

La ecuación (2.90) es válida en la denominada región de difracción de Fraunhofer, o de campo lejano [6].

2.2.7- Propagación del espectro angular:

Si realizamos el análisis de Fourier de la distribución del campo complejo en un plano cualquiera, las diferentes componentes espaciales de Fourier pueden identificarse como ondas planas viajando en diferentes direcciones. El campo en cualquier otro plano paralelo puede calcularse sumando las contribuciones de estas ondas planas, considerando los desplazamientos de fase producidos durante la propagación.

Para comenzar el análisis, supongamos una onda que se propaga en la dirección positiva del eje z e incide en el plano (x,y). El campo complejo en el plano z = 0 (que coincide con el plano (ξ,η) de los puntos anteriores), puede representarse por U(x,y,0). Esta función posee una transformada de Fourier dada por

$$A(f_x, f_y, 0) = \iint_{-\infty}^{\infty} U(x, y, 0) \exp[-j2\pi(f_x x + f_y y)] dx dy \quad (2.91)$$

Siendo A(f_x, f_y, 0) su espectro de frecuencias espaciales.

A su vez, podemos admitir a la función U(x,y,0) como la transformada inversa de su espectro

$$U(x, y, 0) = \iint_{-\infty}^{\infty} A(f_x, f_y, 0) \exp[j2\pi(f_x x + f_y y)] df_x df_y \quad (2.92)$$

Si consideramos la forma de una onda plana simple propagándose según el vector de onda k, con sus cosenos directores (α,β,γ) como se indica en la Figura 2.8

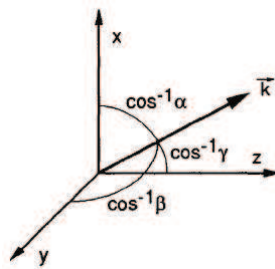


Figura 2.8. El vector de onda k (extraído de Goodman, 2008).

a esta onda la podemos expresar de la siguiente forma

$$P(x, y, z) = \exp(j\vec{k} \cdot \vec{r}) = \exp\left[j\frac{2\pi}{\lambda}(\alpha x + \beta y)\right] \cdot \exp\left(j\frac{2\pi}{\lambda}\gamma z\right) \quad (2.93)$$

donde

$$\vec{r} = x\hat{x} + y\hat{y} + z\hat{z} \quad y \quad \vec{k} = \frac{2\pi}{\lambda}(\alpha\hat{x} + \beta\hat{y} + \gamma\hat{z}) \quad (2.94)$$

Tenemos en cuenta que los cosenos directores están relacionados entre si a través de la siguiente expresión

$$\gamma = \sqrt{1 - \alpha^2 - \beta^2} \quad (2.95)$$

Por lo tanto, en la ecuación (2.92), para el plano $z = 0$ la función exponencial puede considerarse como una onda plana que se propaga en la dirección de los cosenos directores

$$\alpha = \lambda f_x, \quad \beta = \lambda f_y, \quad \gamma = \sqrt{1 - (\lambda f_x)^2 - (\lambda f_y)^2} \quad (2.96)$$

Así, en la ecuación (2.92) podemos considerar al término $A(f_x, f_y, 0)$ como la amplitud compleja de la onda plana, evaluada en $f_x = \alpha/\lambda$ y $f_y = \beta/\lambda$, por ello a la expresión

$$A\left(\frac{\alpha}{\lambda}, \frac{\beta}{\lambda}, 0\right) = \iint_{-\infty}^{\infty} U(x, y, 0) \exp\left[-j2\pi\left(\frac{\alpha}{\lambda}x + \frac{\beta}{\lambda}y\right)\right] dx dy \quad (2.97)$$

se la conoce como espectro angular de la perturbación $U(x, y, 0)$.

Consideremos ahora a la perturbación U en un plano paralelo al (x, y) , pero a una distancia z del mismo, de la ecuación (2.92) tenemos que

$$U(x, y, z) = \iint_{-\infty}^{\infty} A\left(\frac{\alpha}{\lambda}, \frac{\beta}{\lambda}, z\right) \exp\left[j2\pi\left(\frac{\alpha}{\lambda}x + \frac{\beta}{\lambda}y\right)\right] d\frac{\alpha}{\lambda} d\frac{\beta}{\lambda} \quad (2.98)$$

Por otro lado, U debe satisfacer la ecuación de Helmholtz (dada por (2.53)), en todo punto donde no haya fuentes.

Aplicando directamente la ecuación (2.53) a la (2.98), se tiene que la función A debe satisfacer la siguiente ecuación diferencial, con el vector k propagándose en la dirección de z

$$\frac{d^2 A\left(\frac{\alpha}{\lambda}, \frac{\beta}{\lambda}, z\right)}{dz^2} + \left(\frac{2\pi}{\lambda}\right)^2 (1 - \alpha^2 - \beta^2) A\left(\frac{\alpha}{\lambda}, \frac{\beta}{\lambda}, z\right) = 0 \quad (2.99)$$

Cuya solución puede escribirse como

$$A\left(\frac{\alpha}{\lambda}, \frac{\beta}{\lambda}, z\right) = A\left(\frac{\alpha}{\lambda}, \frac{\beta}{\lambda}, 0\right) \exp\left(j\frac{2\pi}{\lambda}z\sqrt{1 - \alpha^2 - \beta^2}\right) \quad (2.100)$$

Cuando los cosenos directores α y β satisfacen la relación

$$\alpha^2 + \beta^2 < 1 \quad (2.101)$$

el efecto de la propagación es un cambio de fases relativas de las distintas componentes del espectro angular. Dado que cada componente de la onda plana se propaga en una dirección

diferente, cada una de ellas recorre una distancia diferente entre dos planos paralelos, introduciéndose los desfases relativos.

Sin embargo, cuando se tiene la siguiente relación

$$\alpha^2 + \beta^2 > 1 \quad (2.102)$$

aquí α y β no pueden seguir considerándose cosenos directores, en este caso la raíz cuadrada de la ecuación (2.100) es imaginaria, y se puede escribir como

$$j\mu = j \frac{2\pi}{\lambda} \sqrt{\alpha^2 + \beta^2 - 1} \quad (2.103)$$

Ahora reescribimos la ecuación (2.100) de la siguiente forma

$$A\left(\frac{\alpha}{\lambda}, \frac{\beta}{\lambda}, z\right) = A\left(\frac{\alpha}{\lambda}, \frac{\beta}{\lambda}, 0\right) \exp(-\mu z) \quad (2.104)$$

asi vemos que estas componentes de las ondas son atenuadas por la propagación, y son denominadas ondas evanescentes.

Podemos decir que la propagación está limitada a la región dentro de la cual se cumple la relación (2.101), entonces expresamos la perturbación observada en (x,y,z) de la ecuación (2.98) como sigue

$$U(x, y, z) = \iint_{-\infty}^{\infty} A\left(\frac{\alpha}{\lambda}, \frac{\beta}{\lambda}, 0\right) \cdot \exp\left(j \frac{2\pi}{\lambda} z \sqrt{1 - \alpha^2 - \beta^2}\right) \cdot \text{circ}\left(\sqrt{\alpha^2 + \beta^2}\right) \cdot \exp\left[j 2\pi\left(\frac{\alpha}{\lambda} x + \frac{\beta}{\lambda} y\right)\right] d\frac{\alpha}{\lambda} d\frac{\beta}{\lambda} \quad (2.105)$$

De esta expresión observamos que ninguna componente del espectro angular, mas allá de la frecuencia de corte para las ondas evanescentes, contribuye al campo U(x,y,z), esta es la razón principal por la que ningún sistema convencional formador de imagen resuelve una estructura periódica que posea un período de longitud menor a la longitud de onda de la luz utilizada.

A continuación reescribamos la solución dada por la ecuación (2.100), como sigue

$$A(f_x, f_y, z) = A(f_x, f_y, 0) \cdot \text{circ}\left[\sqrt{(\lambda f_x)^2 + (\lambda f_y)^2}\right] \exp\left(j \frac{2\pi}{\lambda} z \sqrt{1 - (\lambda f_x)^2 - (\lambda f_y)^2}\right) \quad (2.106a)$$

$$A(f_x, f_y, z) = A(f_x, f_y, 0) \cdot H(f_x, f_y) \quad (2.106b)$$

Donde H es una función de transferencia igual a la siguiente expresión

$$H(f_x, f_y) = \begin{cases} \exp\left(j \frac{2\pi}{\lambda} z \sqrt{1 - (\lambda f_x)^2 - (\lambda f_y)^2}\right), & \text{para } \sqrt{(f_x)^2 + (f_y)^2} < \frac{1}{\lambda} \\ 0, & \text{para el resto} \end{cases} \quad (2.107)$$

De esta forma la propagación del espectro puede considerarse como un filtro espacial dispersivo, con una transmitancia nula en el exterior de la región circular de radio $1/\lambda$, en el plano de frecuencias espaciales. En el interior de dicha región, el modulo de la función de transferencia es la unidad, pero introduce variaciones de fase dependientes de la frecuencia. La dispersión de fase es notable para frecuencias espaciales altas, y se anula si f_x y f_y tienden a cero. Además, para un par de frecuencias fijas, la dispersión de fase aumenta con la distancia de propagación z .

Ahora comparemos la función transferencia de la propagación del espectro angular, ecuación (2.107), con la dada por la aproximación de Fresnel, ecuación (2.87). Para ello obtengamos la transformada de Fourier de esta última

$$H(f_x, f_y) = \mathcal{F}[h(x, y)] = \mathcal{F}\left[\frac{\exp(jkz)}{j\lambda z} \exp\left[j\frac{\pi}{\lambda z}(x^2 + y^2)\right]\right] = \exp(jkz) \cdot \exp[-j\pi\lambda z(f_x^2 + f_y^2)] \quad (2.108)$$

Así en la aproximación de Fresnel, la dispersión de la fase espacial se reduce a una dispersión de fase cuadrática. La expresión (2.108) es una aproximación de la (2.107), dada por

$$z\sqrt{1 - (\lambda f_x)^2 - (\lambda f_y)^2} \approx z\left[1 - \frac{(\lambda f_x)^2}{2} - \frac{(\lambda f_y)^2}{2}\right] \quad (2.109)$$

Siempre y cuando $|\lambda f_x| \ll 1$ y $|\lambda f_y| \ll 1$, siendo estas simplemente restricciones para ángulos de difracción pequeños, como sucede en una aproximación paraxial [1].

2.2.8- Difracción de una lente delgada:

Decimos que una lente es delgada si un rayo de luz que entra por un punto de coordenadas (x, y) de una cara de la misma, sale por un punto de la cara opuesta que tiene aproximadamente las mismas coordenadas, o sea que es despreciable la traslación del rayo dentro de la lente. Por lo tanto, la lente delgada solo retrasa la fase de un frente de onda incidente, en una cantidad proporcional al espesor en cada punto de la misma.

En la Figura 2.9 tenemos que Δ_0 es el máximo espesor de la lente sobre su eje y $\Delta(x, y)$ el espesor en un punto (x, y) .

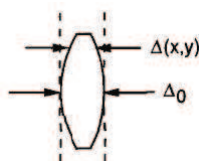


Figura 2.9. La función espesor (extraído de Goodman, 2008).

El desfase total sufrido por la onda al atravesar la lente en el punto (x,y) es igual a

$$\Phi(x,y) = k \cdot n \cdot \Delta(x,y) + k[\Delta_0 - \Delta(x,y)] \quad (2.110)$$

donde n es el índice de refracción del material de la lente, $k \cdot n \cdot \Delta(x,y)$ el desfase provocado por la lente y $k[\Delta_0 - \Delta(x,y)]$ el provocado por el espacio de aire que queda entre los dos planos (ver Figura 2.9). Análogamente, el efecto provocado por la lente se puede representar por una transformación de fase multiplicativa, dada por

$$t_l(x,y) = \exp(jk\Delta_0) \cdot \exp[jk(n-1)\Delta(x,y)] \quad (2.111)$$

El campo complejo sobre un plano inmediatamente detrás de la lente está relacionado con el campo complejo incidente sobre un plano inmediatamente delante de esta, por la siguiente expresión

$$U'_l(x,y) = U_l(x,y) \cdot t_l(x,y) \quad (2.112)$$

Para calcular la función espesor $\Delta(x,y)$ dividimos la lente en tres partes, como se muestra en la Figura 2.10, considerando los radios de curvatura positivos para superficies convexas y negativos para superficies cóncavas, observándolas de izquierda a derecha.

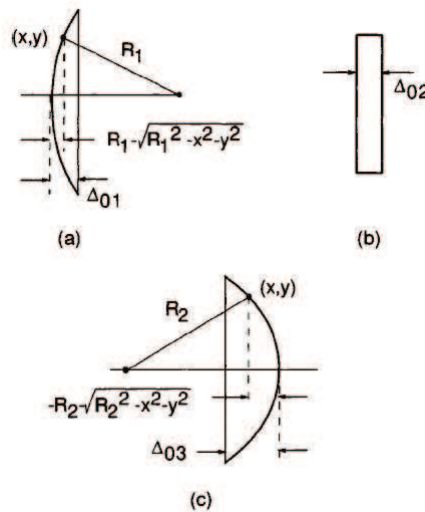


Figura 2.10. Cálculo de la función espesor (extraído de Goodman, 2008).

La función espesor total es la suma de las tres funciones individuales

$$\Delta(x,y) = \Delta_a(x,y) + \Delta_b(x,y) + \Delta_c(x,y) \quad (2.113)$$

De acuerdo con la Figura 2.10, las funciones espesor individuales viene dadas por

$$\Delta_a(x,y) = \Delta_{01} - \left(R_1 - \sqrt{R_1^2 - x^2 - y^2} \right) = \Delta_{01} - R_1 \left(1 - \sqrt{1 - \frac{(x^2+y^2)}{R_1^2}} \right) \quad (2.114a)$$

$$\Delta_b(x, y) = \Delta_{02} \quad (2.114b)$$

$$\Delta_c(x, y) = \Delta_{03} - (-R_2 - \sqrt{R_2^2 - x^2 - y^2}) = \Delta_{03} + R_2 \left(1 - \sqrt{1 - \frac{(x^2+y^2)}{R_2^2}}\right) \quad (2.114c)$$

Sabiendo que $R_2 < 0$. Sumando las tres expresiones obtenemos el espesor total

$$\Delta(x, y) = \Delta_0 - R_1 \left(1 - \sqrt{1 - \frac{(x^2+y^2)}{R_1^2}}\right) + R_2 \left(1 - \sqrt{1 - \frac{(x^2+y^2)}{R_2^2}}\right) \quad (2.115)$$

donde $\Delta_0 = \Delta_{01} + \Delta_{02} + \Delta_{03}$.

Si nos limitamos a los rayos paraxiales, los valores de (x, y) son lo suficientemente pequeños para poder usar la aproximación dada por la ecuación (2.80), entonces

$$\sqrt{1 - \frac{(x^2+y^2)}{R_1^2}} \approx 1 - \frac{(x^2+y^2)}{2R_1^2}, \quad \sqrt{1 - \frac{(x^2+y^2)}{R_2^2}} \approx 1 - \frac{(x^2+y^2)}{2R_2^2} \quad (2.116)$$

Simplificando la ecuación (2.115), de manera que la función espesor resulta igual a

$$\Delta(x, y) = \Delta_0 - \frac{(x^2+y^2)}{2} \left(\frac{1}{R_1} - \frac{1}{R_2}\right) \quad (2.117)$$

Reemplazando esta expresión en la ecuación (2.111), tenemos que la transformación provocada por la lente se define por

$$t_l(x, y) = \exp(jkn\Delta_0) \cdot \exp\left[-jk(n-1)\frac{(x^2+y^2)}{2}\left(\frac{1}{R_1} - \frac{1}{R_2}\right)\right] \quad (2.118)$$

De la definición de distancia focal de la lente, sabemos que

$$\frac{1}{f} = (n-1)\left(\frac{1}{R_1} - \frac{1}{R_2}\right) \quad (2.119)$$

Si omitimos el factor de fase constante en la ecuación (2.118), podemos decir que la transformación viene dada por

$$t_l(x, y) = \exp\left[-j\frac{k}{2f}(x^2 + y^2)\right] \quad (2.120)$$

El significado físico de la transformación de fase provocada por la lente la podemos interpretar si consideramos que sobre esta incide una onda plana de amplitud unitaria, por la ecuación (2.112) vemos que a la salida de la lente tenemos una onda esférica. Si la distancia focal es positiva (en el plano imagen), la onda esférica converge hacia el punto focal imagen. Si la distancia focal es negativa (en el plano objeto), la onda esférica diverge desde el punto focal objeto.

Supongamos que colocamos un objeto plano, con transmitancia $t_A(x,y)$, inmediatamente delante de una lente convergente, de distancia focal f . El objeto es uniformemente iluminado por una onda plana monocromática de amplitud A , que incide en forma normal a este, entonces la perturbación incidente en la lente es

$$U_l(x, y) = A \cdot t_A(x, y) \quad (2.121)$$

y la expresión de la distribución detrás de la lente (ecuación (2.112)), si utilizamos la ecuación (2.120), resulta

$$U'_l(x, y) = U_l(x, y) \cdot t_l(x, y) = U_l(x, y) \cdot \exp\left[-j\frac{k}{2f}(x^2 + y^2)\right] \quad (2.122)$$

Para determinar la distribución $U_P(x,y)$ en el plano focal imagen de la lente, aplicamos la fórmula de difracción de Fresnel (ecuación (2.83)), haciendo $z = f$, obtenemos

$$U_f(u, v) = \frac{\exp(jkf)}{j\lambda f} \exp\left[j\frac{k}{2f}(u^2 + v^2)\right] \iint_{-\infty}^{\infty} \left[U'_l(x, y) \cdot \exp\left[j\frac{k}{2f}(x^2 + y^2)\right] \right] \exp\left[-j\frac{k}{f}(ux + vy)\right] dx dy \quad (2.123)$$

Si omitimos el factor de fase constante, y reemplazamos la ecuación (2.122) en la (2.123), nos queda

$$U_f(u, v) = \frac{\exp\left[j\frac{k}{2f}(u^2 + v^2)\right]}{j\lambda f} \cdot \iint_{-\infty}^{\infty} \left[U_l(x, y) \cdot \exp\left[-j2\pi\left(\frac{u}{\lambda f}x + \frac{v}{\lambda f}y\right)\right] \right] dx dy = \tilde{C} \cdot \mathcal{F}[U_l(x, y)] \quad (2.124)$$

Con este resultado, vemos que la distribución del campo en el plano focal de la lente está dada por la difracción de Fraunhofer (ecuación (2.90)) del campo que incide sobre la lente, aunque la distancia al plano de observación sea la distancia focal, en lugar de ser la distancia de un plano lejano.

Ahora, si consideramos al objeto colocado a una distancia d delante de la lente, y este es iluminado por una onda plana de amplitud A incidiendo normalmente sobre el, y la transmitancia en amplitud de dicho objeto es t_A , tenemos que

$$U_0(x, y) = A \cdot t_A(x, y) \quad (2.125)$$

Si utilizamos el teorema de la convolución de Huygens – Fresnel para el campo del objeto que se propaga la distancia d , obtenemos de la ecuación (2.88) y la (2.108) lo siguiente

$$\mathcal{F}[U_l(x, y)] = \mathcal{F}[U_0(x, y)] \cdot \mathcal{F}[h(x, y)] = F_0(f_x, f_y) \cdot H(f_x, f_y) \quad (2.126a)$$

$$\mathcal{F}[U_l(x, y)] = F_0(f_x, f_y) \cdot \exp(jkd) \cdot \exp\left[-j\pi\lambda d(f_x^2 + f_y^2)\right] \quad (2.126b)$$

Que es la transformada de Fourier del campo inmediatamente antes de la lente, donde $f_x = x/(\lambda d)$ y $f_y = y/(\lambda d)$.

Seguidamente, obtenemos el campo a la distancia focal imagen con la ecuación (2.124), resultando

$$U_f(u, v) = \frac{\exp\left[j\frac{k}{2f}(u^2 + v^2)\right]}{j\lambda f} \mathcal{F}[U_i(x, y)] = \frac{\exp\left[j\frac{\pi}{\lambda f}(u^2 + v^2)\left(1 - \frac{d}{f}\right)\right]}{j\lambda f} F_0\left(\frac{u}{\lambda f}, \frac{v}{\lambda f}\right) \quad (2.127)$$

donde hemos omitido el factor de fase constante de la ecuación (2.126b).

Si desarrollamos la ecuación (2.127), vemos que el campo a la distancia focal imagen, en el plano (u,v), de un objeto situado a una distancia d de la superficie izquierda de la lente, en un plano (ξ,η), considerando la ecuación (2.125), resulta igual a

$$U_f(u, v) = \frac{A}{j\lambda f} \exp\left[j\frac{\pi}{\lambda f}(u^2 + v^2)\left(1 - \frac{d}{f}\right)\right] \iint_{-\infty}^{\infty} t_A(\xi, \eta) \cdot \exp\left[-j\frac{2\pi}{\lambda f}(\xi u + \eta v)\right] d\xi d\eta \quad (2.128)$$

Si observamos el factor de fase cuadrático que precede a la transformada de Fourier, notamos que este desaparece para el caso particular donde d = f. Entonces, cuando el objeto se sitúa en el plano focal objeto de la lente, la curvatura de fase desaparece, obteniéndose una transformada de Fourier exacta en el plano focal imagen [1].

2.2.9- Difracción por una abertura cuadrada:

Dado que los instrumentos detectores de la intensidad luminosa poseen una superficie finita, y generalmente están cubiertos por una ventana de protección, tenemos que un fenómeno de difracción ocurre en esta abertura antes de llegar al fotosensor propiamente dicho.

Pasemos a analizar este fenómeno, considerando una abertura cuadrada de lado 2w, iluminada bajo incidencia normal por una onda monocromática de amplitud unitaria. La distribución del campo inmediatamente después de la abertura es

$$U(\xi, \eta) = \text{rect}\left(\frac{\xi}{2w}\right) \cdot \text{rect}\left(\frac{\eta}{2w}\right) \quad (2.129)$$

Elijamos la integral de convolución de Huygens – Fresnel (ecuación (2.86)) para calcular el campo en el plano (x,y) del fotosensor

$$U(x, y) = \frac{\exp(jkz)}{j\lambda z} \iint_{-\infty}^{\infty} U(\xi, \eta) \cdot \exp\left[j\frac{\pi}{\lambda z}[(x - \xi)^2 + (y - \eta)^2]\right] d\xi d\eta \quad (2.130a)$$

$$U(x, y) = \frac{\exp(jkz)}{j\lambda z} \iint_{-w}^w \exp\left[j\frac{\pi}{\lambda z}[(x - \xi)^2 + (y - \eta)^2]\right] d\xi d\eta \quad (2.130b)$$

Podemos separar la ecuación (2.130b) en el producto de dos integrales unidimensionales, a saber

$$U(x, y) = \frac{\exp(jkz)}{j} G(x) \cdot G(y) \quad (2.131)$$

Donde cada función G es igual a

$$G(x) = \frac{1}{\sqrt{\lambda z}} \int_{-w}^w \exp \left[j \frac{\pi}{\lambda z} (\xi - x)^2 \right] d\xi \quad y \quad G(y) = \frac{1}{\sqrt{\lambda z}} \int_{-w}^w \exp \left[j \frac{\pi}{\lambda z} (\eta - y)^2 \right] d\eta \quad (2.132)$$

Si realizamos el siguiente cambio de variables

$$\alpha = \sqrt{\frac{2}{\lambda z}} (\xi - x) \quad y \quad \beta = \sqrt{\frac{2}{\lambda z}} (\eta - y) \quad (2.133)$$

Tenemos ahora que los límites de integración son iguales a

$$\alpha_1 = -\sqrt{\frac{2}{\lambda z}} (w + x) \quad , \quad \alpha_2 = \sqrt{\frac{2}{\lambda z}} (w - x) \quad , \quad \beta_1 = -\sqrt{\frac{2}{\lambda z}} (w + y) \quad y \quad \beta_2 = \sqrt{\frac{2}{\lambda z}} (w - y) \quad (2.134)$$

Para continuar definimos el número de Fresnel como $N_F = w^2/(\lambda z)$, y también las variables de distancia normalizada en la región de observación, iguales a $X = x/\sqrt{\lambda z}$ y $Y = y/\sqrt{\lambda z}$, transformando los límites de integración de la siguiente manera

$$\alpha_1 = -\sqrt{2}(\sqrt{N_F} + X) \quad , \quad \alpha_2 = \sqrt{2}(\sqrt{N_F} - X) \quad , \quad \beta_1 = -\sqrt{2}(\sqrt{N_F} + Y) \quad y \quad \beta_2 = \sqrt{2}(\sqrt{N_F} - Y) \quad (2.135)$$

Ahora podemos escribir las funciones de la expresión (2.132) de la siguiente forma

$$G(x) = \frac{1}{\sqrt{2}} \int_{\alpha_1}^{\alpha_2} \exp \left(j \frac{\pi}{2} \alpha^2 \right) d\alpha \quad y \quad G(y) = \frac{1}{\sqrt{2}} \int_{\beta_1}^{\beta_2} \exp \left(j \frac{\pi}{2} \beta^2 \right) d\beta \quad (2.136)$$

Estas integrales están relacionadas con las integrales de Fresnel del coseno y del seno, si tenemos en cuenta que

$$\int_{\alpha_1}^{\alpha_2} \exp \left(j \frac{\pi}{2} \alpha^2 \right) d\alpha = \int_0^{\alpha_2} \exp \left(j \frac{\pi}{2} \alpha^2 \right) d\alpha - \int_0^{\alpha_1} \exp \left(j \frac{\pi}{2} \alpha^2 \right) d\alpha = [C(\alpha_2) + j \cdot S(\alpha_2)] - [C(\alpha_1) + j \cdot S(\alpha_1)] \quad (2.137)$$

Podemos reescribir la expresión (2.136) de la forma

$$G(x) = \frac{1}{\sqrt{2}} [C(\alpha_2) - C(\alpha_1) + j[S(\alpha_2) - S(\alpha_1)]] \quad y \quad G(y) = \frac{1}{\sqrt{2}} [C(\beta_2) - C(\beta_1) + j[S(\beta_2) - S(\beta_1)]] \quad (2.138)$$

En la Figura 2.10 tenemos la gráfica de la función $|G(x)|$, denominada función *Chirp*, como puede observarse es casi plana sobre toda la región $(-w,w)$ y prácticamente nula fuera de ella.

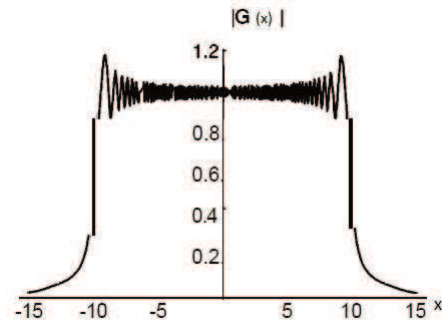


Figura 2.11. Función *Chirp* para $w = 10$ mm y $N_F = 1000$ (extraído de Goodman, 2008).

Finalmente, si sustituimos las ecuaciones dadas en (2.138), en la expresión (2.131), obtenemos el campo en el plano (x,y)

$$U(x, y) = \frac{\exp(jkz)}{2j} [c(\alpha_2) - c(\alpha_1) + j[s(\alpha_2) - s(\alpha_1)]]. [c(\beta_2) - c(\beta_1) + j[s(\beta_2) - s(\beta_1)]] \quad (2.139)$$

Sabemos que la magnitud física medible es la intensidad $I(x,y) = |U(x,y)|^2$, resultando igual a [6]

$$I(x, y) = \frac{1}{4} \left\{ [(c(\alpha_2) - c(\alpha_1))^2 + (s(\alpha_2) - s(\alpha_1))^2] \cdot [(c(\beta_2) - c(\beta_1))^2 + (s(\beta_2) - s(\beta_1))^2] \right\} \quad (2.140)$$

Ahora, consideremos en el plano objeto (ξ, η) solo el borde inferior de la abertura paralelo al eje ξ , y traslademos las coordenadas tal que $-w = 0$ en este nuevo esquema. Además si hacemos a w lo suficientemente grande, podemos alejarnos del borde opuesto en $\xi = 2w$, y concentrarnos solo en la región del borde $\xi = 0$.

Tomando estas consideraciones, a continuación, podemos suponer una función de transmitancia en amplitud para el objeto, dada por

$$t_A(\xi, \eta) = \begin{cases} 0, & \text{para } \xi < 0 \\ 1, & \text{para } \xi \geq 0 \end{cases} \quad (2.141)$$

Esta función de transmitancia es una función escalón, y la respuesta teórica del sistema a la difracción, dada por la ecuación (2.140), la podemos observar en la Figura 2.12, para una intensidad normalizada.

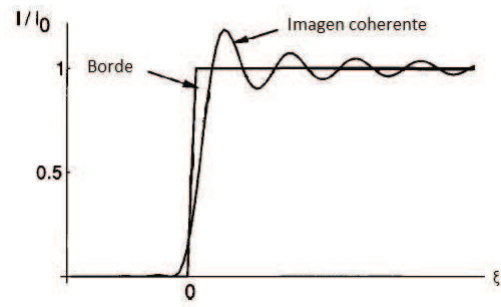


Figura 2.12. Respuesta a una función escalón (extraído de Goodman, 2008).

En la Figura 2.13 se muestra una fotografía de la imagen de un borde iluminado con luz coherente, donde se destacan las franjas provocadas por la difracción [7].

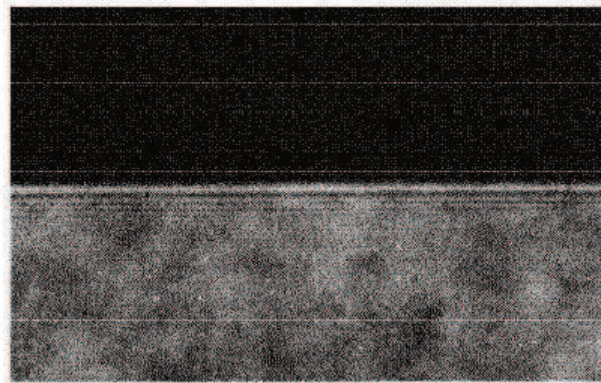


Figura 2.13. Fotografía de un borde iluminado con láser (extraído de Goodman, 2008. Considine, Technical Operations Inc., Burlington Mass.).

Referencias:

- [1] J Goodman (2008). Introduction to Fourier Optics, 2th edition, McGraw – Hill ed.
- [2] E Hecht (2015). Optics, 5th edition, Pearson ed.
- [3] G Chartier (2005). Introduction to Optics, Springer N.Y.
- [4] C Bennet (2008). Principles of Physical Optics, Wiley N.Y.
- [5] L Mandel, E Wolf (1995). Optical Coherence and Quantum Optics, Cambridge.
- [6] M Born, E Wolf (1999). Principles of Optics: Electromagnetic Theory of Propagation, Interference and Diffraction of Light, University Press, Cambridge.
- [7] P Cosidine (1966). Effects of Coherence on Imaging Systems, Journal of Optics.

Capítulo III

Holografía e interferometría holográfica

3.1- Holografía:

Como ya vimos en la ecuación (2.49) un campo de ondas ópticas consiste en una distribución tanto de amplitud como de fase, pero como todos los fotodetectores y dispositivos de almacenamiento de imágenes solo registran intensidades la fase se pierde en el proceso de registro.

Ahora, como también vimos en la ecuación (2.13), si dos ondas de la misma frecuencia interfieren, resulta en una distribución de la intensidad que depende de la diferencia de fase $\Delta\Phi$. Esto es lo que utiliza la holografía, donde la información de fase está codificada por el registro del patrón de intensidades de la interferencia.

Obviamente que para lograr una distribución de intensidad estable durante el proceso de registro, la diferencia de fase $\Delta\Phi$ debe ser estacionaria durante el intervalo de tiempo en que se registra la información, y para esto es fundamental la condición de coherencia mutua entre las ondas.

Fue D. Gabor quien demostró que iluminando un patrón de interferencia previamente registrado, por una de las dos ondas que interfirieron, se puede reconstruir el campo de la otra onda registrada, y este campo reconstruido consiste en una distribución de amplitud y fase, no solo de intensidad [1].

3.1.1- Registro del holograma:

La técnica típica para obtener un holograma consiste en hacer interferir, en el plano de registro, una onda llamada de referencia, que proviene directamente de la fuente luminosa (un láser), con una onda que proviene del objeto a holografear, el cual refleja o transmite la luz que recibe de la misma fuente luminosa mencionada, tal como observamos en la Figura 3.1

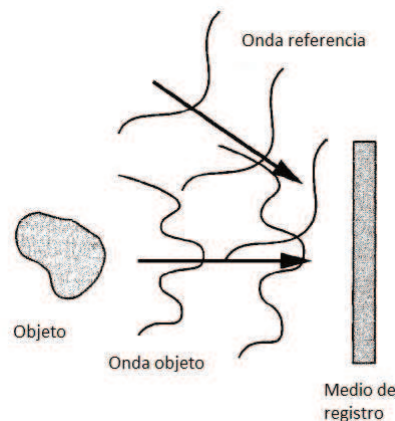


Figura 3.1. Registro de un holograma (extraído de Goodman, 2008).

La intensidad de la suma de los campos complejos depende entonces de la amplitud y de la fase del campo objeto. En efecto, si la onda de referencia está dada por

$$A(x, y) = |A(x, y)| \exp[-j\Psi(x, y)] \quad (3.1)$$

y el frente de onda objeto es igual a

$$a(x, y) = |a(x, y)| \exp[-j\Phi(x, y)] \quad (3.2)$$

Entonces la distribución de la intensidad viene dada por

$$I(x, y) = |A(x, y)|^2 + |a(x, y)|^2 + 2 \cdot |A(x, y)| \cdot |a(x, y)| \cdot \cos[\Psi(x, y) - \Phi(x, y)] \quad (3.3)$$

Aquí vemos que los dos primeros términos de la ecuación (3.3) dependen solo de las intensidades individuales de las ondas, pero el tercero depende tanto de las amplitudes de ambas como de sus fases relativas. De esta forma se registra información de la amplitud y la fase de $a(x, y)$ que caracteriza al objeto [2].

Durante el tiempo de exposición t_A el dispositivo fotosensor recibe la energía

$$E(x, y) = \int_0^{t_A} I(x, y, t) dt \quad (3.4)$$

En el proceso la energía es transmitida al mencionado dispositivo para producir un holograma de amplitud, cuya función de transmisión viene dada por

$$T_A(x, y) = \alpha - \beta \cdot E(x, y) = \alpha - \beta \cdot t_A \cdot I(x, y) = \alpha + \beta' \cdot I(x, y) \quad (3.5)$$

para una intensidad $I(x, y)$ constante en el tiempo. En la ecuación (3.5) α representa la transmitancia de fondo constante y β es el valor del gradiente de la transmitancia en amplitud, cuyo valor es negativo. Finalmente, también asumimos que la intensidad $|A(x, y)|^2$ de la onda de referencia es uniforme sobre toda el área de registro. De manera que la transmitancia del holograma es igual a

$$T_A(x, y) = \alpha + \beta' \cdot [|A(x, y)|^2 + |a(x, y)|^2 + A(x, y) \cdot a^*(x, y) + A^*(x, y) \cdot a(x, y)]$$

$$T_A(x, y) = T_b + \beta' \cdot [|a(x, y)|^2 + A(x, y) \cdot a^*(x, y) + A^*(x, y) \cdot a(x, y)] \quad (3.6)$$

Donde T_b es la transmitancia uniforme dada por la onda de referencia constante [1].

3.1.2- Reconstrucción de la onda objeto:

Una vez registrado el holograma podemos reconstruir la onda objeto. Para ello, supongamos que el holograma está iluminado con una onda coherente $B(x,y)$, de esta manera la luz transmitida por dicho holograma es

$$B(x,y).T_A(x,y) = B.T_b + \beta'.a.a^*.B + \beta'.A^*.B.a + \beta'.A.B.a^* = U_1 + U_2 + U_3 + U_4 \quad (3.7)$$

Vemos que si $B(x,y)$ es igual a la onda de referencia $A(x,y)$, el tercer término en la ecuación (3.7) resulta igual a

$$U_3(x,y) = \beta'.A(x,y).A^*(x,y).a(x,y) = \beta'.|A(x,y)|^2.a(x,y) \quad (3.8)$$

Dado que la intensidad del haz de referencia es uniforme, queda claro que el término U_3 es, salvo un factor constante, una reproducción exacta del frente de onda objeto $a(x,y)$, como vemos en la Figura 3.2a.

Igualmente, si hacemos que $B(x,y)$ sea igual a la conjugada de la onda de referencia, $A^*(x,y)$, el cuarto término de la ecuación (3.7) resulta igual a

$$U_4(x,y) = \beta'.A(x,y).A^*(x,y).a^*(x,y) = \beta'.|A(x,y)|^2.a^*(x,y) \quad (3.9)$$

que es proporcional a la conjugada del frente de onda objeto, como se observa en la Figura 3.2b.

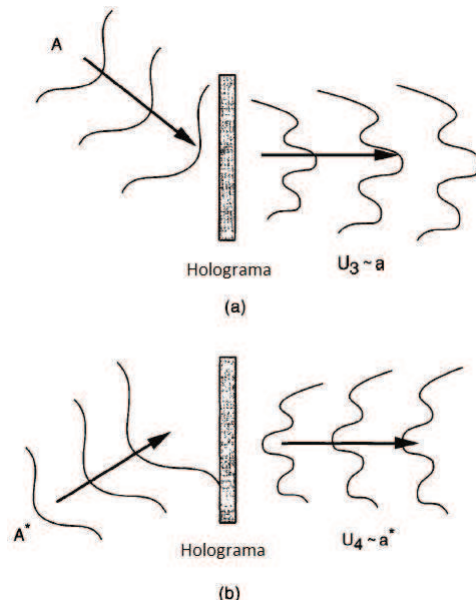


Figura 3.2. Reconstrucción del frente de onda objeto (extraído de Goodman, 2008).

Notemos que en ambos casos el campo de interés (U_3 o U_4) está acompañado por otros tres componentes adicionales, cada uno de los cuales puede considerarse como un efecto de interferencia parásita.

Ahora, observemos que en la ecuación (3.6) el proceso de registro es una transformación no lineal de los campos incidentes. Pero, aunque la transformación global introducida por el holograma es no lineal, la transformación del campo objeto $a(x,y)$ en la componente del campo transmitido $U_3(x,y)$ es lineal, como vemos en la ecuación (3.8), y de la misma manera para $U_4(x,y)$. Por lo tanto, si el campo objeto $a(x,y)$ es la señal de entrada, y el campo $U_3(x,y)$ la señal de salida, el sistema así definido es lineal.

Cuando, en el proceso de reconstrucción, se ilumina al holograma con la onda de referencia $A(x,y)$, la componente $U_3(x,y)$ genera una imagen virtual del objeto original, dado que parece al observador como divergiendo desde la posición del objeto, a pesar de que este ya no se encuentre en ese lugar. Este caso se ilustra en las Figuras 3.3a y b. Del mismo modo, si utilizamos en la reconstrucción la conjugada $A^*(x,y)$ de la onda de referencia, la componente $U_4(x,y)$ genera una imagen real, como vemos en la Figura 3.3c.

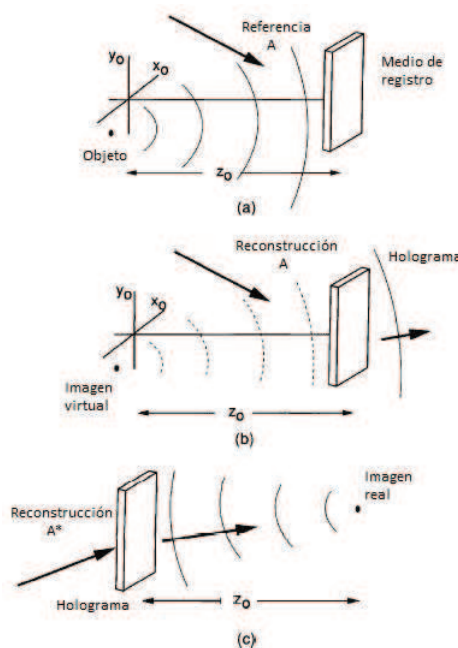


Figura 3.3. Formación de la imagen a través del holograma. a) registro, b) imagen virtual y c) imagen real (extraído de Goodman, 2008).

Para demostrar esto último, consideremos un objeto reducido a un único punto, ya que el resultado para un objeto más complicado se encuentra por superposición lineal de las soluciones de las componentes puntuales. Entonces en el holograma tenemos la suma de la onda de referencia $A(x,y)$ y una onda esférica objeto igual a

$$a(x, y) = a_0 \cdot \exp[jk\sqrt{z_0^2 + (x - x_0)^2 + (y - y_0)^2}] \quad (3.10)$$

donde (x_0, y_0) son las coordenadas del punto objeto y z_0 su distancia normal al plano de registro. Iluminando al holograma con la onda $A^*(x, y)$, obtenemos la componente de la onda transmitida

$$U_4(x, y) = \beta' \cdot |A|^2 \cdot a^* = \beta' \cdot |A|^2 \cdot a_0^* \cdot \exp[-jk\sqrt{z_0^2 + (x - x_0)^2 + (y - y_0)^2}] \quad (3.11)$$

Resultando una onda esférica que converge hacia un foco real, a una distancia z_0 a la derecha del holograma, como se puede observar en la Figura 3.3c [2].

3.1.3- Holograma de Gabor:

Mostramos el esquema de montaje para registrar un holograma de Gabor en la Figura 3.4, tal como lo propuso en su trabajo original.

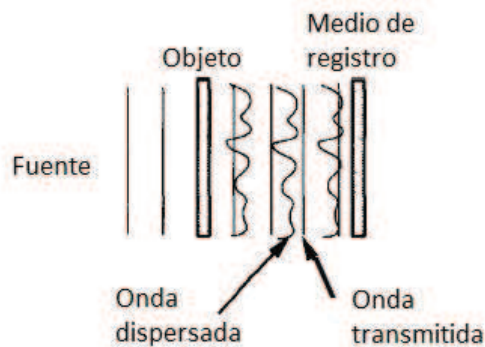


Figura 3.4. Registro del holograma de Gabor (extraído de Goodman, 2008).

Suponemos que el objeto es muy buen transmisor de las ondas ópticas, y que su transmitancia en amplitud está dada por

$$T(x_0, y_0) = T_0 + \Delta T(x_0, y_0) \quad (3.12)$$

donde T_0 es un valor medio elevado de la transmitancia y ΔT la variación alrededor de este valor medio, con

$$|\Delta T| \ll |T_0| \quad (3.13)$$

Cuando el objeto se ilumina con una onda coherente colimada, la luz transmitida está formada por dos componentes, a saber:

- 1- Una onda plana intensa dada por el término T_0
- 2- Una onda difundida débil, generada por la variaciones ΔT de la transmitancia.

Entonces, la intensidad de la luz incidente en el plano de registro, situado a una distancia z_0 del objeto, también viene dada por la ecuación (3.3), donde $A(x,y)$ es la amplitud de la onda plana y $a(x,y)$ la amplitud de la onda difundida.

Por lo tanto, el objeto proporciona la onda de referencia, debido al valor elevado de la transmitancia media T_0 , y la interferencia de esta con la luz difundida da lugar a una distribución de intensidad que depende de la amplitud y la fase de esta misma onda difundida $a(x,y)$.

Si suponemos que el holograma obtenido de esta manera también posee una transmitancia en amplitud dada por la ecuación (3.6), y lo iluminamos con una onda plana de amplitud $B(x,y)$ que incide normalmente, el campo resultante es la suma de los cuatro términos dados en la ecuación (3.7), donde el primer término es una onda plana que pasa a través del holograma, sufriendo una atenuación uniforme, pero no difusión. Al segundo término lo podemos despreciar en virtud de la ecuación (3.13), lo que implica que

$$|a(x,y)| \ll |A(x,y)| \quad (3.14)$$

El tercer término es la componente de un campo proporcional a la onda original difundida $a(x,y)$. Esta onda parece provenir de la imagen virtual del objeto, ubicada a una distancia z_0 a la izquierda del holograma, como vemos en la Figura 3.5. Análogamente, el cuarto término es proporcional a $a^*(x,y)$, y conduce a la formación de una imagen real a una distancia z_0 a la derecha del holograma, como muestra la Figura 3.5.

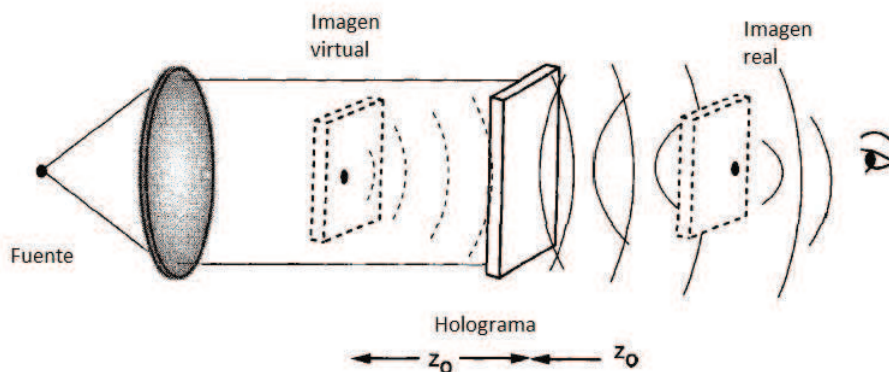


Figura 3.5. Formación de imágenes en el holograma de Gabor (extraído de Goodman, 2008).

Así pues, vemos que el holograma de Gabor genera una imagen real y una imagen virtual, dadas por las variaciones ΔT de la transmitancia del objeto, y que ambas están centradas sobre el eje del holograma. Estas imágenes, denominadas imágenes gemelas, están separadas una distancia axial de $2z_0$, y son acompañadas por un fondo coherente dado por $B(x,y) \cdot T_b$.

La limitación más importante en la aplicación de este esquema de Gabor es la inherente a la necesaria alta transparencia del objeto (ecuación (3.13)). Una segunda limitación está dada por el hecho de que las imágenes gemelas no se pueden separar, si enfocamos la imagen real, aparece

siempre la imagen virtual desenfocada, y viceversa, lo que reduce en gran medida la calidad de las imágenes [2].

3.1.4- Holograma de Leith - Upatnieks:

Consiste en una modificación del esquema de Gabor que resuelve el problema de la imagen gemela. En este esquema se introduce una onda de referencia separada, distinta de la onda que ilumina al objeto, por lo que la onda de referencia no depende de la luz transmitida por el objeto. Además, el haz de referencia se introduce con un ángulo de inclinación con respecto al eje que une los planos objeto y de registro.

La Figura 3.6 muestra el montaje para obtener este holograma.

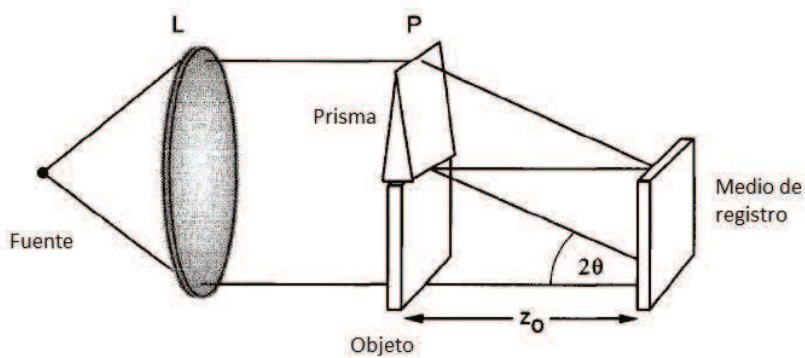


Figura 3.6. Registro del holograma de Leith - Upatnieks (extraído de Goodman, 2008).

La luz procedente de una fuente puntual es colimada por la lente L, parte de la onda plana resultante incide sobre el objeto transparente, y otra parte incide sobre el prisma P, situado arriba del objeto. Este haz es desviado hacia abajo con un ángulo de 2θ respecto a la normal al plano de registro. De esta manera, se produce sobre dicho plano de registro la suma de dos ondas mutuamente coherentes, una transmitida por el objeto, y la otra una onda plana inclinada. Entonces, la distribución de amplitud en el plano de registro se escribe como

$$U(x, y) = |A(x, y)| \cdot \exp[-j2\pi\alpha y] + a(x, y) \quad (3.15)$$

donde α es la frecuencia espacial de la onda de referencia, y viene dada por

$$\alpha = \frac{\text{sen}(2\theta)}{\lambda} \quad (3.16)$$

La distribución de intensidad es igual a

$$I(x, y) = |A|^2 + |a|^2 + |A| \cdot \exp[j2\pi\alpha y] \cdot a + |A| \cdot \exp[-j2\pi\alpha y] \cdot a^* \quad (3.17)$$

Si expresamos la onda objeto explícitamente como una distribución de amplitud y fase

$$a(x, y) = |a(x, y)| \cdot \exp[-j\Phi(x, y)] \quad (3.18)$$

Combinando las dos últimas expresiones, obtenemos

$$I(x, y) = |A(x, y)|^2 + |a(x, y)|^2 + 2 \cdot |A(x, y)| \cdot |a(x, y)| \cdot \cos[2\pi\alpha y - \Phi(x, y)] \quad (3.19)$$

Lo que demuestra que la amplitud y la fase de la luz proveniente del objeto es registrada bajo la forma de modulaciones de amplitud y fase de un portador espacial de frecuencia α .

A continuación, pasamos a describir la reconstrucción de la imagen holográfica del objeto. Para ello, tenemos el holograma que posee una transmitancia en amplitud dada por

$$T_A(x, y) = T_b + \beta' \cdot [|a(x, y)|^2 + A^* \cdot \exp(j2\pi\alpha y) \cdot a(x, y) + A \cdot \exp(-j2\pi\alpha y) \cdot a^*(x, y)] \quad (3.20)$$

y supongamos que se ilumina el holograma con una onda plana que incide normalmente, de amplitud $B(x, y)$, como observamos en la Figura 3.7

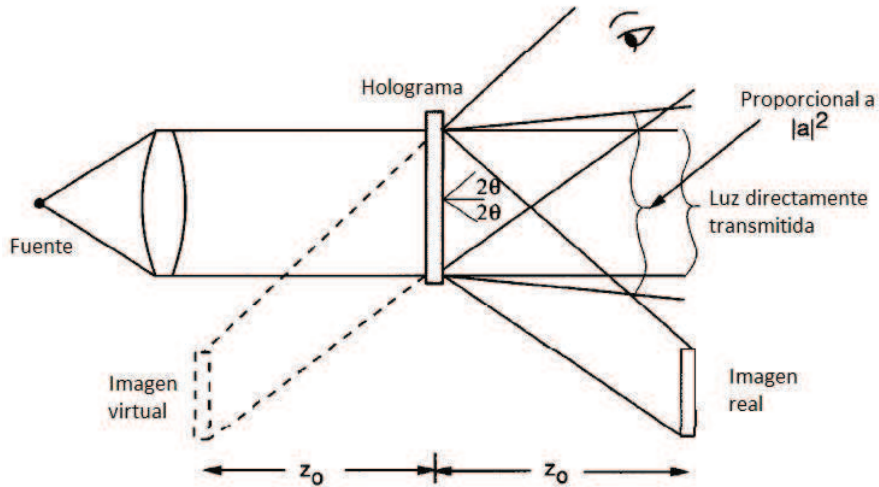


Figura 3.7. Reconstrucción del holograma de Leith - Upatnieks (extraído de Goodman, 2008).

El campo transmitido por el holograma tiene cuatro componentes, cada uno generado por uno de los términos de la transmitancia en la ecuación (3.20), a saber

$$U_1(x, y) = T_1 \cdot B(x, y) = T_b \cdot B(x, y) \quad (3.21a)$$

$$U_2(x, y) = T_2 \cdot B(x, y) = \beta' \cdot B(x, y) \cdot |a(x, y)|^2 \quad (3.21b)$$

$$U_3(x, y) = T_3 \cdot B(x, y) = \beta' \cdot B(x, y) \cdot A^* \cdot a(x, y) \cdot \exp(j2\pi\alpha y) \quad (3.21c)$$

$$U_4(x, y) = T_4 \cdot B(x, y) = \beta' \cdot B(x, y) \cdot A \cdot a^*(x, y) \cdot \exp(-j2\pi\alpha y) \quad (3.21d)$$

La componente del campo $U_1(x,y)$ es una versión atenuada del haz incidente en la reconstrucción, y por lo tanto, es una onda plana que viaja a lo largo del eje óptico.

El segundo término $U_2(x,y)$ varía espacialmente, dado que está compuesto de ondas planas propagándose en direcciones diferentes con respecto al eje óptico.

La componente $U_3(x,y)$ es proporcional al frente de onda objeto $a(x,y)$ multiplicado por un factor exponencial lineal. La proporcionalidad a $a(x,y)$ implica que este término genera una imagen virtual del objeto, a una distancia z_0 a la izquierda del holograma, mientras que el factor exponencial lineal $\exp(j2\pi\alpha y)$ indica que la imagen está desviada un ángulo 2θ con respecto al eje óptico, como vemos en la Figura 3.7.

Análogamente, la componente $U_4(x,y)$ es proporcional al frente de onda conjugado $a^*(x,y)$ del objeto, lo que implica que se forma una imagen real a la distancia z_0 a la derecha del holograma. El factor exponencial lineal $\exp(-j2\pi\alpha y)$ indica que la imagen real se encuentra desviada un ángulo -2θ respecto del eje óptico, como también vemos en la Figura 3.7.

La observación más importante de todo esto, es que, aunque se siguen generando las imágenes gemelas, estas han sido separadas angularmente de las componentes $U_1(x,y)$ y $U_2(x,y)$ de la onda. Esta separación es causada por la utilización de una onda de referencia inclinada, durante el registro.

Para que las imágenes gemelas estén separadas una de la otra, y también del haz próximo al eje óptico, el ángulo de inclinación del haz de referencia con respecto al haz objeto, debe ser mayor que un determinado ángulo mínimo $2\theta_{\min}$.

Para encontrar este mínimo debemos determinar la frecuencia portadora mínima α , para la cual los espectros de frecuencias espaciales de T_3 y T_4 (o sea, la transmitancia del holograma correspondiente a las imágenes virtual y real) no se superponen entre si, ni con los espectros de T_1 y T_2 .

Ahora consideremos los espectros de frecuencias espaciales de los diferentes términos de la transmitancia dados en la ecuación (3.21). Si no tenemos en cuenta la superficie finita del holograma, podemos obtener las siguientes transformadas

$$G_1(f_x, f_y) = \mathcal{F}[T_1(x, y)] = T_b \cdot \delta(x, y) \quad (3.22)$$

$$G_2(f_x, f_y) = \mathcal{F}[T_2(x, y)] = \beta' \cdot G_a(f_x, f_y) * G_a(f_x, f_y) \quad (3.23)$$

En la ecuación (3.23) se utilizó la operación autocorrelación $*$, donde $G_a(f_x, f_y) = \mathcal{F}[a(x, y)]$. Y finalmente tenemos

$$G_3(f_x, f_y) = \mathcal{F}[T_3(x, y)] = \beta' \cdot A * G_a(f_x, f_y - \alpha) \quad (3.24a)$$

$$G_4(f_x, f_y) = \mathcal{F}[T_4(x, y)] = \beta' \cdot A \cdot G_a^*(-f_x, -f_y - \alpha) \quad (3.24b)$$

Destaquemos que el ancho de banda de la función $G_a(f_x, f_y)$ es igual al ancho de banda del objeto, dado que los espectros no difieren más que por la función transferencia de la propagación, que es la función de fase pura dada en la ecuación (2.107).

Si suponemos que el objeto no tiene componentes de frecuencias espaciales mayores que B [ciclos/mm], podemos representar el espectro de $|G_a(f_x, f_y)|$ como vemos en la Figura 3.8a. El espectro correspondiente a la transmitancia completa del holograma se muestra en la Figura 3.8b.

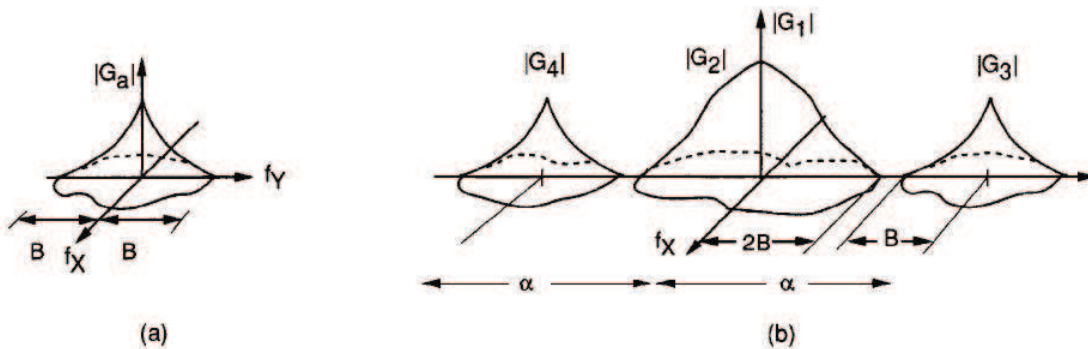


Figura 3.8. Espectros. a) del objeto y b) del holograma (extraído de Goodman, 2008).

El término $|G_1(f_x, f_y)|$ es una función δ centrada en el origen del plano (f_x, f_y) . El término $|G_2(f_x, f_y)|$ es proporcional a la función autocorrelación de $|G_a(f_x, f_y)|$, y se extiende hasta la frecuencia $2B$. $|G_3(f_x, f_y)|$ es proporcional a $|G_a(f_x, f_y)|$ pero está centrada en el punto $(0, \alpha)$, mientras que $|G_4(f_x, f_y)|$ es proporcional a la función simétrica de $|G_a(f_x, f_y)|$, pero centrada en $(0, -\alpha)$.

En la Figura 9.8b vemos que $|G_3|$ y $|G_4|$ pueden separarse de $|G_2|$ si $\alpha \geq 3B$ o, lo que es igual si

$$\text{sen}(2\theta) \geq 3B\lambda \quad (3.25)$$

Por lo que el valor mínimo del ángulo para el cual se produce la separación es

$$(2\theta)_{\min} = \text{sen}^{-1}(3B\lambda) \quad (3.26)$$

Dado que el término G_2 es generado por la interferencia de las ondas que provienen de cada punto del objeto con las que provienen de los otros puntos del mismo, y por el contrario, G_3 y G_4 son generados por la interferencia entre la onda de referencia y la onda objeto, cuando esta onda objeto es mas débil que la de referencia (es decir $|a| < |A|$), G_2 es mucho menor que G_3 o G_4 , pudiéndose descartarse, en este caso el ángulo mínimo es aquel que separa solo a G_3 de G_4 , es decir [2]

$$(2\theta)_{\min} = \text{sen}^{-1}(B\lambda) \quad (3.27)$$

3.1.5- Otros tipos de hologramas:

3.1.5.1- Holograma de Fourier:

Si tenemos un objeto delgado podemos registrar la transformada de Fourier del objeto y del haz de referencia, este es el llamado holograma de Fourier.

Supongamos un objeto ubicado en el plano focal, a la izquierda de una lente, iluminado por una luz coherente, como en la Figura 3.9. Si la amplitud compleja del objeto es $a(x,y)$, entonces la amplitud compleja del campo objeto en el plano de registro, que se encuentra localizado en el plano focal a la derecha de la lente, es la transformada de Fourier de $a(x,y)$

$$\varepsilon_p = \mathcal{F}[a(x, y)] \quad (3.28)$$

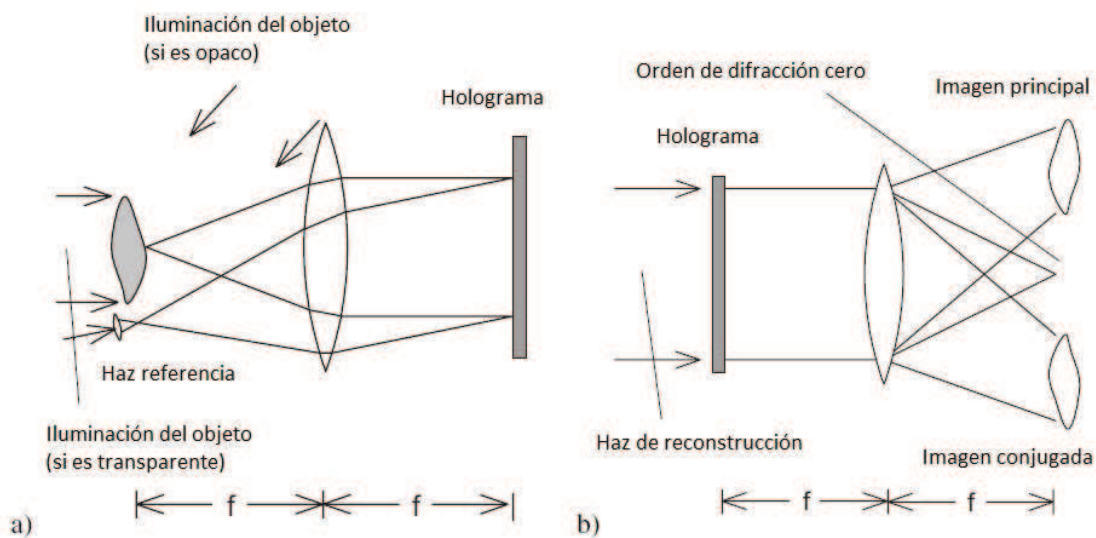


Figura 3.9. Holograma de Fourier. a) registro y b) reconstrucción (extraído de Kreis, 2005).

La onda de referencia es esférica, y es emitida desde una fuente situada en el punto (x_0, y_0) en el plano focal a la izquierda de la lente. Si asumimos para esta onda una amplitud unitaria, tenemos que la distribución de amplitud de la onda de referencia es igual a $A(x,y) = \delta(x-x_0, y-y_0)$, por lo que la amplitud compleja del campo de referencia en el plano de registro es igual a

$$\varepsilon_R = \mathcal{F}[\delta(x - x_0, y - y_0)] = \exp[-j2\pi(\xi x_0 + \eta y_0)] \quad (3.29)$$

Luego, en el plano de registro, obtenemos el holograma caracterizado por la siguiente transmisión en amplitud

$$T(\xi, \eta) = \beta'[\varepsilon_R \varepsilon_R^* + \varepsilon_P \varepsilon_P^* + \varepsilon_R^* \varepsilon_P + \varepsilon_R \varepsilon_P^*](\xi, \eta)$$

$$T(\xi, \eta) = T_b(\xi, \eta) + \beta'[\varepsilon_P \varepsilon_P^* + \varepsilon_R^* \varepsilon_P + \varepsilon_R \varepsilon_P^*](\xi, \eta) \quad (3.30)$$

Para reconstruir la imagen del objeto, iluminamos este holograma de Fourier con una onda plana igual a la de referencia, obteniendo

$$T \cdot \varepsilon_R = T_b \cdot \varepsilon_R + \beta' [\varepsilon_P \varepsilon_P^* \varepsilon_R + \varepsilon_R \varepsilon_R^* \varepsilon_P + \varepsilon_R \varepsilon_R \varepsilon_P^*] \quad (3.31)$$

Si en el proceso de reconstrucción el holograma se sitúa en el plano focal, a la izquierda de una lente de reconstrucción, como se aprecia en la Figura 3.9b, esto provoca que en el plano focal de la derecha (x', y') se obtenga la transformada de Fourier de $T \cdot \varepsilon_R$, dada por la expresión

$$\mathcal{F}[T(\xi, \eta) \cdot \varepsilon_R(\xi, \eta)](x', y') = T_b \cdot \mathcal{F}[\varepsilon_R] + \beta' \mathcal{F}[\varepsilon_P \varepsilon_P^* \varepsilon_R] + \beta' \mathcal{F}[\varepsilon_R \varepsilon_R^* \varepsilon_P] + \beta' \mathcal{F}[\varepsilon_R \varepsilon_R \varepsilon_P^*]$$

$$\mathcal{F}[T \cdot \varepsilon_R] = C_1 \cdot \delta(x' + x_0, y' + y_0) + C_2 [(a_p * a_p)(x' + x_0, y' + y_0)] + C_2 \cdot a_p(-x', -y') + C_2 \cdot a_p^*(x' + 2x_0, y' + 2y_0) \quad (3.32)$$

El primer término del campo de la reconstrucción, dado por la ecuación (3.32), es un punto en el plano focal (x', y'), denominado término de continua. El segundo término está caracterizado por formar un halo alrededor del punto de continua, similar al señalado en la Figura 3.7 para $|a|^2$. El tercer término es proporcional al frente de onda original del objeto, pero invertido. Finalmente, el cuarto término es el conjugado del frente de onda objeto, desplazado una cantidad $(-2x_0, -2y_0)$. Ambas imágenes, la real y la virtual, se sitúan en el plano focal a la derecha de la lente [1].

3.1.5.2- Holograma de Fourier sin lentes:

Si en el esquema del holograma de Fourier, a través de otra lente colimamos los haces objeto y referencia, podemos interpretar que el objeto y la fuente de referencia se encuentran ambos a una distancia cualquiera (incluso infinita) del plano de registro.

En el esquema del holograma de Fourier sin lentes esto es posible si el objeto y la fuente se encuentran situados ambos a la misma distancia, finita, del plano de registro, como se muestra en la Figura 3.10.

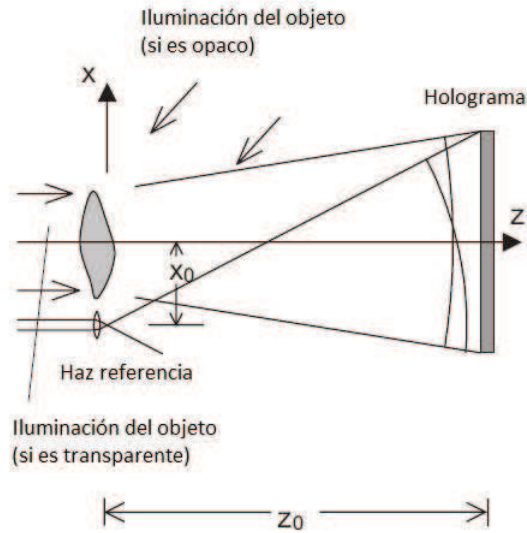


Figura 3.10. Registro del holograma de Fourier sin lentes (extraído de Kreis, 2005).

Consideremos el haz que proviene del objeto $a_P(x,y)$, aplicando la integral de Fresnel – Kirchhoff, la distribución de amplitud compleja en el plano de registro es igual a

$$\varepsilon_P(\xi, \eta) = \frac{\exp(jkz)}{j\lambda z} \exp[j\pi z \lambda (\xi^2 + \eta^2)] \iint_{-\infty}^{\infty} a_P(x, y) \cdot \exp\left[\frac{j\pi}{\lambda z} (x^2 + y^2)\right] \cdot \exp[-2j\pi(\xi x + \eta y)] dx dy$$

$$\varepsilon_P(\xi, \eta) = C \cdot \exp[j\pi z \lambda (\xi^2 + \eta^2)] \cdot \mathcal{F}\left[a_P(x, y) \cdot \exp\left[\frac{j\pi}{\lambda z} (x^2 + y^2)\right]\right] = C \cdot \exp[j\pi z \lambda (\xi^2 + \eta^2)] \cdot \varepsilon'_P(\xi, \eta) \quad (3.33)$$

Para el haz de referencia, igual a $A_R(x,y) = \delta(x-x_0, y-y_0)$ aplicamos la misma integral, obteniendo

$$\varepsilon_R(\xi, \eta) = \frac{\exp(jkz)}{j\lambda z} \exp[j\pi z \lambda (\xi^2 + \eta^2)] \iint_{-\infty}^{\infty} \delta(x - x_0, y - y_0) \cdot \exp\left[\frac{j\pi}{\lambda z} (x^2 + y^2)\right] \cdot \exp[-2j\pi(\xi x + \eta y)] dx dy$$

$$\varepsilon_R(\xi, \eta) = C \cdot \exp[j\pi z \lambda (\xi^2 + \eta^2)] \cdot \varepsilon'_R(\xi, \eta) \quad (3.34)$$

Resultando un holograma con la transmisión en amplitud siguiente

$$T(\xi, \eta) = \beta' [|\varepsilon_R|^2 + |\varepsilon_P|^2 + \varepsilon_R^* \varepsilon_P + \varepsilon_R \varepsilon_P^*] = \beta' [|\varepsilon'_R|^2 + |\varepsilon'_P|^2 + \varepsilon'_R{}^* \varepsilon'_P + \varepsilon'_R \varepsilon'_P{}^*]$$

$$T(\xi, \eta) = T_b(\xi, \eta) + \beta' [\varepsilon'_P \varepsilon'_P{}^* + \varepsilon'_R{}^* \varepsilon'_P + \varepsilon'_R \varepsilon'_P{}^*](\xi, \eta) \quad (3.35)$$

Obteniéndose una expresión equivalente a la (3.30). El hecho crucial es que la distancia z en las expresiones (3.33) y (3.34) debe ser la misma para que el factor exponencial cuadrático $\exp[j\pi z \lambda (\xi^2 + \eta^2)]$ y su conjugado resulten igual a la unidad en la expresión (3.35).

El efecto del factor de fase esférico de la onda objeto, asociado con la difracción de Fresnel, es anulado por la onda de referencia esférica que posee la misma curvatura [1].

3.1.5.3- Hologramas imagen, de Fresnel y de Fraunhofer:

Si el objeto es enfocado por una lente sobre el plano de registro, Figura 3.11, el resultado se denomina holograma imagen.

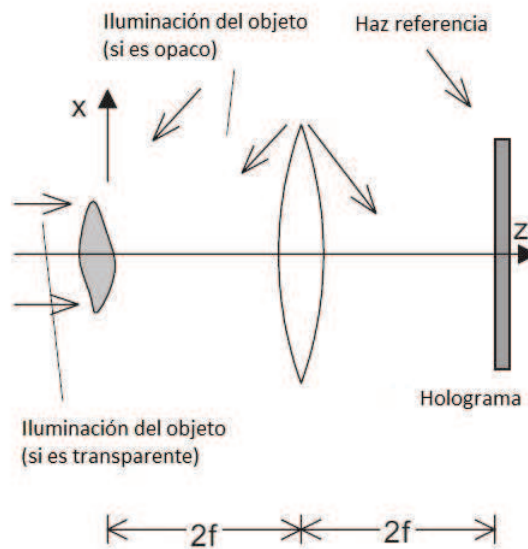


Figura 3.11. Registro del holograma imagen (extraído de Kreis, 2005).

Aquí la imagen real del objeto es registrada a través del campo producido por la reflexión (o transmisión) de la luz proveniente de dicho objeto. La ventaja de este holograma imagen reside en el hecho de que la reconstrucción puede realizarse por una fuente de luz con un bajo grado de coherencia, produciendo imágenes holográficas con una definición bastante aceptable. Una desventaja es que el ángulo de observación del objeto está limitado por la apertura angular de la lente utilizada.

Para esta clase de holograma la configuración mas generalizada es la ubicar el plano de registro a una distancia del objeto dentro del campo cercano, o región de difracción de Fresnel, por lo que en ese caso se lo denomina holograma de Fresnel.

Si la distancia entre el objeto y el plano de registro se incrementa hasta la región de difracción de Fraunhofer, o de campo lejano, tenemos el llamado holograma de Fraunhofer.

También cabe mencionar que estos hologramas, de Fresnel y de Fraunhofer, pueden ser obtenidos sin la necesidad de utilizar lentes [1].

3.1.6- Posiciones de las imágenes y aumentos:

Supongamos que la onda de referencia es generada por una fuente puntual ubicada en las coordenadas (x_r, y_r, z_r) , como vemos en la Figura 3.12a. Dado que la transformación del campo de amplitud objeto en el campo de amplitud imagen es lineal, y si tenemos que el ángulo de inclinación del haz de referencia es lo suficientemente grande para separar las imágenes gemelas entre sí, y de los otros términos de continua de la luz transmitida, es suficiente considerar solo un punto del objeto, como una simple fuente puntual objeto, situada en las coordenadas (x_o, y_o, z_o) . Observemos que en la Figura 3.12a, tanto z_r como z_o son negativos, dado que están a la izquierda del plano de registro (o sea que tenemos ondas esféricas divergentes).

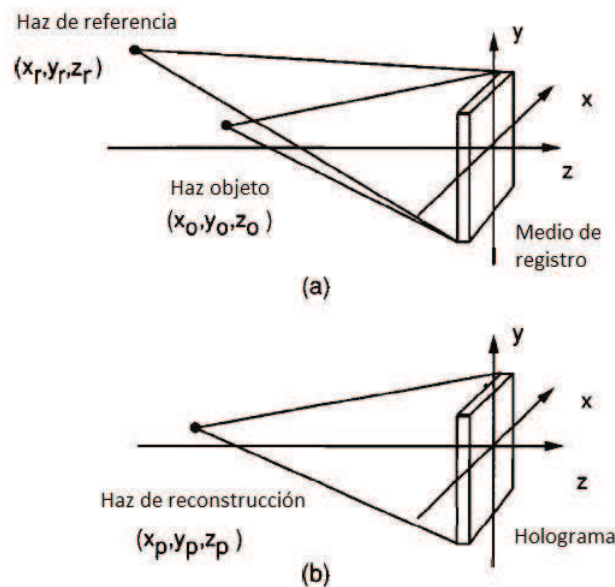


Figura 3.12. Posiciones geométricas para a) registro y b) reconstrucción (extraído de Goodman, 2008).

En la reconstrucción, mostrada en la Figura 3.12b, el holograma está iluminado por una onda esférica originada en una fuente puntual ubicada en las coordenadas (x_p, y_p, z_p) . Aquí también z_p es negativa, dado que la onda de reconstrucción es divergente.

Para generalizar, suponemos que la luz utilizada en el registro del holograma posee una longitud de onda λ_1 , y la luz utilizada en la reconstrucción de la imagen del objeto posee una longitud de onda λ_2 .

Si utilizamos las aproximaciones de fase cuadrática en las ondas esféricas de referencia y objeto, el campo total que incide sobre el plano de registro resulta igual a

$$U(x, y) = A(x_r, y_r) \cdot \exp\left[\frac{-j\pi}{z_r \lambda_1} [(x - x_r)^2 + (y - y_r)^2]\right] + a(x_o, y_o) \cdot \exp\left[\frac{-j\pi}{z_o \lambda_1} [(x - x_o)^2 + (y - y_o)^2]\right] \quad (3.36)$$

donde $A(x_r, y_r)$ y $a(x_o, y_o)$ son constantes complejas que representan las amplitudes y las fases relativas de ambas ondas esféricas.

La distribución de intensidad que origina el patrón de interferencia es

$$I = |A|^2 + |a|^2 + A^* a \cdot \exp \left[\frac{-j\pi}{\lambda_1} \left[\frac{(x-x_0)^2}{z_0} + \frac{(y-y_0)^2}{z_0} - \frac{(x-x_r)^2}{z_r} - \frac{(y-y_r)^2}{z_r} \right] \right] + A a^* \cdot \exp \left[\frac{-j\pi}{\lambda_1} \left[-\frac{(x-x_0)^2}{z_0} - \frac{(y-y_0)^2}{z_0} + \frac{(x-x_r)^2}{z_r} + \frac{(y-y_r)^2}{z_r} \right] \right] \quad (3.37)$$

De esta manera obtenemos un holograma con una determinada transmitancia en amplitud, entonces considerando solo los dos términos importantes de esta transmitancia, obtenemos

$$T_3 = \beta' A a^* \cdot \exp \left[\frac{j\pi}{\lambda_1} \left[\frac{(x-x_0)^2}{z_0} + \frac{(y-y_0)^2}{z_0} - \frac{(x-x_r)^2}{z_r} - \frac{(y-y_r)^2}{z_r} \right] \right] \quad (3.38a)$$

$$T_4 = \beta' A^* a \cdot \exp \left[\frac{-j\pi}{\lambda_1} \left[\frac{(x-x_0)^2}{z_0} + \frac{(y-y_0)^2}{z_0} - \frac{(x-x_r)^2}{z_r} - \frac{(y-y_r)^2}{z_r} \right] \right] \quad (3.38b)$$

Si iluminamos el holograma con una onda esférica, que en la aproximación paraxial, es igual a

$$U_p(x, y) = B(x_p, y_p) \cdot \exp \left[\frac{-j\pi}{\lambda_2 z_p} [(x - x_p)^2 + (y - y_p)^2] \right] \quad (3.39)$$

Entonces los dos frentes de onda de interés, se determinan de la siguiente manera

$$U_3 = T_3 \cdot U_p = \beta' A a^* B \cdot \exp \left[-j\pi \left[-\frac{(x-x_0)^2}{z_0 \lambda_1} - \frac{(y-y_0)^2}{z_0 \lambda_1} + \frac{(x-x_r)^2}{z_r \lambda_1} + \frac{(y-y_r)^2}{z_r \lambda_1} + \frac{(x-x_p)^2}{z_p \lambda_2} + \frac{(y-y_p)^2}{z_p \lambda_2} \right] \right] \quad (3.40a)$$

$$U_4 = T_4 \cdot U_p = \beta' A^* a B \cdot \exp \left[-j\pi \left[\frac{(x-x_0)^2}{z_0 \lambda_1} + \frac{(y-y_0)^2}{z_0 \lambda_1} - \frac{(x-x_r)^2}{z_r \lambda_1} - \frac{(y-y_r)^2}{z_r \lambda_1} + \frac{(x-x_p)^2}{z_p \lambda_2} + \frac{(y-y_p)^2}{z_p \lambda_2} \right] \right] \quad (3.40b)$$

Dado que en (3.40) solo tenemos los términos lineales y cuadráticos, las dos expresiones de U_3 y U_4 podemos considerarlas como aproximaciones de fase cuadrática de las ondas esféricas que salen del holograma. La presencia de los términos lineales nos indica que las ondas son convergentes o divergentes, hacia o desde puntos que no están sobre el eje z. De esta manera podemos obtener las coordenadas de las imágenes (x_i, y_i, z_i) , si comparamos las ecuaciones (3.40) con otra exponencial de fase cuadrática de la forma

$$U_i(x, y) = K \cdot \exp \left[\frac{-j\pi}{\lambda_2 z_i} [(x - x_i)^2 + (y - y_i)^2] \right] \quad (3.41)$$

A partir de las expresiones que contienen los términos cuadráticos de x e y en (3.40) y (3.41), obtenemos la distancia axial z_i de los puntos imagen

$$\frac{1}{z_i} = \frac{1}{z_p} \pm \frac{\lambda_2}{\lambda_1 z_0} \mp \frac{\lambda_2}{\lambda_1 z_r} \quad (3.42)$$

Donde los signos superiores se aplican a la onda imagen U_3 y los inferiores a U_4 . Si iluminamos el holograma desde la izquierda, cuando z_i es negativa la imagen es virtual y se ubica a la izquierda del holograma, si z_i es positiva la imagen es real y está a la derecha del holograma.

Las coordenadas x_i e y_i de los puntos imagen se encuentran igualando los términos lineales en x e y de las ecuaciones (3.40) y (3.41), obteniéndose lo siguiente

$$x_i = \pm \frac{\lambda_2 z_i}{\lambda_1 z_0} x_0 \mp \frac{\lambda_2 z_i}{\lambda_1 z_r} x_r + \frac{z_i}{z_p} x_p \quad (3.43a)$$

$$y_i = \pm \frac{\lambda_2 z_i}{\lambda_1 z_0} y_0 \mp \frac{\lambda_2 z_i}{\lambda_1 z_r} y_r + \frac{z_i}{z_p} y_p \quad (3.43b)$$

Las expresiones (3.42), (3.43a) y (3.43b) proporcionan las relaciones que nos permiten predecir la posición de las imágenes creadas por el proceso holográfico.

Notemos que para el holograma de Fourier sin lente las distancias axiales del objeto y de la fuente de referencia son las mismas, $z_0 = z_r$, por lo que ambas imágenes, virtual y real, se encuentran en el mismo plano a una distancia $z_i = z_p$.

Los aumentos axial y transversal del proceso holográfico pueden ser encontrados a partir de las ecuaciones (3.42) y (3.43). Sabiendo que el aumento transversal viene dado por la siguiente expresión

$$M_t = \left| \frac{\partial x_i}{\partial x_0} \right| = \left| \frac{\partial y_i}{\partial y_0} \right| = \left| \frac{\lambda_2 z_i}{\lambda_1 z_0} \right| = \left| \left(1 - \frac{z_0}{z_r} \pm \frac{\lambda_1 z_0}{\lambda_2 z_p} \right)^{-1} \right| \quad (3.44)$$

Análogamente, el aumento axial es igual a

$$M_a = \left| \frac{\partial z_i}{\partial z_0} \right| = \left| \frac{\lambda_1}{\lambda_2} \frac{1}{\left(1 - \frac{z_0}{z_r} \pm \frac{\lambda_1 z_0}{\lambda_2 z_p} \right)^2} \right| = \frac{\lambda_1}{\lambda_2} M_t^2 \quad (3.45)$$

Observemos que, por lo general, los aumentos axial y transversal no son iguales, esto es importante cuando consideramos la formación de imágenes de objetos tridimensionales, dado que la diferencia entre estos aumentos crea una distorsión tridimensional de la imagen de dicho objeto [2].

3.2- Interferometría holográfica:

Dado que, como vimos anteriormente, consideramos a la holografía como un método para almacenar y reconstruir campos ópticos, este concepto puede ser ampliado al utilizar la interferencia de dos campos ópticos almacenados, lo que se denomina interferometría holográfica.

Entonces, se define a la interferometría holográfica como la comparación interferométrica de dos o más campos ópticos, donde cada uno de ellos puede ser reconstruido holográficamente. Debemos aclarar que solo pequeñas diferencias entre los campos ópticos pueden ser comparadas por la interferometría holográfica.

Los requerimientos para obtener un interferograma holográfico podemos resumirlos de la siguiente manera:

- 1) En lo concerniente al objeto bajo observación, este debe conservar la misma microestructura durante todo el proceso.
- 2) La forma geométrica del objeto debe ser aproximadamente la misma para todos los campos ópticos registrados, dado que estos luego han de compararse.
- 3) Deben cumplirse los requerimientos de ancho de longitud de onda y coherencia para la fuente laser utilizada.
- 4) Los cambios en el objeto deben ser en un rango pequeño, donde solo varíe la fase de la onda de luz dispersada por este.

Si el cambio en el campo objeto es espacialmente homogéneo, donde solo varíe la fase de este en forma muy suave, de punto a punto de dicho objeto, se puede visualizar un patrón de interferencia al superponer dos hologramas. Este patrón es llamado interferograma holográfico, y es utilizado en muchas aplicaciones como un método de medición no destructivo [3].

3.2.1- Interferograma holográfico de doble exposición:

En este método, dos frentes de onda dispersados por el mismo objeto son registrados de manera consecutiva en el mismo medio de registro. Estos dos frentes de onda corresponden a dos estados diferentes del objeto, uno en las condiciones iniciales y el otro luego de efectuarse un cambio en algún parámetro físico (por ejemplo ejerciendo una carga sobre el objeto).

Consideremos la amplitud compleja del primer frente de onda objeto, que emana de un punto P del mismo, cuya expresión puede escribirse como

$$a_1(P) = a_{01}(P) \cdot \exp[j\Phi(P)] \quad (3.46)$$

Este frente de onda es registrado holográficamente. En la expresión (3.46) a_{01} es la amplitud real y $\Phi(P)$ es la distribución de fase, esta última varía espacialmente de manera aleatoria, debido a la microestructura del objeto que refleja (o refracta) la luz de forma difusa.

Si se produce la variación de un parámetro físico del objeto (por ejemplo la deformación de un objeto opaco, o el cambio en la distribución del índice de refracción de un objeto transparente), esto provoca una variación en la distribución de la fase del punto P, dado por $\Delta\Phi(P)$. Entonces la amplitud compleja del segundo frente de onda registrado holográficamente por el mismo medio de registro, es igual a

$$a_2(P) = a_{02}(P) \cdot \exp[j[\Phi(P) + \Delta\Phi(P)]] \quad (3.47)$$

Luego el holograma registrado con los dos frentes de onda objeto es reconstruido, esto produce la interferencia de ambos frentes objeto, provocando una distribución de intensidad estacionaria dada por

$$I(P) = |a_1(P) + a_2(P)|^2$$

$$I(P) = [a_{01}(P) \cdot \exp[j\Phi(P)] + a_{02}(P) \cdot \exp[j[\Phi(P) + \Delta\Phi(P)]]] \cdot [a_{01}(P) \cdot \exp[-j\Phi(P)] + a_{02}(P) \cdot \exp[-j[\Phi(P) + \Delta\Phi(P)]]]$$

$$I(P) = I_1(P) + I_2(P) + \sqrt{I_1(P) \cdot I_2(P)} \cdot [\exp[j\Delta\Phi(P)] + \exp[-j\Delta\Phi(P)]]$$

$$I(P) = I_1(P) + I_2(P) + 2 \cdot \sqrt{I_1(P) \cdot I_2(P)} \cdot \cos[\Delta\Phi(P)] \quad (3.48)$$

Para amplitudes iguales, estos es $a_{01}(P) = a_{02}(P)$, tenemos

$$I(P) = 2 \cdot I_1(P) \{1 + \cos[\Delta\Phi(P)]\} \quad (3.49)$$

similar a la ecuación (2.13).

La variación en la fase $\Delta\Phi$ es denominada diferencia de fase de interferencia. Si la variación espacial de la fase de interferencia, sobre la superficie reconstruida, es pequeña, la distribución de intensidad (3.49) representa la irradiancia del objeto modulada por un patrón de franjas cosenoidales.

Los centros brillantes de estas franjas son los contornos de la superficie del objeto donde la fase de interferencia es siempre un múltiplo entero par de π , los centros oscuros de las franjas corresponden a múltiplos enteros impares de π .

Si la fase de interferencia cambia demasiado rápido desde un punto observado hasta su vecino próximo, esto es en un valor mayor que π , entonces solo se observará una distribución de intensidad aleatoria.

En la Figura 3.13 vemos el proceso de registro de los dos hologramas, Figuras a y b, y la reconstrucción del interferograma en c, por el método de la doble exposición [1].

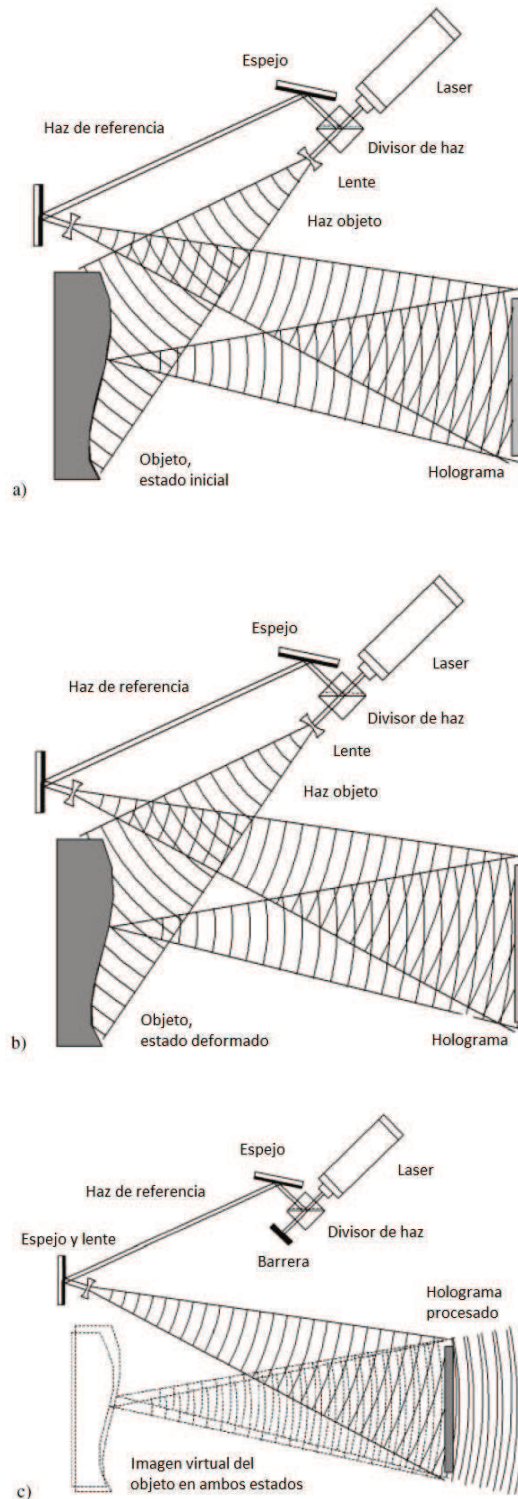


Figura 3.13. Método de la doble exposición, a) y b) registro, c) reconstrucción (extraído de Kreis, 2005).

3.2.2- Variación de la fase de interferencia:

Supongamos un objeto opaco que refleja difusamente luz monocromática, siendo este deformado mecánicamente, lo que provoca el desplazamiento de un punto P de su superficie. La medida del desplazamiento del punto de la superficie puede ser considerada como una diferencia de camino óptico $\delta(P)$. Esta es la diferencia entre los caminos desde el punto de la fuente de iluminación S, hasta el punto de la superficie P (en el estado inicial y luego en el estado deformado), y de este hacia el punto de observación B, como se aprecia en la Figura 3.14.

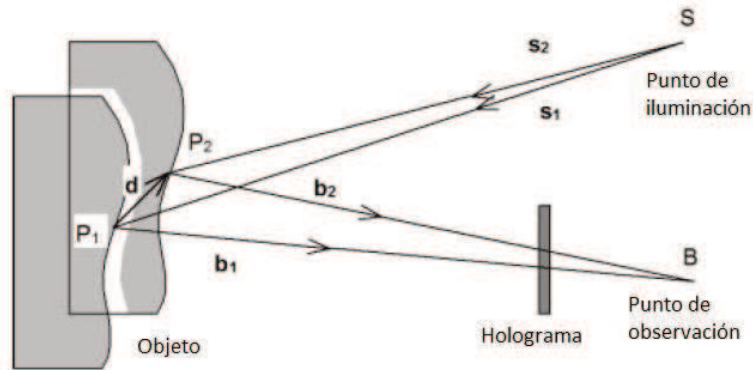


Figura 3.14. Esquema holográfico para medir la deformación de un objeto opaco (extraído de Kreis, 2005).

La fase de interferencia está relacionada con la diferencia de camino óptico a través de la siguiente ecuación

$$\Delta\phi(P) = \frac{2\pi}{\lambda} \delta(P) \quad (3.50)$$

En el esquema holográfico de la Figura 3.14 tenemos la iluminación que diverge del punto $S = (x_S, y_S, z_S)$, llamado punto de iluminación, y tenemos el punto de observación en el medio de registro dado por $B = (x_B, y_B, z_B)$.

Cuando el objeto es deformado, un punto de su superficie se desplaza de la posición $P_1 = (x_{P1}, y_{P1}, z_{P1})$ hacia su nueva posición $P_2 = (x_{P2}, y_{P2}, z_{P2})$, este movimiento queda definido por el vector desplazamiento

$$\vec{d}(P) = [d_x(P), d_y(P), d_z(P)] = P_2 - P_1 \quad (3.51)$$

Aclaremos que los puntos P_1 y P_2 son diferentes en una escala microscópica, pero macroscópicamente pueden ser indistinguibles.

La diferencia de camino óptico puede ser expresada por

$$\delta(P) = \overline{SP_1} + \overline{P_1B} - (\overline{SP_2} + \overline{P_2B}) = \hat{s}_1 \cdot \overline{SP_1} + \hat{b}_1 \cdot \overline{P_1B} - \hat{s}_2 \cdot \overline{SP_2} - \hat{b}_2 \cdot \overline{P_2B} \quad (3.52)$$

donde \hat{s}_1 y \hat{s}_2 son versores en la dirección de iluminación, \hat{b}_1 y \hat{b}_2 son versores en la dirección de observación.

Ahora designemos por $\hat{s}(P_1, P_2)$ como la bisectriz de los versores \hat{s}_1 y \hat{s}_2 , y $\hat{b}(P_1, P_2)$ como la bisectriz de los versores \hat{b}_1 y \hat{b}_2 . Además llamemos como $\Delta\hat{s}(P_1, P_2)$ y $\Delta\hat{b}(P_1, P_2)$ a la mitad de la diferencia entre estos versores. Por lo tanto, podemos escribir

$$\hat{s}(P_1, P_2) = \frac{1}{2}[\hat{s}_1 + \hat{s}_2] \quad , \quad \hat{b}(P_1, P_2) = \frac{1}{2}[\hat{b}_1 + \hat{b}_2] \quad (3.53a)$$

$$\Delta\hat{s}(P_1, P_2) = \frac{1}{2}[\hat{s}_1 - \hat{s}_2] \quad , \quad \Delta\hat{b}(P_1, P_2) = \frac{1}{2}[\hat{b}_1 - \hat{b}_2] \quad (3.53b)$$

Por definición, podemos escribir el vector desplazamiento dado en (3.51) igual a

$$\overrightarrow{P_1B} - \overrightarrow{P_2B} = d(P) \quad \text{y} \quad \overrightarrow{SP_2} - \overrightarrow{SP_1} = d(P) \quad (3.54)$$

Insertando las ecuaciones (3.53) y (3.54) en la ecuación (3.52) obtenemos

$$\delta(P) = (\hat{s} + \Delta\hat{s}) \cdot \overrightarrow{SP_1} + (\hat{b} + \Delta\hat{b}) \cdot \overrightarrow{P_1B} - (\hat{s} - \Delta\hat{s}) \cdot \overrightarrow{SP_2} - (\hat{b} - \Delta\hat{b}) \cdot \overrightarrow{P_2B}$$

$$\delta(P) = \hat{b} \cdot \vec{d} - \hat{s} \cdot \vec{d} + \Delta\hat{b}(\overrightarrow{P_1B} + \overrightarrow{P_2B}) + \Delta\hat{s}(\overrightarrow{SP_1} + \overrightarrow{SP_2}) \quad (3.55)$$

Dado que los desplazamientos son muy pequeños, comparados con las distancias del esquema holográfico montado, tenemos que $|\vec{d}|$ está en el rango de los micrómetros, y $\overrightarrow{SP_1}$ y $\overrightarrow{P_1B}$ en el rango de los centímetros. Además el versor $\Delta\hat{s}$ es aproximadamente ortogonal a $(\overrightarrow{SP_1} + \overrightarrow{SP_2})$, y el versor $\Delta\hat{b}$ es casi ortogonal a $(\overrightarrow{P_1B} + \overrightarrow{P_2B})$, por lo tanto sus productos escalares son aproximadamente cero, y pueden ser omitidos en la ecuación (3.55), por lo que ya no hay distinción entre los dos puntos P_1 y P_2 . Entonces la ecuación (3.55) puede reescribirse de la forma

$$\delta(P) = \vec{d}(P) \cdot [\hat{b}(P) - \hat{s}(P)] \quad (3.56)$$

Considerando las coordenadas del punto de iluminación, de observación y de la superficie del objeto, podemos obtener las expresiones de los versores iluminación y observación

$$\hat{s}(P) = \begin{bmatrix} s_x(P) \\ s_y(P) \\ s_z(P) \end{bmatrix} = \frac{1}{\sqrt{(x_P - x_S)^2 + (y_P - y_S)^2 + (z_P - z_S)^2}} \begin{bmatrix} x_P - x_S \\ y_P - y_S \\ z_P - z_S \end{bmatrix} \quad (3.57)$$

$$\hat{b}(P) = \begin{bmatrix} b_x(P) \\ b_y(P) \\ b_z(P) \end{bmatrix} = \frac{1}{\sqrt{(x_B - x_P)^2 + (y_B - y_P)^2 + (z_B - z_P)^2}} \begin{bmatrix} x_B - x_P \\ y_B - y_P \\ z_B - z_P \end{bmatrix} \quad (3.58)$$

Estos versores junto con el factor $2\pi/\lambda$ de la ecuación (3.50) constituyen el denominado vector sensibilidad, dado por

$$\vec{e}(P) = \frac{2\pi}{\lambda} [\hat{b}(P) - \hat{s}(P)] \quad (3.59)$$

Entonces la fase de interferencia resulta igual a

$$\Delta\Phi(P) = \vec{d}(P) \cdot \vec{e}(P) \quad (3.60)$$

Vemos que la fase de interferencia, en un punto dado, se determina por el producto escalar del vector desplazamiento, de dicho punto, y del vector sensibilidad.

El vector sensibilidad está definido solamente por la geometría del esquema holográfico montado, y este a su vez, debe realizarse para obtenerse la máxima sensibilidad. De esta manera lo que se mide en la fase de interferencia es la proyección del vector desplazamiento sobre el vector sensibilidad. Entonces para desplazamientos ortogonales al vector sensibilidad, la fase de interferencia es cero, independientemente de la magnitud del desplazamiento.

La ecuación (3.60) es la base, utilizada por la interferometría holográfica, para todas las mediciones cuantitativas de la deformación en objetos opacos [1].

3.2.3- Ambigüedad del signo de la fase de interferencia:

Cuando tratamos de extraer la fase de interferencia $\Delta\Phi(x,y)$ de la distribución de intensidad $I(x,y)$ (como en la ecuación (3.49)), el problema con que nos encontramos es que la función coseno es periódica y par

$$\cos(\Delta\Phi) = \cos(\pm\Delta\Phi + 2\pi n) \quad \text{con } n=0,1,2,\dots \quad (3.61)$$

Entonces una distribución de intensidad posee una indefinición dada por la ambigüedad del signo y la adición de un múltiplo entero de 2π en la función coseno.

Una manera de deshacerse de la ambigüedad de signo es utilizar información de las condiciones del esquema experimental, donde ocurre el cambio en el camino óptico que provoca la distribución de la fase de interferencia.

En muchas aplicaciones se puede asumir que la distribución de fase de interferencia es una función continua y diferenciable, y varía de forma gradual. Por ejemplo, conociendo la dirección de

la carga que actúa sobre el objeto, que puede ser de tensión o compresión, se puede asignar el signo adecuado a la fase de interferencia, sabiendo en que dirección esta se incrementa o disminuye.

A toda función trigonométrica inversa la podemos expresar por una función del tipo tag^{-1} , a saber

$$\cos^{-1}(\Delta\Phi) = \text{tag}^{-1} \left[\frac{\sqrt{1-(\Delta\Phi)^2}}{\Delta\Phi} \right] \quad (3.62)$$

La función tag^{-1} de una sola variable posee su valor principal en el intervalo $(-\pi/2, \pi/2)$, pero en muchos algoritmos de procesamiento (ver capítulo IV) la función tag^{-1} está compuesta de un coeficiente entre el seno del argumento y el coseno del mismo argumento, entonces si consideramos los signos del numerador y el denominador por separado, en este caso el valor principal de la función tag^{-1} está determinado de forma consistente en el intervalo $(-\pi, \pi)$.

La ambigüedad en el término $2\pi n$ de la ecuación (3.61) se manifiesta en la distribución de la fase de interferencia por el denominado envolvimiento de fase de módulo 2π . En este caso solo el valor principal de la función tag^{-1} , en el intervalo $(-\pi, \pi)$, está determinado como un valor pico, y cuando este valor es alcanzado, la fase salta hacia otro valor pico, aunque en realidad la fase puede variar lentamente.

El efecto del módulo 2π puede corregirse por el procesamiento llamado demodulación o desenvolvimiento de fase, mediante la suma o resta de múltiplos enteros de 2π los saltos en la fase son eliminados.

Para agregar el término correcto, en algunas aplicaciones, es necesario conocer el valor exacto del desplazamiento, o lo que es lo mismo la fase de interferencia, en un punto predeterminado P del patrón de distribución de interferencia, siendo preferible que este valor sea cero, $\Delta\Phi(P) = 0$. Si asumimos que la variación de la fase de interferencia es continua, el múltiplo de 2π para un punto en particular, se determina contando los saltos de 2π desde el punto P antedicho, hasta este punto, a lo largo de una trayectoria continua. Esta técnica solo es posible si el vector sensibilidad se mantiene constante.

En la Figura 3.15 vemos un ejemplo de una ménsula sujeta en el extremo izquierdo, y con una carga aplicada en el extremo derecho que varía en los dos registros holográficos tomados.

La Figura 3.15a es una gráfica del desplazamiento en mm, provocado por la variación de la carga, calculado para el esquema holográfico del experimento. La Figura 3.15b es el interferograma holográfico, donde las franjas observadas son los saltos en módulos 2π de la variación de la fase.

Observemos que en el extremo de la izquierda, que está fijo, el desplazamiento es menor y las franjas son mas anchas. Por el contrario, en el extremo derecho el desplazamiento es mayor y las franjas son muy finas.

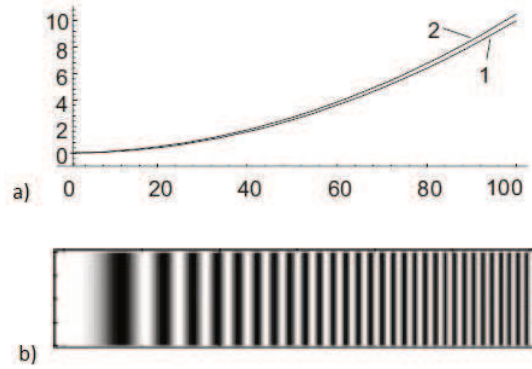


Figura 3.15. Efecto del múltiplo de 2π en la distribución de la fase de interferencia para una ménsula de 100 mm de longitud, con la ubicación de $S=(-200,0,250)$ mm, $B=(50,0,250)$ mm y $\lambda=514.5$ nm. a) Desplazamiento medido en mm. b) Patrón de interferencia (extraído de Kreis, 2005).

Si evaluamos la distribución de intensidad dada en la ecuación (3.49) por medio de la transformada de Fourier podemos obtener la distribución de la fase de interferencia envuelta, o lo que se denomina el mapa de fase envuelto, para ello reescribamos la ecuación (3.49) de la forma

$$I(x, y) = b(x, y) + c(x, y) \cdot \cos[\Delta\Phi(x, y)] \quad (3.63)$$

Haciendo uso de la exponencial compleja obtenemos

$$g(x, y) = \frac{1}{2} c(x, y) \cdot \exp[j\Delta\Phi(x, y)] \quad (3.64)$$

e introduciendo la ecuación (3.64) en la (3.63) resulta

$$I(x, y) = b(x, y) + g(x, y) + g^*(x, y) \quad (3.65)$$

Aplicando la transformada de Fourier a $I(x, y)$ obtenemos

$$\mathcal{J}(f_x, f_y) = B(f_x, f_y) + G(f_x, f_y) + G^*(f_x, f_y) \quad (3.66)$$

donde $\mathcal{J}(f_x, f_y)$ es una distribución hermítica en el dominio de las frecuencias espaciales, por lo tanto

$$\mathcal{J}(f_x, f_y) = \mathcal{J}^*(-f_x, -f_y) \quad (3.67)$$

teniendo en cuenta que la parte real de $\mathcal{J}(f_x, f_y)$ es par, y la parte imaginaria impar.

Si graficamos el espectro de amplitud $|\mathcal{J}(f_x, f_y)|$, vemos que este posee simetría respecto del punto $\mathcal{J}(0,0)$, el término de continua, como se observa en la Figura 3.16b. Por otro lado, el término

$B(f_x, f_y)$ contiene el pico en el punto $\mathcal{I}(0,0)$, y las variaciones de baja frecuencia dadas por la dispersión alrededor del término de continua.

Observemos que los términos $G(f_x, f_y)$ y $G^*(f_x, f_y)$ contienen la misma información, como se evidencia por la ecuación (3.67).

Utilizando un filtro pasa banda en el dominio de las frecuencias espaciales, el término $B(f_x, f_y)$ y uno de los dos términos $G(f_x, f_y)$, o $G^*(f_x, f_y)$, pueden ser eliminados, como vemos en la Figura 3.16c. El espectro que queda, $G(f_x, f_y)$ o $G^*(f_x, f_y)$, no es hermitico, entonces aplicando la transformada de Fourier inversa por ejemplo a $G(f_x, f_y)$, obtenemos la función compleja $g(x,y)$ sin perder su parte real e imaginaria. De este modo la fase de interferencia puede calcularse por

$$\Delta\Phi(x, y) = \text{tag}^{-1} \left[\frac{\text{Im } g(x,y)}{\text{Re } g(x,y)} \right] \quad (3.68)$$

Utilizando la transformada inversa en $G^*(f_x, f_y)$, la fase de interferencia resulta igual a $-\Delta\Phi(x,y)$, lo que manifiesta la ambigüedad del signo.

Un ejemplo, en una dimensión, de la aplicación de la transformada de Fourier es mostrada en la Figura 3.16

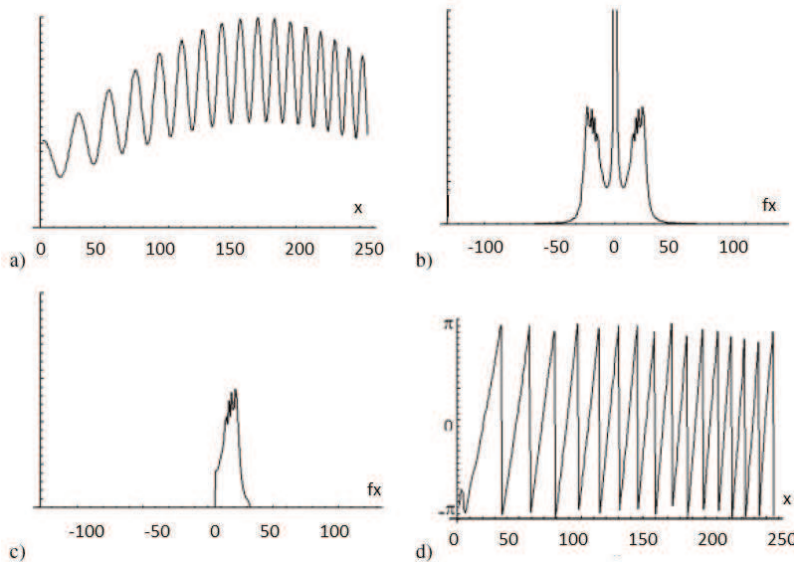


Figura 3.16. Evaluación mediante transformada de Fourier. a) Distribución de intensidad.

b) Espectro de amplitud. c) Espectro filtrado. d) Mapa de fase envuelta (extraído de Kreis, 2005).

La distribución de la intensidad, con la variación del contraste y el término de continua, se muestra en la Figura 3.16a. El espectro de amplitud resultante de aplicar la transformada de Fourier a la función de la Figura 3.16a se muestra en la Figura 3.16b. Luego de aplicar el filtro pasa banda, el espectro de amplitud que queda lo vemos en la Figura 3.16c, en este caso

$G(f_x, f_y)$. Realizando la transformada inversa de Fourier a la función de la Figura 3.16c obtenemos el gráfico de la fase de interferencia envuelta, como vemos en la Figura 3.16d [1].

3.3- Elementos utilizados en el esquema holográfico:

No trataremos de ahondar aquí en los fundamentos de los diversos elementos utilizados en holografía, estos pueden ser estudiados en muchos textos específicos [4], solo realizaremos una breve descripción de los componentes mas comunes que forman parte de un esquema óptico holográfico.

3.3.1- El láser:

La implementación de la holografía y la interferometría, en el espectro visible, solo fue posible gracias a la invención de una fuente de radiación de luz coherente, el láser.

El principio básico de funcionamiento del láser es la emisión estimulada de radiación, donde la emisión de fotones es disparada por una onda electromagnética. Todos los fotones generados de esta manera poseen la misma frecuencia, polarización, fase y dirección que la onda estimuladora.

Normalmente en un grupo de átomos, en equilibrio térmico, la mayoría de estos se encuentran en la configuración más baja de energía, o estado base. Si se puede lograr que una gran parte de los átomos llegue al estado excitado (o de más alta energía), entonces estos al volver a su estado base emiten ondas que pueden estimular a los otros átomos, provocando una avalancha de ondas, todas de la misma fase y propagándose en la misma dirección.

Para provocar la excitación de los átomos se debe proveer de energía al sistema mediante un proceso llamado bombeo. Entonces, el láser puede ser considerado como un amplificador, donde una onda introducida en el sistema genera muchas ondas con la misma dirección, frecuencia y fase.

Para prevenir que el dispositivo amplifique solo interferencia se instalan dos espejos planos en lados opuestos del recinto que contiene los átomos a excitar, denominado medio activo. Estos espejos son ajustados para que queden exactamente paralelos uno de otro, entonces los fotones son reflejados hacia adelante y hacia atrás, así se forman ondas estables entre los dos espejos. Si uno de estos espejos es semitransparente algunos fotones logran escapar, formándose el denominado haz laser de radiación coherente.

El láser de haz continuo más utilizado en aplicaciones holográficas es el de gas helio – neón (láser He-Ne), dado que es económico. Este emite luz roja coherente de 633 nm de longitud de onda, posee una gran vida útil y no necesita agua u otro elemento para refrigerar el sistema, como

otros dispositivos laser. La fuente de excitación de este tipo de laser es eléctrica, con un rango de potencias entre 2 mW a 50 mW [4].

3.3.2- Componentes ópticos:

Para obtener hologramas de buena calidad los instrumentos ópticos y mecánicos deben estar montados fijamente en una mesa aislada de vibraciones y movimientos externos de cualquier tipo.

La propagación del haz laser que ilumina al objeto puede ser mejorada colocando un *pin hole* en el punto focal de la lente magnificadora o colimadora. Debido a que el polvo y las marcas en las superficies ópticas producen interferencia espacial, el *pin hole* actúa como un filtro espacial que deja pasar solo el término de continua del espectro del haz, esto resulta en una iluminación limpia, con características aproximadamente gaussianas.

Los polarizadores tienen un rol menor en la holografía interferométrica. Con ellos podemos garantizarnos que la onda objeto y la onda de referencia estén polarizadas en el mismo ángulo, y así obtener un buen contraste de las franjas en un interferograma.

Por último, cabe mencionar al divisor de haz formado por un semiespejo, de donde obtenemos el haz de referencia y el haz objeto, ver la Figura 3.13 [3].

3.3.3- Cámara CCD:

La holografía digital se basa en la grabación digitalizada de los hologramas utilizando una matriz CCD (*charge couple device*) como medio de registro. Las cámaras CCD han reemplazado casi completamente a los otros medios de registro, que utilizaban la denominada holografía convencional.

Una cámara CCD funciona transfiriendo la imagen almacenada a través de la misma matriz de fotodetectores.

El bloque de construcción básico de este dispositivo es el semiconductor de metal – óxido, o MOS, configurado como capacitor, como se muestra en la Figura 3.17

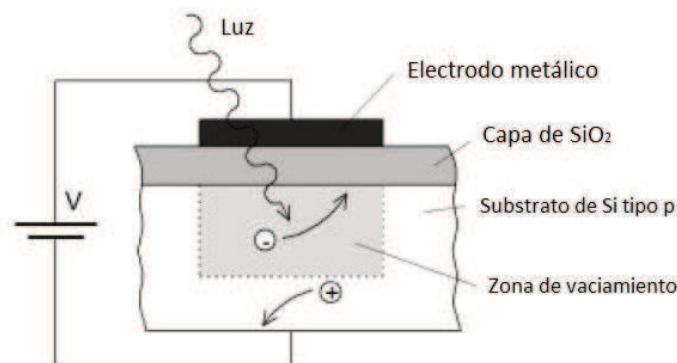


Figura 3.17. Esquema de capacitor MOS (extraído de Kreis, 2005).

Un capacitor MOS está formado comúnmente, por un sustrato de silicio intrínseco con una capa aislante de dióxido de silicio sobre él. Un electrodo metálico delgado, llamado compuerta, es depositado sobre la capa de dióxido de silicio.

Si aplicamos un voltaje positivo al electrodo, esto causa que los huecos en el silicio tipo p se muevan lejos de la interface Si – SiO₂, dejando una región de vaciamiento de cargas positivas, esta región de vaciamiento varía en forma gradual, con la distancia, desde la interface, y actúa como una diferencia de potencial para los electrones móviles. Aumentando el voltaje aplicado a la compuerta, se aumenta la profundidad de la región de vaciamiento.

Si un fotón con suficiente energía es absorbido en la región de vaciamiento, se genera un par electrón – hueco, a continuación el hueco se mueve hacia el electrodo negativo, y el electrón se queda en la zona de vaciamiento. Si más luz es absorbida por el silicio, más electrones son almacenados en la zona de vaciamiento, esto puede continuar hasta que el silicio quede saturado, o el voltaje en el electrodo deje de aplicarse.

La cantidad de carga negativa que puede ser almacenada es proporcional al voltaje aplicado, el espesor del óxido y el área del electrodo. El número total de electrones que pueden ser almacenados es denominado capacidad. De esta forma una matriz bidimensional de estos capacitores MOS puede almacenar imágenes en la forma de electrones atrapados bajo la capa de dióxido de silicio, en la zona de vaciamiento, que denominamos celda.

La manera en que se mueven las cargas acumuladas fuera de la matriz CCD es ajustando los voltajes de cada compuerta, para lograr transferir las cargas de una celda hacia su vecina. Un ejemplo de este proceso está ilustrado en la Figura 3.18 para tres voltajes desfasados temporalmente.

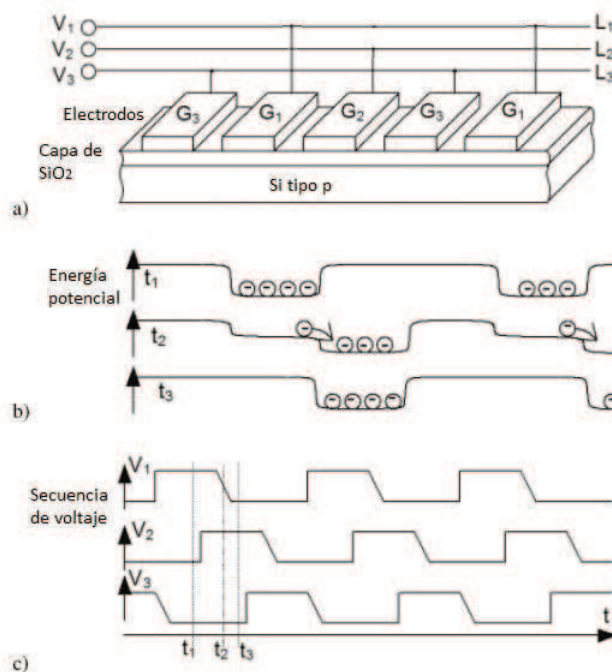


Figura 3.18. Transferencia de carga en un CCD (extraído de Kreis, 2005).

Dado que una matriz CCD consiste en un grupo de fotodetectores ordenados en filas y columnas, podemos considerar una sola columna de esta matriz para ejemplificar su funcionamiento.

Para el voltaje trifásico esquematizado en la Figura 3.18, las compuertas de la columna están conectadas en líneas de voltaje separadas L_1 , L_2 y L_3 , en grupos contiguos de tres compuertas G_1 , G_2 y G_3 . Inicialmente, en el instante t_1 , un voltaje V_1 es aplicado a las compuertas G_1 , de manera que los fotoelectrones son recolectados en todas las celdas bajo las compuertas G_1 , Figura 3.18b. El período de tiempo en que se aplica el voltaje es llamado tiempo de integración. Antes que V_1 se haga cero, el voltaje V_2 aumenta, y en el instante t_2 , cuando V_1 es menor que V_2 , las compuertas G_2 dan mayor potencial a sus celdas, de modo que los electrones se mueven de las celdas de G_1 a las celdas de G_2 . La misma secuencia de voltajes es repetida para las líneas L_2 y L_3 , ahora los electrones son transferidos desde las celdas de G_2 hacia las celdas de G_3 . Esta secuencia de voltajes se repite hasta mover todos los electrones de la matriz.

Los electrones son registrados fuera de la matriz, como una señal en serie, por un detector MOS lineal [5].

Referencias:

- [1] T Kreis (2005). Handbook of Holographic Interferometry, Wiley – VCH ed.
- [2] J Goodman (2008). Introduction to Fourier Optics, 2th edition, McGraw – Hill ed.
- [3] P Hariharan (2007). Optical Interferometry, Academic Press, San Diego.
- [4] L Mandel, E Wolf (1995). Optical Coherence and Quantum Optics, Cambridge.
- [5] G Holst (1996). CCD Arrays, Cameras and Displays. Optical Engineering, SPIE.

Capítulo IV

Holografía digital

La técnica de la holografía digital se basa en la grabación digital del holograma mediante cámaras fotográficas CCD o CMOS, y la posterior reconstrucción, por métodos numéricos, de los campos ópticos utilizando un software de procesamiento. Esta técnica no debe confundirse con la generación numérica de un holograma, que luego puede ser impreso en algún medio físico, para lograr reconstruir ópticamente el campo de un objeto, obteniéndose así la imagen de algo que no existe físicamente. Veremos un ejemplo de un holograma generado por ordenador en el próximo capítulo que tratará sobre la aplicación experimental de la holografía digital [1].

4.1- Teorema del muestreo:

Para el tratamiento de datos de forma digital, una función continua $g(x,y)$ es representada por un conjunto discreto de puntos en el plano (x,y) . Si estos puntos son tomados suficientemente próximos entre sí, los datos muestreados pueden representar fielmente la función original.

Para una clase particular de funciones, llamadas de banda limitada en frecuencias, la reconstrucción se puede lograr exactamente con la única condición de que el intervalo entre dos muestras no supere cierto límite.

Por funciones de banda limitada se refieren a aquellas cuyas transformadas de Fourier son distintas de cero solo en una región finita del espacio de frecuencias.

Ahora pasemos a definir una función denominada peine, igual a

$$\text{comb}(x) = \sum_{n=-\infty}^{\infty} \delta(x - n) \quad (4.1)$$

Por lo que si consideramos una red rectangular de muestras para la función $g(x,y)$, podemos obtener la función de muestreo dada por

$$g_s(x, y) = \text{comb}\left(\frac{x}{X}\right) \cdot \text{comb}\left(\frac{y}{Y}\right) \cdot g(x, y) \quad (4.2)$$

La función muestreo consiste en una matriz de funciones δ , separadas por intervalos de longitud X y longitud Y sobre los ejes x e y respectivamente, como se muestra en la Figura 4.1

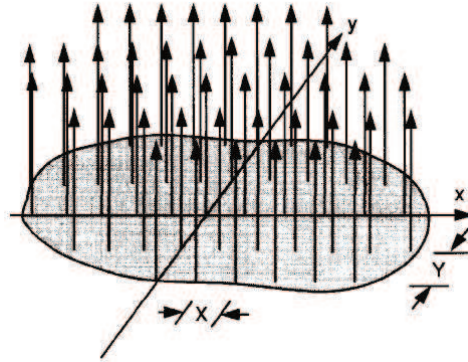


Figura 4.1. Función muestreo (extraído de Goodman, 2008).

El área bajo cada función δ es proporcional al valor de la función $g(x,y)$ en cada punto particular de la red de muestreo.

Si aplicamos el teorema de la convolución, se puede hallar el espectro $G_s(f_x, f_y)$ de $g_s(x,y)$ mediante la expresión

$$G_s(f_x, f_y) = \mathcal{F} \left[\text{comb} \left(\frac{x}{X} \right) \cdot \text{comb} \left(\frac{y}{Y} \right) \right] \otimes G(f_x, f_y) \quad (4.3)$$

En esta expresión vemos que por un lado tenemos

$$\mathcal{F} \left[\text{comb} \left(\frac{x}{X} \right) \cdot \text{comb} \left(\frac{y}{Y} \right) \right] = X \cdot Y \cdot \text{comb}(X \cdot f_x) \cdot \text{comb}(Y \cdot f_y) = \sum_{n,m=-\infty}^{\infty} \delta \left(f_x - \frac{n}{X}, f_y - \frac{m}{Y} \right) \quad (4.4)$$

Por lo que el espectro de la función muestreada viene dado por

$$G_s(f_x, f_y) = \sum_{n,m=-\infty}^{\infty} G \left(f_x - \frac{n}{X}, f_y - \frac{m}{Y} \right) \quad (4.5)$$

Entonces se obtiene el espectro de g_s construyendo el espectro de g alrededor de cada punto $\left(\frac{n}{X}, \frac{m}{Y} \right)$ del plano (f_x, f_y) , como vemos en la Figura 4.2a.

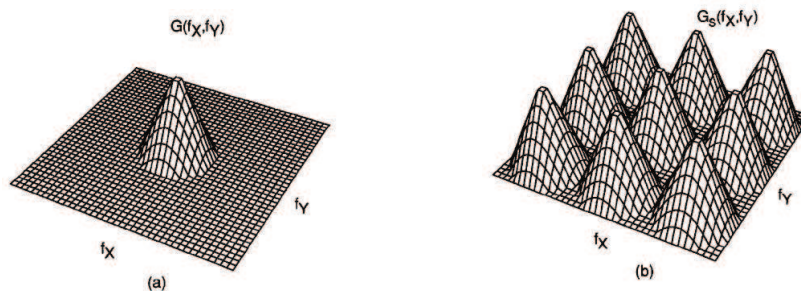


Figura 4.2. a) Espectro de la función original. b) Espectro de los datos muestreados (solo vemos tres períodos en cada dirección, pero se extienden al infinito) (extraído de Goodman, 2008).

Si suponemos que la función $g(x,y)$ es de banda limitada, su espectro $G(f_x, f_y)$ es distinto de cero solo en una región finita del espacio de frecuencias. Se puede encontrar la región donde el espectro de la función muestreada es distinto de cero, considerando solo la región alrededor de cada punto $\left(\frac{n}{X}, \frac{m}{Y}\right)$ en el plano de frecuencias. Si X e Y son suficientemente pequeños (por que las muestras son tomadas en puntos suficientemente próximos entre si), los intervalos $\frac{1}{X}$ y $\frac{1}{Y}$ entre las diversas regiones espectrales consideradas son lo suficientemente grandes para que las regiones vecinas no se solapen, como observamos en la Figura 4.2b.

Por lo tanto podemos encontrar el espectro original $G(f_x, f_y)$ en forma exacta a partir de $G_s(f_x, f_y)$, haciendo pasar la función muestreada $g_s(x,y)$ a través de un filtro lineal invariante, que transmita sin distorsión solo el término donde $n = 0$ y $m = 0$ de la expresión (4.5), y que elimine totalmente los demás términos. De este modo se obtiene una réplica exacta de los datos originales $g(x,y)$.

Ahora, para determinar la distancia máxima permitida entre dos muestras, llamemos $2B_x$ y $2B_y$ a las dimensiones del rectángulo mas pequeño que contenga todo el dominio elegido en el plano de frecuencias, donde existe la función $G(f_x, f_y)$. Dado que los términos del espectro en (4.5), de la función muestreada, están separados por las distancias $\frac{1}{X}$ y $\frac{1}{Y}$, la separación de estas regiones espectrales está asegurada si

$$X \leq \frac{1}{2B_x} \quad \text{e} \quad Y \leq \frac{1}{2B_y} \quad (4.6)$$

Por último, nos queda precisar la función de transferencia exacta del filtro por el cual se deben hacer pasar los datos muestreados. Notemos que si la relación (4.6) se satisface, existe una función de transferencia que siempre dará el resultado buscado, cualquiera sea la forma de la región en el dominio de frecuencias, donde existe $G(f_x, f_y)$, esta es

$$H(f_x, f_y) = \text{rect}\left(\frac{f_x}{2B_x}\right) \cdot \text{rect}\left(\frac{f_y}{2B_y}\right) \quad (4.7)$$

De modo que obtenemos la reconstrucción exacta de $G(f_x, f_y)$, a partir de $G_s(f_x, f_y)$, dado que el espectro de la señal de salida aplicando el filtro es

$$G_s(f_x, f_y) \cdot \text{rect}\left(\frac{f_x}{2B_x}\right) \cdot \text{rect}\left(\frac{f_y}{2B_y}\right) = G(f_x, f_y) \quad (4.8)$$

Si a la expresión (4.8) la trasladamos al dominio espacial, tenemos

$$\left[\text{comb}\left(\frac{x}{X}\right) \cdot \text{comb}\left(\frac{y}{Y}\right) \cdot g(x, y) \right] \otimes h(x, y) = g(x, y) \quad (4.9)$$

donde $h(x,y)$ es la respuesta al impulso del filtro aplicado, y se escribe como

$$h(x, y) = \mathcal{F}^{-1} \left[\text{rect} \left(\frac{f_x}{2B_x} \right) \cdot \text{rect} \left(\frac{f_y}{2B_y} \right) \right] = 4B_x B_y \cdot \text{sinc}(2B_x x) \cdot \text{sinc}(2B_y y) \quad (4.10)$$

teniendo en cuenta que

$$\text{comb} \left(\frac{x}{X} \right) \cdot \text{comb} \left(\frac{y}{Y} \right) \cdot g(x, y) = X \cdot Y \cdot \sum_{n,m=-\infty}^{\infty} g(nX, mY) \cdot \delta(x - nX, y - mY) \quad (4.11)$$

La expresión (4.9) la podemos transformar en

$$g(x, y) = 4 \cdot B_x \cdot B_y \cdot X \cdot Y \cdot \sum_{n,m=-\infty}^{\infty} g(nX, mY) \cdot \text{sinc}[2B_x(x - nX)] \cdot \text{sinc}[2B_y(y - mY)] \quad (4.12)$$

Finalmente, si los intervalos de muestreo X e Y elegidos son los máximos permitidos, en la expresión (4.6), podemos reescribir la ecuación (4.12) y transformarla en

$$g(x, y) = \sum_{n,m=-\infty}^{\infty} g \left(\frac{n}{2B_x}, \frac{m}{2B_y} \right) \cdot \text{sinc} \left[2B_x \left(x - \frac{n}{2B_x} \right) \right] \cdot \text{sinc} \left[2B_y \left(y - \frac{m}{2B_y} \right) \right] \quad (4.13)$$

Esta última expresión representa el resultado fundamental del llamado Teorema del Muestreo de Whittaker – Shannon [2].

4.2- Digitalización del holograma:

La habilidad de una matriz CCD o CMOS para resolver los detalles de una imagen se denomina resolución espacial. Los parámetros principales que determinan la resolución espacial son la cantidad y tamaño de los píxeles.

Normalmente estos dispositivos poseen una matriz rectangular de píxeles fotosensibles (descritos en el apartado 3.3), organizados en N líneas y M columnas. La distancia centro a centro de los píxeles, o medida de pixel, es $\Delta\xi$ y $\Delta\eta$ en las dos direcciones normales de la matriz. Si existe un espacio entre píxeles adyacentes, la dimensión efectiva de pixel es $\alpha\Delta\xi$ y $\beta\Delta\eta$, con $\alpha \leq 1$ y $\beta \leq 1$, como se observa en la Figura 4.3.

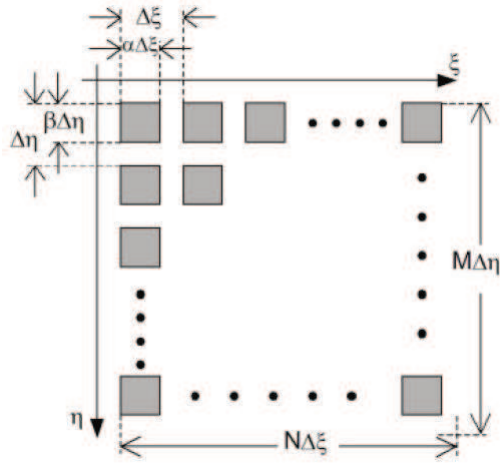


Figura 4.3. Parámetros de una matriz CCD (extraído de Kreis, 2005).

Los términos α y β se denominan factores de llenado (*fill factors*) en las direcciones de ξ y η respectivamente. Para píxeles cuadrados tenemos que $\alpha\Delta\xi = \beta\Delta\eta$, y si también tenemos igual espacio vertical y horizontal entre ellos, resulta $\Delta\xi = \Delta\eta$. Entonces, el factor de llenado al cuadrado es el área del fotodetector, dividida por el área total de la matriz de píxeles.

De acuerdo con el teorema del muestreo, que vimos anteriormente, la señal recibida puede ser reproducida y almacenada fielmente por el dispositivo, si las frecuencias espaciales son superiores a

$$f_{N\xi} = \frac{1}{2\Delta\xi} \quad \text{y} \quad f_{N\eta} = \frac{1}{2\Delta\eta} \quad (4.14)$$

donde $f_{N\xi}$ y $f_{N\eta}$ son las frecuencias de Nyquist en ambas direcciones.

Una herramienta para analizar la capacidad de un dispositivo para grabar y reproducir señales que varían espacialmente (como el caso de las imágenes), es la función de transferencia de modulación (MTF, *modulation transfer function*), que está definida como el cociente entre la modulación de la señal de salida y la modulación de la de entrada, a una frecuencia dada.

$$MTF = \frac{M_o}{M_i} \quad (4.15)$$

Donde la modulación de la señal viene dada por

$$M = \frac{V_{max} - V_{min}}{V_{max} + V_{min}} \quad (4.16)$$

con V_{max} y V_{min} referidos a los valores máximo y mínimo, respectivamente, de la señal. La MTF indica como, variando la frecuencia, el sistema afecta a la señal de salida, esto se puede apreciar gráficamente en la Figura 4.4.

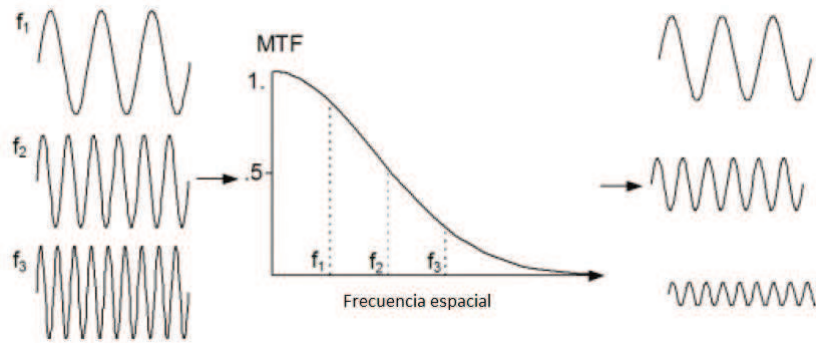


Figura 4.4. MTF aplicado a tres frecuencias (extraído de Kreis, 2005).

La configuración de la matriz de píxeles (tamaño y separación entre ellos) influencia en la modulación de las señales de entrada y salida, como observamos en la Figura 4.5. En 4.5a tenemos un píxel puntual, idealmente pequeño, donde el máximo y el mínimo de las amplitudes de las señales de entrada y salida coinciden, obteniéndose así una $MTF = 1$. En la Figura 4.5b tenemos que los píxeles poseen un ancho igual a la mitad de la distancia de centro a centro entre los mismos. Aquí, las líneas gruesas representan la señal de salida de los píxeles, y podemos ver que la amplitud máxima de salida es menor que la amplitud máxima de entrada, entonces tenemos que la $MTF < 1$. Para un factor de llenado igual a uno, como en la Figura 4.5c, la MTF es aún menor que en el caso anterior. Por todo esto, podemos observar que si aumentamos el ancho del píxel, la MTF disminuye [3].

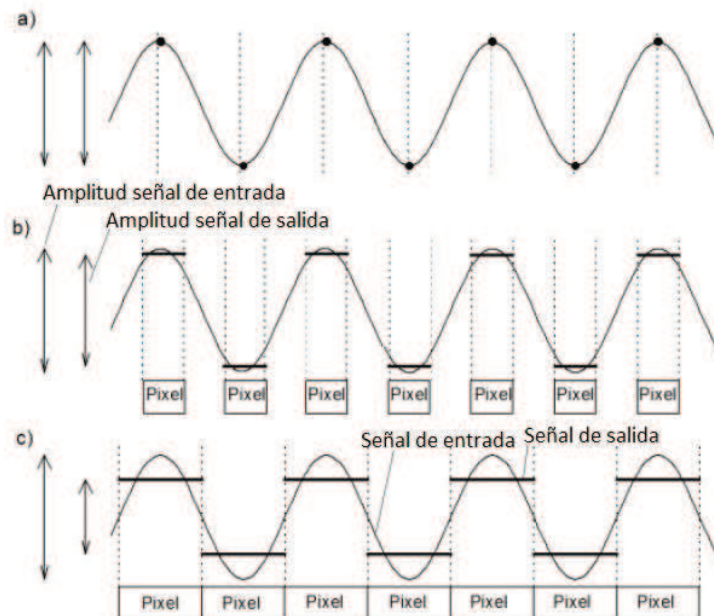


Figura 4.5. Muestreo con diferentes tamaños de píxeles (extraído de Kreis, 2005).

A continuación determinaremos la frecuencia espacial del holograma. Para ello, consideremos una onda de referencia y una onda objeto planas, incidentes sobre el medio de registro, caracterizado este por poseer un espesor determinado.

En la Figura 4.6a, los dos vectores de onda forman cada uno un ángulo de inclinación θ respecto a la normal a la superficie de registro. Con líneas de trazo, en la Figura 4.6a, tenemos los frentes de onda, donde dos frentes sucesivos están separados por una distancia igual a una longitud de onda λ . A lo largo de las rectas de intersección de los dos frentes de onda, las ondas están en fase y sus amplitudes se suman, proporcionando la intensidad máxima. Cuando transcurre un intervalo de tiempo, los frentes de onda se propagan en la dirección de sus vectores de onda, y las rectas de interferencia constructiva se desplazan a través del medio de registro, generando planos de intensidad máxima. Estos planos son bisectores del ángulo 2θ formado por los dos vectores de onda, y aparecen dispuestos periódicamente en todo el medio de registro.

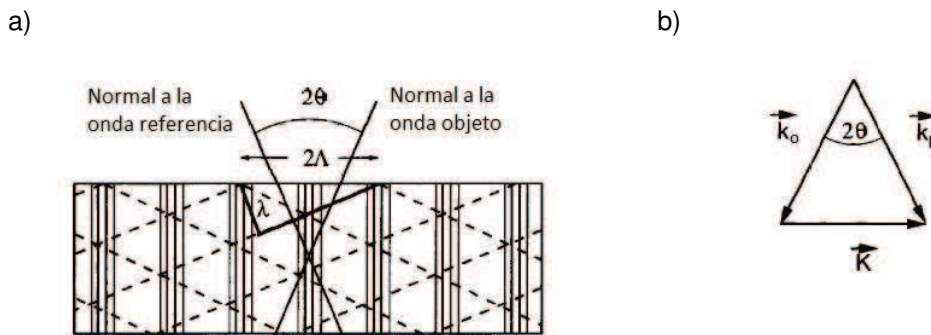


Figura 4.6. a) Registro de un holograma. b) Diagrama de vectores de onda y vector de planos de interferencia (extraído de Goodman, 2008).

Para describir matemáticamente el proceso de interferencia, como vimos en el capítulo 2, las amplitudes complejas de las dos ondas la podemos representar mediante

$$U_R(\vec{r}) = A \cdot \exp(j\vec{k}_R \cdot \vec{r}) \quad \text{y} \quad U_O(\vec{r}) = a \cdot \exp(j\vec{k}_O \cdot \vec{r} - \Phi) \quad (4.17)$$

donde \vec{r} es el vector posición con componentes (x,y,z) , \vec{k}_R y \vec{k}_O los vectores de onda de referencia y objeto respectivamente, y Φ la diferencia de fase entre las ondas.

La distribución de intensidad resultante de la interferencia de estas ondas viene dada, como ya vimos en el capítulo 3, por

$$I(\vec{r}) = |A|^2 + |a|^2 + 2 \cdot |A| \cdot |a| \cdot \cos[(\vec{k}_R - \vec{k}_O) \cdot \vec{r} + \Phi] \quad (4.18)$$

Si consideramos al holograma, generado por esta distribución de intensidad, como una red de difracción, podemos definir un vector red de difracción \vec{K} como la diferencia entre los dos vectores de onda

$$\vec{K} = \vec{k}_R - \vec{k}_O \quad (4.19)$$

El vector \vec{K} , de módulo $2\pi/\Lambda$, donde Λ es el período de las franjas, apunta en la dirección de la diferencia entre los vectores \vec{k}_R y \vec{k}_O , Figura 4.6b. Si observamos las Figuras 4.6a y b, podemos deducir que el período Λ del holograma viene dado por la siguiente expresión [2]

$$\Lambda = \frac{2\pi}{|\vec{K}|} = \frac{\lambda}{2 \cdot \text{sen}\theta} \quad (4.20)$$

Para continuar con nuestro análisis, asumimos que la matriz de pixeles del CCD es cuadrada, con $N = M$ líneas y columnas, y $\Delta\xi = \Delta\eta$ es la longitud del pixel.

Consideremos una onda plana de referencia que incide normalmente en el plano del CCD, como vemos en la Figura 4.7, donde 2θ es el ángulo, en un punto H, entre la onda de referencia y la onda objeto emitida desde un punto P del mismo. Como vimos, Λ es la distancia entre dos franjas de interferencia consecutivas que forman el holograma a partir del punto H.

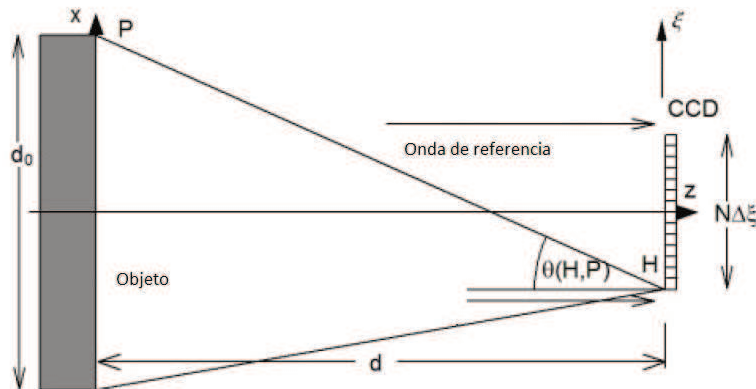


Figura 4.7. Geometría para la adquisición de un holograma digital (extraído de Kreis, 2005).

El teorema del muestreo requiere que el período Λ sea mayor que la distancia de dos pixeles, esto es

$$\Lambda > 2\Delta\xi \quad (4.21)$$

O si lo queremos expresar utilizando la frecuencia espacial de las franjas del holograma, podemos escribir que

$$f < \frac{1}{2\Delta\xi} \quad (4.22)$$

En los casos prácticos el ángulo θ resulta ser muy pequeño, por lo que podemos considerar a $\text{sen}\theta = \text{tag}\theta = \theta$ en los cálculos. Entonces, combinando las expresiones (4.20) y (4.22) obtenemos

$$\theta \leq \frac{\lambda}{4\Delta\xi} \quad (4.23)$$

cuyo límite superior es $\theta_{\max} = \lambda/(4\Delta\xi)$, por lo que si tomamos la frecuencia espacial dada en (4.22), resulta

$$f < \frac{2}{\lambda} \text{sen}(\theta_{\max}) \quad (4.24)$$

siendo esta la inversa del período del holograma, dado en (4.20).

En conclusión, podemos utilizar un CCD para grabar hologramas, siempre y cuando el ángulo entre el haz de referencia y el haz objeto sea pequeño, condición necesaria para cumplir con el teorema del muestreo.

Esta restricción en el ángulo se obtiene si utilizamos objetos de dimensiones laterales pequeñas, o por objetos de dimensiones mayores ubicados lejos del plano del CCD.

Para la configuración típica dada en la Figura 4.7, con un objeto plano de dimensión lateral d_0 en la dirección x , ubicado simétricamente respecto del eje óptico, y con una onda de referencia plana propagándose a lo largo del eje óptico e interceptando al plano del CCD de forma ortogonal, podemos calcular la máxima longitud lateral del objeto, ubicado a una distancia d del CCD, por medio de la Figura 4.7, donde podemos observar que

$$\text{tag}(\theta) = \frac{\frac{d_0 + N\Delta\xi}{2}}{d} \quad (4.25)$$

Si consideramos la restricción en el ángulo impuesta por (4.23), obtenemos

$$\frac{(d_0 + N\Delta\xi)}{2d} < \frac{\lambda}{4\Delta\xi} \quad (4.26)$$

Resolviendo la expresión anterior para d_0 , tenemos que el límite en la extensión lateral del objeto, ubicado a una distancia d , está dada por

$$d_0 < \frac{\lambda d}{2\Delta\xi} - N\Delta\xi \quad (4.27)$$

Análogamente, si tenemos un objeto de dimensión lateral d_0 , resolviendo la expresión (4.26) para d , obtenemos la mínima distancia entre dicho objeto y el CCD, por medio de la siguiente expresión [1]

$$d > \frac{2\Delta\xi}{\lambda} (d_0 + N\Delta\xi) \quad (4.28)$$

4.3- Reconstrucción numérica de la imagen del objeto:

En la holografía digital la reconstrucción es realizada por medio de la multiplicación de la distribución de intensidad grabada digitalmente (o sea, el holograma digital), por un modelo digital de la onda de referencia, obteniéndose numéricamente el campo imagen, a una distancia definida.

4.3.1- Onda de referencia numérica:

La onda de referencia más fácil de procesar numéricamente es la onda plana que incide normalmente en el CCD, Figura 4.8. En el esquema de la Figura 4.8 el objeto está ubicado en el plano (x,y) normal al eje óptico z, en z = 0, y el CCD está ubicado en la posición z = d sobre el plano (ξ,η), paralelo al anterior.

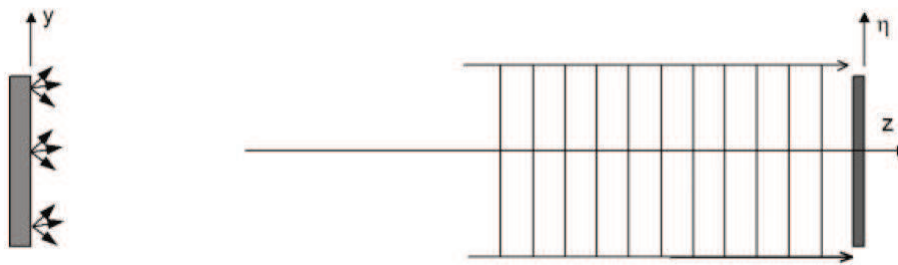


Figura 4.8. Onda de referencia incidiendo normalmente en el plano del CCD (extraído de Kreis, 2005).

La descripción numérica de esta onda plana es

$$A(\xi, \eta) = A_R + j.0 \quad (4.29)$$

donde tenemos una distribución puramente real, con amplitud constante A_R . Entonces, cualquier factor de intensidad constante no tiene relevancia en el procesamiento numérico y puede ser omitido, utilizando solo $A(\xi, \eta) = 1$ como onda de referencia. Una multiplicación del holograma por esta onda de referencia no cambia dicho holograma, por lo tanto la multiplicación puede ser obviada. Por esta razón, la onda de referencia plana que incide normalmente, es la más utilizada en holografía digital.

Por otro lado, una onda plana inclinada con respecto al plano (η,z) un ángulo θ , como vemos en la Figura 4.9, posee un vector de onda dado por

$$\vec{K} = \frac{2\pi}{\lambda}(0, \text{sen}\theta, \text{cos}\theta) \quad (4.30)$$

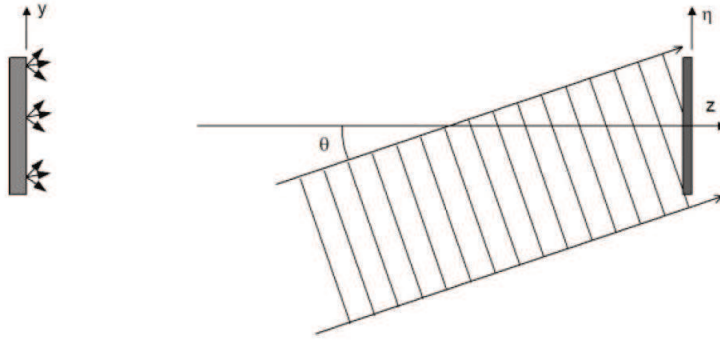


Figura 4.9. Onda de referencia plana con incidencia oblicua (extraído de Kreis, 2005).

En el punto determinado por las coordenadas de $\vec{r} = (\xi, \eta, z)$ en el plano del holograma, tenemos que la onda de referencia es igual a

$$A(\xi, \eta) = A_R \cdot \exp[j(\vec{K} \cdot \vec{r} + \Phi)] = A_R \cdot \exp\left[j\left(\frac{2\pi}{\lambda} \eta \cdot \text{sen}\theta + \frac{2\pi}{\lambda} z \cdot \text{cos}\theta + \Phi\right)\right]$$

$$A(\xi, \eta) = A_R \cdot \exp\left[j\left(\frac{2\pi}{\lambda} z \cdot \text{cos}\theta + \Phi\right)\right] \cdot \exp\left(j\frac{2\pi}{\lambda} \eta \cdot \text{sen}\theta\right) \quad (4.31)$$

Aquí vemos que el primer término exponencial de (4.31) es un factor de fase que permanece constante sobre todo el holograma, por lo que puede ser omitido en el proceso de reconstrucción numérica, entonces reescribimos la expresión (4.31) como

$$A(\xi, \eta) = A_R \cdot \exp\left(j\frac{2\pi}{\lambda} \eta \cdot \text{sen}\theta\right) \quad (4.32)$$

que describe una onda plana de referencia que incide con un ángulo oblicuo sobre el plano del CCD. Sin embargo, debemos observar que un ángulo θ grande provoca una gran densidad de franjas de variación de fase, no pudiendo lograrse la reconstrucción, dado que no se cumple con el teorema del muestreo.

Por último tenemos a la onda de referencia esférica, que diverge desde una fuente puntual ubicada en (x_R, y_R, z_R) , como vemos en la Figura 4.10

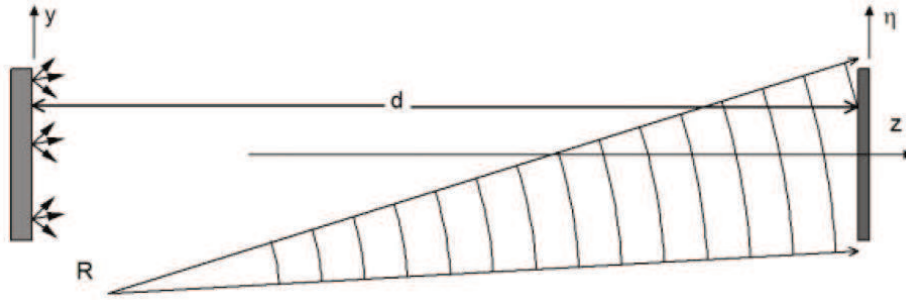


Figura 4.10. Onda de referencia esférica (extraído de Kreis, 2005).

Esta onda podemos describirla mediante la siguiente expresión

$$A(\xi, \eta) = A_R \cdot \exp[j(k \cdot \rho + \Phi)] \quad (4.33)$$

$$\text{donde } \rho = \sqrt{(\xi - x_R)^2 + (\eta - y_R)^2 + (d - z_R)^2} \quad \text{y} \quad k = \frac{2\pi}{\lambda} \quad (4.34)$$

Con una onda esférica podemos realizar el proceso de reconstrucción utilizando una onda de referencia divergente, como la (4.33), o su conjugada convergente, dada por

$$A^*(\xi, \eta) = A_R \cdot \exp[-j(k \cdot \rho + \Phi)] \quad (4.35)$$

Si multiplicamos al holograma digital por la onda de referencia divergente (4.33), el resultado es un campo óptico que diverge desde la imagen virtual en el plano objeto. Multiplicando el holograma por la onda de referencia convergente (4.35), el campo óptico reconstruido converge hacia la imagen real en el plano imagen.

Un caso especial se da cuando $d = z_R$, esto es si la fuente de la onda de referencia está sobre el plano objeto, correspondiendo a la llamada holografía de Fourier sin lentes, descrita en el apartado 3.5.2 del capítulo anterior [1].

4.3.2- Reconstrucción numérica por la transformada de Fresnel:

Consideremos un objeto opaco que es iluminado por una onda de luz coherente, entonces la reflexión de su superficie puede ser descrita por la siguiente función compleja

$$b(x, y, z) = |b_0(x, y, z)| \cdot \exp[j\beta(x, y, z)] \quad (4.36)$$

donde b_0 indica la variación de amplitud y β la variación de fase de la superficie del objeto. La fase $\beta(x, y, z)$ es aleatoria, debido a que la superficie dispersa, en forma difusa, la luz que recibe.

Sin embargo, para nuestro análisis, podemos suponer que un punto de la superficie del objeto emite una onda esférica $b(x,y,z)$. Y para simplificar el desarrollo, también suponemos que el objeto es delgado, así la variable z de $b(x,y,z)$ se hace aproximadamente constante, de este modo, la microestructura rugosa de la superficie está solo descrita por la fase aleatoria β .

Situando al objeto a una distancia d del medio de registro, como vemos en la Figura 4.11, donde el plano del medio de registro, o plano del holograma, tiene las coordenadas (ξ,η)

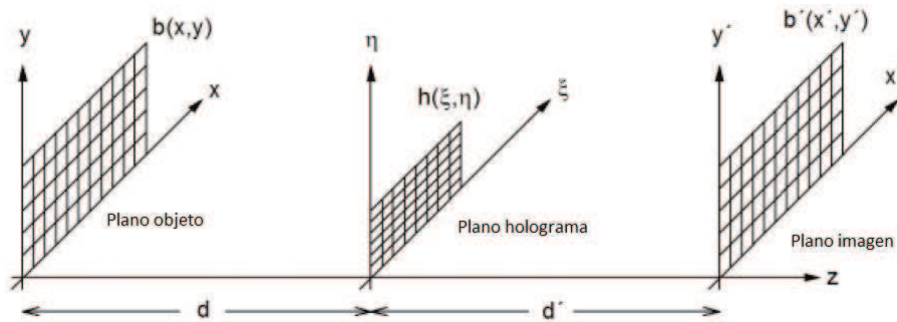


Figura 4.11. Geometría de la holografía digital (extraído de Kreis, 2005).

A una distancia d' del plano del holograma tenemos el plano imagen, o plano de observación, de coordenadas (x',y') , donde la imagen real del objeto puede ser reconstruida.

Si al campo objeto lo definimos con $b(x,y)$, ecuación (4.36), su correspondiente campo en el plano holograma es $B(\xi,\eta)$. Superponiendo este campo con el campo de referencia $A(\xi,\eta)$, la interferencia entre ambos produce una distribución de intensidad $h(\xi,\eta)$, que es grabada en el CCD. Luego, esta distribución real de intensidad $h(\xi,\eta)$ es digitalizada para procesarla numéricamente, y producir una distribución compleja $b'(x',y')$ que representa la imagen reconstruida del objeto.

En la mayoría de las aplicaciones holográficas la distancia entre el objeto y el plano del holograma se encuentra dentro del campo cercano, por lo que podemos aplicar la aproximación de Fresnel a la difracción del campo objeto en el plano holograma, dada por (2.83), obteniéndose

$$B(\xi,\eta) = \frac{\exp(jkd)}{j\lambda d} \exp\left[j\frac{\pi}{\lambda d}(\xi^2 + \eta^2)\right] \iint_{-\infty}^{\infty} \left[b(x,y) \cdot \exp\left[j\frac{\pi}{\lambda d}(x^2 + y^2)\right] \right] \exp\left[-j\frac{2\pi}{\lambda d}(x\xi + y\eta)\right] dx dy \quad (4.37)$$

Siendo las frecuencias espaciales, en el plano del holograma, iguales a

$$f_{\xi} = \frac{\xi}{\lambda d} \quad y \quad f_{\eta} = \frac{\eta}{\lambda d} \quad (4.38)$$

El factor constante $\frac{\exp(jkd)}{j\lambda d}$ no depende de las coordenadas de los planos objeto y holograma, entonces puede ser omitido.

Ahora bien, dada una matriz CCD, tenemos definidas la cantidad de pixeles, $N \times M$, y las distancias entre pixeles $\Delta\xi$ y $\Delta\eta$, por lo que en el plano del holograma tenemos las coordenadas discretas iguales a

$$\xi = n \cdot \Delta\xi \quad y \quad \eta = m \cdot \Delta\eta \quad \text{con } n = 1, 2, \dots, N \quad y \quad m = 1, 2, \dots, M \quad (4.39)$$

Considerando las expresiones dadas en (4.6), si las dimensiones del dominio en el plano de frecuencias es igual a $(N \cdot \Delta f_\xi) \times (M \cdot \Delta f_\eta)$, la separación de los puntos muestreados en el plano objeto resulta igual a

$$\Delta x = \frac{1}{N \cdot \Delta f_\xi} = \frac{\lambda d}{N \cdot \Delta\xi} \quad y \quad \Delta y = \frac{1}{M \cdot \Delta f_\eta} = \frac{\lambda d}{M \cdot \Delta\eta} \quad (4.40)$$

En el plano (ξ, η) el modelo discreto de la onda de referencia $A(n\Delta\xi, m\Delta\eta)$ se superpone a la onda discreta objeto, resultando en la distribución discreta de intensidad $h(n\Delta\xi, m\Delta\eta)$, que es almacenada como un holograma digital

$$h = (B + A)(B + A)^* \quad (4.41)$$

Luego, el proceso de reconstrucción es modelado numéricamente, multiplicando el holograma digital h por la onda de referencia. Entonces, para obtener la imagen real debemos multiplicar el holograma por la conjugada $A^*(\xi, \eta)$ de la onda de referencia.

La imagen real, en el plano (x', y') está determinada por la inversa de la transformada de Fresnel.

Notemos que para obtener una imagen clara necesitamos que $d' = d$, así la transformada de Fresnel en el plano imagen resulta

$$b'(x', y') = \exp\left[j \frac{\pi}{\lambda d} (x'^2 + y'^2)\right] \iint \left[h(\xi, \eta) \cdot A^*(\xi, \eta) \cdot \exp\left[j \frac{\pi}{\lambda d} (\xi^2 + \eta^2)\right] \right] \exp\left[-j \frac{2\pi}{\lambda d} (x'\xi + y'\eta)\right] d\xi d\eta \quad (4.42)$$

Si hacemos las siguientes sustituciones

$$f_{x'} = \frac{x'}{\lambda d} \quad y \quad f_{y'} = \frac{y'}{\lambda d} \quad (4.43)$$

y reemplazamos estas expresiones en la (4.40), obtenemos

$$\Delta f_{x'} = \frac{1}{N \cdot \Delta\xi} = \frac{\Delta x'}{\lambda d} \quad y \quad \Delta f_{y'} = \frac{1}{M \cdot \Delta\eta} = \frac{\Delta y'}{\lambda d} \quad (4.44)$$

donde vemos que el espacio de pixel en la imagen real coincide con el del plano objeto ($\Delta x' = \Delta x$, $\Delta y' = \Delta y$), de esta manera podemos escribir la versión discreta de la expresión (4.42), igual a

$$b'(n\Delta f_{x'}, m\Delta f_{y'}) = \exp[j\pi\lambda d(n^2\Delta f_{x'}^2 + m^2\Delta f_{y'}^2)].$$

$$\cdot \sum_{k=0}^{N-1} \sum_{l=0}^{M-1} h(k\Delta\xi, l\Delta\eta). A^*(k\Delta\xi, l\Delta\eta). \exp\left[j\frac{\pi}{\lambda d}(k^2\Delta\xi^2 + l^2\Delta\eta^2)\right]. \exp\left[-j2\pi\left(\frac{kn}{N} + \frac{lm}{M}\right)\right] \quad (4.45)$$

Podemos expresar la fórmula anterior con los parámetros fundamentales $\Delta\xi$ y $\Delta\eta$, obteniendo lo siguiente

$$b'(n\Delta x', m\Delta y') = \exp\left[j\pi\lambda d\left(\frac{n^2}{N^2\Delta\xi^2} + \frac{m^2}{M^2\Delta\eta^2}\right)\right].$$

$$\cdot \sum_{k=0}^{N-1} \sum_{l=0}^{M-1} h(k\Delta\xi, l\Delta\eta). A^*(k\Delta\xi, l\Delta\eta). \exp\left[j\frac{\pi}{\lambda d}(k^2\Delta\xi^2 + l^2\Delta\eta^2)\right]. \exp\left[-j2\pi\left(\frac{kn}{N} + \frac{lm}{M}\right)\right] \quad (4.46)$$

Siendo esta la fórmula central en la reconstrucción para la holografía digital.

El resultado de $b'(n\Delta x', m\Delta y')$ es la representación numérica de un campo óptico complejo, en el cual tenemos

$$I(n\Delta x', m\Delta y') = |b'(n\Delta x', m\Delta y')|^2 \quad (4.47a)$$

$$\Phi(n\Delta x', m\Delta y') = \text{tag}^{-1} \left[\frac{\text{Im } b'(n\Delta x', m\Delta y')}{\text{Re } b'(n\Delta x', m\Delta y')} \right] \quad (4.47b)$$

De esta manera las distribuciones de intensidad y fase pueden ser determinadas. Esta es la gran ventaja de la reconstrucción digital, comparada con la óptica, dado que en la reconstrucción óptica solo se obtiene una distribución de intensidad.

En la reconstrucción digital tenemos acceso a la distribución de fase de módulo 2π , donde, a primera vista, no notaremos un patrón definido, dado que la fase varía aleatoriamente por la rugosidad de la superficie del objeto. Pero esta distribución de fase se vuelve importante en la interferometría holográfica digital, como veremos mas adelante.

Si observamos el último termino exponencial en (4.46), $\exp\left[-j2\pi\left(\frac{kn}{N} + \frac{lm}{M}\right)\right]$, podemos hacer uso del algoritmo denominado la transformada rápida de Fourier o FFT (*Fast Fourier Transform*) para calcular el campo óptico imagen.

Notemos que con los parámetros de la matriz del CCD, N , M , $\Delta\xi$ y $\Delta\eta$, o sea, la cantidad y tamaño de los pixeles, para el caso de la reconstrucción de Fresnel obtenemos una imagen con el siguiente tamaño

$$N\Delta x \times M\Delta y = \frac{\lambda d}{\Delta\xi} \times \frac{\lambda d}{\Delta\eta} \quad (4.48)$$

Un hecho remarcable es que el tamaño de pixel de la imagen depende de la longitud de onda λ del láser utilizado y de la distancia d de reconstrucción [1].

La reconstrucción numérica, así como la óptica, produce campos convergentes (la imagen real) y campos divergentes (la imagen virtual) simultáneamente. La reconstrucción en el plano imagen, situado en $d = d'$, produce una imagen real clara (en foco), acompañada por una imagen virtual difusa (fuera de foco). Pero se puede reconstruir numéricamente la imagen virtual, podemos analizar como hacerlo de la siguiente forma. Si imaginamos una lente convexa, de distancia focal f , ubicada directamente detrás del holograma, de acuerdo con la ecuación de las lentes, tenemos

$$\frac{1}{f} = \frac{1}{d} + \frac{1}{d'} \quad (4.49)$$

Para un aumento igual a la unidad, obtenemos $d = d'$, entonces $f = d/2$, y la transformación de fase que realiza la lente, de acuerdo con (2.120), es igual a

$$t_l = \exp \left[-j \frac{2\pi}{\lambda d} (\xi^2 + \eta^2) \right] \quad (4.50)$$

Ahora calculamos el campo en el plano (x', y') , considerando la expresión anterior de la lente en el plano holograma, incluida dentro del término del campo incidente en (4.42), resultando

$$h(\xi, \eta) \cdot A^*(\xi, \eta) \cdot \exp \left[-j \frac{2\pi}{\lambda d} (\xi^2 + \eta^2) \right] \quad (4.51)$$

lo que multiplica a la función exponencial cuadrática de la transformada de Fresnel (*chirp function*). Si pasamos todos estos términos a su forma discreta, asumiendo la existencia de la lente, de la expresión (4.46), obtenemos

$$\begin{aligned} h(k\Delta\xi, l\Delta\eta) \cdot A^*(k\Delta\xi, l\Delta\eta) \cdot \exp \left[-j \frac{2\pi}{\lambda d} (k^2\Delta\xi^2 + l^2\Delta\eta^2) \right] \cdot \exp \left[j \frac{\pi}{\lambda d} (k^2\Delta\xi^2 + l^2\Delta\eta^2) \right] = \\ = h(k\Delta\xi, l\Delta\eta) \cdot A^*(k\Delta\xi, l\Delta\eta) \cdot \exp \left[-j \frac{\pi}{\lambda d} (k^2\Delta\xi^2 + l^2\Delta\eta^2) \right] \quad (4.52) \end{aligned}$$

Aquí vemos que solo cambia d por $-d$ en la reconstrucción, dada en (4.46), dentro de la doble sumatoria, entonces la reconstrucción de la imagen virtual puede interpretarse como el cálculo numérico del campo en el plano $z = -d$, que es el plano objeto, de acuerdo con la Figura 4.11.

El signo de la distancia d que precede la doble sumatoria en (4.46) no cambia, dado que no altera la fase, y tampoco tiene influencia sobre la intensidad.

En la Figura 4.12 tenemos un ejemplo de reconstrucción numérica de un dado, donde se han variado las distancias, en a) $d = 105.4$ cm, en b) $d = 79$ cm, en c) $d = 52.7$ cm, en d) $d = -52.7$ cm, en e) $d = -79$ cm y en f) $d = -105.4$ cm [1].

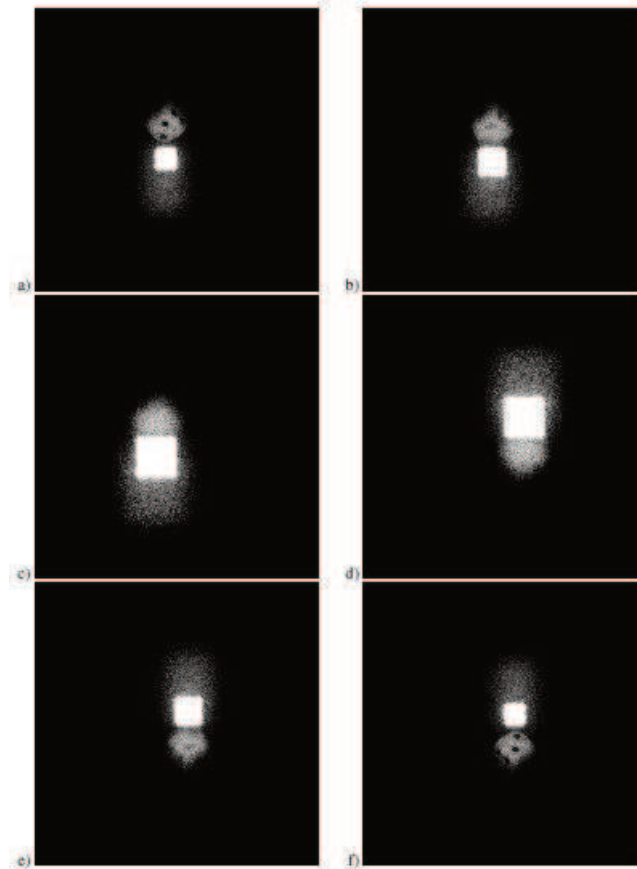


Figura 4.12. Variación de la imagen con la distancia de reconstrucción (extraído de Kreis, 2005).

4.3.3- Eliminación del término de continua:

En las imágenes holográficas de intensidad de la Figura 4.12 observamos un cuadrado central brillante, más brillante que las imágenes, real o virtual, reconstruidas. Físicamente este cuadrado representa el orden cero de la difracción de la onda de referencia. Desde un punto de vista óptico, es la parte no difractada de la onda de referencia en el proceso de reconstrucción. Y desde un punto de vista computacional, este es el término de continua (*D.C.*) de la transformada de Fresnel.

Si no consideramos el factor exponencial que está antes de la doble suma en la expresión (4.46), la transformada de Fresnel es la transformada de Fourier del producto formado por el holograma, la onda de referencia conjugada y la *chirp function*. De acuerdo al teorema de la convolución, esto es equivalente a la convolución de las transformadas de Fourier de las funciones individuales.

La transformada de Fourier del holograma multiplicado por la onda de referencia conjugada, dado por

$$H(n\Delta f_{x'}, m\Delta f_{y'}) = FFT[h(k\Delta\xi, l\Delta\eta) \cdot A^*(k\Delta\xi, l\Delta\eta)] \quad (4.53)$$

posee un pico de gran amplitud en la frecuencia espacial (0,0), este término de continua puede ser calculado por

$$H(0,0) = \sum_{k=0}^{N-1} \sum_{l=0}^{M-1} h(k\Delta\xi, l\Delta\eta) \cdot A^*(k\Delta\xi, l\Delta\eta) \quad (4.54)$$

Pudiendo ser modelada esta expresión por una delta de Dirac.

Entonces, el término de continua de la transformada de Fresnel, es ahora el término de continua de la transformada de Fourier del holograma multiplicado por la onda de referencia conjugada, y esta convolucionada con la transformada de Fourier de la *chirp function*, restringida dentro del plano finito del holograma.

Recordando que la *chirp function* de la expresión (4.46) es igual a

$$\exp\left[j\frac{\pi}{\lambda d}(k^2\Delta\xi^2 + l^2\Delta\eta^2)\right] = \exp\left(j\frac{\pi}{\lambda d}k^2\Delta\xi^2\right) \cdot \exp\left(j\frac{\pi}{\lambda d}l^2\Delta\eta^2\right) \quad (4.55)$$

Podemos definir la función unidimensional finita

$$g(k\Delta\xi) = \exp\left(j\frac{\pi}{\lambda d}k^2\Delta\xi^2\right) \cdot \text{rect}\left(\frac{k\Delta\xi}{N}\right) \quad (4.56)$$

Entonces, la transformada discreta de Fourier de esta *chirp function* finita unidimensional, si nos basamos en la expresión (2.139), resulta

$$G(n\Delta f_{x'}) = \sum_{k=0}^{N-1} \exp\left(j\frac{\pi}{\lambda d}k^2\Delta\xi^2\right) \cdot \text{rect}\left(\frac{k\Delta\xi}{N}\right) \cdot \exp\left(-j2\pi\frac{kn}{N}\right) \quad (4.57a)$$

$$G(n\Delta f_{x'}) = \frac{\exp(-j\pi\frac{k^2}{\beta})}{\sqrt{2\beta}} \left\{ C\left[\sqrt{2\beta}\left(\frac{N}{2} - \frac{k}{\beta}\right)\right] - C\left[\sqrt{2\beta}\left(-\frac{N}{2} - \frac{k}{\beta}\right)\right] + jS\left[\sqrt{2\beta}\left(\frac{N}{2} - \frac{k}{\beta}\right)\right] - jS\left[\sqrt{2\beta}\left(-\frac{N}{2} - \frac{k}{\beta}\right)\right] \right\} \quad (4.57b)$$

donde definimos el término $\beta = \frac{\Delta\xi^2}{\lambda d}$, y como recordamos, C y S son las funciones del coseno y seno de Fresnel.

Vemos entonces, que el término de continua tiene un ancho de $4\pi\frac{N}{2}\beta = 2\pi N\frac{\Delta\xi^2}{\lambda d}$.

Si contamos las frecuencias espaciales $\frac{2\pi n}{N}$ por el número correspondiente de píxeles $n=0,1,\dots,N-1$, en el intervalo $[0,2\pi]$, tenemos que el ancho del término de continua es $\frac{N^2\Delta\xi^2}{\lambda d}$ píxeles, y en dos dimensiones el área del término de continua es igual a

$$\frac{N^2 \Delta \xi^2}{\lambda d} \times \frac{M^2 \Delta \eta^2}{\lambda d} \text{ pixeles al cuadrado} \quad (4.58)$$

Observamos que el ancho del término de continua se incrementa si el ancho del pixel o el número de pixeles aumentan, y disminuye si se aumenta la distancia d . Esto último lo podemos observar en la Figura 4.12. Para el caso límite de una distancia d tendiendo a infinito, el término de continua ocuparía un solo pixel.

Reconstruyendo la imagen a partir de la expresión (4.46), el holograma está definido en el intervalo $[0, N\Delta\xi] \times [0, M\Delta\eta]$, entonces la *chirp function* dada en (4.55) contiene las frecuencias desde 0 a $\frac{N\Delta\xi}{\pi\lambda d}$ en la dirección ξ , y desde 0 a $\frac{M\Delta\eta}{\pi\lambda d}$ en la dirección η , por lo que el cuadrado del término de continua está localizado totalmente en el primer cuadrante, empezando en (0,0). La imagen es normalmente reordenada para que el término de continua aparezca en el centro, cosa que sucede con las imágenes reconstruidas en holografía convencional, como vemos en la Figura 4.12, donde el término de continua fue movido a la zona central de la imagen.

Para mover la *chirp function* finita una distancia dada $k_0\Delta\xi$ y $l_0\Delta\eta$, en ambas direcciones, debemos reescribir la expresión (4.55) de la forma siguiente

$$\exp\left[j\frac{\pi}{\lambda d}(k - k_0)^2 \Delta \xi^2\right] \cdot \exp\left[j\frac{\pi}{\lambda d}(l - l_0)^2 \Delta \eta^2\right] \quad (4.59)$$

que contienen las frecuencias desde $-\frac{k_0\Delta\xi}{\pi\lambda d}$ hasta $\frac{N\Delta\xi - k_0\Delta\xi}{\pi\lambda d}$, y desde $-\frac{l_0\Delta\eta}{\pi\lambda d}$ hasta $\frac{M\Delta\eta - l_0\Delta\eta}{\pi\lambda d}$, en las dos direcciones. Si k_0 se encuentra entre $-N$ y $+N$, y l_0 entre $-M$ y $+M$, el término de continua, en la reconstrucción, está dividido en las cuatro esquinas de la imagen antes de reubicarlo.

Ahora, analicemos la Figura 4.13. La parte real de la *chirp function* sin desplazar, o sea con $k_0 = l_0 = 0$, es mostrada en 4.13a, por lo que la imagen virtual resultante posee un término de continua ubicado totalmente en la esquina superior derecha, antes de reubicarlo, Figura 4.13b. Luego se lo reubica en el tercer cuadrante, como se ve en la Figura 4.13c. La parte real de la *chirp function* desplazada, con $k_0 = N/2$ y $l_0 = M/2$, es mostrada en la Figura 4.13d. Aquí el término de continua es dividido en partes iguales, ubicándose en las cuatro esquinas de la imagen, antes de reubicarlo, Figura 4.13e. Luego se reubica en el centro de la imagen, Figura 4.13f.

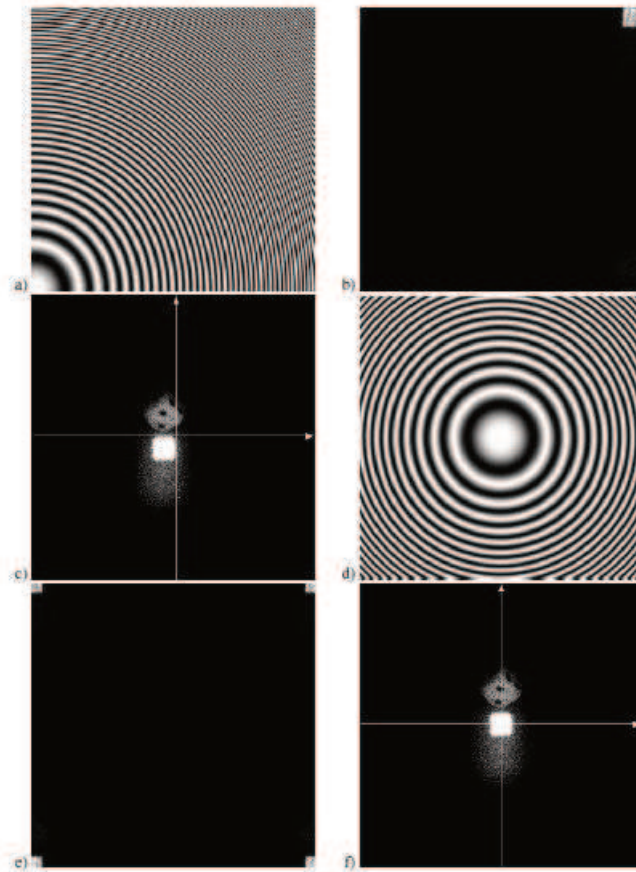


Figura 4.13. Desplazamiento de la *chirp function* y localización de término D.C. (extraído de Kreis, 2005).

El término de continua no tiene un uso práctico, y posee una gran intensidad comparada con la imagen real, lo que degrada la calidad del campo óptico de la imagen reconstruida. En holografía digital existen métodos numéricos efectivos para eliminar el término de continua [4].

Una vez obtenido el holograma digital, y multiplicado este por la onda de referencia conjugada, como en la expresión (4.54), podemos calcular la intensidad promedio

$$h_m(k\Delta\xi, l\Delta\eta) = \frac{1}{N.M} \sum_{k=0}^{N-1} \sum_{l=0}^{M-1} h(k\Delta\xi, l\Delta\eta) \cdot A^*(k\Delta\xi, l\Delta\eta) \quad (4.60)$$

Luego, si sustraemos esta intensidad promedio h_m del holograma multiplicado por la referencia conjugada, obtenemos

$$h'(k\Delta\xi, l\Delta\eta) = h(k\Delta\xi, l\Delta\eta) \cdot A^*(k\Delta\xi, l\Delta\eta) - h_m(k\Delta\xi, l\Delta\eta) \quad (4.61)$$

para $k = 0, 1, \dots, N-1$ y $l = 0, 1, \dots, M-1$

En consecuencia, el espectro de Fourier del término de continua, calculado por (4.54), es cero. La convolución del término de continua (igual a cero) con la transformada de la *chirp function*

finita, también resulta cero. Con el holograma modificado h' podemos obtener las imágenes real y virtual, como describimos anteriormente, pero con el término de continua eliminado.

Si asumimos que la onda de referencia es plana e incide normalmente en el plano holograma, y además la amplitud $A = 1$, entonces solo al holograma h , que únicamente posee intensidades positivas, se le resta la intensidad promedio h_m , por lo que el holograma modificado h' puede poseer intensidades negativas. Esto es ópticamente imposible, pero numéricamente este concepto, de intensidades negativas, está permitido.

Podemos interpretar la eliminación del término de continua como la aplicación de un filtro pasa altos, con una frecuencia de corte muy próxima a cero, donde solo la frecuencia espacial (0,0) es eliminada.

Se pueden obtener buenos resultados con un filtro pasa altos definido por la substracción del promedio de intensidad sobre un cuadrado de 3x3 pixeles en el centro del holograma digital original, resultando la siguiente expresión

$$h'(k, l) = h(k, l) - \frac{1}{3.3} [h(k-1, l-1) + h(k-1, l) + h(k-1, l+1) + h(k, l-1) + h(k, l) + h(k, l+1) + h(k+1, l-1) + h(k+1, l) + h(k+1, l+1)] \quad (4.62)$$

$$\text{para } k = 0, 1, \dots, N-1 \text{ y } l = 0, 1, \dots, M-1$$

donde los factores $\Delta\xi$ y $\Delta\eta$ se omitieron para simplificar la expresión [1].

4.3.4- Eliminación de la imagen gemela:

Si el objeto está ubicado completamente fuera del eje óptico, entonces las imágenes gemelas, en el plano de reconstrucción, no se superpondrán. Así entonces, el espectro de amplitud del holograma digital estará formado por dos espectros parciales ubicados simétricamente. Uno de estos contiene las frecuencias que forman la imagen real, el otro contiene las de la imagen virtual. Lo que deseamos es tener solo la imagen real en el plano de reconstrucción.

Un ejemplo de la eliminación de la imagen gemela lo tenemos en la Figura 4.14

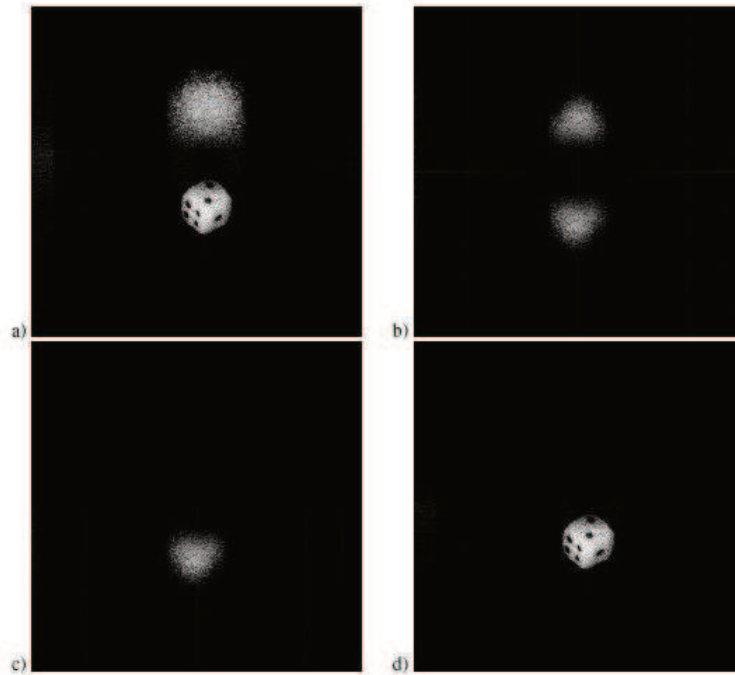


Figura 4.14. Eliminación de la imagen gemela (extraído de Kreis, 2005).

En la Figura 4.14a vemos la imagen reconstruida luego de aplicar el filtro pasa altos, definido en la expresión (4.62), en esta figura podemos observar la imagen gemela difusa. El espectro de amplitud completo, del holograma filtrado, se muestra en la Figura 4.14b, donde observamos que aplicando el filtro pasa altos, antes de la transformación, no tenemos el pico de frecuencia cero en el centro, dado que las bajas frecuencias fueron eliminadas.

Luego, si eliminamos la mitad del espectro, dejamos solo un espectro parcial, como vemos en la Figura 4.14c. Esta operación es realizada en el espectro complejo por una transformada inversa de Fourier del holograma filtrado.

Por último, la reconstrucción de la distribución de intensidad, para este holograma procesado, solo muestra la imagen real, como observamos en la Figura 4.14d [1].

4.3.5- Reconstrucción numérica por la convolución de Huygens - Fresnel:

Como vimos en la sección 2-5, una distribución de campo dada por $U(\xi, \eta) = h(\xi, \eta) \cdot A^*(\xi, \eta)$ en el plano holograma, es difractada, y su distribución de campo en un plano imagen (x', y') puede ser calculada por la integral de convolución (2.86), que en este caso la reescribimos como

$$b(x', y') = \iint [h(\xi, \eta) \cdot A^*(\xi, \eta)] \cdot g[(x' - \xi), (y' - \eta)] d\xi d\eta = [h(\xi, \eta) \cdot A^*(\xi, \eta)] \otimes g(\xi, \eta) \quad (4.63)$$

donde ahora g es la respuesta al impulso, dada por

$$g(x' - \xi, y' - \eta) = \frac{\exp(jkr)}{j\lambda|r|} = \frac{\exp\left(jk\sqrt{(x'-\xi)^2+(y'-\eta)^2+d^2}\right)}{j\lambda\sqrt{(x'-\xi)^2+(y'-\eta)^2+d^2}} \quad (4.64)$$

El teorema de la convolución reduce enormemente el proceso de cálculo numérico, reemplazando la convolución en el dominio espacial, por la multiplicación de los espectros complejos en el dominio de las frecuencias espaciales, seguido por la transformada inversa de Fourier de dicho producto, llevándolo nuevamente al dominio espacial, para obtener la distribución del campo en el plano imagen (x', y')

$$b'(x', y') = \mathcal{F}^{-1}\{\mathcal{F}[h(\xi, \eta) \cdot A^*(\xi, \eta)] \cdot \mathcal{F}[g(\xi, \eta)]\}(x', y') \quad (4.65)$$

Donde utilizamos el algoritmo de la transformada rápida de Fourier, FFT, para obtener las transformadas en el producto, y la transformada inversa.

Para el cálculo numérico definimos la respuesta al impulso en forma discreta, reescribiendo la expresión (4.64) como sigue

$$g(k\Delta\xi, l\Delta\eta) = \frac{\exp\left(j\frac{2\pi}{\lambda}\sqrt{(k-1)^2\Delta\xi^2+(l-1)^2\Delta\eta^2+d^2}\right)}{j\lambda\sqrt{(k-1)^2\Delta\xi^2+(l-1)^2\Delta\eta^2+d^2}} \quad (4.66)$$

Observando que debemos utilizar $k = 1, 2, \dots, N$ y $l = 1, 2, \dots, M$.

Luego, utilizando el algoritmo de la FFT obtenemos la función de transferencia discreta $G(n\Delta f_x, m\Delta f_y)$ [1].

4.3.6- Reconstrucción numérica por la propagación del espectro angular:

Si calculamos la función de transferencia $G(f_x, f_y)$ directamente, reescribiendo la expresión dada en (2.107), obteniendo

$$G(f_x, f_y) = \begin{cases} \exp\left(j\frac{2\pi}{\lambda}d\sqrt{1 - (\lambda f_x)^2 - (\lambda f_y)^2}\right), & \text{para } \sqrt{(f_x)^2 + (f_y)^2} < \frac{1}{\lambda} \\ 0, & \text{para el resto} \end{cases} \quad (4.67)$$

Y a continuación pasamos a discretizar los valores de las frecuencias espaciales, tenemos

$$n\Delta f_x = \frac{(n-1)}{N\Delta\xi} \quad \text{y} \quad m\Delta f_y = \frac{(m-1)}{M\Delta\eta} \quad (4.68)$$

para $n = 1, 2, \dots, N$ y $m = 1, 2, \dots, M$.

Entonces, la función de transferencia finita discreta, viene dada por

$$G(n\Delta f_x, m\Delta f_y) = \exp\left(j\frac{2\pi}{\lambda}d\sqrt{1 - \left(\frac{\lambda(n-1)}{N\Delta\xi}\right)^2 - \left(\frac{\lambda(m-1)}{M\Delta\eta}\right)^2}\right) \quad (4.69a)$$

$$G(n\Delta f_{x'}, m\Delta f_{y'}) = \exp\left(j\frac{2\pi d}{N\Delta\xi}\sqrt{\frac{N^2\Delta\xi^2}{\lambda^2} - (n-1)^2 - (m-1)^2}\right) \quad (4.69b)$$

Donde la expresión (4.69b) es para el caso frecuente donde $N = M$ y $\Delta\xi = \Delta\eta$.

De esta manera la reconstrucción dada por la expresión (4.65) se convierte de la siguiente forma

$$b'(x', y') = \mathcal{F}^{-1}\{\mathcal{F}[h(\xi, \eta) \cdot A^*(\xi, \eta)] \cdot G(f_{x'}, f_{y'})\}(x', y') \quad (4.70)$$

Debemos notar algunas diferencias en los métodos de reconstrucción de Fresnel, de convolución y de propagación del espectro angular. Si consideramos el plano holograma (ξ, η) en el dominio espacial y utilizamos la transformada de Fresnel para la reconstrucción, el resultado es el espectro de Fourier de un producto en el dominio espacial. Este espectro se encuentra en el dominio de las frecuencias espaciales, esto queda de manifiesto por la existencia del término de continua, y por el cambio en la resolución en el plano imagen, dado que en el plano holograma en el dominio espacial tenemos una resolución de $\Delta\xi$, y en el plano de la imagen real tenemos una resolución igual a $\Delta x' = \frac{\lambda d}{N\Delta\xi}$. Por lo que el tamaño de la imagen reconstruida depende de la distancia de reconstrucción d y de la longitud de onda λ utilizada.

Si realizamos la reconstrucción con el método de la convolución, transformamos al holograma multiplicado por la onda de referencia conjugada, desde el dominio espacial hacia el dominio de las frecuencias espaciales, donde luego multiplicamos este espectro por la función de transferencia, por último, aplicando la transformada inversa de Fourier a este producto, obtenemos una distribución en el dominio espacial nuevamente. El resultado es una imagen reconstruida con la misma resolución $\Delta\xi$ del holograma original. Lo mismo podemos observar en el método de propagación del espectro angular, con la ventaja de que en este último método solo necesitamos realizar dos cálculos, la transformada en el plano holograma y la transformada inversa en el plano imagen, para obtener la reconstrucción digital [5].

Pero la ventaja más importante del método de propagación del espectro angular consiste en que podemos reconstruir la imagen incluso para una distancia $d = 0$, o sea en el mismo plano holograma, donde $x' = \xi$ e $y' = \eta$, cosa que es imposible con los otros dos métodos descritos anteriormente. Esta robustez del método de propagación del espectro angular ante la variación de la distancia de reconstrucción lo hace apropiado para poder aplicarlo en el trabajo experimental desarrollado en esta tesis de maestría, como lo detallaremos en el siguiente capítulo.

En la Figura 4.15 vemos un ejemplo de reconstrucción numérica utilizando los tres métodos analizados. El objeto tiene un tamaño de $256 \times 256 \mu\text{m}^2$, y las imágenes fueron reconstruidas a distancias (de izquierda a derecha) de $d = 0 \mu\text{m}$, $d = 125 \mu\text{m}$, $d = 250 \mu\text{m}$, $d = 375 \mu\text{m}$ y $d = 500 \mu\text{m}$. Observamos como en el método de Fresnel la imagen cambia de tamaño según la distancia

de reconstrucción, cosa que no sucede con los otros dos métodos. También como en los métodos de Fresnel y de convolución la imagen no puede resolverse para una distancia $d = 0$, siendo el método de propagación del espectro angular el que obtiene una imagen más nítida para las distancias mas pequeñas.

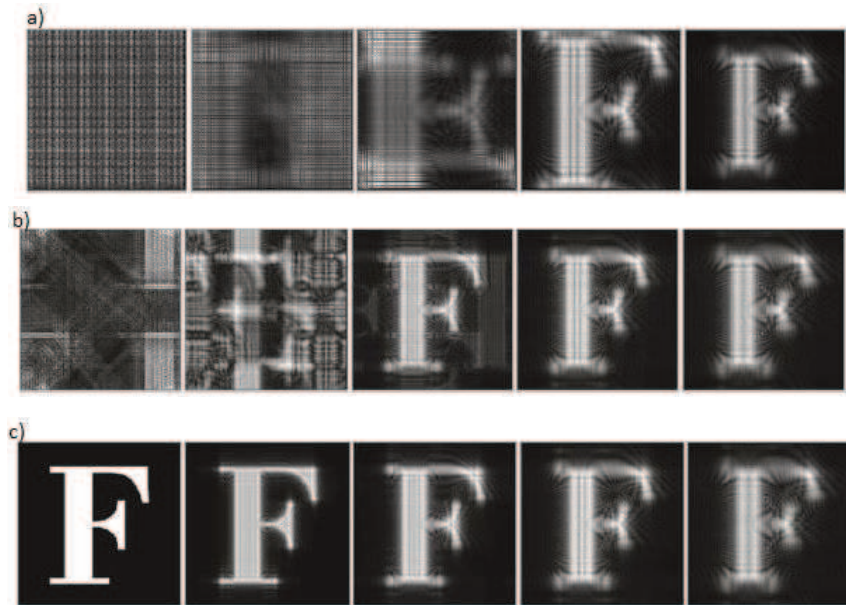


Figura 4.15. Imágenes de distribución de amplitud con los tres métodos de reconstrucción. a) Fresnel, b) convolución y c) propagación del espectro angular (extraído de Kim, 2011).

4.4- Interferometría holográfica digital:

Comenzaremos por considerar el método de la doble exposición, ya descrito en la sección 2.1, pero ahora desde el punto de vista del procesamiento digital. Recordemos que los dos hologramas digitales almacenados representan dos estados, uno inicial y otro deformado, de la superficie del objeto.

Los hologramas digitales se pueden sumar, y esta suma ser reconstruida numéricamente, aplicando cualquiera de los tres métodos descritos anteriormente, esto resulta en un campo complejo de distribución de intensidad, con el correspondiente patrón de interferencia, en concordancia con lo que puede obtenerse con la interferometría holográfica óptica. Luego, un método para evaluar la fase de interferencia, como el descrito en la sección 2-3, puede emplearse para determinar la distribución de la fase de interferencia, o mapa de fase, partiendo del patrón de intensidad. Un ejemplo de esta técnica lo podemos ver en la Figura 4.16, donde se grabaron dos hologramas digitales de una pieza de aluminio, el primero sin deformar y el segundo con la aplicación de una carga en el centro de la pieza. A continuación, se sumaron digitalmente dichos hologramas, se aplicó un filtro pasa altos para eliminar el termino de continua, y se reconstruyó el interferograma, obteniéndose la imagen que muestra la Figura 4.16, que sería

similar a la obtenida por medios ópticos convencionales. Los parámetros fundamentales de este ejemplo son $N = M = 2048$, $\lambda = 632.8 \text{ nm}$ y $\Delta\xi = \Delta\eta = 9 \text{ }\mu\text{m}$.

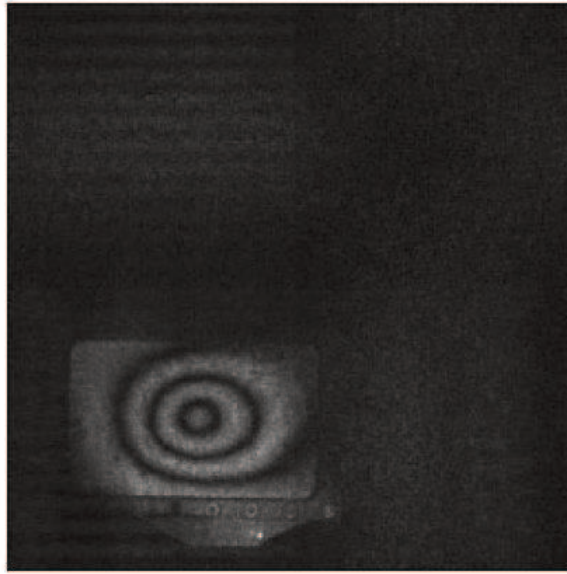


Figura 4.16. Reconstrucción de la intensidad a partir de la suma de hologramas (extraído de Kreis, 2005).

Pero la verdadera ventaja de la interferometría holográfica digital reside en el hecho de que podemos acceder a la distribución de fase de cada uno de los hologramas, como vimos en la expresión (4.47b), sin la necesidad de utilizar técnicas como las descritas anteriormente.

Entonces, si los dos hologramas digitales son reconstruidos individualmente, y la distribución de fase de cada uno de ellos es calculada, la distribución de la fase de interferencia puede ser obtenida mediante la substracción de las fases de módulo 2π de dichos hologramas.

Si consideramos que los campos del objeto, correspondientes a los dos estados del mismo, los reconstruimos individualmente, podemos obtener las siguientes expresiones

$$b'_1(n\Delta x', m\Delta y') = b'_{1R}(n\Delta x', m\Delta y') + j \cdot b'_{1I}(n\Delta x', m\Delta y') \quad (4.71a)$$

$$y \quad b'_2(n\Delta x', m\Delta y') = b'_{2R}(n\Delta x', m\Delta y') + j \cdot b'_{2I}(n\Delta x', m\Delta y') \quad (4.71b)$$

donde los subíndices R e I indican la parte real y la parte imaginaria de los campos complejos reconstruidos en el plano imagen (x', y') .

Luego obtenemos las dos distribuciones de fase mediante la expresión (4.47b), siendo estas igual a

$$\Phi_1(n\Delta x', m\Delta y') = \text{tag}^{-1} \left[\frac{\text{Im}[b'_1(n\Delta x', m\Delta y')]}{\text{Re}[b'_1(n\Delta x', m\Delta y')]} \right] = \text{tag}^{-1} \left[\frac{b'_{1I}(n\Delta x', m\Delta y')}{b'_{1R}(n\Delta x', m\Delta y')} \right] \quad (4.72a)$$

$$y \quad \Phi_2(n\Delta x', m\Delta y') = \text{tag}^{-1} \left[\frac{\text{Im}[b'_2(n\Delta x', m\Delta y')]}{\text{Re}[b'_2(n\Delta x', m\Delta y')]} \right] = \text{tag}^{-1} \left[\frac{b'_{2I}(n\Delta x', m\Delta y')}{b'_{2R}(n\Delta x', m\Delta y')} \right] \quad (4.72b)$$

La fase de interferencia de módulo 2π puede ser calculada ahora por la siguiente expresión

$$\Delta\Phi(n\Delta x', m\Delta y') = \begin{cases} \Phi_2(n\Delta x', m\Delta y') - \Phi_1(n\Delta x', m\Delta y') + 2\pi & , \text{si } (\Phi_2 - \Phi_1) < -\pi \\ \Phi_2(n\Delta x', m\Delta y') - \Phi_1(n\Delta x', m\Delta y') - 2\pi & , \text{si } (\Phi_2 - \Phi_1) \geq +\pi \\ \Phi_2(n\Delta x', m\Delta y') - \Phi_1(n\Delta x', m\Delta y') & , \text{en cualquier otro caso} \end{cases} \quad (4.73)$$

Aquí, en las expresiones dadas en (4.73) asumimos que las fases calculadas en los términos de (4.72), varían dentro del rango de $[-\pi, +\pi]$, por eso es necesario discriminar los tres casos en la diferencia de fase de interferencia.

Ahora veamos el ejemplo de la pieza de la Figura 4.16, pero reconstruyendo los hologramas separadamente, mediante el método de Fourier. Así obtenemos las dos distribuciones de fase $\Phi_1(n\Delta x', m\Delta y')$ en la Figura 4.17a, y $\Phi_2(n\Delta x', m\Delta y')$ en la 4.17b, donde observamos patrones aleatorios, sin poder ver alguna información útil. Pero si substraemos las dos fases obtenemos la diferencia de fase de interferencia de módulo 2π , como resulta en la Figura 4.17c [1].

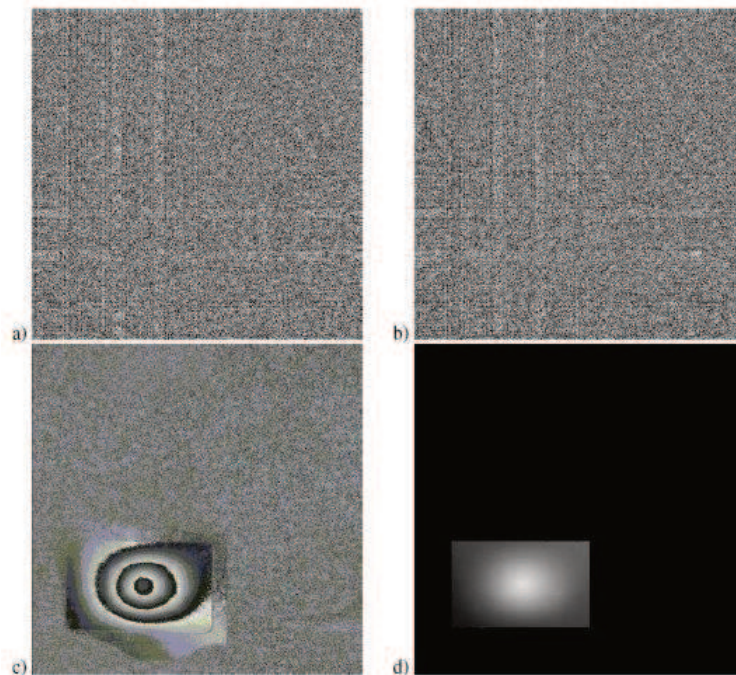


Figura 4.17. Fases reconstruidas a) y b), c) diferencia de fase de módulo 2π y d) diferencia de fase demodulada (extraído de Kreis, 2005).

Si sabemos que la carga está aplicada en el centro de la pieza, es paralela al eje z, y en dirección al plano de registro (z positivo), el desplazamiento de los puntos de la superficie ocurre en esa misma dirección, y el punto central es el de mayor desplazamiento. Aplicando un método de desenvolvimiento, o demodulación de fase, que veremos a continuación, obtenemos el

diagrama de fase desenvuelta, dado en la Figura 4.17d, donde vemos que la fase varía suavemente, desde el mayor desplazamiento en el centro, hacia los bordes de la pieza que se encuentran fijos.

4.4.1- Demodulación de la fase de interferencia:

En la mayoría de las aplicaciones prácticas se desea observar una distribución de fase de interferencia con variaciones suaves y continuas, pero lo que obtenemos en la distribución de la fase de interferencia es una imagen con zonas de discontinuidades o saltos de fase, denominada mapa de fase envuelto de módulo 2π , como observáramos en la Figura 4.17c.

El proceso para resolver las discontinuidades en 2π se llama desenvolvimiento de fase o demodulación, y consiste en sumar una función escalón, formada solo por escalones de altura 2π , a la distribución de fase envuelta.

Este proceso de demodulación asume una distribución de la fase de interferencia continua, que está modulada (o envuelta) dentro de un intervalo de longitud 2π , como observamos en la Figura 4.18a para una dimensión. Por lo tanto, las discontinuidades, como escalones demasiado altos, o agujeros demasiado profundos, o el borde mismo del objeto, deben ser evitados para poder aplicar este proceso. De este modo, se debe definir una región en la superficie de la pieza, donde se pueda ejecutar el proceso de demodulación.

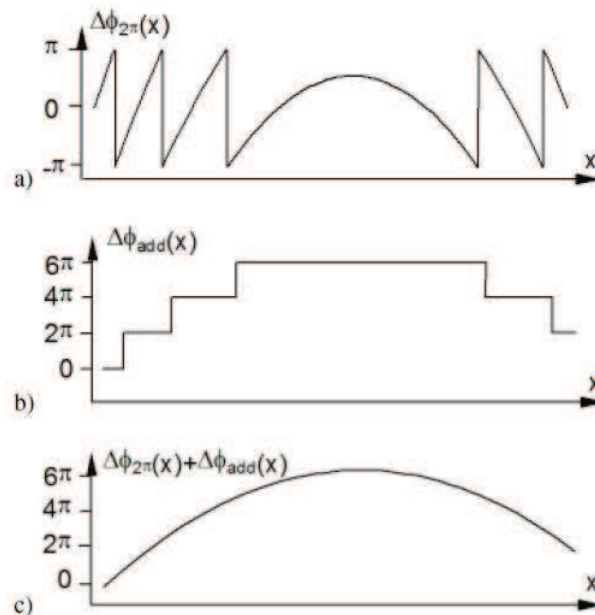


Figura 4.18. a) Fase de interferencia de módulo 2π , b) función escalón y c) distribución de fase de interferencia desenvuelta (extraído de Kreis, 2005).

Veamos cómo podemos demodular una distribución de fase de interferencia unidimensional $\Delta\Phi(x)$, considerando la fase de los píxeles adyacentes, y obteniendo la diferencia de fase

$\Delta\Phi(x+1) - \Delta\Phi(x)$. Si esta diferencia es menor a $-\pi$, una cantidad adicional 2π es sumada a la fase $\Delta\Phi$ desde $x+1$ en adelante, si la diferencia es mayor a $+\pi$, una cantidad 2π es restada de la fase $\Delta\Phi$ comenzando por el pixel $x+1$. De esta manera, al terminar de recorrer el mapa de fase envuelta, se pueden acumular una cantidad de estos términos 2π , como vemos en la Figura 4.18b, donde cada pixel es un múltiplo entero de 2π , y debe ser sumado a la fase de interferencia para que esta sea demodulada, como resulta en la Figura 4.18c.

Una desventaja de este método se evidencia si, por alguna interferencia no deseada (ruido) en el mapa de fase envuelto, obtenemos un error en la diferencia de fase entre pixeles adyacentes, llevando a una suma o resta errónea del término 2π , luego este error es arrastrado a los siguientes pixeles, si no tenemos la suerte de que el error sea anulado por otro error con el signo contrario.

En la práctica, generalmente, el objeto es iluminado por un haz laser que posee una expansión con perfil gaussiano, por lo que en el patrón de interferencia registrado, el contraste en los márgenes de la imagen disminuye notablemente, de manera que en esas zonas se tenga una alta probabilidad de cometer un error en el cálculo del valor de la fase de interferencia, debido a errores en la diferencia de fases de pixeles adyacentes. Por tal motivo es recomendable que el proceso de demodulación comience con los pixeles de la región central del mapa de fase envuelto.

El proceso de demodulación unidimensional descrito puede ser ejecutado en dos dimensiones, procesando por ejemplo la línea superior de pixeles del mapa de fase envuelto, y luego continuar con la línea siguiente, hasta completar toda la matriz imagen de la fase de interferencia [6].

4.5- Topografía superficial con holografía digital:

La idea principal en la topografía holográfica es modular la imagen de un objeto tridimensional mediante franjas que corresponden a niveles de elevación constantes con respecto a un plano de referencia. En esta técnica el objeto no es desplazado, de modo que el cambio de estado que produce el interferograma, generando el patrón de franjas, es provocado por la variación de otros factores, como pueden ser la variación de longitud de onda del laser, la variación del índice de refracción del medio que rodea al objeto, o la variación de la dirección de iluminación del objeto.

A continuación describiremos brevemente cada uno de estos tres métodos [1].

4.5.1- Variación de la longitud de onda:

La longitud del vector sensibilidad depende de la longitud de onda λ , como vemos en la expresión (3.59), entonces un cambio en la longitud de onda varía el vector sensibilidad, y con este también varía la longitud de camino óptico. La longitud de onda se puede variar utilizando,

por ejemplo, un láser de ión argón, que posee saltos discretos predefinidos en un rango limitado de longitudes de onda.

Entonces, si registramos dos hologramas, con diferentes longitudes de onda en cada uno de ellos, las franjas observadas en el interferograma pueden ser interpretadas como la intersección de la superficie del objeto con un conjunto de elipsoides. Esto se esquematiza en la Figura 4.19a, donde se muestra la intersección del objeto con los elipsoides en el plano (x,z). Las líneas de contorno resultante de esta intersección, en el plano (x,y), se esquematizan en la Figura 4.19b.

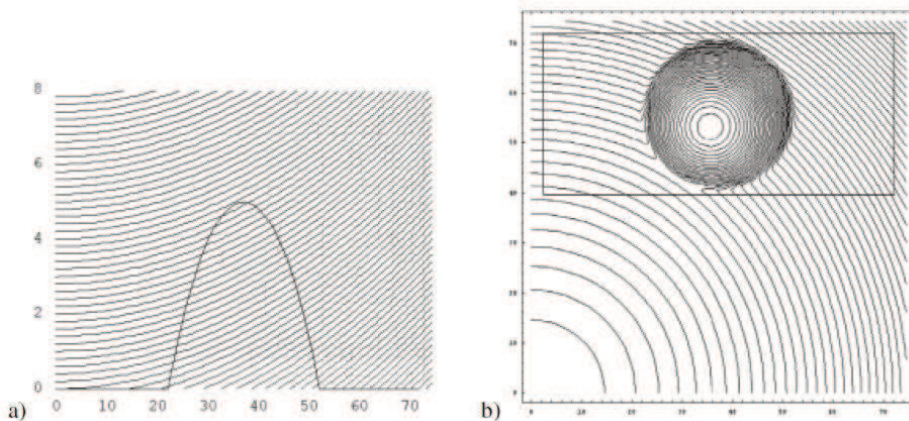


Figura 4.19. Intersección del objeto con un conjunto de elipsoides. a) plano (x,z) y b) plano (x,y) (extraído de Kreis, 2005).

La distribución de la fase de interferencia de módulo 2π , obtenida luego de la reconstrucción y la substracción de las fases de los dos hologramas digitales, es mostrada en la Figura 4.20.

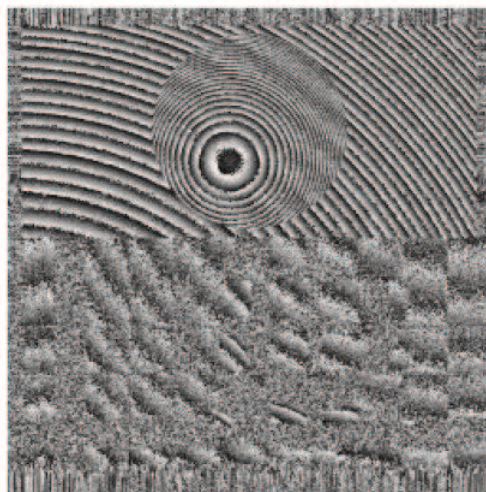


Figura 4.20. Distribución de la fase de interferencia de módulo 2π con el método de las dos longitudes de onda (extraído de Kreis, 2005).

Donde los parámetros utilizados en este ejemplo fueron $N \times M = 1024 \times 1024$ píxeles, distancia centro a centro entre píxeles $\Delta\xi = \Delta\eta = 6,8 \mu\text{m}$. La fuente de iluminación divergente se situó en el punto $S = (0;0;40)$ cm, la distancia desde la parte plana del objeto hasta el CCD es de 80 cm, y las dos longitudes de onda utilizadas fueron $\lambda_1 = 632,8 \text{ nm}$ y $\lambda_2 = 633,8 \text{ nm}$. La reconstrucción se realizó por el método de Fresnel, con la eliminación del término de continua [1].

4.5.2- Variación del índice de refracción:

Este método es también denominado método de inmersión. Aquí la longitud de onda es variada por el cambio en la velocidad de la luz, dado que $\lambda = n \cdot c/v$, donde n es el índice de refracción del medio transparente que rodea al objeto.

Para realizar esta técnica, el objeto es colocado en un recipiente cerrado de vidrio, que contiene un gas o un líquido transparente, con un índice de refracción n_1 . Acto seguido, se captura el primer holograma. Luego el medio es reemplazado por otro, de índice de refracción n_2 , para obtener así el segundo holograma.

Un ejemplo de este método utiliza un líquido cuyo índice de refracción varía de $n_1 = 1,330$ a $n_2 = 1,331$. El objeto se encuentra separado de la pared frontal de vidrio una distancia de 20 mm. Las curvas de fase constante resultantes en el espacio, y sus intersecciones con la superficie del objeto, en el plano (x,z) , están esquematizadas en la Figura 4.21a, y las líneas de contorno, en el plano (x,y) , están esquematizadas en la Figura 4.21b.

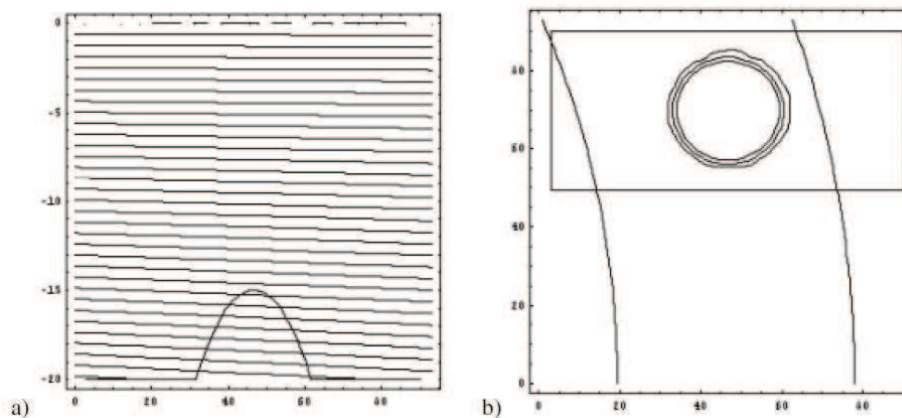


Figura 4.21. Intersección del objeto con líneas de contorno, método de inmersión. a) plano (x,z) y b) plano (x,y) (extraído de Kreis, 2005).

La Figura 4.22 muestra la distribución de fase de interferencia de módulo 2π , resultante del interferograma, donde la reconstrucción de los hologramas digitales se realizó con el método de Fresnel [1].

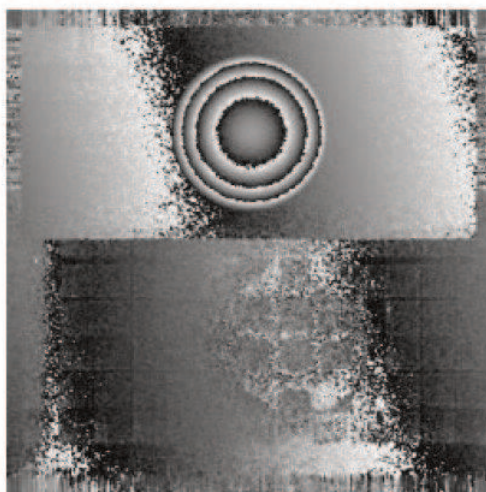


Figura 4.22. Distribución de la fase de interferencia de módulo 2π con el método de inmersión (extraído de Kreis, 2005).

4.5.3- Variación de la dirección de iluminación:

De la expresión (3.59) podemos notar que el vector sensibilidad puede variarse también alterando la dirección de iluminación o de observación. Si el punto de iluminación S_1 , de la grabación del primer holograma, es cambiado al punto S_2 en la grabación del segundo holograma, la longitud del camino óptico cambia, por lo que también cambia la fase de interferencia, de acuerdo con la expresión (3.50).

En esta técnica el objeto es interceptado por un conjunto de hiperboloides, con focos en los dos puntos de iluminación, formando franjas en la superficie de dicho objeto. Este patrón de franjas es independiente del punto de observación.

Un ejemplo de esta técnica lo podemos ver en las Figuras 4.23 y 4.24, donde observamos un esquema de la intersección de los hiperboloides con el objeto en el plano (x,z) , en la Figura 4.23a, y las líneas de contorno del objeto, en el plano (x,y) , en la Figura 4.23b.

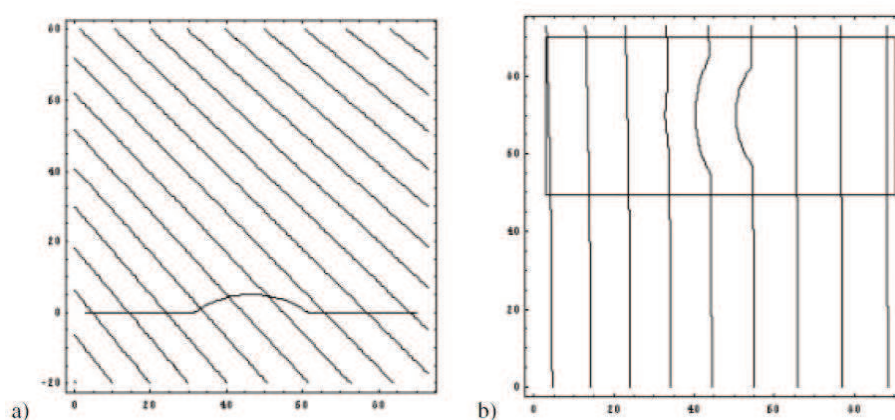


Figura 4.23. Intersección del objeto con hiperboloides, método de los dos puntos de iluminación. a) plano (x,z) y b) plano (x,y) (extraído de Kreis, 2005).

La Figura 4.24 muestra la fase de interferencia de módulo 2π del interferograma, resultante de la reconstrucción de los dos hologramas digitales.

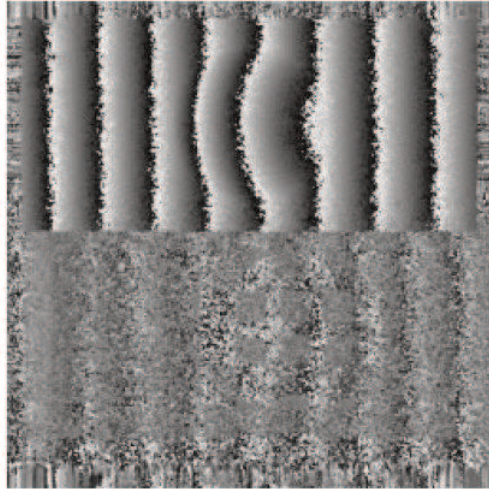


Figura 4.24. Distribución de la fase de interferencia de módulo 2π con el método de los dos puntos de iluminación (extraído de Kreis, 2005).

En este ejemplo el punto de iluminación fue cambiado de posición de $S_1 = (-300;0;400)$ mm a $S_2 = (-299,97;0;400,03)$ mm.

Observamos que la iluminación desde una posición oblicua es mejor, porque si las ondas de iluminación tienen su vector de onda ortogonal a la superficie del objeto, también los hiperboloides intersectarán a la superficie de forma ortogonal, entonces la variación de altura de la parte central del objeto de este ejemplo podría no ser detectada, ya que no tendríamos una variación lateral de las franjas de interferencia.

También podemos remarcar, que la variación del punto de iluminación determina la densidad de las franjas de interferencia observadas [1].

Referencias:

- [1] T Kreis (2005). Handbook of Holographic Interferometry, Wiley – VCH ed.
- [2] J Goodman (2008). Introduction to Fourier Optics, 2th edition, McGraw – Hill ed.
- [3] G Holst (1996). CCD Arrays, Cameras and Displays. Optical Engineering, SPIE.
- [4] T Kreis, W Juptner (1997). Suppression of the d.c. term in digital holography, Optical Engineering.
- [5] M Kim (2011). Digital Holographic Microscopy, Springer N.Y.
- [6] D Bone (1991). Fourier fringe analysis: the two-dimensional phase unwrapping problem, Applied Optics.

Capítulo V

Microscopía holográfica digital: trabajo experimental, procesamientos y resultados.

En este capítulo trataremos con una aplicación específica de la holografía digital, utilizada como una técnica para realizar observaciones microscópicas, dicha técnica es denominada microscopía holográfica digital o DHM (*Digital Holographic Microscopy*), y es utilizada para medir espesores de películas delgadas, que es el objetivo propuesto en nuestro trabajo de tesis.

5.1- Esquema óptico para la DHM:

Un esquema típico para realizar DHM consiste en una fuente de iluminación, un interferómetro con lentes objetivo de microscopio óptico, una cámara digital y una computadora con el software de procesamiento [1].

La fuente de iluminación utilizada es el láser, como ya vimos en la sección 3.3.1. En nuestro trabajo utilizamos un láser de He-Ne, marca Melles-Griot, con una potencia de 30 mW y una longitud de onda $\lambda = 632,8\text{ nm}$.

Los tipos de interferómetros empleados en microscopía se describieron en la sección 1.3.1, de estos los más utilizados en DHM son el Mach-Zehnder y el Michelson. El primero de ellos es utilizado para objetos transparentes, que transmiten la luz hacia la cámara digital. El segundo es más apropiado para objetos opacos, aunque es posible observar objetos transparentes (o semitransparentes) colocando al objeto sobre una superficie especular, con lo que logramos una configuración de doble paso de la luz a través del objeto. Los interferómetros pueden incluir varios elementos ópticos, como pueden ser *pinholes*, atenuadores y polarizadores para ajustar la relación de intensidades entre el haz de referencia y el haz objeto.

En nuestro trabajo adoptamos una configuración de interferómetro de Michelson, con un cubo divisor de haz no polarizador, un filtro óptico formado por un *pinhole* de $10\text{ }\mu\text{m}$ de diámetro y lentes colimadoras, aptas para un rango de longitudes de onda desde 450 nm hasta 650 nm , con lo que se consigue una fuente de iluminación que emite ondas planas. Este esquema puede observarse en la Figura 5.1.

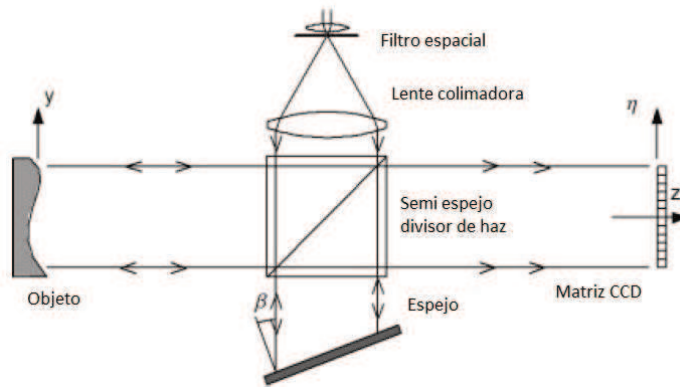


Figura 5.1. Esquema de interferómetro de Michelson para DHM (extraído de Kreis, 2005).

La ventaja principal del esquema implementado en este trabajo es que no se necesita de lentes objetivo para obtener el holograma, como sí ocurre con los dispositivos para DHM construidos comercialmente, esto resulta en una configuración más económica y versátil, con la que podríamos conseguir los mismos resultados que con estos equipos estandarizados.

La cámara digital, con una configuración CCD o CMOS, descrita en la sección 3.3.3 es la encargada de capturar y digitalizar el holograma. Como ya vimos, la cantidad y tamaño de los pixeles son los parámetros principales que limitan la resolución del dispositivo para realizar DHM, y condiciona el rango de aplicación de la misma. Es de esperar que, con el avance de la tecnología, disminuya el tamaño de los pixeles y por lo tanto se incremente la cantidad de los mismos en la matriz de fotosensores, mejorando aún más la resolución en la DHM.

La cámara digital utilizada en el presente trabajo consiste en una matriz CMOS de 4096×2160 pixeles (aproximadamente 8,9 megapixeles), con un tamaño de pixel de $1,55 \mu m$, marca FlirSystems®, modelo BlackFly Flea3.

Por último, el software de procesamiento empleado es Matlab®, y las rutinas de procesamiento fueron desarrolladas ad-hoc.

5.2- Microscopía de fase cuantitativa:

En microscopía holográfica la imagen de fase, formada por la luz transmitida a través de un objeto transparente, o la reflejada por un objeto opaco, puede convertirse en información cuantitativa relacionada con el espesor, el índice de refracción o la topografía superficial de dicho objeto, como se mencionó en la sección 1.2.

Por otra parte, podemos decir que la holografía digital es una herramienta muy efectiva para lograr una microscopía de fase cuantitativa de alta precisión. Dado que la imagen de fase se obtiene directamente procesando el holograma digital, sólo es necesario obtener un holograma en esta técnica de DHM, a diferencia del método de los dos hologramas en interferometría holográfica digital visto en la sección 4.4 que requieren de dos o más hologramas para poder comparar distintos estados del objeto de estudio y así determinar variaciones de fase en dichos estados.

Además, a diferencia de los dispositivos para microscopía óptica, en DHM no se necesitan elementos mecánicos de focalización.

Sin embargo, la característica más importante de la DHM es que la imagen de fase permite obtener una representación de la morfología superficial del objeto en el caso de un objeto opaco y reflectante, con una precisión nanométrica, como se verifica en este trabajo de tesis.

Como vimos en la sección 3.2.2, la fase de interferencia $\Delta\Phi$ está relacionada con la diferencia de camino óptico δ punto a punto sobre la superficie del objeto con la que interactúa el haz objeto; y, dado que el haz objeto recorre dos veces estos caminos, como se aprecia en la Figura 5.2, tenemos que este espesor Δz es igual a la mitad de la diferencia de camino óptico, o $\delta = 2\Delta z$.

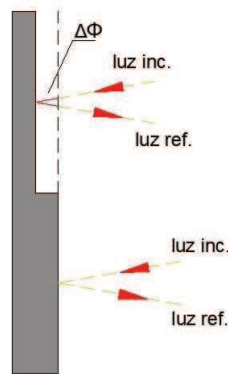


Figura 5.2. Variación de fase por diferencia de camino óptico que provoca el escalón.

Entonces, si reescribimos la ecuación (3.50) obtenemos la distribución de diferencias de altura del objeto, en el plano holograma, dada por

$$\Delta z(\xi, \eta) = \frac{\lambda}{4\pi} \Delta\Phi(\xi, \eta) \quad (5.1)$$

Pero, como vimos en la sección 3.2.3, los valores de la fase están restringidos al intervalo $[-\pi, +\pi]$, entonces observamos en la expresión (5.1) que diferencias abruptas de altura mayores a media longitud de onda, $\lambda/2$, quedan indeterminadas, lo cual impone ciertas restricciones para que dichas variaciones puedan ser resueltas con técnicas de desenvolvimiento de fase de módulo 2π (sección 4.4.1) [2].

La microscopía de fase cuantitativa o QPM (*Quantitative Phase Microscopy*) es tal vez la aplicación más importante de la DHM, ya que se pueden obtener mapas bidimensionales de una resolución nanométrica en profundidad con la captura de un solo holograma, sin tener las dificultades, como mencionamos, que poseen las técnicas de doble exposición para obtener la topografía superficial vistas en la sección 4.5 [1].

Es necesario advertir que para obtener imágenes correctamente enfocadas (ver Figuras 4.12 y 4.15), debemos medir el valor de la distancia de reconstrucción con una precisión del orden del

milímetro, como veremos más adelante. Además, tenemos que considerar las aberraciones, siempre presentes, propias del sistema óptico que degradan la calidad de la imagen de fase, las cuales pueden ser corregidas digitalmente.

Por último, como mencionamos en la sección 4.3.6, se eligió para la reconstrucción el método de propagación del espectro angular, por su robustez frente a la variación de la distancia de reconstrucción, y porque el ángulo β (ver Figura 5.1) entre el haz de referencia y el haz objeto, necesario para separar el término de continua, puede ser incluso más pequeño que el requerido para aplicar el método de Fresnel [3].

Podemos calcular este ángulo $\beta = \theta_{max}$ de la expresión (4.23), para que cumpla con el teorema del muestreo, como vimos en la sección 4.2

$$\theta_{max} = \beta = \frac{\lambda}{4\Delta\xi} = \frac{632,8 \text{ nm}}{4 \cdot 1,55 \mu\text{m}} \cong 5,85^\circ \quad (5.2)$$

Resultando en una frecuencia espacial, cuyo límite máximo viene dado por la expresión (4.24), como podemos ver en el siguiente valor hallado

$$f_{max} = \frac{2}{\lambda} \text{sen}(\theta_{max}) = \frac{2 \cdot \text{sen}(5,85^\circ)}{632,8 \text{ nm}} \cong 3,22 \times 10^5 \frac{1}{\text{m}} \quad (5.3)$$

Por lo tanto, el período mínimo del holograma que puede ser resuelto, es el valor inverso del resultado dado en (5.3)

$$\Lambda_{min} = \frac{1}{f_{max}} \cong 3 \mu\text{m} \quad (5.4)$$

Que corresponde aproximadamente a la longitud de dos pixeles del sensor utilizado, como es requerido por el teorema del muestreo.

5.3- DHM de una muestra de silicio cristalino:

El primer objeto analizado es una pieza rectangular de silicio cristalino (c-Si) de 2 cm de ancho por 1,7 cm de alto, que podemos apreciar en la Figura 5.3, posee una superficie especular, que la hace apta para calibrar el montaje óptico del interferómetro.



Figura 5.3. Muestra de silicio cristalino de 2 cm de ancho por 1,7 cm de alto.

Dado que el haz que llega al objeto prácticamente no es alterado por la superficie del mismo, tenemos en el plano holograma dos ondas planas, la de referencia y la objeto, interfiriendo y formando el patrón de franjas predicho en la sección 2.1.1.

Con las dimensiones del objeto podemos estimar la distancia de ubicación mínima de este con respecto al plano holograma, para que la matriz CMOS capture, por ejemplo, toda la longitud vertical de la muestra. Utilizando la expresión (4.28) obtenemos, con $d_0 = 17 \text{ mm}$

$$d_{min} = \frac{2\Delta\xi}{\lambda} (d_0 + N\Delta\xi) = \frac{2.1,55 \times 10^{-6} \text{ m}}{632,8 \times 10^{-9} \text{ m}} (17 \times 10^{-3} \text{ m} + 2160.1,55 \times 10^{-6} \text{ m}) \cong 10 \text{ cm} \quad (5.5)$$

En la Figura 5.4 tenemos, como ejemplo, dos hologramas de la muestra de c-Si, capturados a una distancia del objeto al plano CMOS de 31 cm , donde podemos apreciar la diferencia entre los períodos de cada imagen holográfica. Esto es provocado al variar la inclinación del espejo del interferómetro entre una captura y la otra, con lo que alteramos el valor del ángulo β .

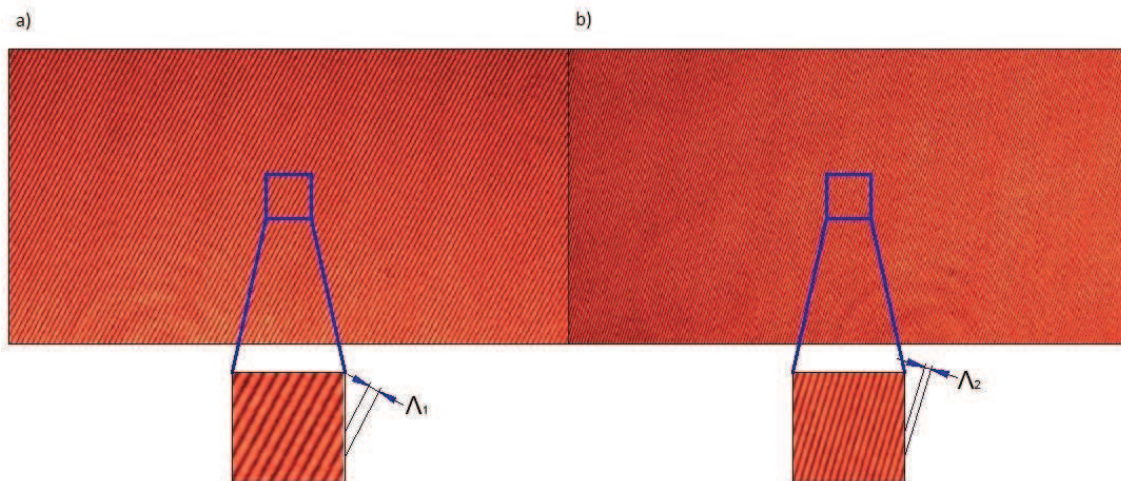


Figura 5.4. Hologramas de la muestra de c-Si con diferentes períodos. Tamaño de las imágenes 4096×2160 pixeles².

El período del holograma de la Figura 5.4a es igual a

$$\Lambda_1 = 22 \text{ pixeles} \cdot 1,55 \mu\text{m}/\text{pixel} \cong 34 \mu\text{m} \quad (5.6)$$

Y el período del holograma de la Figura 5.4b es igual a

$$\Lambda_2 = 14 \text{ pixeles} \cdot 1,55 \mu\text{m}/\text{pixel} \cong 22 \mu\text{m} \quad (5.7)$$

Luego, el ángulo β puede ser aproximado por una expresión derivada de la (5.3), siendo igual a

$$\beta \cong \frac{\lambda}{2.A} \quad (5.8)$$

Entonces tenemos que el ángulo entre el haz de referencia y el haz objeto, para estos dos hologramas, resulta

$$\beta_1 = \frac{\lambda}{2.A_1} = \frac{632,8 \times 10^{-9} \text{ m}}{2.34 \times 10^{-6} \text{ m}} \cong 0,53^\circ \quad \text{y} \quad \beta_2 = \frac{\lambda}{2.A_2} = \frac{632,8 \times 10^{-9} \text{ m}}{2.22 \times 10^{-6} \text{ m}} \cong 0,82^\circ \quad (5.9)$$

Como vemos ambos ángulos se encuentran muy lejos del valor impuesto para θ_{max} en la expresión (5.2). Llegar a obtener un período de franjas igual a dos píxeles es muy dificultoso en la práctica, debido a vibraciones en el montaje óptico, como así también a la definición y al ruido propio del dispositivo CMOS.

Una vez obtenido el holograma lo procesamos con un algoritmo escrito en Matlab®, que en este trabajo llamamos `reconstruccion.m` (se puede ver completo en el apéndice), cuyo empleo pasamos a describir a continuación.

Se lee la imagen en formato bmp correspondiente al holograma a procesar

```
filename = 'holograma1b.bmp';
holo_ini = imread(filename);
```

Convertimos la imagen color del holograma a escala de grises

```
if size(holo_ini,3) == 3    (si es a color)
    holo_ini = rgb2gray(imread(filename));
else
holo_ini = imread(filename);
end
```

Mostramos los hologramas de la Figura 5.4 en escala de grises, como vemos en las Figuras 5.5a y 5.5b

```
figure(1)
imshow(holo_ini);
```

Aplicamos la FFT para obtener el espectro de frecuencias espaciales y lo mostramos

```
holo_iniffft = fftshift(fft2(fftshift(holo_ini)));
figure(2)
imagesc(log(abs(holo_iniffft)));
```

donde el comando `fftshift` traslada el punto $(0,0)$ del plano de frecuencias espaciales al centro de la imagen del espectro. Podemos observar en las Figuras 5.5c y 5.5d que a menor período mayor

es la separación entre máximos de frecuencia espacial, lo que facilita el filtrado del espectro, como veremos más adelante.

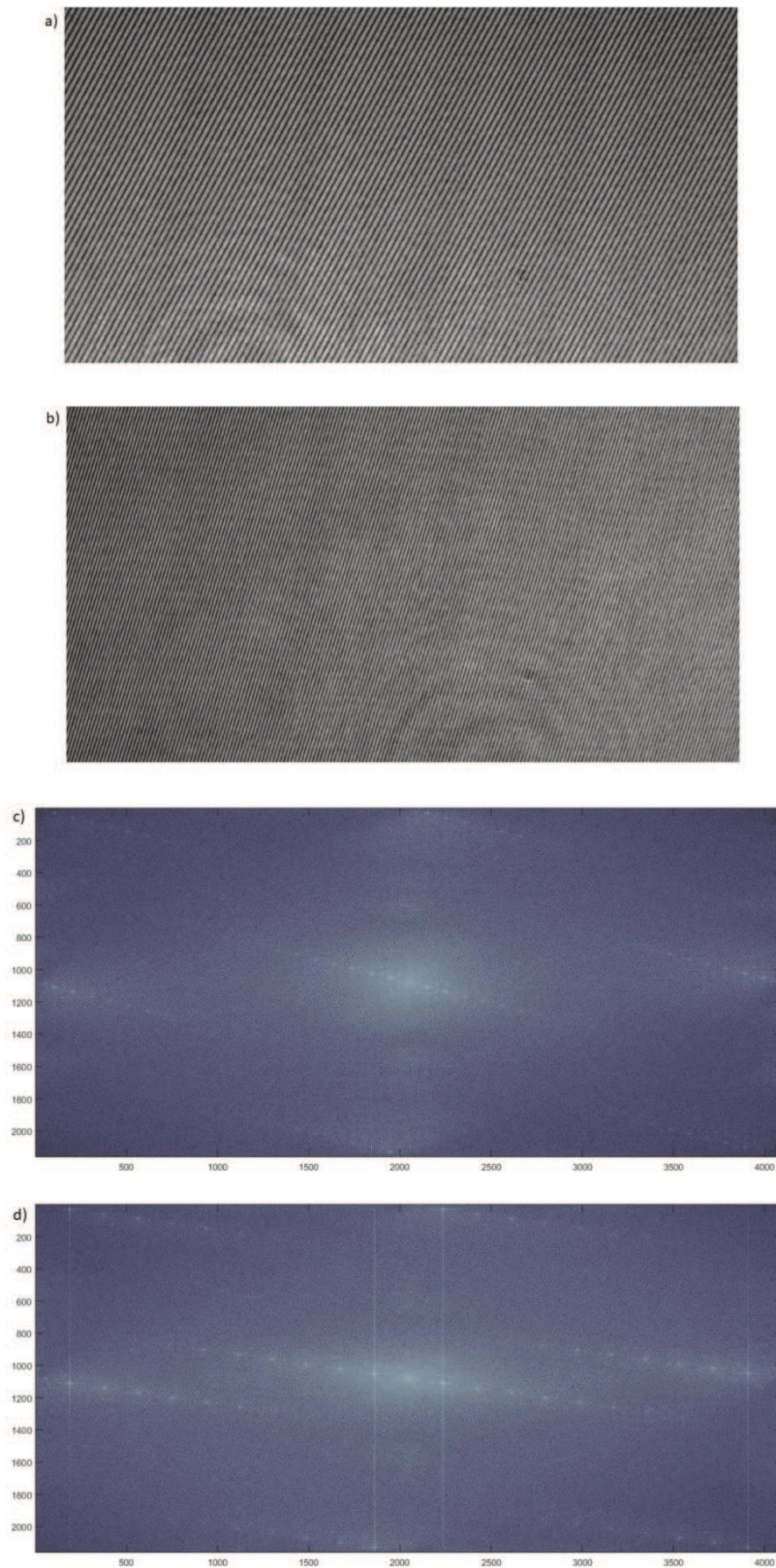


Figura 5.5. a) y b) Hologramas en escala de grises. c) y d) Espectros de frecuencia.

A continuación se lee el tamaño de la matriz que compone el holograma

```
[nfil,ncol] = size(holo_ini);
```

Definiéndose en este caso los valores $nfil = 2160$ y $ncol = 4096$.

El paso siguiente es implementar el filtro para eliminar el término de continua

```
filtlevel = 50;
holo_filtDC = double(holo_ini) - conv2(double(holo_ini), ...
ones(filtlevel), 'same')/filtlevel^2;
```

donde el comando *filtlevel* define el valor de la frecuencia de corte del filtro pasa altos, a mayor valor de este es menor la frecuencia de corte.

El comando *ones(filtlevel)* es una matriz compuesta enteramente de unos, de tamaño *filtlevel* x *filtlevel*, en este caso de 50×50 .

El comando *conv2* realiza la convolución, en dos dimensiones, entre *double(holo_ini)* y *ones(filtlevel)*, con la extensión 'same' que provoca que el resultado de la convolución tenga el mismo tamaño que *double(holo_ini)*, una matriz de 2160×4096 .

Podemos mostrar como resulta el holograma, filtrado el término de continua, y su espectro de frecuencias espaciales, mediante la siguiente línea de comandos

```
figure(3)
imshow(holo_filtDC);
holo_fft = fftshift(fft2(fftshift(holo_filtDC)));
figure(4)
imagesc(log(abs(holo_fft)));
```

En la Figura 5.6a vemos el holograma 1b filtrado y su espectro en la figura 5.6b, donde notamos que en el centro del mismo se ha quitado el término de continua.

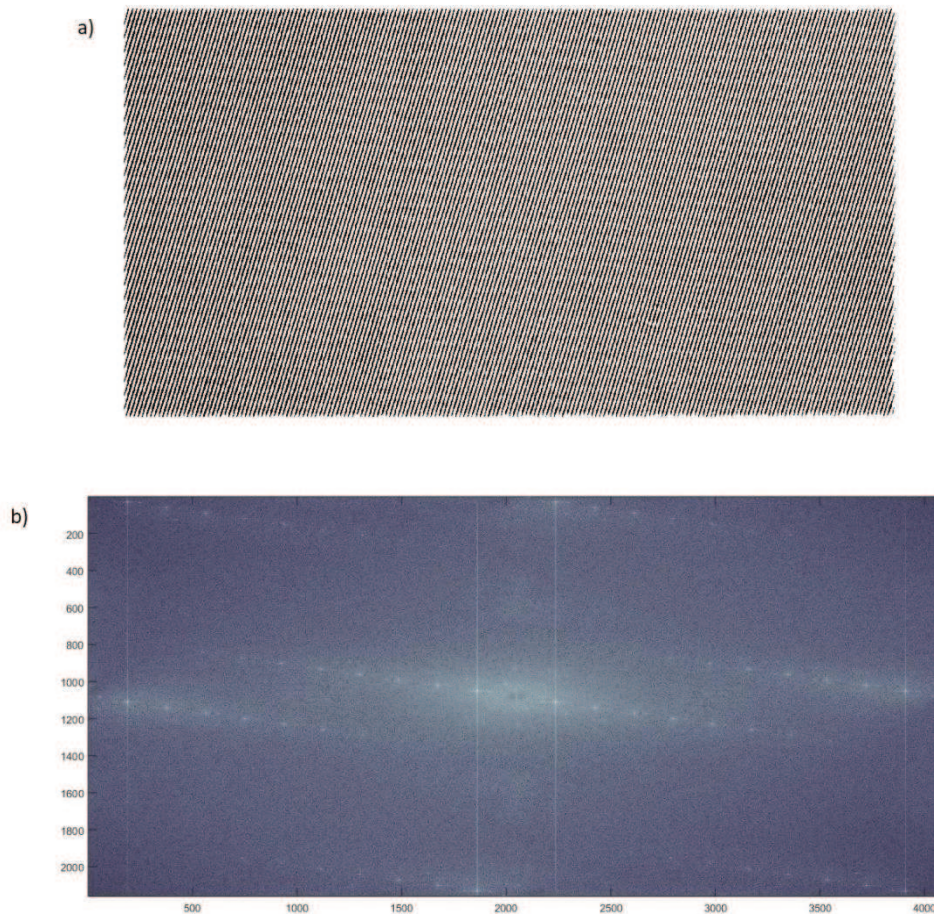


Figura 5.6. a) Holograma filtrado. b) Espectro de frecuencia.

El paso siguiente es encontrar la ubicación de la frecuencia espacial donde se produce el primer pico máximo, desde el origen de coordenadas, denominada frecuencia portadora. Para ello, utilizamos los siguientes comandos

```
[filMax,colMax] = find(abs(holo_fft) == max(max(abs(holo_fft))));
```

Donde se define un rango de valores máximos del espectro, mediante el comando *find* que encuentra elementos de la matriz que no valgan cero, y deben coincidir con los valores máximos de la matriz *holo_fft*. Resultando en un rango de valores para la columna de la matriz igual a $[colMax] = [1862,2236]$, para este holograma 1b seleccionado, o sea, que dentro de esos valores de la coordenada *x* de los pixeles se encuentra el pico máximo del espectro, la frecuencia portadora.

Seguidamente, definimos un círculo con centro en el origen del plano del espectro, que contenga a esta frecuencia portadora, cuyo radio se encuentra definido por

```
r = colMax(2)-ncol/2;
```

en nuestro ejemplo el valor del radio resulta igual a $r = 2236 - 4096/2 = 188$ en pixeles.

El paso siguiente es utilizar un filtro pasa bajos, para eliminar las altas frecuencias y solo quedarnos con las frecuencias que rodean a la frecuencia portadora de interés. Esto se realiza con el algoritmo que llamamos *creamascara.m*, que puede verse en el apéndice.

Este algoritmo es llamado por el algoritmo que estamos corriendo, *reconstruccion.m*, y trabaja en el plano de frecuencias, creando un círculo (o máscara) centrado en los valores medios de los vectores *filMax* y *colMax*, con el radio *r* que definimos anteriormente. La línea de comando de este filtro es la siguiente

```
creamascara(holo_fft, filMax(2), colMax(2), r)
```

Este sub algoritmo crea una ventana donde se observa el espectro de frecuencias, y sobre este, el círculo definido anteriormente, como podemos apreciar en la Figura 5.7. En la ejecución de esta rutina podemos seleccionar manualmente el radio de la máscara, que aparece con el valor definido previamente (en el ejemplo $r = 188$), por medio del cursor, ajustándolo a nuestra elección; una vez conformes con el valor del radio terminamos la ejecución de este sub algoritmo.

Como mencionamos anteriormente, que los máximos de frecuencia estén más separados posibilita una mayor área abarcada por la máscara, ya que tenemos menos probabilidades de incluir algún otro máximo secundario, dado que con mayor cantidad de frecuencias seleccionadas mejor reconstruiremos la imagen.

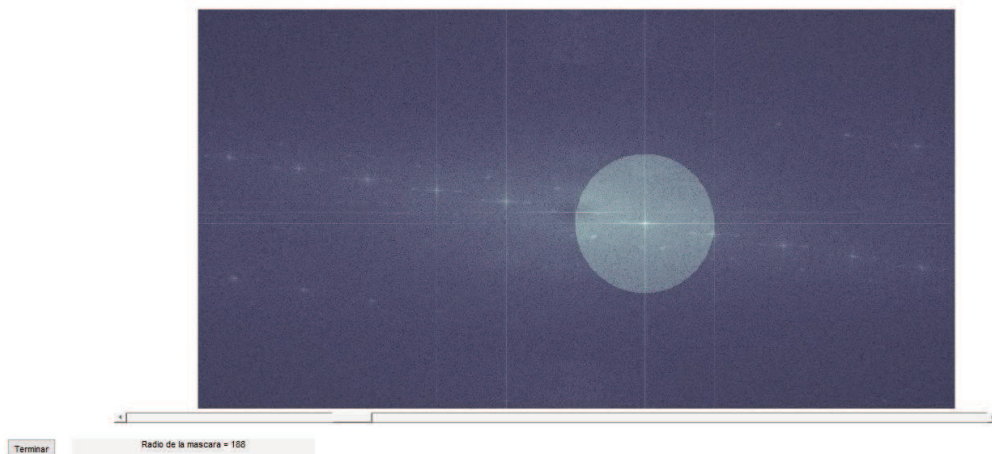


Figura 5.7. Ventana del sub algoritmo *creamascara.m*.

Una vez definido este filtro pasa bajos, llamado *maskfinal*, se lo multiplica por el espectro de frecuencias espaciales del holograma

```
holo_fft_mask = holo_fft.*maskfinal;
```

Así obtenemos el espectro de frecuencias filtrado que observamos en la Figura 5.8a, mostrado a partir de las siguientes líneas de comando

```
figure(5)
imagesc(log(abs(holo_fft_mask)));
axis image
axis off
```

Como vemos, se encuentra desplazado a una distancia de r pixeles del origen.

Para obtener el espectro centrado en el origen del plano de frecuencias, necesitamos desplazar la frecuencia portadora hacia dicho punto, esto se logra con los comandos siguientes

```
rowsToShift = round(nfil/2 - filMax(2));
columnsToShift = round(ncol/2 - colMax(2));
```

Obteniéndose el espectro filtrado y desplazado, que observamos en la Figura 5.8b, mediante la siguiente sentencia

```
holo_fft_mask_shift = circshift(holo_fft_mask,...
[rowsToShift columnsToShift]);
```

Luego puede mostrarse este último espectro con las siguientes líneas

```
figure(6)
imagesc(log(abs(holo_fft_mask_shift)));
```

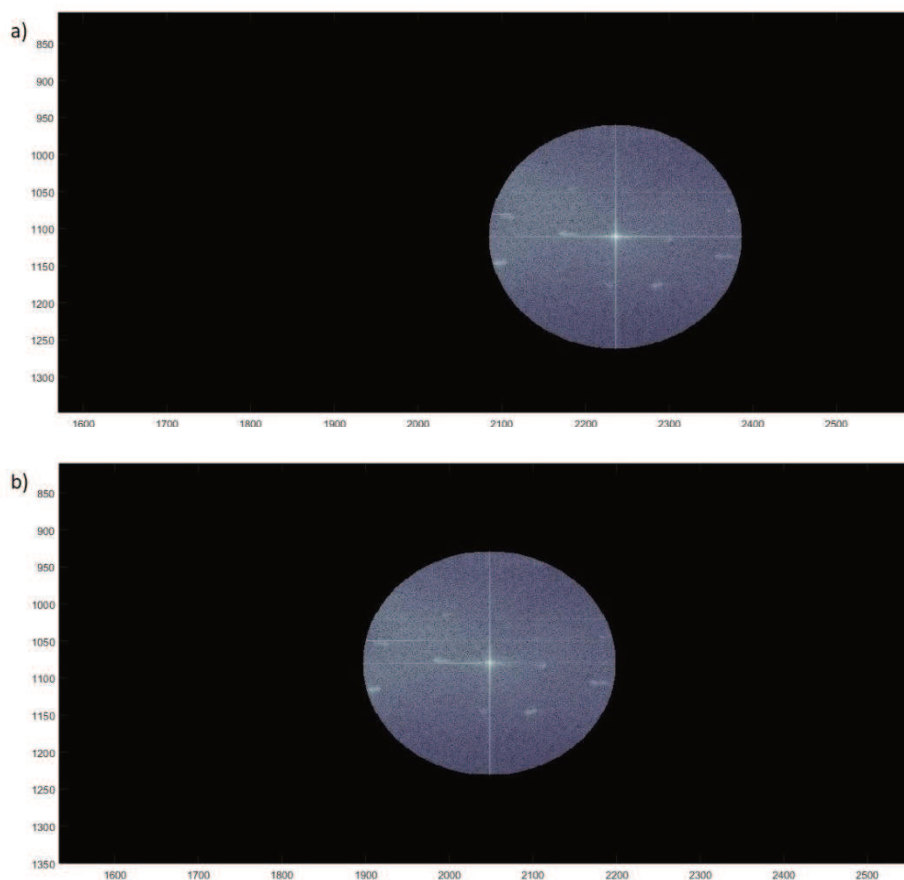


Figura 5.8. Espectro de frecuencia filtrado. a) Original. b) Centrado.

Una vez obtenido y procesado el espectro de frecuencias espaciales del holograma, es necesario definir los parámetros de la propagación desde el plano holograma hacia el plano imagen, definiendo los siguientes valores, a saber:

El tamaño horizontal y vertical del pixel, en metros

$$\begin{aligned} \text{pixh} &= 1.55\text{e-}6; \\ \text{pixv} &= 1.55\text{e-}6; \end{aligned}$$

La longitud de onda λ del láser utilizado, en metros

$$\text{lambd} = 632.8\text{e-}9;$$

El número de onda, en m^{-1} , definido por

$$k = 2*\text{pi}/\text{lambd};$$

El rango de coordenadas en metros, en el plano (x,y) , donde existe la imagen, de acuerdo a la cantidad de pixeles de la matriz CMOS

$$\begin{aligned} \text{ximg} &= \text{linspace}(-\text{ncol}/2, \text{ncol}/2, \text{ncol}) * \text{pixh}; \\ \text{yimg} &= \text{linspace}(-\text{nfil}/2, \text{nfil}/2, \text{nfil}) * \text{pixv}; \end{aligned}$$

Utilizando este rango de coordenadas definimos una grilla en el plano imagen (x,y) , dada por el siguiente comando

$$[\text{X}, \text{Y}] = \text{meshgrid}(\text{ximg}, \text{yimg});$$

A continuación definimos lo que se denomina paso de frecuencia, que es la frecuencia espacial dividida por la longitud del pixel, cuya expresión, en m^{-2} , la podemos obtener partiendo de las ecuaciones (4.40)

$$\frac{\Delta f_x}{\Delta \xi} = \frac{1}{N \cdot \Delta \xi} \frac{1}{\Delta \xi} = \frac{1}{N \cdot \Delta \xi^2} \quad \text{y} \quad \frac{\Delta f_y}{\Delta \eta} = \frac{1}{M \cdot \Delta \eta} \frac{1}{\Delta \eta} = \frac{1}{M \cdot \Delta \eta^2} \quad (5.10)$$

Entonces los pasos de frecuencia, en las direcciones x e y respectivamente, se obtienen gracias a las siguientes expresiones

$$\begin{aligned} \text{fn} &= 1 / (\text{ncol} * \text{pixh}^2); \\ \text{fm} &= 1 / (\text{nfil} * \text{pixv}^2); \end{aligned}$$

Como ya mencionamos, el método de reconstrucción de la imagen que utilizamos es el de propagación del espectro angular, por lo que primeramente debemos obtener la variación de la fase. De la expresión (4.67) esta fase Φ es igual a

$$\Phi(f_x, f_y) = \frac{2\pi}{\lambda} \sqrt{1 - (\lambda f_x)^2 - (\lambda f_y)^2} \quad (5.11)$$

De este modo, el cálculo de la variación de la fase en la propagación hacia el plano imagen, se encuentra determinado por

$$fi = k * \text{sqrt}(1 - (\text{lambda} * fn * X) .^2 - (\text{lambda} * fm * Y) .^2);$$

Luego necesitamos saber la distancia de propagación, o reconstrucción, que obtenemos estimando, con una regla metálica milimetrada, la separación entre la muestra y la cámara CMOS. En este caso ubicamos la muestra a *31 cm* de la cámara, y a este dato lo volcamos al algoritmo mediante la siguiente variable, dada en metros

$$dcent = 31e-2;$$

Ahora estamos en condiciones de obtener la función de transferencia finita discreta de la expresión (4.69a), reescrita aquí de la siguiente manera

$$G(f_x, f_y) = \exp\left(j \frac{2\pi}{\lambda} d \sqrt{1 - (\lambda f_x)^2 - (\lambda f_y)^2}\right) \quad (5.12)$$

Que trasladada al algoritmo se transforma en la siguiente línea

$$G = \exp(1i * fi * dcent);$$

Entonces la reconstrucción en el plano imagen podemos realizarla a través de la transformada inversa de Fourier, como en la expresión (4.70), considerando a la onda plana de referencia conjugada $A^* = 1$, por lo que la reescribimos de la siguiente forma

$$b(x, y) = \mathcal{F}^{-1}\{\mathcal{F}[h(\xi, \eta)] \cdot G(f_x, f_y)\}(x, y) \quad (5.13)$$

Cuya ejecución en el algoritmo la realizamos con la FFT⁻¹ centrada, en dos dimensiones, de esta manera

$$\text{img_rec} = \text{ifftshift}(\text{ifft2}(\text{ifftshift}(\text{holo_fft_mask_shift}.*G)));$$

Resultando *img_rec* una función de dos dimensiones compleja, donde pueden extraerse la distribución de intensidad, o módulo, y la distribución de la fase de esta, por medio de los siguientes comandos

$$\begin{aligned} \text{mod} &= \text{abs}(\text{img_rec}); \\ \text{fase} &= \text{angle}(\text{img_rec}); \end{aligned}$$

Por último se obtienen las imágenes del módulo y la fase del objeto a través de las siguientes líneas de comando

```

figure(7)
imagesc(imadjust(mod/max(max(mod))))
axis image
axis off
title(['Reconstruccion del modulo @ ',num2str(dcent*100),' cm']);
colormap gray

figure(8)
imagesc(fase)
axis image
axis off
title(['Reconstruccion de la fase @ ',num2str(dcent*100),' cm']);
colormap gray

```

En la Figura 5.9 observamos las imágenes del módulo y la fase del objeto. En la Figura 5.9a estas se obtuvieron sin eliminar el término de continua y sin el filtro pasa bajos, en la 5.9b eliminando el término de continua pero sin el filtro pasa bajos, y en la 5.9c utilizando ambos filtros.

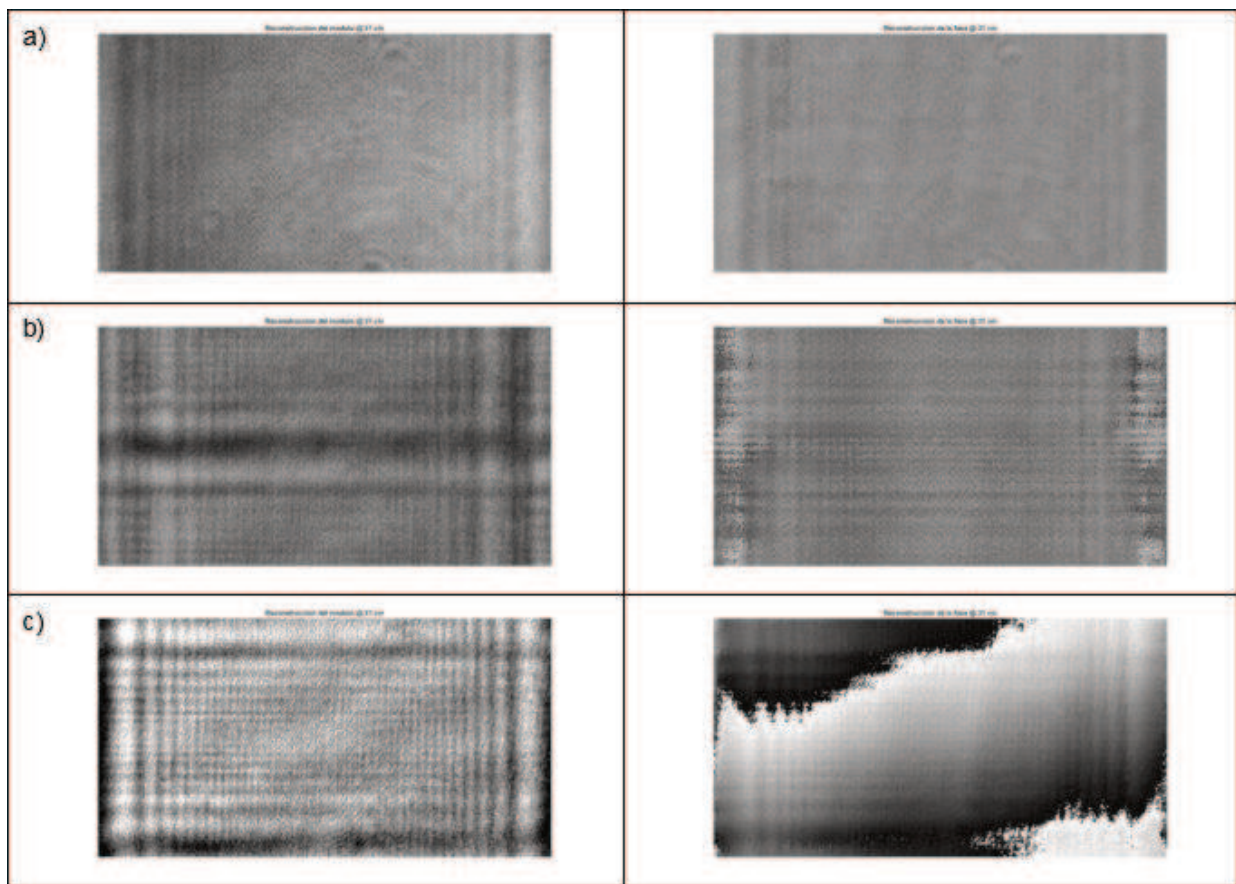


Figura 5.9. Imágenes de módulo y fase del objeto. a) Sin filtro, b) con filtro de DC y c) con filtros de DC y pasa bajos.

Podemos notar, en la Figura 5.9a, que si no utilizamos ningún filtro, el módulo se ve velado por el término de continua, siendo este la parte no difractada de la onda de referencia, por lo tanto aún seguimos observando el patrón de interferencia de ambas ondas, no lográndose distinguir al objeto. En cuanto a la fase, solo logramos obtener la fase de interferencia de módulo 2π , caracterizada por las franjas, que muestran la inclinación del espejo con respecto al objeto, ya que ambos frentes de onda son planos. Salvo por el valor de las intensidades relativas, no podemos distinguir mayor diferencia entre las imágenes de módulo y fase en este caso sin filtrar.

En la Figura 5.9b podemos distinguir borrosamente al objeto en la imagen del módulo, pero seguimos observando las franjas de interferencia, sobre todo en las zonas de altas frecuencias espaciales, que degradan mucho la imagen. Notamos también las líneas verticales y horizontales provocadas por la difracción de borde, que vimos en la sección 2.2.9. En la imagen de fase vemos que esta presenta zonas claras y oscuras, con diferencias de fase entre ellas de π radianes, en los extremos izquierdo y derecho, aunque aún persisten las franjas de la fase de interferencia de módulo 2π , sobre todo en las zonas de altas frecuencias; además se observa una zona central muy velada por las franjas de la fase de interferencia, lo que también deteriora mucho la imagen.

Por último, en la Figura 5.9c, tenemos la imagen del módulo, donde se puede apreciar al objeto prácticamente plano, con las líneas de difracción de bordes, y sin ninguna franja de interferencia. Aunque esta imagen holográfica del objeto podría decirse que es de menor calidad que una imagen óptica, la gran ventaja de este método es la obtención de la imagen de fase, como mencionamos en la sección 5.2. En dicha imagen de fase podemos observar que esta posee una distribución bastante uniforme, como es de esperar en un objeto plano, salvo en las zonas superior izquierda e inferior derecha, con una diferencia de fase de π radianes entre ambas. Estas diferencias en la fase son provocadas por las aberraciones del esquema óptico utilizado, debido a la inclinación del objeto con respecto al eje óptico, la imprecisión en el valor de la distancia de reconstrucción y las curvaturas de los frentes de onda, que distan de ser frentes de onda planos perfectos.

5.4- DHM de una grilla de plata depositada sobre silicio cristalino:

El segundo objeto analizado es una pieza rectangular formada por una película de plata, depositada en forma de grilla, sobre silicio cristalino, como podemos apreciar en la Figura 5.10. Esta pieza posee un tamaño de $3,8\text{ cm}$ de ancho por $1,6\text{ cm}$ de alto, cada cuadrado de plata mide alrededor de $1,5\text{ mm}$ de longitud. La misma fue utilizada en un trabajo previo [3] donde se obtuvo una medida aproximada del espesor del depósito de plata.



Figura 5.10. Muestra de plata sobre silicio cristalino de 3,8 cm de ancho por 1,6 cm de alto.

El objetivo que nos propusimos, utilizando esta muestra, fue lograr reproducir la medida del espesor hallada previamente, determinar la distancia de reconstrucción óptima y corregir digitalmente las aberraciones en la imagen de fase.

En la Figura 5.11 tenemos el holograma de una zona de esta muestra, ya que la imagen se encuentra limitada por el tamaño de la matriz CMOS, de unos 6,3 mm de ancho por 3,3 mm de alto. Podemos observar claramente las franjas de interferencia, con un período $\Lambda = 12$ píxeles, pero lo más importante es que si nos detenemos en las regiones límites de los depósitos (detalle aumentado) logramos apreciar cómo estas franjas se desvían, debido al cambio de nivel entre la base de c-Si y el escalón de Ag.

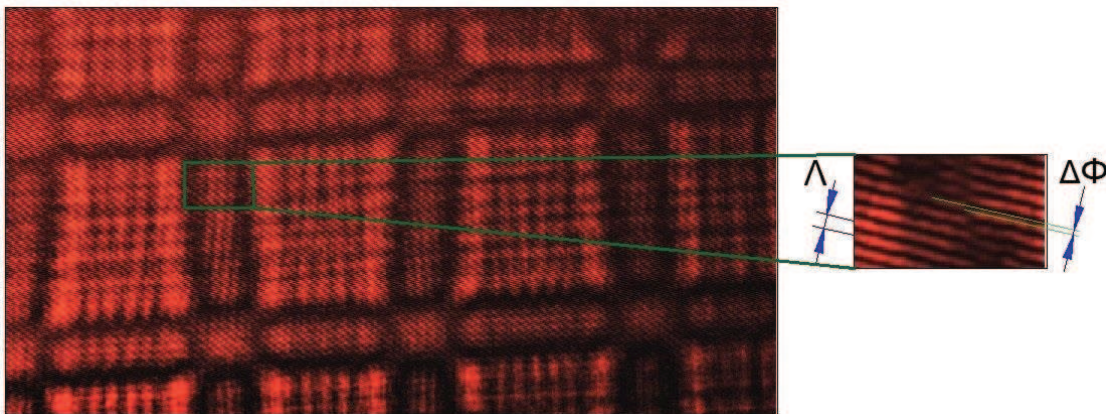


Figura 5.11. Holograma de la muestra. Tamaño de la imagen 4096 x 2160 píxeles².

Este desplazamiento de las franjas de interferencia es provocado por la diferencia de camino óptico, lo que resulta en una diferencia de fase que podemos medir en la imagen de fase, y así obtener el espesor del escalón por medio de la expresión (5.1), como veremos más adelante.

5.4.1- Determinación de la distancia focal de reconstrucción:

Cuando procesamos el holograma con el algoritmo *reconstruccion.m*, utilizamos la distancia de reconstrucción que nosotros medimos, pero como mencionamos en la sección 5.2, necesitamos obtener una imagen de fase correctamente enfocada para lograr una medición aceptable del espesor de material depositado, y esto se vuelve difícil si lo queremos realizar seleccionando

a mano esta distancia de reconstrucción óptima. En la Figura 5.12 podemos apreciar como varía la definición en las imágenes del módulo y la fase en función de la distancia de reconstrucción.

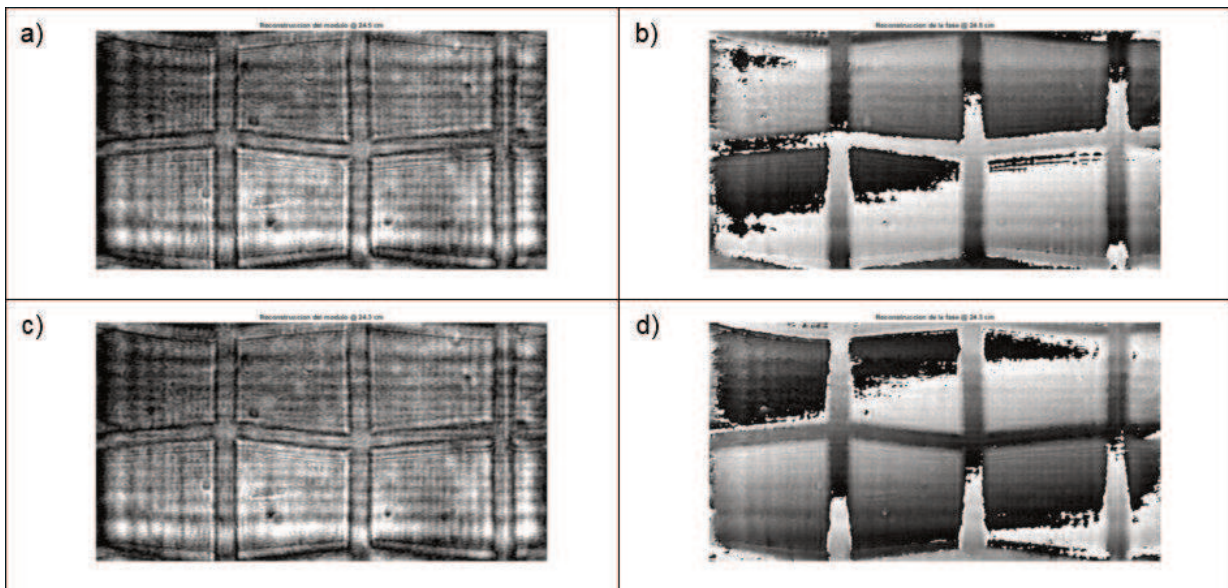


Figura 5.12. Imágenes del módulo y la fase a diferentes distancias de reconstrucción.
a) y b) $d = 24,5 \text{ cm}$. c) y d) $d = 24,3 \text{ cm}$.

En las Figuras 5.12a y b la distancia de reconstrucción utilizada fue la medida en el esquema óptico, en cambio en las Figuras 5.12c y d se eligió una distancia de $24,3 \text{ cm}$. En ambos casos el procesamiento se realizó utilizando el filtro pasa bajos con un radio $r \approx 100$ píxeles. Si comparamos las imágenes del módulo observamos que la imagen 5.12b se encuentra mejor enfocada que la 5.12a, dado que los bordes de los depósitos de Ag (sobre todo en la zona central) los vemos mejor definidos. Lo mismo sucede con las imágenes de fase, donde observamos en la imagen 5.12d que los bordes de los depósitos no se ven tan esfumados, como si sucede en la imagen 5.12c, lo que resultaría en una medición errónea de la topografía del objeto. Otra cosa que podemos observar es el cambio significativo en las regiones de la imagen de fase, dado que ésta varía en estos 2 mm de diferencia en la distancia de reconstrucción, aproximadamente unos 20 mil radianes.

Ahora, como vemos, necesitamos contar con algún método que permita determinar esta distancia de reconstrucción focal de alguna manera más objetiva y consistente que simplemente probando a mano diferentes distancias, hasta suponer que se logró focalizar la imagen reconstruida. Para ello recurrimos a un algoritmo, denominado *autofoco.m*, que puede verse en el apéndice, propuesto para analizar los píxeles de la imagen del módulo reconstruida, a diferentes distancias dentro de un intervalo previamente definido, y así lograr determinar la distancia focal de reconstrucción de la misma [4]. Estos algoritmos de autofocus son muy utilizados en cámaras digitales, por citar un ejemplo.

Para poder ejecutar este sub algoritmo debemos agregar, luego de la línea de la distancia de reconstrucción *dcent*, las siguientes líneas

```
dcent = 24.5e-2; % distancia estimada
rango = 5e-3; % semi intervalo de exploracion
paso = 1e-3; % paso de busqueda
```

donde definimos el intervalo de búsqueda y el paso, o salto, en cada valor de la distancia, ambos valores dados en metros. Luego de declaradas las variables anteriores, estamos en condiciones de llamar al sub algoritmo propiamente dicho, mediante el siguiente comando

```
[dopt,FM] = autofocus(dcent,rango,paso,fi,holo_fft_mask_shift);
```

Este trabaja con la imagen del módulo reconstruida a la distancia medida *dcent*, con los parámetros *fi* y *holo_fft_mask_shift*, y solicita que seleccionemos en dicha imagen una región de análisis. La región debe ser suficientemente pequeña, dado que si la misma abarca una gran cantidad de pixeles podemos correr el riesgo de obtener resultados erróneos, debido a incluir zonas degradadas por aberraciones y ruido. En la Figura 5.13 mostramos un ejemplo de la zona seleccionada.

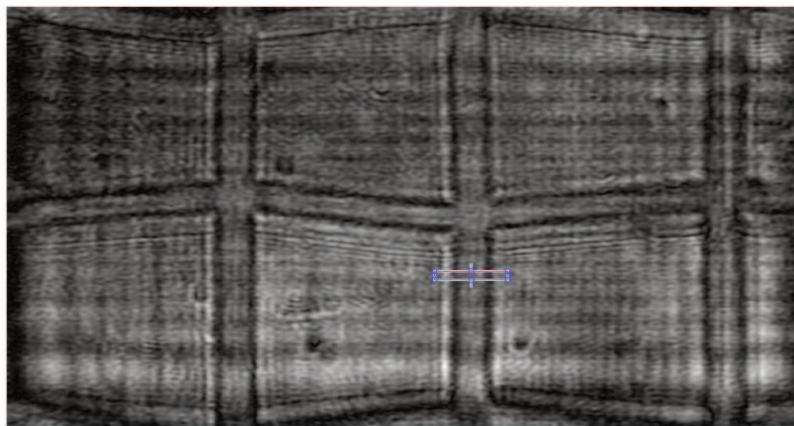


Figura 5.13. Zona seleccionada en la imagen del módulo para el algoritmo *autofoco.m*.

Luego el programa trabaja con una secuencia de imágenes, cuya cantidad depende del rango y paso adoptados, cada una con sus correspondiente distancia de reconstrucción, estas son analizadas, determinado el grado de enfoque relativo de todos los pixeles incluidos dentro del área de interés seleccionada previamente, utilizando un operador de medición del enfoque.

Los operadores de medición del enfoque, o FM (*focus measured*), se encuentran declarados en el sub algoritmo *fmeasure.m* [4], incluido en el apéndice, y trabajan en base a la suposición de

que los píxeles vecinos a un dado pixel se encuentran todos en un mismo plano de la superficie del objeto.

A continuación, el algoritmo *autofoco.m* realiza una gráfica del valor dado por el operador FM, en función de la distancia, donde el máximo valor dado por dicho operador se encuentra precisamente en la distancia focal.

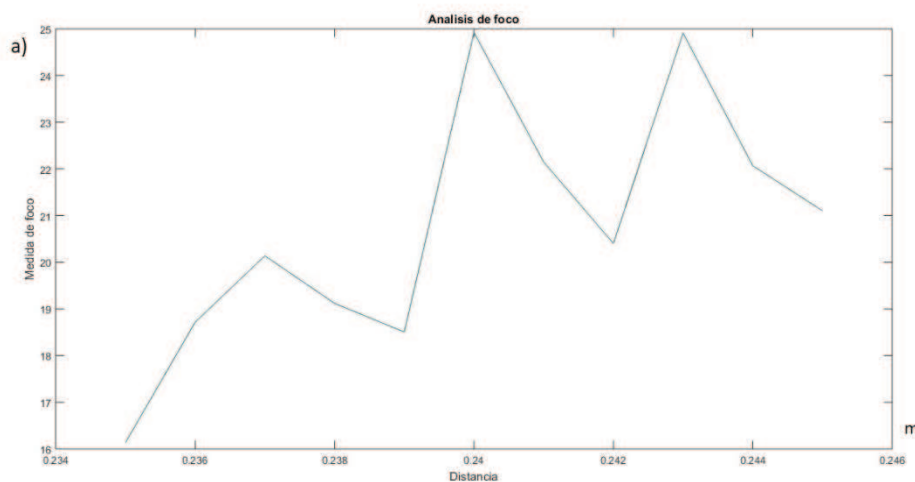
Para conocer, en nuestro caso de aplicación, cuál de estos operadores logra el mejor desempeño en la determinación de la distancia focal, todos ellos fueron evaluados a diferentes distancias de reconstrucción.

Se consideró valorar cualitativamente si el método arrojaba un máximo único en la función de medición del enfoque FM, que se aprecie bien definido dentro del intervalo. También se consideró valorar cuantitativamente que este máximo se encuentre dentro de la desviación estándar del histograma de los valores de distancias focales arrojados por todos los operadores, confeccionado para cada distancia de reconstrucción.

Otro parámetro a tener en cuenta es si el operador mantiene una variación lineal en la distancia focal, de la misma forma que varía linealmente la distancia medida.

Para realizar esta selección se tomaron hologramas de la muestra a siete diferentes distancias, desde *24 cm* hasta *39 cm*. Luego se procesaron cada uno de los hologramas utilizando los 28 operadores de medición del foco que posee el sub algoritmo *fmeasure.m*.

Como ejemplo tenemos en la Figura 5.14 tres graficas de la función FM, obtenidas con tres operadores diferentes, para la misma distancia medida, rango, paso y área seleccionada de la imagen del módulo.



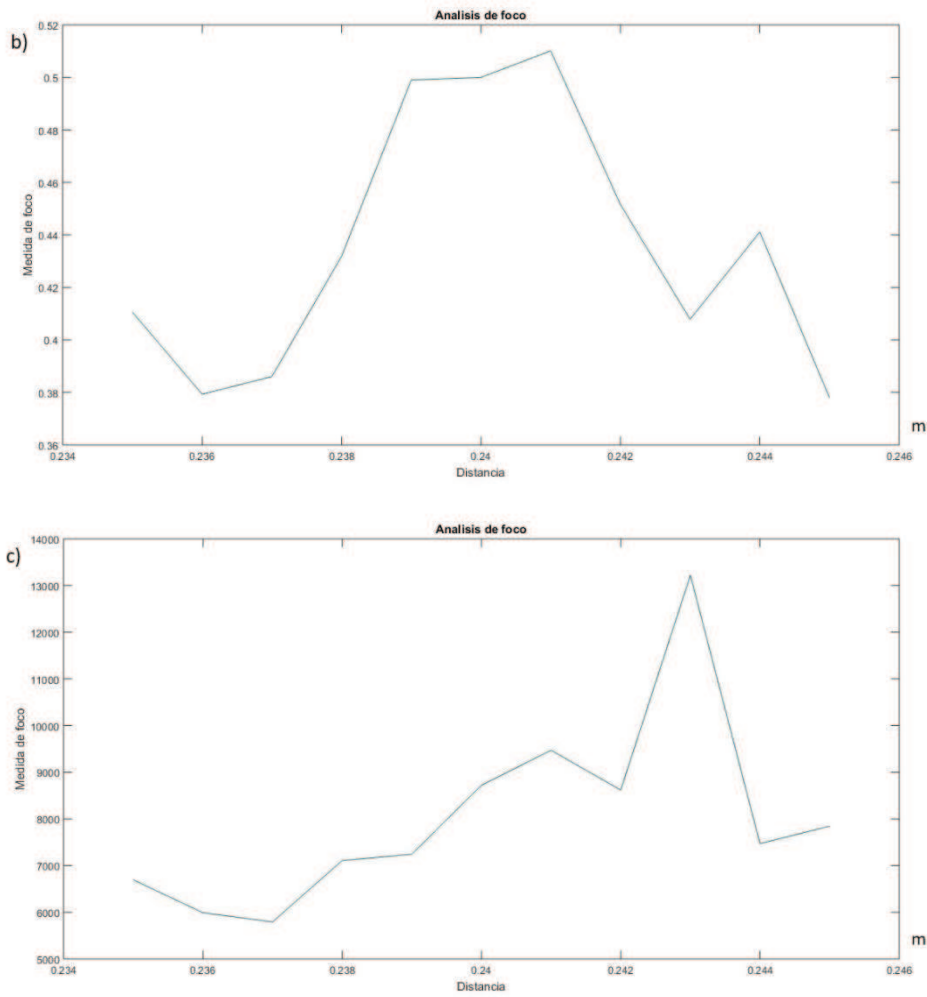


Figura 5.14. Gráfica de FM en función de la distancia, para $d_{cent} = 24 \text{ cm}$, $rango = 0,5 \text{ cm}$ y $paso = 0,1 \text{ cm}$.
a) Operador GLLV, b) LAPE y c) TENV.

La Figura 5.15 nos muestra el histograma, con los valores arrojados por los 28 operadores utilizados, para la distancia de reconstrucción $d_{cent} = 24 \text{ cm}$.

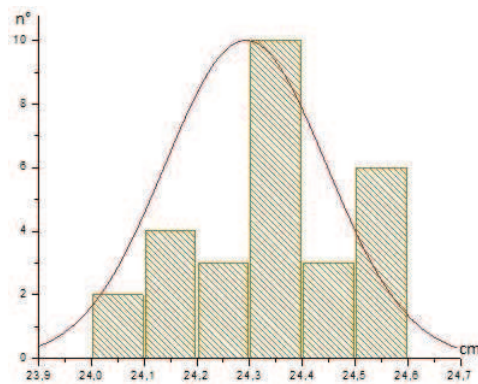


Figura 5.15. Histograma para $d_{cent} = 24 \text{ cm}$.

De este histograma podemos calcular la distancia focal media igual a $24,29\text{ cm}$, y su desviación estándar igual a $0,15\text{ cm}$. En la Tabla 5.1 se observan las distancias focales medias, las desviaciones estándar y los valores máximos y mínimos para todas las distancias de reconstrucción evaluadas, según los histogramas obtenidos.

Tabla 5.1. Valores medios y desviaciones estándar de las distancias focales.

d cm	24	26,5	29	31,5	34	36,5	39
$d_{focal\ media}\text{ cm}$	24,29	26,56	28,98	31,57	34,32	36,63	39,21
Desv. est. cm	0,15	0,35	0,31	0,37	0,21	0,37	0,28
$d_{focal\ min}\text{ cm}$	24,14	26,21	28,66	31,20	34,11	36,26	38,94
$d_{focal\ max}\text{ cm}$	24,45	26,92	29,29	31,94	34,53	37,01	39,49

En la Figura 5.16 observamos tres ejemplos en donde podemos comparar la linealidad en la variación de la distancia medida, con respecto a la variación en la distancia focal calculada por el operador utilizado.

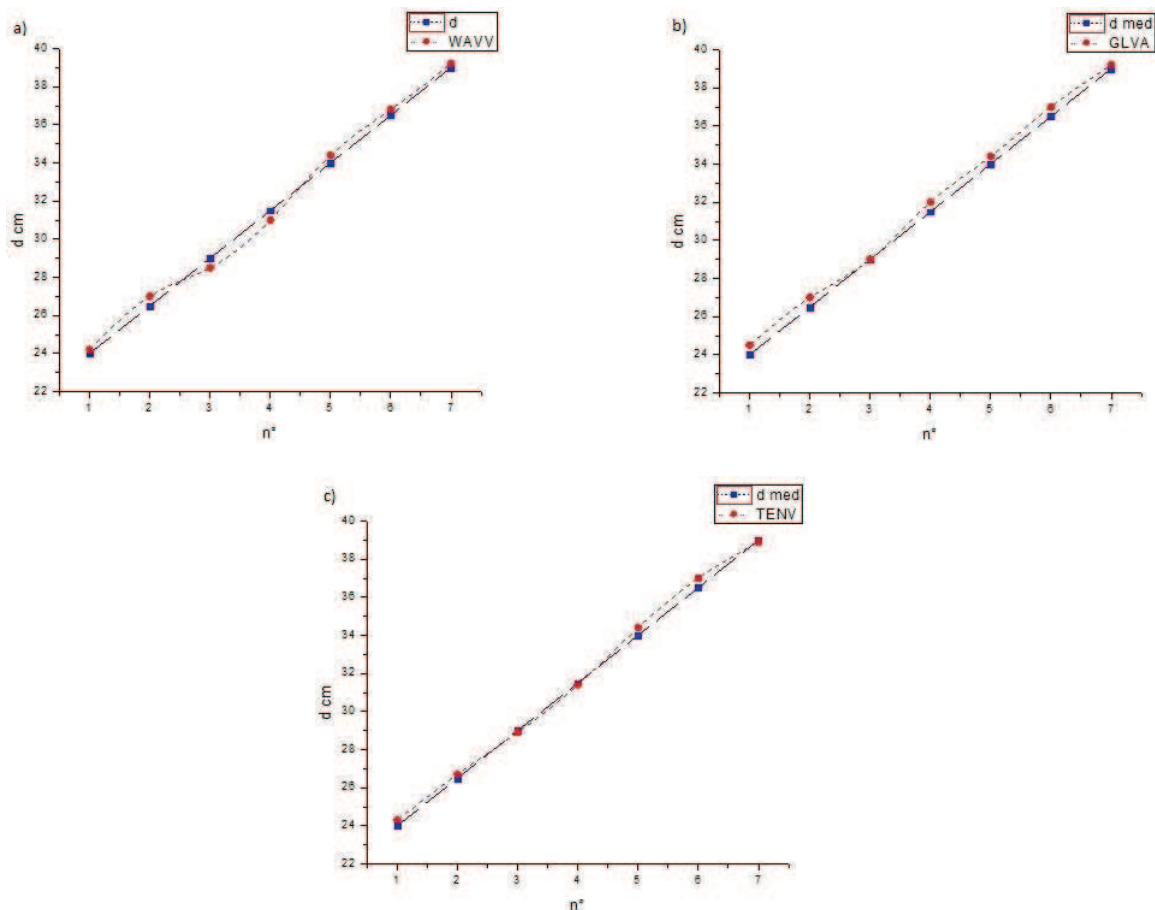


Figura 5.16. Comparación de d_{cent} vs d_{foco} . a) Operador WAVV , b) GLVA y c) TENV.

Con la información de las 196 gráficas de la función FM (ver los ejemplos en la Figura 5.13), los siete histogramas (ver el ejemplo en la Figura 5.14) y las 28 gráficas de la variación de la distancia focal (ver los ejemplos en la Figura 5.15), podemos componer la Tabla 5.2 para los 28 operadores, que señala si un método en particular cumple con los tres requisitos impuestos, a saber:

- Posee un máximo único bien definido en el intervalo
- El máximo obtenido se encuentra dentro de la desviación estándar
- Los máximos obtenidos, para cada distancia de reconstrucción, mantienen una relación aproximadamente lineal.

Tabla 5.2. Desempeño de los operadores de medición de foco.

Método	Def.							Desv.							Lin.
	24	26,5	29	31,5	34	36,5	39	24	26,5	29	31,5	34	36,5	39	
ACMO	o	o	o	x	x	x	x	o	o	o	o	x	o	x	O
BREN	x	o	o	x	o	o	o	x	x	o	o	o	o	o	X
CONT	x	o	o	x	o	o	x	o	x	x	x	o	o	x	X
CURV	o	o	x	x	x	x	x	x	o	o	x	o	o	o	X
DCTE	o	x	x	o	o	o	o	o	o	x	x	o	o	o	X
DCRT	x	x	o	x	o	o	o	o	o	o	o	o	o	o	O
GDER	x	x	x	x	o	o	o	o	o	o	o	o	o	o	O
GLLV	x	x	x	x	x	o	x	x	o	x	x	o	o	x	X
GLVA	o	o	x	o	x	x	o	x	x	o	x	o	o	o	O
GLVN	o	x	x	x	x	o	o	x	o	o	o	x	o	x	O
GRAE	o	o	o	o	x	o	o	x	x	o	x	o	x	o	X
GRAS	o	x	x	o	o	o	x	x	x	x	o	o	o	x	X
GRAT	o	o	o	x	x	o	o	o	x	x	o	o	o	o	X
HELM	o	x	o	x	o	o	o	o	x	x	x	o	o	x	X
HISE	o	o	o	o	o	x	o	x	x	o	o	o	o	o	O
HISR	o	x	x	x	x	x	x	x	o	x	x	o	x	x	X
LAPD	x	o	o	x	o	x	x	x	o	x	o	o	x	x	X
LAPE	x	o	x	o	x	x	o	x	o	x	o	o	o	o	O
LAPM	x	o	x	x	o	o	x	x	o	x	o	o	o	o	O
LAPV	x	o	x	o	o	x	o	o	o	o	o	o	x	o	X
SFIL	o	x	o	o	o	o	o	o	o	x	o	x	o	x	O
SFRQ	x	o	o	x	o	o	x	o	x	o	o	o	o	x	O
TENG	x	o	o	x	o	o	o	o	x	x	x	o	o	o	X
TENV	o	o	o	o	o	o	o	o	o	o	o	o	o	x	O
VOLA	o	o	x	x	x	x	x	o	x	o	o	o	o	x	X
WAVR	o	x	o	o	o	x	x	o	o	o	x	o	o	o	O
WAVS	o	x	o	x	o	o	x	o	o	x	x	o	x	x	X
WAVV	o	o	o	o	x	x	x	o	x	x	x	o	o	o	X

Claramente podemos concluir, según la Tabla 5.2, que el operador de mejor desempeño para esta aplicación es el denominado TENV.

Este operador utiliza la variancia del gradiente de la imagen para obtener un valor numérico de la función FM. Es utilizado generalmente en los procesos de control de los sistemas de autoenfoco de las cámaras digitales aplicadas a microscopía óptica. La expresión para calcular el valor de la función FM se encuentra definida por la siguiente ecuación [7]

$$FM(x, y) = \sum_{i,j \in \Omega(x,y)} [G_{i,j}(x, y) - \bar{G}]^2 \quad (5.14)$$

donde $G_{i,j}(x, y)$ es la magnitud del gradiente, calculada por

$$G_{i,j}(x, y) = \sqrt{G_{i,j}(x)^2 + G_{i,j}(y)^2} \quad (5.15)$$

siendo $G_{i,j}(x)$ y $G_{i,j}(y)$ los gradientes de la imagen, en las direcciones x e y respectivamente, para cada pixel seleccionado, y \bar{G} es el valor medio de la magnitud de este gradiente, dentro de área seleccionada $\Omega(x, y)$.

Si volvemos a observar la Figura 5.14c, notamos en la gráfica de la función FM, que tenemos un máximo bien definido para una distancia focal de $24,3 \text{ cm}$.

En la Figura 5.17 podemos observar como varía la focalización de la imagen de una zona del módulo en función de la distancia de reconstrucción, contrastada con la gráfica de la función FM, donde notamos la imagen focalizada en el pico máximo de la misma, y como esta imagen se va desenfocando hacia los extremos de la gráfica.

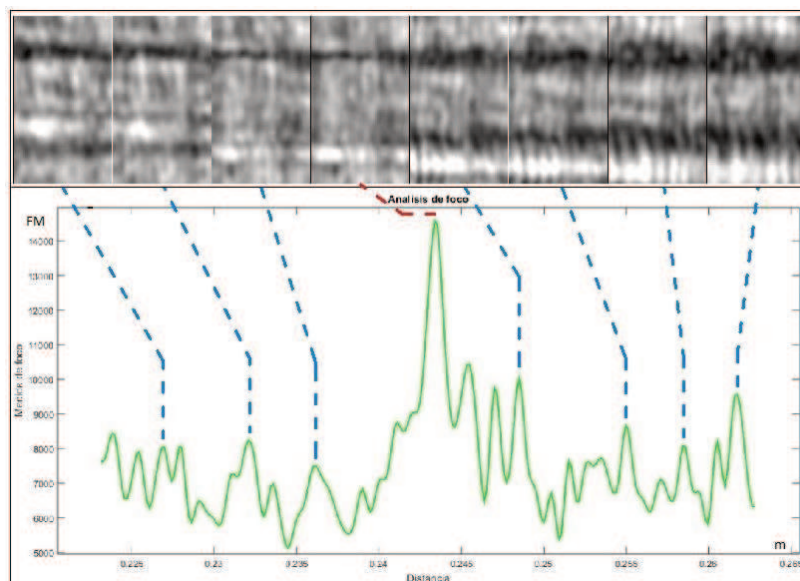


Figura 5.17. Variación de focalización de la imagen del módulo y gráfica de FM en función de la distancia, para $d_{cent} = 24,3 \text{ cm}$, $rango = 2 \text{ cm}$ y $paso = 0,05 \text{ cm}$, operador TENV.

Una ventaja de utilizar este método digital de enfoque es que podemos aumentar la precisión en el valor de la distancia focal con el solo hecho de disminuir la longitud del intervalo de rango y paso utilizados anteriormente. Por ejemplo, tomando como base la gráfica de la Figura 5.14c, utilizamos ahora una $d_{cent} = 24,3 \text{ cm}$, un rango de $0,1 \text{ cm}$ y un paso de $0,01 \text{ cm}$, obtenemos una distancia focal $d_f = 24,35 \text{ cm}$, como observamos en la gráfica de la función FM de la Figura 5.18.

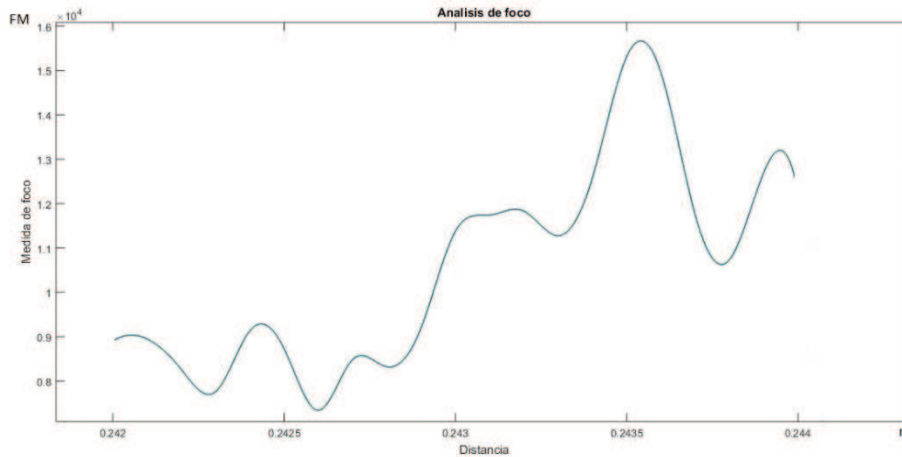


Figura 5.18. Gráfica de FM en función de la distancia, para $d_{cent} = 24,3 \text{ cm}$, $rango = 0,1 \text{ cm}$ y $paso = 0,01 \text{ cm}$, operador TENV.

5.4.2- Corrección digital de aberraciones en la imagen de fase:

Como se mencionó, al comienzo de este capítulo, necesitamos obtener una imagen de fase con la mínima aberración posible a fin de lograr la mejor precisión en la determinación del espesor de la película delgada depositada.

A continuación desarrollemos un breve concepto de las aberraciones presentes en un sistema óptico. Si nos basamos en la teoría de la difracción, vista en el capítulo II, tenemos que la única imperfección en el sistema óptico es el tamaño finito de la abertura, siendo la onda esférica ideal. Las aberraciones alteran este frente de onda ideal, introduciendo típicamente errores en la fase de esta onda. Para incluir dichas aberraciones, el campo óptico en un punto P_0 , a la salida de la pantalla difractora, puede ser modificado por la siguiente expresión

$$U_{A,P_0}(x,y) = U_{P_0}(x,y) \cdot \exp[jk\Phi_A(x,y)] \quad (5.16)$$

donde $U_{P_0}(x,y)$ es el campo óptico sin aberraciones en el punto P_0 , dado por la ecuación (2.74), y $\Phi_A(x,y)$ es el error en la fase introducido por las aberraciones.

A la función de fase $\Phi_A(x,y)$ es conveniente escribirla en términos de coordenadas polares, como $\Phi_A(\rho,\theta)$, dado que en general las aberturas Σ son circulares, y las aberraciones pueden ser representadas en términos de los polinomios de Zernike. Estos polinomios son ortogonales, y se encuentran normalizados dentro de un círculo de radio unitario. Entonces, tenemos que la función

de fase $\Phi_A(\rho, \theta)$ resulta de una expansión de los polinomios de Zernike, $z_k(\rho, \theta)$, donde ρ es la coordenada radial dentro del círculo unitario, y θ es el ángulo polar. Los polinomios de Zernike se expresan de la siguiente forma

$$z_k(\rho, \theta) = R_n^m(\rho) \cdot \cos(m\theta) \quad (5.17)$$

donde n y m son enteros positivos, $R_n^m(\rho)$ es un polinomio de grado n , el cual no contiene ninguna potencia de ρ que sea menor a m . Por lo que $R(\rho)$ es par cuando m es par, y viceversa. Estos polinomios pueden consultarse en las tablas de la bibliografía de referencia. En la Figura 5.19 vemos la representación gráfica de la expresión $z_k(\rho, \theta)$ para los primeros ocho términos de estos polinomios.

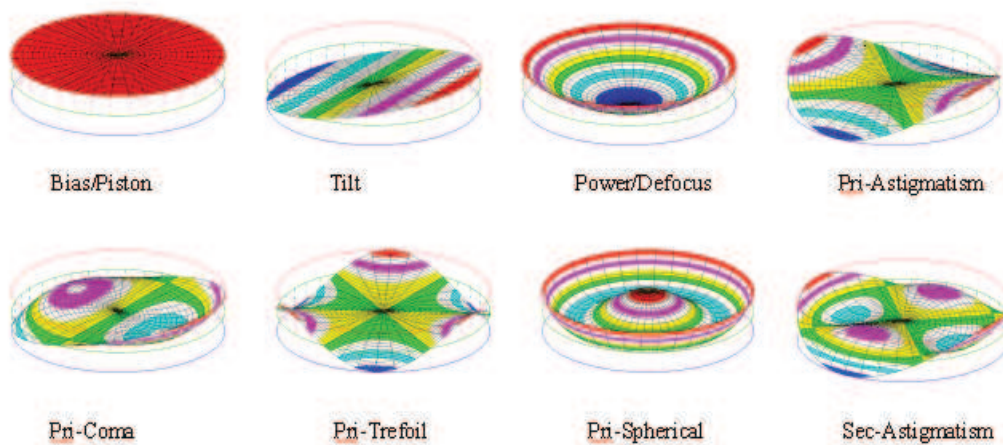


Figura 5.19. Representación de los polinomios de Zernike (extraído de Grenberg, Michels, 2017).

Definidos estos polinomios de Zernike, la representación de $\Phi_A(\rho, \theta)$ puede escribirse de la siguiente forma

$$\Phi_A(\rho, \theta) = \sum_{k=1}^K w_k z_k(\rho, \theta) \quad (5.18)$$

siendo K un entero (que en las tablas llega al valor 37). El coeficiente w_k se determina, para valores finitos de n y m , por mínimos cuadrados [5].

Comentamos al principio de este párrafo que las aberraciones se originan cuando el frente de onda óptico de un sistema formador de imágenes se encuentra deformado, de manera involuntaria. La reconstrucción del frente de onda por métodos holográficos necesita de la habilidad para controlar y manipular este frente de onda, de manera de compensar tales aberraciones. Como veremos, la holografía digital ofrece mucha flexibilidad y versatilidad para detectar y controlar las aberraciones.

Por ejemplo, supongamos un sistema de imagen que transmite el campo óptico de un objeto como a_o , pero por las imperfecciones en el sistema, se agregan aberraciones a_A al campo objeto,

impidiendo la formación de una imagen perfecta. Vimos que interfiriendo con el campo de referencia A_R se forma un holograma, donde el término de la imagen gemela tiene la forma $(a_0 + a_A)A_R^*$. El término de la aberración puede ser eliminado si se adquiere otro holograma, donde no está presente el objeto, para lograr obtener el término correspondiente a $a_A.A_R^*$. A continuación este holograma es restado del primero, resultando el holograma digital del campo objeto $a_0.A_R^*$ sin aberraciones. En un principio esto resulta en la eliminación completa de la aberración, sin necesidad de conocer algún detalle sobre la misma. De esta manera, la aberración del sistema óptico puede ser obtenida, y entonces utilizada para obtener la imagen corregida de cualquier objeto, siempre y cuando los parámetros ópticos del sistema no sean modificados. Una posible desventaja de este método es que requiere de, al menos, la obtención de dos hologramas.

Existen técnicas que logran compensar la aberración con la obtención de un solo holograma. En estas usualmente se presume que el objeto es delgado, como los depósitos de película delgada utilizados en este trabajo, o dispositivos MEMS.

Como un primer ejemplo de estas técnicas podemos mencionar que se puede compensar parcialmente la aberración, sin tener conocimiento detallado de los parámetros que definen la misma, aplicando un filtro pasa altos sobre una pequeña región del espectro de frecuencias espaciales, centrada en la frecuencia portadora. Pero se debe tener en cuenta que este método asume que la aberración está, mayoritariamente, definida por los polinomios de Zernike de bajo orden. Como desventaja debemos mencionar que se pierden en la imagen las características que posee el objeto en las bajas frecuencias espaciales.

Ahora, teniendo en consideración las aberraciones caracterizadas por los polinomios de Zernike, podemos comentar brevemente algunas características de los tres primeros tipos de aberraciones. Si observamos en la Figura 5.19, el orden más bajo de los polinomios de Zernike es el pistón, y define la aberración provocada por la diferencia entre la distancia de reconstrucción y la distancia focal del objeto, que provoca la imagen desenfocada que estudiamos en el apartado 5.4.1.

Continuando con las gráficas de la Figura 5.19, el orden siguiente en los polinomios de Zernike es el que define la aberración provocada por la inclinación del frente de onda (*tilt*), que proviene de la desalineación del haz objeto y el haz referencia. Esto podría corregirse girando, en la forma correspondiente, el haz de la onda de referencia plana (o sea, girando el espejo). Pero, recordemos que necesitamos un ángulo de inclinación para obtener el patrón de franjas de interferencia, y además, este ángulo debe tener la magnitud suficiente como para lograr acercarse, lo más posible, al valor teórico dado por el teorema del muestreo. Esto hace inevitable que la imagen de fase obtenida contenga la aberración provocada por la inclinación, la que debemos corregir posteriormente por métodos numéricos, como veremos más adelante.

El siguiente orden en las aberraciones (Figura 5.19) es la esférica, que también provoca un desenfoque en las imágenes. Dicha aberración se origina por imperfecciones en los dispositivos

que conforman el sistema óptico. A menudo, la aberración esférica se observa si los frentes de onda objeto y referencia poseen curvaturas muy dispares entre sí. Esto puede ser compensado digitalmente utilizando una onda de referencia, en la reconstrucción, que posea una curvatura cuadrática [1].

Ahora veamos la corrección de la inclinación, o *tilt*, en la imagen de fase por medio de un sub algoritmo que denominamos *corrigefase_plano.m*, el cual podemos apreciar en el apéndice. Este sub algoritmo crea un plano definido por la siguiente ecuación

$$\Gamma(x, y, z) = k_0 \cdot x + l_0 \cdot y + z \quad (5.19)$$

donde los parámetros k_0 y l_0 definen los valores iniciales de los vectores en las direcciones x e y respectivamente. El parámetro z se define dentro del intervalo de una longitud de onda, tomando valores en radianes, entre $-\pi$ y π , para minimizar los saltos de fase, moviendo verticalmente al plano Γ . Este valor de z podemos pensarlo como un ajuste fino de la distancia focal, determinada por el sub algoritmo *autofoco.m*, visto en el apartado anterior.

Una vez obtenida la imagen de fase a la distancia de reconstrucción focal definimos los parámetros del plano Γ de corrección, mediante los siguientes comandos

```
k0 = 0.003; % vector en x
l0 = 0.003; % vector en y
z0 = 0; % distancia en z entre -pi y pi
```

en este ejemplo se adoptaron esos valores para los parámetros k_0 y l_0 como una primera aproximación a los valores necesarios para corregir el *tilt*, siendo cero el valor de z , o sea que comenzamos en la posición del plano a la distancia focal. Notemos que los vectores poseen valores pequeños, debido a que, como calculáramos en las expresiones (5.9), los ángulos de inclinación también son pequeños.

La matriz formada por el plano así obtenido, luego debe sumarse digitalmente a la matriz de la imagen de fase, dando como resultado una matriz con la imagen de fase corregida, minimizando de esta forma las aberraciones.

Una vez declarados los parámetros del plano, llamamos al sub algoritmo de corrección de la aberración (que puede verse en el apéndice) mediante la siguiente línea de comando

```
corrigefase_plano (fasefoco, k0, l0, z0);
```

Este trabaja con la imagen de la fase a la distancia focal, y con los parámetros del plano que definidos anteriormente. Abre una ventana, que podemos ver en la Figura 5.20a, donde tenemos a la izquierda la imagen de fase original, en el centro la imagen del plano a sumar, y a la derecha la

imagen de fase corregida. Debajo tenemos las barras, con los botones desplazables, que controlan los valores de los tres parámetros del plano. A medida que vamos modificando los valores de estos, podemos llegar a obtener la imagen de fase corregida, como apreciamos en la Figura 5.20b.

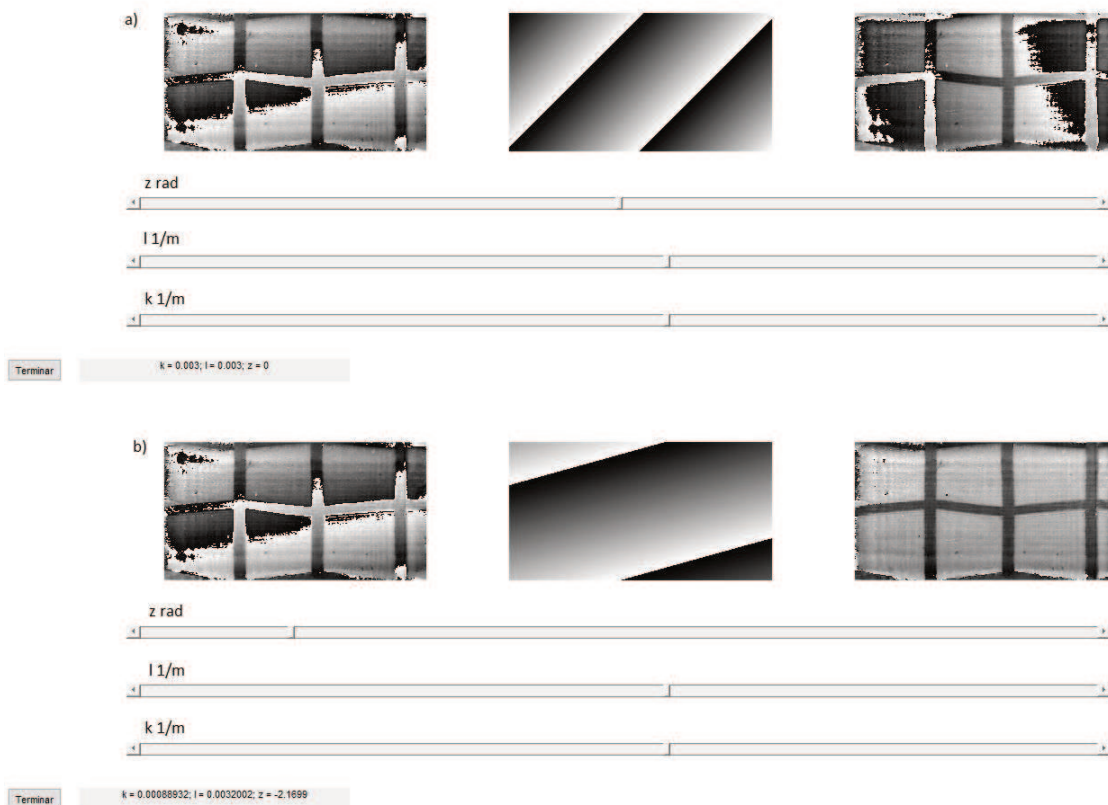


Figura 5.20. Corrección de la aberración de pistón y *tilt* en la imagen de fase.

a) Parámetros iniciales y b) Parámetros finales.

Los valores finales de los parámetros del plano a sumar, para obtener la imagen de fase corregida, resultaron iguales a $k = 0,00088932$ 1/m , $l = 0,0032002$ 1/m y $z = - 2,1699$ radianes.

En la Figura 5.21 podemos apreciar esta imagen de fase corregida, donde las regiones de los depósitos de Ag se encuentran claramente diferenciadas de la base de c-Si, y cada una posee valores de fase bastante uniformes, si bien notamos en las zonas de los bordes que aún permanecen algunas aberraciones.

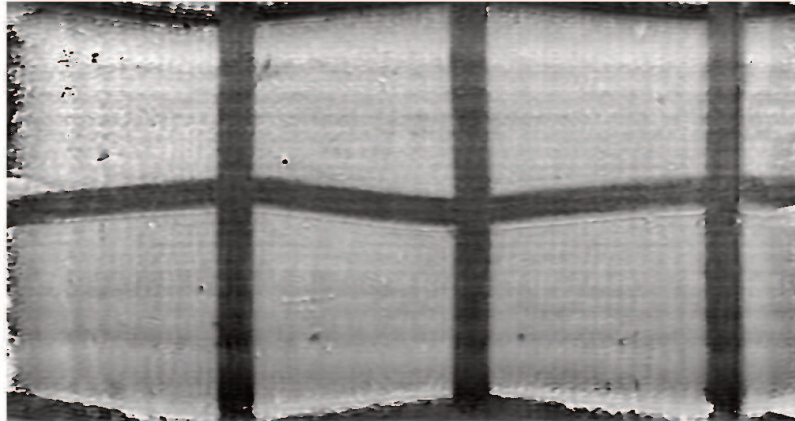


Figura 5.21. Imagen de fase corregida en pistón y *tilt*.

5.4.3- Determinación del espesor de película de Ag mediante la imagen de fase:

Una vez obtenida la imagen de fase corregida estamos en condiciones de determinar los valores de la fase de interferencia, tanto en las zonas del depósito de Ag como en las zonas del sustrato de c-Si. Como mencionamos en la sección 5.2, las diferencias en los valores de las fases entre el depósito y el sustrato, como se observa claramente en la figura 5.21, se deben a las diferencias entre los caminos ópticos de los haces reflejados, provocadas por el espesor de dicho depósito de Ag.

Pero, además de la diferencia de fases dadas por las distintas alturas en la superficie del objeto, tenemos que considerar las variaciones de fase provocadas por los coeficientes de reflexión de cada material. Este coeficiente de reflexión se define por la relación entre las amplitudes de la onda reflejada e incidente

$$r = \frac{E_r}{E_i} = \frac{(1-\tilde{n})}{(1+\tilde{n})} \quad (5.20)$$

donde \tilde{n} es el índice de refracción complejo, dado por la siguiente expresión [6]

$$\tilde{n} = n - j\kappa \quad (5.21)$$

A la longitud de onda del láser de He-Ne utilizado, $\lambda = 632,8 \text{ nm}$, tomando datos de [7], para la Ag tenemos un índice complejo igual a

$$\tilde{n}_{Ag} = 0,135 - j3,99 \quad (5.22)$$

y para el c-Si el índice complejo es igual a

$$\tilde{n}_{c-Si} = 3,856 - j0,019 \quad (5.23)$$

Utilizando estos índices en la ecuación (5.20) calculamos el coeficiente de reflexión de cada material, resultando las siguientes expresiones

$$r_{Ag} = -0,868 + j0,464 \quad (5.24a)$$

$$r_{c-Si} = -0,588 + j0,002 \quad (5.24b)$$

Como vemos, ambos coeficientes de reflexión son también complejos, dado que existe una diferencia de fases entre la onda incidente y la reflejada. Esta diferencia de fases se calcula por medio de la siguiente expresión

$$\Delta\Phi_{mat} = 180^\circ + tg^{-1} \left(\frac{Imr}{Re_r} \right) \quad (5.25)$$

Entonces, para estos materiales, tenemos las siguientes variaciones en las fases

$$\Delta\Phi_{Ag} = 180^\circ + tg^{-1} \left(\frac{0,464}{-0,868} \right) = 151,87^\circ = 2,65 \text{ rad} \quad (5.26a)$$

$$\Delta\Phi_{c-Si} = 180^\circ + tg^{-1} \left(\frac{0,002}{-0,588} \right) = 179,84^\circ = 3,14 \text{ rad} \quad (5.26b)$$

a la longitud de onda del láser de He-Ne. En la Figura 5.22a tenemos un esquema, donde observamos cómo se vería una imagen de fase de dos piezas de superficies especulares de Ag y c-Si, destacando que, por más que ambas estén situadas al mismo nivel, seguiríamos notando una diferencia de fase. En la Figura 5.22b se esquematizan las variaciones de fase por reflexión de ambos materiales.

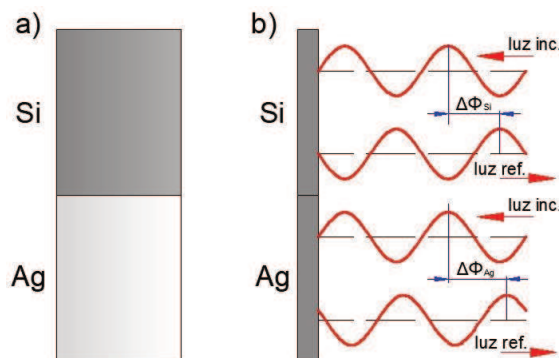


Figura 5.22. a) Esquema de imagen de fase. b) Variaciones de fase por reflexión en c-Si y Ag.

Considerando lo anterior, debemos reescribir la expresión (5.1) para calcular la altura del objeto, dependiendo del material del mismo, restando la variación de fase por reflexión, de la siguiente manera

$$\Delta z_{mat}(\xi, \eta) = \frac{\lambda}{4\pi} [\Delta\Phi(\xi, \eta) - \Delta\Phi_{mat}] \quad (5.27)$$

Por lo que debemos tener en cuenta sobre que región de la imagen de fase se calcula la altura dada por la expresión anterior. En nuestro caso, observando la imagen de fase de la Figura 5.21, notamos que la zonas claras, de mayor fase, corresponden a Ag, y las zonas oscuras, de menor fase, a c-Si.

Para realizar un cálculo estadístico del espesor del depósito de Ag, seleccionamos cuatro segmentos en la imagen de fase, como se aprecia en la Figura 5.23, dos correspondientes a la región de c-Si, y dos a la región de Ag, de los cuales obtuvimos los valores de la fase en cada punto.

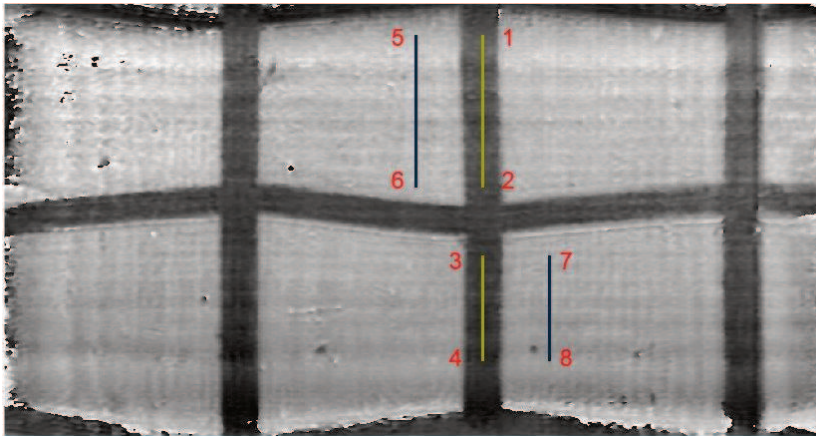


Figura 5.23. Segmentos para el cálculo del espesor del depósito. Los puntos señalados en pixeles para c-Si $P_1=(2380,200)$, $P_2=(2380,900)$, $P_3=(2400,1270)$ y $P_4=(2400,1800)$. Para Ag $P_5=(2000,200)$, $P_6=(2000,900)$, $P_7=(2800,1270)$ y $P_8=(2800,1800)$.

De este modo, seleccionando las columnas en la matriz de la imagen de fase, determinadas por estos segmentos previamente señalados, tenemos 1234 valores de fase en cada una de las regiones de Ag y c-Si. Luego estos valores de fase deben ser corregidos, restando a las columnas de cada región la fase correspondiente, dada por las expresiones (5.26), así podemos obtener el valor de la altura de la superficie para cada elemento de estas dos columnas, aplicando la ecuación (5.27). La diferencia en estas alturas es lo que nos da, finalmente, el espesor Δh del depósito de Ag, como lo escribimos en la siguiente expresión

$$\Delta h(\eta) = \Delta z_{Ag}(\xi_{Ag}, \eta) - \Delta z_{c-Si}(\xi_{c-Si}, \eta) \quad (5.28)$$

Realizando un histograma de los 1234 valores obtenidos de $\Delta h(\eta)$, como se observa en la Figura 5.24, resulta el siguiente valor medio para el espesor del depósito de Ag, con su desviación estándar

$$\Delta h = 156 \text{ nm} \pm 11 \text{ nm} \quad (5.29)$$

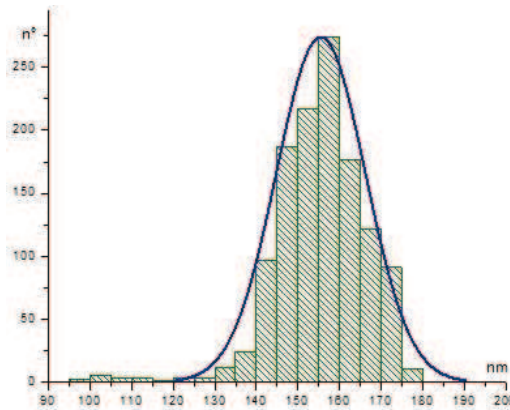


Figura 5.24. Histograma para los espesores de depósito de Ag.

Este valor para el espesor de la película de Ag es aproximado al valor declarado en el trabajo señalado en la referencia [3].

En la Figura 5.25 tenemos una imagen de la fase, en la cual se indicaron un segmento donde se obtuvo la gráfica del perfil que podemos apreciar en la Figura 5.26, y una zona seleccionada para elaborar la imagen de la Figura 5.27, donde se aprecia un arreglo tridimensional de la superficie del objeto.

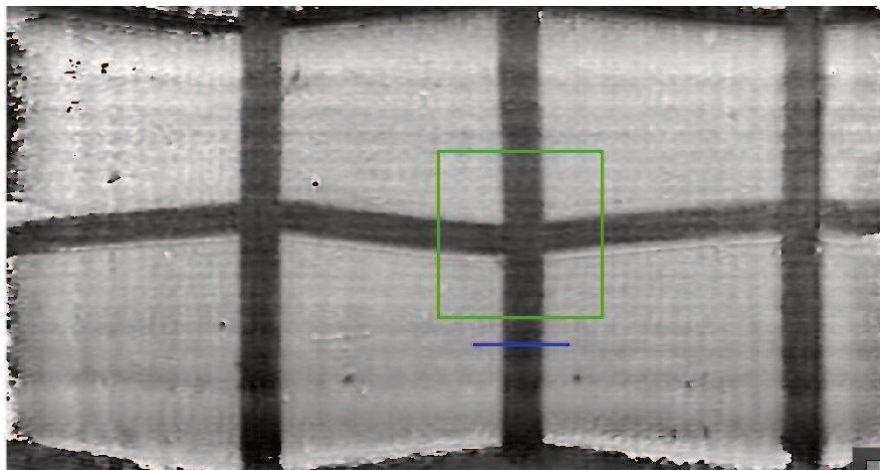


Figura 5.25. Imagen de fase donde se señalan el segmento para la gráfica del perfil y el área para la gráfica tridimensional.

Notemos en la imagen del perfil de la Figura 5.26, las ondulaciones del mismo en la parte superior, donde se encuentra la superficie de Ag, como así también en la parte inferior, correspondiente a la superficie de c-Si, estas se deben a los efectos de la difracción provocada por la abertura rectangular del dispositivo CMOS, como vimos en el apartado 2.2.9. Este efecto es observable también en la imagen tridimensional de la Figura 5.27.

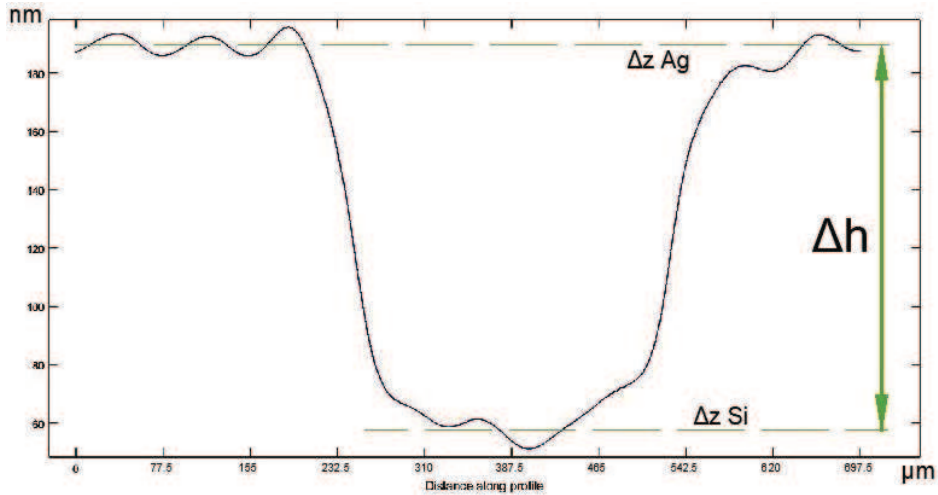


Figura 5.26. Gráfica del perfil del segmento indicado en la figura 5.24, donde se indican las alturas medias de las superficies de Ag y c-Si, y la diferencia entre las mismas, que nos señala el espesor Δh del depósito de Ag en ese segmento.

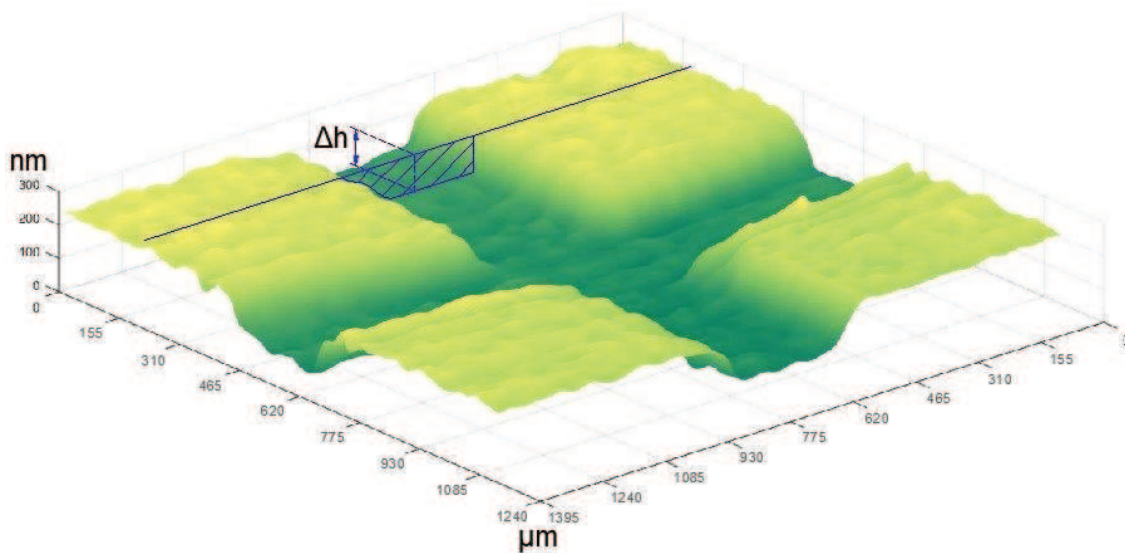


Figura 5.27. Gráfica tridimensional del área señalada en la figura 5.24, donde se indica el espesor Δh del depósito de Ag para un determinado segmento.

5.5- DHM de una película de SiO₂ sobre silicio cristalino:

Con la muestra utilizada en los apartados anteriores, de espesor de película aproximadamente conocido, logramos comprobar la eficacia del método de DHM propuesto, ahora lo utilizaremos para medir el espesor de película delgada en una muestra donde no se conoce dicha medida, esto constituye el objetivo propuesto en este trabajo de tesis.

El objeto analizado es una pieza rectangular de silicio cristalino con una zona que contiene una película delgada de dióxido de silicio. Esta película fue formada por un flujo de aire a alta temperatura que entró en contacto con una zona expuesta de la superficie de c-Si, produciéndose la reacción que formó la película de SiO₂, como podemos apreciar en la Figura 5.28. Dicha pieza posee un tamaño de 2,3 cm de ancho por 1,3 cm de alto, y la película de SiO₂ mide alrededor de 15 mm de ancho por 6 mm de alto.



Figura 5.28. Muestra de silicio cristalino, de 2,3 cm de ancho por 1,3 cm de alto, con película delgada de SiO₂, de 15 mm de ancho por 6 mm de alto.

En la Figura 5.29 tenemos el holograma de esta muestra, donde podemos observar las franjas de interferencia, con un período $\Lambda = 14$ píxeles, y en el detalle aumentado de la región límite de la película, logramos apreciar el desvío de dichas franjas, provocado por el cambio de nivel entre la base de c-Si y el escalón de SiO₂.

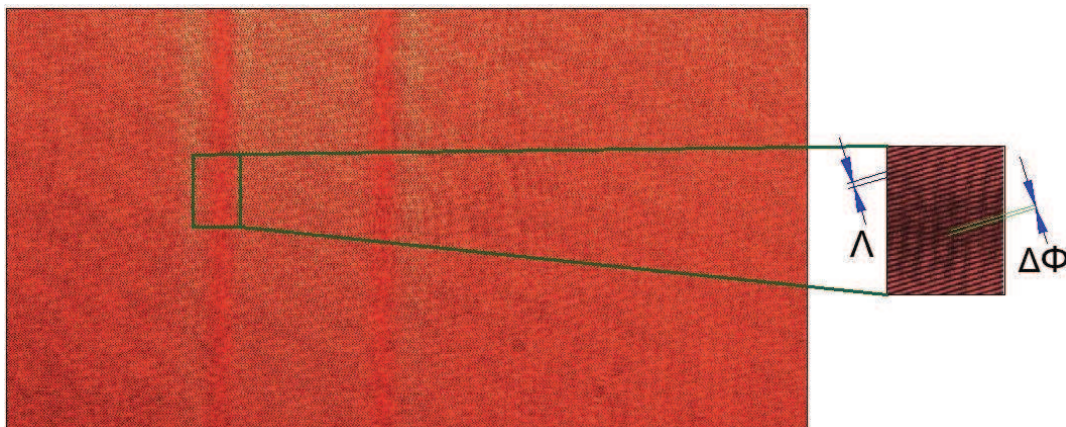


Figura 5.29. Holograma de la muestra. Tamaño de las imagen 4096 x 2160 píxeles².

El procesamiento del holograma se realizó utilizando el filtro pasa bajos con un radio de máscara $r \approx 60$ píxeles.

La distancia de reconstrucción, para una primera aproximación, fue la medida en el esquema óptico, siendo $d_c = 10 \text{ cm}$. Para encontrar la distancia focal, con el operador TENV, se utilizaron un rango de 1 cm y un paso de 1 mm , con lo que se obtuvo la distancia de reconstrucción focal $d_f = 9,4 \text{ cm}$. En una segunda aproximación, con $d_c = 9,4 \text{ cm}$, un rango de 10 mm y un paso de $0,1 \text{ mm}$, hallamos una distancia de reconstrucción focal igual a $d_f = 9,42 \text{ cm}$, como se aprecia en la gráfica de la función FM de la Figura 5.30.

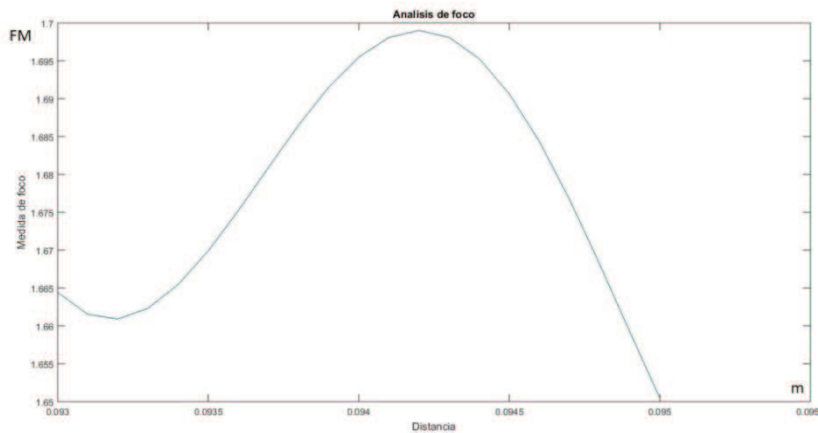


Figura 5.30. Gráfica de la función FM para $d_c = 9,4 \text{ cm}$, $r = 10 \text{ mm}$ y $p = 0,1 \text{ mm}$.

Con esta distancia de reconstrucción focal se obtuvieron las imágenes de módulo y fase que podemos apreciar en las Figuras 5.31a y 5.31b respectivamente.

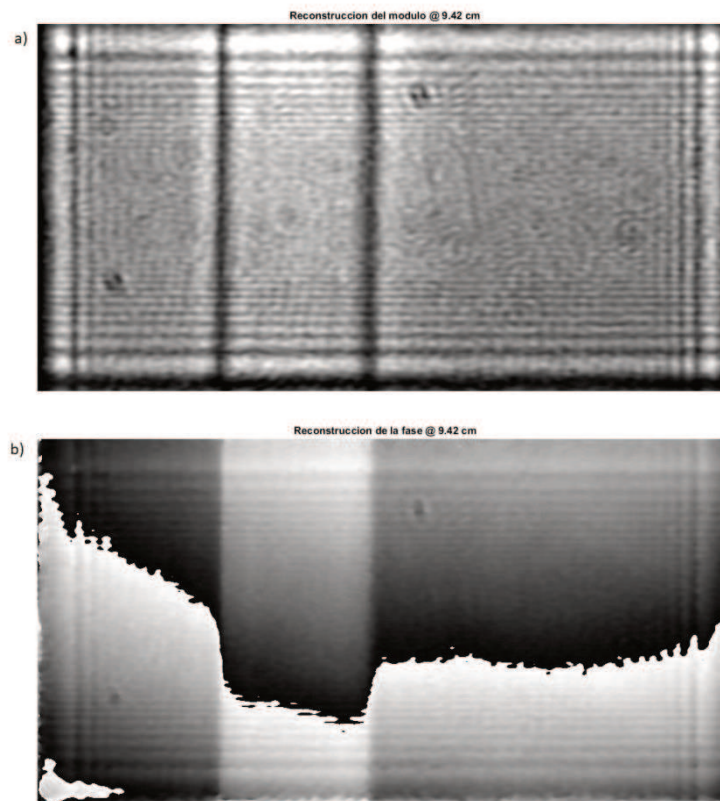


Figura 5.31. a) Imagen de módulo. b) Imagen de fase. Para $d_f = 9,42 \text{ cm}$.

Aquí notamos claramente, en ambas imágenes, las franjas debidas a los efectos de la difracción provocada por la abertura rectangular del dispositivo CMOS (sección 2.2.9). En la imagen de módulo pueden distinguirse perfectamente los límites de la película de SiO₂ (líneas verticales), y en la imagen de fase observamos las diferencias entre las zonas de c-Si y SiO₂, además de los cambios de fase ocasionados por la inevitable aberración presente en la imagen holográfica.

Para lograr disminuir la aberración, damos al plano de corrección los parámetros iniciales con los siguientes valores $k_0 = 0,003$, $l_0 = 0,003$ y $z_0 = 0$, como lo hicimos en el caso anterior. Luego de trabajar con este procesamiento, los valores finales de los parámetros del plano a sumar, para obtener la imagen de fase corregida, resultaron iguales a $k = -0,00083076 /m$, $l = 0,0038814 /m$ y $z = -0,0091797$ radianes.

En la Figura 5.32 vemos la imagen obtenida para la fase corregida, donde la región de la película de SiO₂ se encuentra claramente diferenciada de la base de c-Si, y además, cada región posee valores de fase bastante uniformes, si bien notamos en las zonas de los bordes de esta imagen que aún persisten algunas aberraciones.

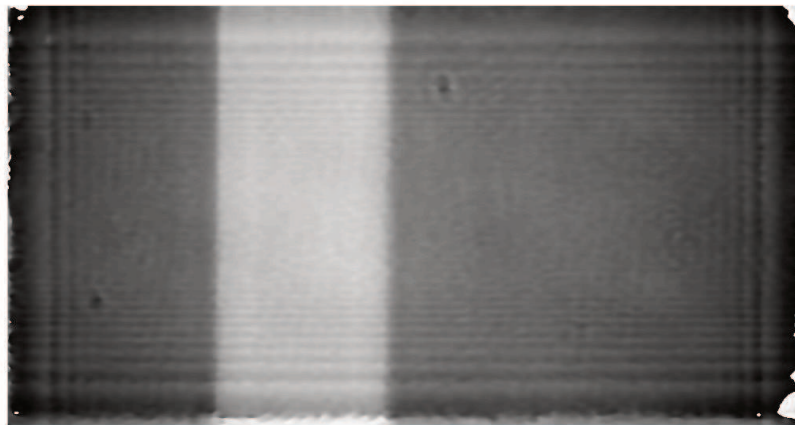


Figura 5.32. Imagen de fase corregida en pistón y *tilt*.

Con la imagen de fase corregida, estamos en condiciones de determinar los valores de la fase de interferencia en la zona de SiO₂ y en las zonas de c-Si.

Dado que el SiO₂ es transparente en la longitud de onda del láser de He-Ne, se produce la refracción de la onda luminosa dentro de dicho material, por lo que debemos considerar en la expresión 5.27 la longitud de la onda que lo atraviesa, definida por el índice de refracción

$$\lambda = \frac{\lambda_0}{n_{SiO_2}} \quad (5.30)$$

A la longitud de onda $\lambda_0 = 632,8 \text{ nm}$, tomando datos de [7], para el SiO_2 tenemos un índice de refracción aproximadamente igual a

$$n_{\text{SiO}_2} \approx 1,54 \quad (5.31)$$

Entonces, para calcular la altura del objeto en la zona del SiO_2 , además de restar la variación de fase por reflexión, debemos considerar el cambio en la longitud de onda provocado por la refracción, de la siguiente manera

$$\Delta z_{\text{SiO}_2}(\xi, \eta) = \frac{\lambda_0}{4\pi n_{\text{SiO}_2}} [\Delta\Phi(\xi, \eta) - \Delta\Phi_{c-\text{Si}}] \quad (5.32)$$

En la Figura 5.33a observamos un esquema de cómo se vería la imagen de fase de una pieza de c-Si con dos zonas, una con un recubrimiento de SiO_2 , y la otra sin recubrir, donde notaríamos una diferencia de fase. En la Figura 5.33b se esquematizan las variaciones de fase por reflexión, en la zona sin recubrir, y por refracción y reflexión en la zona recubierta.

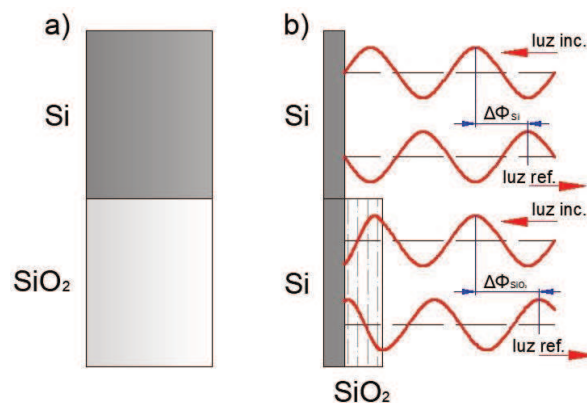


Figura 5.33. a) Esquema de imagen de fase. b) Variaciones de fase por reflexión en c-Si, y por reflexión y refracción en c-Si cubierto con SiO_2 .

Si observamos la imagen de fase de la Figura 5.32, notamos que la zona clara, de mayor fase, corresponden a SiO_2 , y las zonas oscuras, de menor fase, a c-Si.

Para realizar el cálculo estadístico del espesor de la película de SiO_2 , seleccionamos los cuatro segmentos en la imagen de fase que apreciamos en la Figura 5.34, dos correspondientes a la región de c-Si, y dos a la región de SiO_2 , de los cuales obtuvimos los valores de la fase en cada punto.

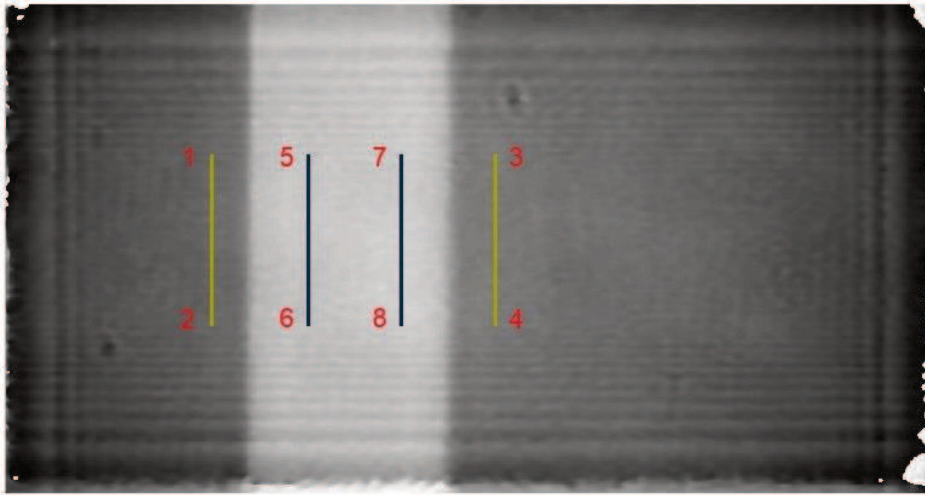


Figura 5.34. Segmentos para el cálculo del espesor de película. Los puntos señalados en pixeles para c-Si $P_1=(750,650)$, $P_2=(750,1420)$, $P_3=(2500,650)$ y $P_4=(2500,1420)$. Para SiO_2 $P_5=(1300,650)$, $P_6=(1300,1420)$, $P_7=(1700,650)$ y $P_8=(1700,1420)$.

Seleccionando las columnas en la matriz de la imagen de fase, determinadas por estos segmentos, tenemos 1544 valores de fase en cada una de las regiones de SiO_2 y c-Si. Luego, estos valores de fase son corregidos, restando a las columnas de cada región la fase correspondiente a la expresión (5.26b). De esta manera obtenemos el valor de la altura de la superficie en la zona de c-Si, para cada elemento de la columna correspondiente, aplicando la ecuación (5.27). Análogamente, la altura de la superficie en la zona de SiO_2 , la obtenemos aplicando la ecuación (5.32) para cada elemento de su correspondiente columna.

La diferencia en estas alturas nos da el espesor Δh de la película de SiO_2

$$\Delta h(\eta) = \Delta z_{\text{SiO}_2}(\xi_{\text{SiO}_2}, \eta) - \Delta z_{\text{c-Si}}(\xi_{\text{c-Si}}, \eta) \quad (5.33)$$

Realizando el histograma de los 1544 valores obtenidos para $\Delta h(\eta)$, como se observa en la Figura 5.35, resulta el siguiente valor medio para el espesor de película delgada de SiO_2 , con su correspondiente desviación estándar

$$\Delta h = 88 \text{ nm} \pm 10 \text{ nm} \quad (5.34)$$

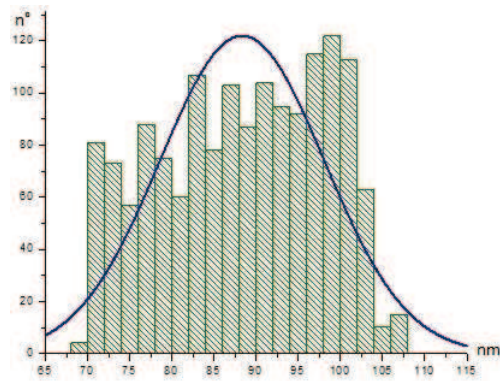


Figura 5.35. Histograma para los espesores de película de SiO₂.

En la Figura 5.36 tenemos la imagen de fase, donde se señalaron el segmento para obtener la gráfica del perfil de la Figura 5.37, y la zona seleccionada para el arreglo tridimensional de la superficie en la Figura 5.38.

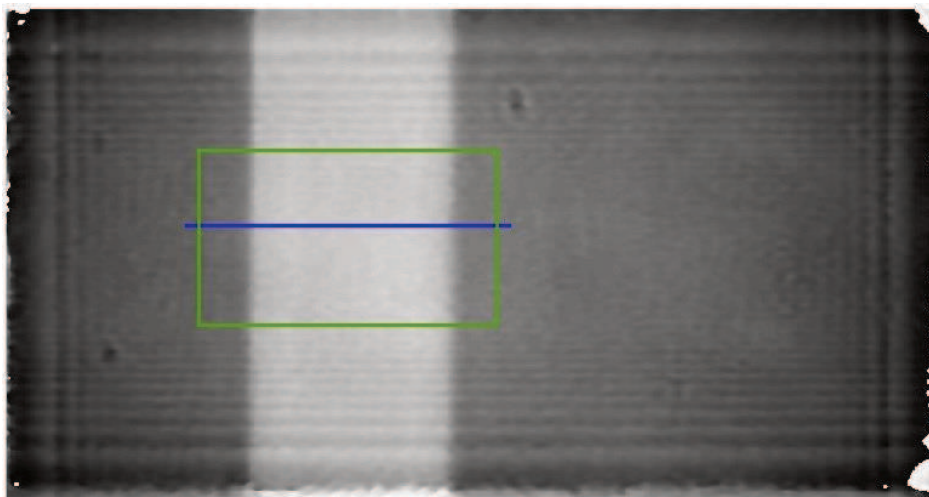


Figura 5.36. Imagen de fase donde se señalan el segmento para la gráfica del perfil y el área para la gráfica tridimensional.

En la imagen del perfil de la Figura 5.37, se evidencian las ondulaciones en la parte superior (la superficie de SiO₂), como así también en la parte inferior (la superficie de c-Si), por los efectos de la difracción, observables también en la imagen tridimensional de la Figura 5.38.

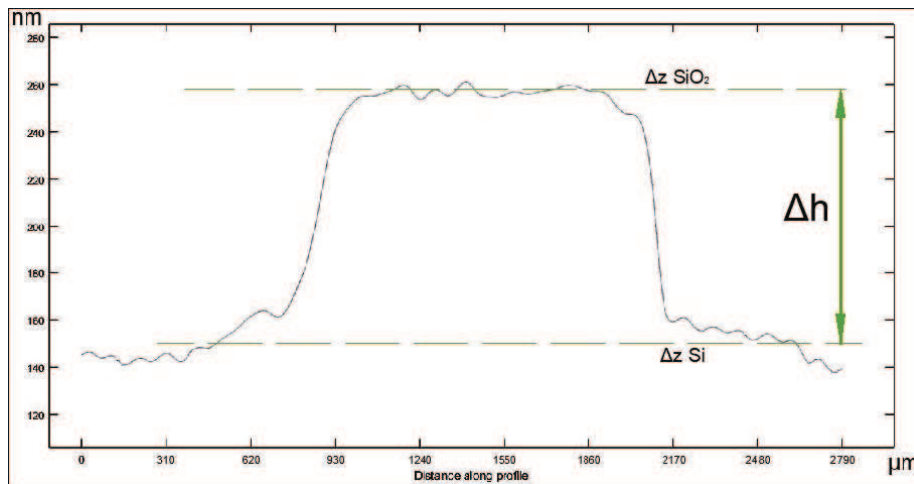


Figura 5.37. Gráfica del perfil del segmento señalado en la figura 5.35, donde se indican las alturas medias de las superficies de SiO₂ y c-Si, y la diferencia entre las mismas que nos señala el espesor Δh de la película de SiO₂ en ese segmento.

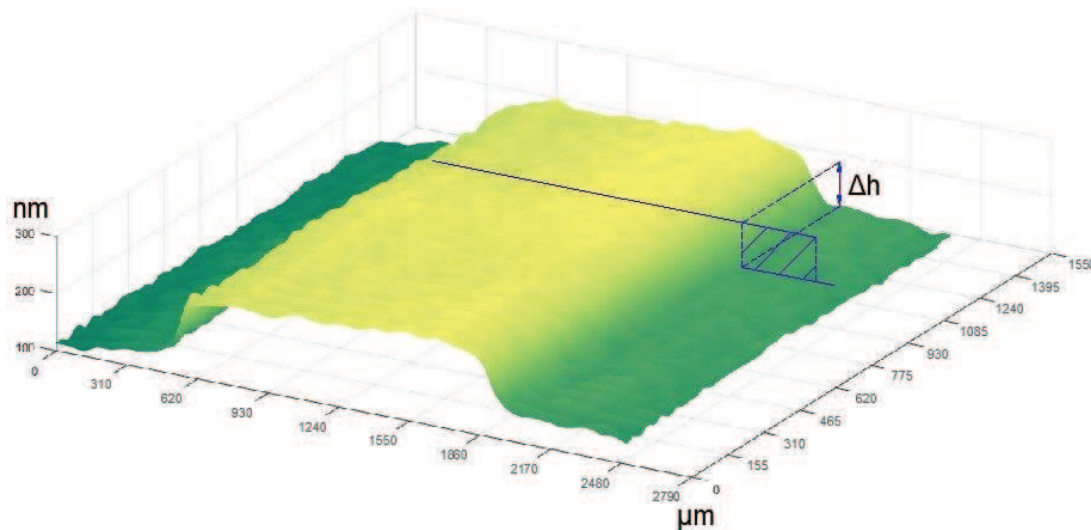


Figura 5.38. Gráfica tridimensional del área señalada en la figura 5.35, donde se indica el espesor Δh de la película de SiO₂ para un determinado segmento.

5.6- Comprobación del procesamiento digital por métodos numéricos:

Podemos simular un holograma en forma numérica, creando una imagen digital que posea un patrón de franjas similar a los obtenidos mediante DHM, como podemos apreciar en la Figura 5.39. De esta forma obtenemos la imagen de objetos inexistentes en la realidad, que poseen dimensiones arbitrarias escogidas de antemano. Dicho holograma fue creado con un período $\Lambda = 6 \text{ pixeles}$, y posee un patrón de franjas verticales para facilitar la visualización del desplazamiento de estas, como vemos en el detalle aumentado de la Figura 5.39. Estos desplazamientos en el patrón de franjas fueron realizados trasladando, una determinada cantidad de pixeles, áreas

seleccionadas específicamente, para simular los tres objetos que alcanzamos a apreciar en el holograma de la Figura 5.39.

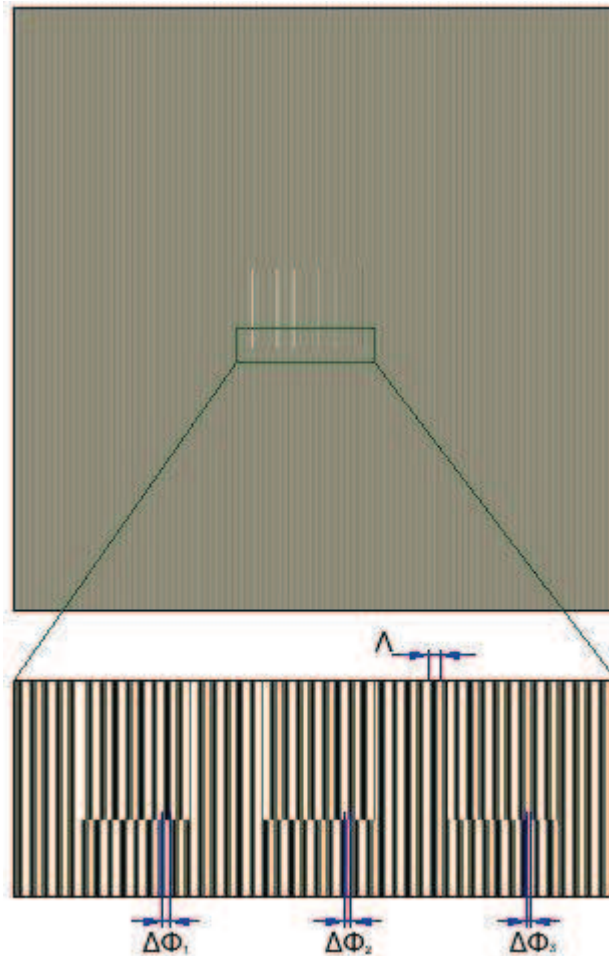


Figura 5.39. Holograma generado numéricamente, donde se simularon tres objetos con distintas diferencias de fase arbitrarias. Tamaño de la imagen 1024 x 1024 píxeles².

El patrón de franjas en cada objeto, de izquierda a derecha, se encuentra desplazado las siguientes cantidades de píxeles, que podemos igualar a radianes de diferencias de fase

$$\Delta\Phi_1 = 5 \text{ píxeles} = \frac{5 \text{ píxel} \cdot 2\pi \text{ rad}}{6 \text{ píxel}} \cong 5,24 \text{ rad} \quad (5.35a)$$

$$\Delta\Phi_2 = 4 \text{ píxeles} = \frac{4 \text{ píxel} \cdot 2\pi \text{ rad}}{6 \text{ píxel}} \cong 4,19 \text{ rad} \quad (5.35b)$$

$$\Delta\Phi_3 = 2 \text{ píxeles} = \frac{2 \text{ píxel} \cdot 2\pi \text{ rad}}{6 \text{ píxel}} \cong 2,09 \text{ rad} \quad (5.35c)$$

Estos radianes de diferencias de fase se transforman en las alturas de los objetos, como vimos mediante la ecuación (5.1), utilizando la longitud de onda del láser de He-Ne, resultando teóricamente iguales a

$$\Delta z_{1teor} = \frac{\lambda}{4\pi} \Delta\Phi_1 = \frac{632,8 \text{ nm}}{4\pi \text{ rad}} 5,24 \text{ rad} \cong 264 \text{ nm} \quad (5.36a)$$

$$\Delta z_{2teor} = \frac{\lambda}{4\pi} \Delta\Phi_2 = \frac{632,8 \text{ nm}}{4\pi \text{ rad}} 4,19 \text{ rad} \cong 211 \text{ nm} \quad (5.36a)$$

$$\Delta z_{3teor} = \frac{\lambda}{4\pi} \Delta\Phi_3 = \frac{632,8 \text{ nm}}{4\pi \text{ rad}} 2,09 \text{ rad} \cong 105 \text{ nm} \quad (5.36a)$$

A continuación procesamos el holograma con el mismo algoritmo utilizado en las muestras anteriormente descritas. Para el filtro pasa bajos se utilizó un radio $r = 80$ pixeles. Por otro lado, debemos aclarar que dado que el holograma es una creación numérica, no existe una distancia de propagación para la onda objeto, por lo que la distancia de reconstrucción se hace igual a cero, $d_c = 0 \text{ cm}$, o sea en el mismo plano del dispositivo CMOS, utilizando una de las ventajas del método de propagación del espectro angular.

De esta forma obtuvimos las imágenes del módulo y la fase que observamos en la Figura 5.40 a y b, respectivamente

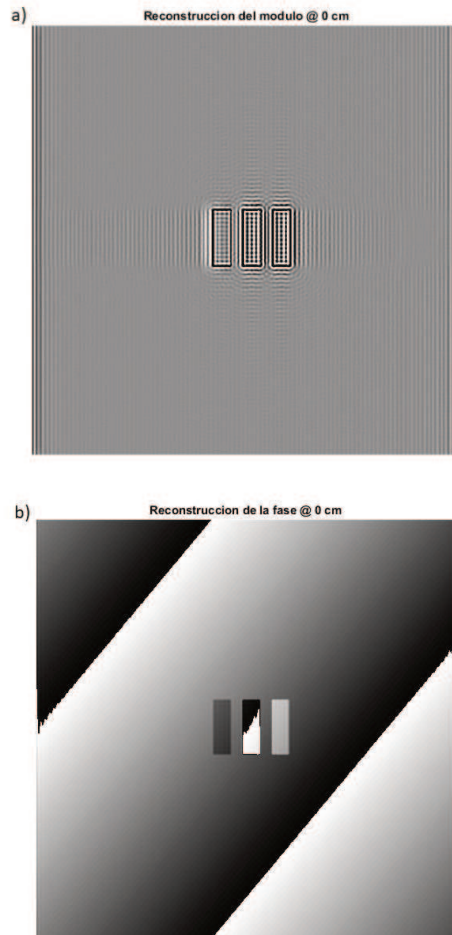


Figura 5.40. a) Imagen del módulo. b) Imagen de fase. Para $d_c = 0 \text{ cm}$.

Como observamos en la imagen de fase, esta posee la inevitable aberración, provocada por la inclinación del supuesto plano objeto con respecto al plano de referencia, necesaria para que exista el patrón de franjas de interferencia que forman el holograma. Para lograr suprimir la aberración, damos al plano de corrección los parámetros iniciales con los siguientes valores $k_0 = 0,003$, $l_0 = 0,003$ y $z_0 = 0$, como en los casos anteriores. Los valores finales para obtener la imagen de fase corregida, resultaron iguales a $k = 0,0075147 \text{ 1/m}$, $l = 0,0055277 \text{ 1/m}$ y $z = -1,9429 \text{ radianes}$.

En la Figura 5.41 vemos la imagen de la fase corregida, donde se distinguen claramente los tres objetos con sus diferentes valores de fase, del fondo de fase uniforme.

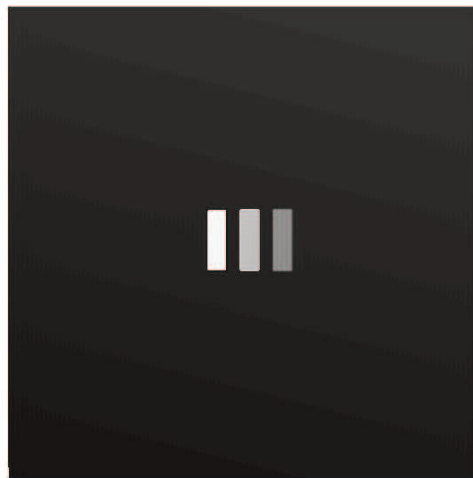


Figura 5.41. Imagen de fase corregida en pistón y *tilt*.

Luego, con esta imagen de fase corregida podemos obtener la gráfica del perfil que observamos en la Figura 5.42, cuya ordenada ya está dada en nm, tomada para un segmento delimitado por los puntos $P_1 = (400,500)$ y $P_2 = (650,500)$ píxeles.

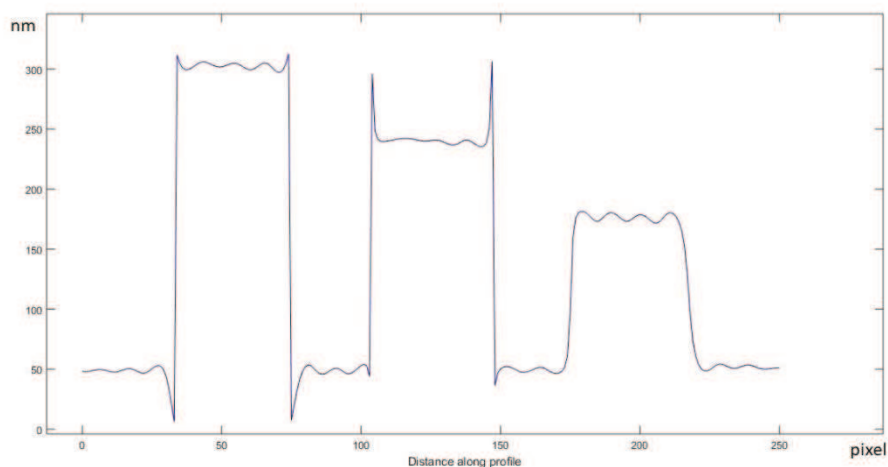


Figura 5.42. Gráfica del perfil del segmento P_1P_2 indicado.

Aquí también notamos en la imagen del perfil las ondulaciones, dado que el procesamiento, como vimos, trabaja con la transformada de Fourier (sección 2.2.9) para obtener las imágenes difractadas sobre el plano CMOS del módulo y la fase, procedentes de un holograma finito. Este efecto es observable también en la imagen tridimensional de la Figura 5.43.

En la gráfica del perfil observamos unos picos en los límites superior e inferior de los objetos izquierdo y central, estos se deben a errores durante el procesamiento de la corrección de la imagen de fase, que produjeron píxeles con mayor intensidad en las zonas superiores, y con menor intensidad en las zonas inferiores de dichos objetos, debido al cambio brusco de fase en los bordes. Esto no sucedió con el objeto de la derecha, que es el de menor altura, dado que el salto de fase es más pequeño. También se observa que el nivel inferior se mantiene en un valor de cota uniforme a lo largo del perfil.

Con los datos del gráfico del perfil podemos calcular las siguientes alturas para los objetos

$$\Delta z_{1med} \cong (300 - 50)nm = 250 \text{ nm} \quad (5.37a)$$

$$\Delta z_{2med} \cong (240 - 50)nm = 190 \text{ nm} \quad (5.37b)$$

$$\Delta z_{3med} \cong (160 - 50)nm = 110 \text{ nm} \quad (5.37c)$$

Las cuales son aproximadas a las alturas predichas por las expresiones (5.36).

En la Figura 5.43 observamos un arreglo tridimensional de los objetos creados numéricamente a través del holograma de la Figura 5.39. En esta imagen los objetos se ven desde el reverso, para una mejor perspectiva, siendo ahora el de la izquierda el de menor altura.

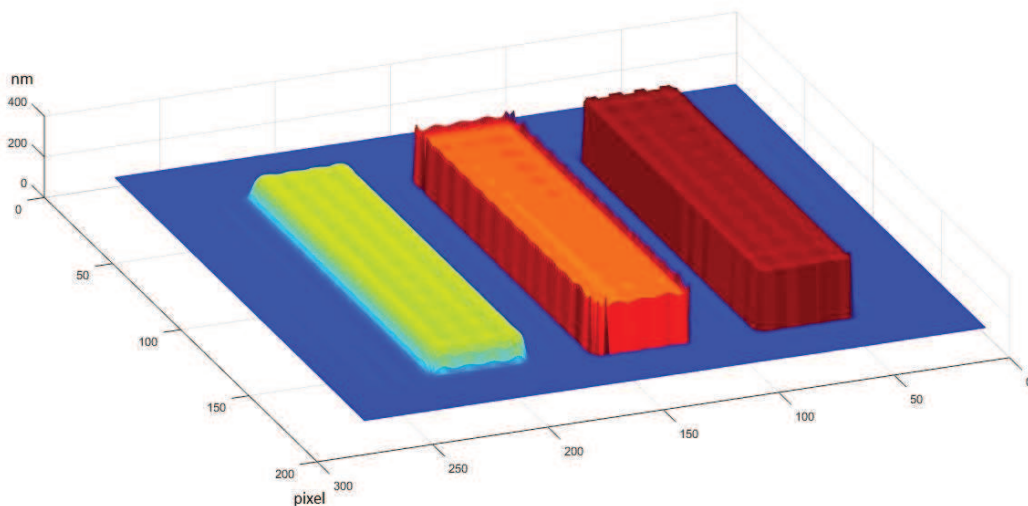


Figura 5.43. Gráfica tridimensional de los objetos generados a partir del holograma digital creado numéricamente.

Referencias:

- [1] M Kim (2011). Digital Holographic Microscopy, Springer N.Y.
- [2] E Cuche, F Bevilacqua, C Depeursinge (1999). Digital Holography for Quantitative Phase-contrast Imaging. Optics Letters, Optical Society of America.
- [3] N Budini, R Urteaga (2016). Determinación Cuantitativa de Fase por Microscopía Holográfica Digital de Reflexión. Informe de Laboratorio. Instituto de Física del Litoral, CONICET.
- [4] S Pertuz, D Puig (2013). Analysis of Focus Measure Operators for Shape-from-Focus. Pattern Recognition, Elsevier.
- [5] V Mahajan (2007). Zernike Polynomials and Wavefront Fitting. Optical Shop Testing, 3th edition, Wiley ed.
- [6] F Wooten (1972). Optical Properties of Solids. Academic Press, NY.
- [7] E Palik (1998). Handbook of Optical Constants of Solids. Academic Press, NY.

Capítulo VI

Conclusiones:

En este trabajo de Tesis de Maestría en Física se propuso implementar la técnica de Microscopía Holográfica Digital (DHM), como un método para medir espesores de películas delgadas, en el orden de los nanómetros.

Se comenzó con un estudio teórico de la óptica física para interpretar el funcionamiento de los sistemas de obtención de imágenes holográficas.

Luego se montó un esquema óptico formando un interferómetro de Michelson, donde se utilizó un láser de He-Ne, con una longitud de onda de 632,8 nm, y una cámara digital CMOS de 8,9 megapíxeles.

Como primer paso fue necesario calibrar el interferómetro, utilizando como objeto una pieza plana de c-Si, con la que se lograron obtener hologramas que poseían un patrón de franjas de interferencia con un período de 14 píxeles. Luego, estos hologramas digitales fueron procesados numéricamente, para obtener las imágenes reconstruidas del módulo y la fase del objeto. Con el objetivo de mejorar la calidad de las imágenes, se emplearon dos filtros digitales, el primero de ellos elimina el término de continua del espectro de frecuencias espaciales, y el segundo se encarga de eliminar los términos de mayor frecuencia, generalmente asociados con el ruido del dispositivo CMOS.

Luego de filtrado el holograma, se procede a la reconstrucción de la imagen del objeto, para lo cual se utilizó el método de propagación del espectro angular, por su robustez frente al ángulo de inclinación del haz objeto con respecto al haz de referencia, y por la ventaja de poder resolver la reconstrucción de la imagen con una distancia de propagación igual a cero. De esta manera, calibrando los elementos del montaje óptico y procesando los hologramas, se obtuvieron imágenes del módulo y la fase del objeto que mostraban una calidad aceptable. Aunque, cabe mencionar, que en la imagen de fase se observaron las aberraciones provocadas, mayormente, por la inclinación, siempre presente, del objeto con respecto al haz de referencia, cosa que fue necesario corregir para lograr nuestro objetivo.

Una vez ajustado el esquema óptico e implementado el algoritmo de procesamiento, se probó esta técnica de DHM en una pieza que poseía una película de Ag, depositada en forma de grilla, sobre un sustrato de c-Si. El espesor de este depósito de Ag había sido medido en un trabajo previo.

Un primer paso para obtener imágenes de fase con buena definición, fue determinar la distancia de reconstrucción focal, por medio de un algoritmo que utiliza un operador de medición de enfoque. Se necesitó trabajar con los 28 operadores disponibles para lograr determinar cuál de todos poseía el mejor desempeño en esta aplicación. Este resultó ser el denominado TENV, que utiliza la variancia del gradiente de la imagen del módulo. Mediante este algoritmo de enfoque se

determinó la distancia de reconstrucción focal de los hologramas, con una precisión de una décima de milímetro. Por ejemplo, para un holograma obtenido a una distancia, medida manualmente, de 24 cm, se obtuvo una distancia focal de 24,35 cm, por medio del algoritmo de análisis, observándose claras diferencias en la definición de las imágenes, tanto de módulo como de fase.

Luego de obtenida la imagen de fase enfocada, se procedió a corregir sus aberraciones, en pistón y *tilt*, por medio de un algoritmo que crea digitalmente un plano, al cual debemos darle la altura en el eje z, y las inclinaciones en los ejes x e y. Seguidamente este plano se suma digitalmente a la imagen de fase. El objetivo es llegar a los valores adecuados de los parámetros en los tres ejes, para lograr minimizar las aberraciones de la imagen de fase. Con este método se lograron obtener imágenes de fase donde se notó una clara uniformidad para los distintos valores de fase de las zonas con Ag y las zonas de c-Si.

Escogiendo una imagen de fase corregida se procedió a determinar el espesor del depósito de Ag, teniendo en cuenta los coeficientes de reflexión de cada material, a la longitud de onda del láser utilizado.

Dado que se tomaron 1234 pixeles de la imagen de fase, tanto en la zona de Ag como en la de c-Si, se pudo determinar estadísticamente la medida de dicho espesor, arrojando un valor medio de 156 nm, con una desviación estándar de 11 nm, resultado que está en acuerdo con el trabajo citado anteriormente, comprobando de esta manera la viabilidad de este método de *DHM* para obtener la medida del espesor de película delgada en el orden de los nanómetros.

Una vez probado el método, procedimos a obtener la medición del espesor de una película delgada de SiO₂ formada sobre un sustrato de c-Si. La muestra fue ubicada a una distancia medida de 10 cm del dispositivo CMOS. Con el procedimiento para determinar la distancia de reconstrucción focal, se obtuvo una distancia óptima de 9,42 cm, lográndose imágenes de módulo y fase bien definidas.

A continuación, se utilizó una imagen de fase para corregir las aberraciones de pistón y *tilt*, logrando una imagen de fase corregida, con una distribución uniforme y bien diferenciada, en los valores de fase, entre las zonas de SiO₂ y c-Si.

Utilizando esta imagen de fase corregida, estuvimos en condiciones de medir el espesor de la película de SiO₂, teniendo en cuenta el índice de refracción de este material, y el coeficiente de reflexión del c-Si, a la longitud de onda del láser de He-Ne. Para ello, se tomaron los valores de fase de 1544 pixeles, en cada una de las zonas de SiO₂ y c-Si, resultando un valor medio para el espesor de la película de 88 nm, con una desviación estándar igual a 10 nm.

La topografía superficial de la muestra se pudo graficar en forma de perfil, de dos dimensiones, y también en una representación tridimensional, evidenciando las diferencias de nivel existente entre el sustrato de c-Si y la película de SiO₂.

Por último, se realizó una comprobación numérica para el procesamiento digital. Con este fin, se creó un holograma digital, que posee tres regiones bien definidas, donde las franjas de interferencia fueron desplazadas en longitudes conocidas. Dichos desplazamientos simulaban diferencias de altura teóricas de 264 nm, 211 nm y 105 nm para cada región respectivamente.

Se procesó la imagen de fase obtenida de este holograma, lográndose una imagen de fase corregida, con la cual se pudieron medir las diferencias de nivel en cada una de las regiones antedichas. Los resultados obtenidos de estas mediciones fueron 250 nm, 190 nm y 110 nm, para cada región, en una aproximación aceptable a los valores teóricos calculados. De esta forma, se logró comprobar la eficacia del procesamiento digital implementado para la medición de espesores nanométricos.

Considerando los resultados obtenidos en este trabajo, podemos decir que se lograron cumplir los siguientes objetivos:

- Montaje y calibración del arreglo óptico en la configuración de interferómetro de Michelson.
- Obtención de hologramas de buena calidad de dos muestras diferentes, la primera de c-Si con un depósito delgado de Ag, y la segunda de c-Si con una película delgada de SiO₂.
- Procesamiento digital de los hologramas, obteniéndose imágenes de módulo y fase filtradas.
- Adopción de un algoritmo de enfoque, para lograr la distancia de reconstrucción óptima, resultando en imágenes de fase con mejor definición.
- Corrección digital de las aberraciones, en *pitch* y *tilt*, de la imagen de fase, consiguiéndose buenas uniformidades en la distribución de los valores de fase, para los distintos materiales y espesores existentes en las muestras.
- Determinación estadística de los espesores de películas delgadas, a partir de las imágenes de fase corregidas.
- confección de gráficos de los perfiles de las superficies de las muestras, a partir de la imagen de fase corregida.
- confección de gráficos tridimensionales de la topografía superficial de las muestras.
- Creación de un holograma numérico, como método de comprobación del procesamiento digital implementado.

Podemos concluir que este método de DHM puede ser utilizado eficazmente en la medición de espesores de películas delgadas, que posean una altura menor a media longitud de onda, siendo la incerteza en las mediciones realizadas de los espesores de un 10 % aproximadamente.

Como objetivos a plantear para futuros trabajos de mejora del método desarrollado en este trabajo, podemos mencionar:

- Corrección digital de aberraciones esféricas en la imagen de fase.
- Desarrollo de algoritmos de demodulación de fase 2π , para medir espesores mayores a media longitud de onda.
- Comparación de resultados con técnicas espectroscópicas.

Apéndice:

Algoritmos de Matlab®

reconstruccion.m

```
% lee el holograma
filename = 'hologramal.bmp';
holo_ini = imread(filename);

% y lo convierte a escala de grises
if size(holo_ini,3) == 3
    holo_ini = rgb2gray(imread(filename));
else
    holo_ini = imread(filename);
end

% muestra el holograma
figure(1)
imshow(holo_ini);

% aplica FFT para ver espectro de frecuencias espaciales
holo_iniff = fftshift(fft2(fftshift(holo_ini)));
figure(2)
imagesc(log(abs(holo_iniff)));

%define tamaño de filas y columnas
[nfil,ncol] = size(holo_ini);

% aplica filtro pasa altos para eliminar la frecuencia 0 (termino de DC)
filtlevel = 50; % mayor valor = menor frecuencia de corte
holo_filtDC = double(holo_ini) - conv2(double(holo_ini),...
    ones(filtlevel),'same')/filtlevel^2;

% muestra el holograma filtrado
figure(3)
imshow(holo_filtDC);

% aplica FFT al holograma filt para ver espectro de frecuencias espaciales
holo_fft = fftshift(fft2(fftshift(holo_filtDC)));
figure(4)
imagesc(log(abs(holo_fft)));

% busca la frecuencia portadora
[filMax,colMax] = find(abs(holo_fft) == max(max(abs(holo_fft))));
r = colMax(2)-ncol/2;

% genera una mascara para filtrar el rango de frecuencias de interes
% llama a la funcion creamascara
creamascara(holo_fft,filMax(2),colMax(2),r)

% espectro con el filtro pasa bajo
holo_fft_mask = holo_fft.*maskfinal;

%imagen del espectro de frecuencias filtrado por la mascara
figure(5)
imagesc(log(abs(holo_fft_mask)));

% centra la frecuencia portadora
rowsToShift = round(nfil/2 - filMax(2));
columnsToShift = round(ncol/2 - colMax(2));
holo_fft_mask_shift = circshift(holo_fft_mask, [rowsToShift columnsToShift]);

%imagen del espectro filtrado centrado
figure(6)
```

```

imagesc(log(abs(holo_fft_mask_shift)));

% se definen los parametros de la propagacion del espectro de frecuencias
pixh = 1.55e-6; % tamaño horizontal de pixel
pixv = 1.55e-6; % tamaño vertical de pixel
lambda = 632.8e-9; % long. de onda
k = 2*pi/lambda; % número de onda
ximg = linspace(-ncol/2,ncol/2,ncol)*pixh; % coordenada x
yimg = linspace(-nfil/2,nfil/2,nfil)*pixv; % coordenada y
[X,Y] = meshgrid(ximg,yimg); % grilla de coordenadas en el plano de propagacion
fn = 1/(ncol*pixh^2); % paso de frecuencia en x
fm = 1/(nfil*pixv^2); % paso de frecuencia en y
fi = k*sqrt(1-(lambda*fn*X).^2-(lambda*fm*Y).^2); % variacion de fase al propagar

% distancia de reconstruccion dada
dcent = 24e-2; % distancia estimada

% se propaga a la distancia dada
G = exp(1i*fi*dcent);

%reconstruccion en el plano imagen
img_rec = ifftshift(ifft2(ifftshift(holo_fft_mask_shift.*G)));

% se extraen modulo y fase a la distancia de propagacion
mod = abs(img_rec); % modulo
fase = angle(img_rec); % fase

% se grafican modulo y fase
figure(7)
imagesc(imadjust(mod/max(max(mod))))
axis image
axis off
title(['Reconstruccion del modulo @ ',num2str(dcent*100),' cm']);
colormap gray

figure(8)
imagesc(fase)
axis image
axis off
title(['Reconstruccion de la fase @ ',num2str(dcent*100),' cm']);
colormap gray

```

creamascara.m

```

function creamascara(holo_fft,fil,col,r0)

[m,n] = size(holo_fft);

[Xmask,Ymask] = meshgrid(-(col-1):(n-col),-(fil-1):(m-fil));

mask0 = ((Xmask.^2+Ymask.^2)<=r0^2);

f = figure;
% colormap gray
h = imagesc((double(mask0~=0)+0.75*double(mask0==0)).*log(abs(holo_fft)));
axis off
axis image
drawnow

button = uicontrol('Style','pushbutton',...
    'String','Terminar',...
    'Position',[5 5 60 25],...
    'Callback',@cierra);

sld1 = uicontrol('Parent',f,...

```



```

    'Style','Slider',...
    'Units','Normalized',...
    'Position',[0.1,0.08,0.8,0.02],...
    'Value',r0,...
    'Min',0,...
    'Max',4*r0,...
    'SliderStep',[0.01 0.05],...
    'Callback',@actualizasld1);

txt = uicontrol('Style','Text',...
    'String',['Radio de la mascara = ',num2str(sld1.Value)],...
    'Position',[85 5 300 25],...
    'Callback',@actualizatxt);

function actualizasld1(sourcesld1,~)
    r = sourcesld1.Value;

    mask = ((Xmask.^2+Ymask.^2)<=r^2);

    % actualiza imagenes
    h.CData = (double(mask~=0)+0.75*double(mask==0)).*log(abs(holo_fft));

    actualizatxt;
end

function cierra(~,~)
    % actualiza variables en el workspace y cierra
    assignin('base','maskfinal',double(((Xmask.^2+Ymask.^2)<=round(sld1.Value)^2)));
    assignin('base','rfinal',round(sld1.Value));
    close(f)
end

function actualizatxt(~,~)
    % actualiza variables en el workspace y cierra
    set(txt,'String',['radio de la mascara = ',num2str(sld1.Value)]);
end

end

```

autofoco.m

```

function [dopt,FM] = autofoco(dcent,rango,paso,fi,holo_fft_mask_shift)
% calculos del rango de distancias a evaluar
dini = dcent-rango;
dfin = dcent+rango;
dist = dini:paso:dfin;

% propagacion a la distancia estimada dcent
G = exp(1i*fi*dcent);
img_rec_dcent = ifftshift(ifft2(ifftshift(G.*holo_fft_mask_shift)));

% seleccion del area de interes
figure(7)
h = imagesc(abs(img_rec_dcent));
colormap gray
axis image
axis off
[~,ROI] = imcrop(h);
close(gcf)

% defino vector donde se guardara la medida de foco
FM = nan(size(dist));
% se prepara la figura
figure
h = plot(dist,FM);

```

```

title('Análisis de foco')
xlabel('Distancia')
ylabel('Medida de foco')
% loop de propagación a las sucesivas distancias y medida del foco
for i = 1:length(dist)
    G = exp(1i*fi*dist(i)); % factor de propagación de fase
    img_rec = ifftshift(iff2(iff2shift(G.*holo_fft_mask_shift)));
    moddist = abs(img_rec); % módulo
    FM(i) = fmeasure(moddist, 'TENV', ROI);
    h.YData = FM;
    drawnow expose
end
dopt = dist(FM==max(FM));
end

```

fmeasure.m (extraído de Pertuz, 2016)

```

function FM = fmeasure(Image, Measure, ROI)

if nargin>2 && ~isempty(ROI)
    Image = imcrop(Image, ROI);
end

WSize = 15; % Size of local window (only some operators)

switch upper(Measure)
    case 'ACMO' % Absolute Central Moment (Shirvaikar2004)
        if ~isinteger(Image), Image = im2uint8(Image);
        end
        FM = AcMomentum(Image);

    case 'BREN' % Brenner's (Santos97)
        [M, N] = size(Image);
        DH = zeros(M, N);
        DV = zeros(M, N);
        DV(1:M-2, :) = Image(3:end, :) - Image(1:end-2, :);
        DH(:, 1:N-2) = Image(:, 3:end) - Image(:, 1:end-2);
        FM = max(DH, DV);
        FM = FM.^2;
        FM = mean2(FM);

    case 'CONT' % Image contrast (Nanda2001)
        ImContrast = @(x) sum(abs(x(:)-x(5)));
        FM = nlfilter(Image, [3 3], ImContrast);
        FM = mean2(FM);

    case 'CURV' % Image Curvature (Helmlil2001)
        if ~isinteger(Image), Image = im2uint8(Image);
        end
        M1 = [-1 0 1; -1 0 1; -1 0 1];
        M2 = [1 0 1; 1 0 1; 1 0 1];
        P0 = imfilter(Image, M1, 'replicate', 'conv')/6;
        P1 = imfilter(Image, M1, 'replicate', 'conv')/6;
        P2 = 3*imfilter(Image, M2, 'replicate', 'conv')/10 ...
            -imfilter(Image, M2, 'replicate', 'conv')/5;
        P3 = -imfilter(Image, M2, 'replicate', 'conv')/5 ...
            +3*imfilter(Image, M2, 'replicate', 'conv')/10;
        FM = abs(P0) + abs(P1) + abs(P2) + abs(P3);
        FM = mean2(FM);

    case 'DCTE' % DCT energy ratio (Shen2006)
        FM = nlfilter(Image, [8 8], @DctRatio);
        FM = mean2(FM);

    case 'DCTR' % DCT reduced energy ratio (Lee2009)
        FM = nlfilter(Image, [8 8], @ReRatio);

```

```

FM = mean2(FM);

case 'GDER' % Gaussian derivative (Geusebroek2000)
    N = floor(WSize/2);
    sig = N/2.5;
    [x,y] = meshgrid(-N:N, -N:N);
    G = exp(-(x.^2+y.^2)/(2*sig^2))/(2*pi*sig);
    Gx = -x.*G/(sig^2);
    Gx = Gx/sum(abs(Gx(:)));
    Gy = -y.*G/(sig^2);
    Gy = Gy/sum(abs(Gy(:)));
    Rx = imfilter(double(Image), Gx, 'conv', 'replicate');
    Ry = imfilter(double(Image), Gy, 'conv', 'replicate');
    FM = Rx.^2+Ry.^2;
    FM = mean2(FM);

case 'GLVA' % Graylevel variance (Krotkov86)
    FM = std2(Image);

case 'GLLV' %Graylevel local variance (Pech2000)
    LVar = stdfilt(Image, ones(WSize,WSize)).^2;
    FM = std2(LVar)^2;

case 'GLVN' % Normalized GLV (Santos97)
    FM = std2(Image)^2/mean2(Image);

case 'GRAE' % Energy of gradient (Subbarao92a)
    Ix = Image;
    Iy = Image;
    Iy(1:end-1,:) = diff(Image, 1, 1);
    Ix(:,1:end-1) = diff(Image, 1, 2);
    FM = Ix.^2 + Iy.^2;
    FM = mean2(FM);

case 'GRAT' % Thresholded gradient (Snatos97)
    Th = 0; %Threshold
    Ix = Image;
    Iy = Image;
    Iy(1:end-1,:) = diff(Image, 1, 1);
    Ix(:,1:end-1) = diff(Image, 1, 2);
    FM = max(abs(Ix), abs(Iy));
    FM(FM<Th)=0;
    FM = sum(FM(:))/sum(sum(FM~=0));

case 'GRAS' % Squared gradient (Eskicioglu95)
    Ix = diff(Image, 1, 2);
    FM = Ix.^2;
    FM = mean2(FM);

case 'HELM' %Helml's mean method (Helml2001)
    MEANF = fspecial('average',[WSize WSize]);
    U = imfilter(Image, MEANF, 'replicate');
    R1 = U./Image;
    R1(Image==0)=1;
    index = (U>Image);
    FM = 1./R1;
    FM(index) = R1(index);
    FM = mean2(FM);

case 'HISE' % Histogram entropy (Krotkov86)
    FM = entropy(Image);

case 'HISR' % Histogram range (Firestone91)
    FM = max(Image(:))-min(Image(:));

```

```

case 'LAPE' % Energy of laplacian (Subbarao92a)
    LAP = fspecial('laplacian');
    FM = imfilter(Image, LAP, 'replicate', 'conv');
    FM = mean2(FM.^2);

case 'LAPM' % Modified Laplacian (Nayar89)
    M = [-1 2 -1];
    Lx = imfilter(Image, M, 'replicate', 'conv');
    Ly = imfilter(Image, M, 'replicate', 'conv');
    FM = abs(Lx) + abs(Ly);
    FM = mean2(FM);

case 'LAPV' % Variance of laplacian (Pech2000)
    LAP = fspecial('laplacian');
    ILAP = imfilter(Image, LAP, 'replicate', 'conv');
    FM = std2(ILAP)^2;

case 'LAPD' % Diagonal laplacian (Thelen2009)
    M1 = [-1 2 -1];
    M2 = [0 0 -1;0 2 0;-1 0 0]/sqrt(2);
    M3 = [-1 0 0;0 2 0;0 0 -1]/sqrt(2);
    F1 = imfilter(Image, M1, 'replicate', 'conv');
    F2 = imfilter(Image, M2, 'replicate', 'conv');
    F3 = imfilter(Image, M3, 'replicate', 'conv');
    F4 = imfilter(Image, M1, 'replicate', 'conv');
    FM = abs(F1) + abs(F2) + abs(F3) + abs(F4);
    FM = mean2(FM);

case 'SFIL' %Steerable filters (Minhas2009)
    % Angles = [0 45 90 135 180 225 270 315];
    N = floor(WSize/2);
    sig = N/2.5;
    [x,y] = meshgrid(-N:N, -N:N);
    G = exp(-(x.^2+y.^2)/(2*sig^2))/(2*pi*sig);
    Gx = -x.*G/(sig^2);Gx = Gx/sum(Gx(:));
    Gy = -y.*G/(sig^2);Gy = Gy/sum(Gy(:));
    R(:,:,1) = imfilter(double(Image), Gx, 'conv', 'replicate');
    R(:,:,2) = imfilter(double(Image), Gy, 'conv', 'replicate');
    R(:,:,3) = cosd(45)*R(:,:,1)+sind(45)*R(:,:,2);
    R(:,:,4) = cosd(135)*R(:,:,1)+sind(135)*R(:,:,2);
    R(:,:,5) = cosd(180)*R(:,:,1)+sind(180)*R(:,:,2);
    R(:,:,6) = cosd(225)*R(:,:,1)+sind(225)*R(:,:,2);
    R(:,:,7) = cosd(270)*R(:,:,1)+sind(270)*R(:,:,2);
    R(:,:,8) = cosd(315)*R(:,:,1)+sind(315)*R(:,:,2);
    FM = max(R,[],3);
    FM = mean2(FM);

case 'SFRQ' % Spatial frequency (Eskicioglu95)
    Ix = Image;
    Iy = Image;
    Ix(:,1:end-1) = diff(Image, 1, 2);
    Iy(1:end-1,:) = diff(Image, 1, 1);
    FM = mean2(sqrt(double(Iy.^2+Ix.^2)));

case 'TENGR' % Tenengrad (Krotkov86)
    Sx = fspecial('sobel');
    Gx = imfilter(double(Image), Sx, 'replicate', 'conv');
    Gy = imfilter(double(Image), Sx, 'replicate', 'conv');
    FM = Gx.^2 + Gy.^2;
    FM = mean2(FM);

case 'TENV' % Tenengrad variance (Pech2000)
    Sx = fspecial('sobel');
    Gx = imfilter(double(Image), Sx, 'replicate', 'conv');
    Gy = imfilter(double(Image), Sx, 'replicate', 'conv');
    G = Gx.^2 + Gy.^2;
    FM = std2(G)^2;

```

```

case 'VOLA' % Vollaath's correlation (Santos97)
    Image = double(Image);
    I1 = Image; I1(1:end-1,:) = Image(2:end,:);
    I2 = Image; I2(1:end-2,:) = Image(3:end,:);
    Image = Image.*(I1-I2);
    FM = mean2(Image);

case 'WAVS' %Sum of Wavelet coeffs (Yang2003)
    [C,S] = wavedec2(Image, 1, 'db6');
    H = wrcoef2('h', C, S, 'db6', 1);
    V = wrcoef2('v', C, S, 'db6', 1);
    D = wrcoef2('d', C, S, 'db6', 1);
    FM = abs(H) + abs(V) + abs(D);
    FM = mean2(FM);

case 'WAVV' %Variance of Wav...(Yang2003)
    [C,S] = wavedec2(Image, 1, 'db6');
    H = abs(wrcoef2('h', C, S, 'db6', 1));
    V = abs(wrcoef2('v', C, S, 'db6', 1));
    D = abs(wrcoef2('d', C, S, 'db6', 1));
    FM = std2(H)^2+std2(V)+std2(D);

case 'WAVR'
    [C,S] = wavedec2(Image, 3, 'db6');
    H = abs(wrcoef2('h', C, S, 'db6', 1));
    V = abs(wrcoef2('v', C, S, 'db6', 1));
    D = abs(wrcoef2('d', C, S, 'db6', 1));
    A1 = abs(wrcoef2('a', C, S, 'db6', 1));
    A2 = abs(wrcoef2('a', C, S, 'db6', 2));
    A3 = abs(wrcoef2('a', C, S, 'db6', 3));
    A = A1 + A2 + A3;
    WH = H.^2 + V.^2 + D.^2;
    WH = mean2(WH);
    WL = mean2(A);
    FM = WH/WL;
otherwise
    error('Unknown measure %s', upper(Measure))
end
end

function fm = AcMomentum(Image)
[M, N] = size(Image);
Hist = imhist(Image)/(M*N);
Hist = abs((0:255)-mean2(Image)).*Hist;
fm = sum(Hist);
end

function fm = DctRatio(M)
MT = dct2(M).^2;
fm = (sum(MT(:))-MT(1,1))/MT(1,1);
end

function fm = ReRatio(M)
M = dct2(M);
fm = (M(1,2)^2+M(1,3)^2+M(2,1)^2+M(2,2)^2+M(3,1)^2)/(M(1,1)^2);
end

```

corrigefase_plano.m

```
function corrigefase_plano(fasefoco,k0,l0,z0)
[m,n] = size(fasefoco);
x = 1:n;
y = 1:m;
[X,Y] = meshgrid(x,y);
offset0 = mod(z0 + k0*X + l0*Y,2*pi);
fase_final0 = mod(fasefoco + pi + offset0,2*pi);

f = figure;
colormap gray
s1 = subplot(131);
h1 = imagesc(fasefoco);
axis(s1,'off')
axis(s1,'image')
s2 = subplot(132);
h2 = imagesc(offset0);
axis(s2,'off');
axis(s2,'image');
s3 = subplot(133);
h3 = imagesc(fase_final0);
axis(s3,'off');
axis(s3,'image');

button = uicontrol('Style','pushbutton',...
    'String','Terminar',...
    'Position',[5 5 60 25],...
    'Callback',@cierra);

sld1 = uicontrol('Parent',f,...
    'Style','Slider',...
    'Units','Normalized',...
    'Position',[0.1,0.2,0.8,0.02],...
    'Value',k0,...
    'Min',-10*k0,...
    'Max',10*k0,...
    'SliderStep',[1 5]*1e-3,...
    'Callback',@actualizasld1);
sld2 = uicontrol('Parent',f,...
    'Style','Slider',...
    'Units','Normalized',...
    'Position',[0.1,0.1,0.8,0.02],...
    'Value',l0,...
    'Min',-10*l0,...
    'Max',10*l0,...
    'SliderStep',[1 5]*1e-3,...
    'Callback',@actualizasld2);
sld3 = uicontrol('Parent',f,...
    'Style','Slider',...
    'Units','Normalized',...
    'Position',[0.1,0.3,0.8,0.02],...
    'Value',z0,...
    'Min',-pi,...
    'Max',pi,...
    'SliderStep',[1 5]*1e-3,...
    'Callback',@actualizasld3);

txt = uicontrol('Style','Text',...
    'String',['k = ',num2str(sld1.Value),'; l = ',num2str(sld2.Value),'; z = ',num2str(sld3.Value)],...
    'Position',[85 5 300 25],...
    'Callback',@actualizatxt);

function actualizasld1(sourcesld1,callbackdatasld1)
    k = sourcesld1.Value;
    l = get(sld2,'Value');
```

```

z = get(sld3, 'Value');

offset = mod(k*X + l*Y + z, 2*pi);
fase_final = mod(fasefoco + pi + offset, 2*pi);

% actualiza imagenes
h2.CData = offset;
h3.CData = fase_final;

actualizatxt;
end

function actualizasld2(sourcesld2, callbackdatasld2)
k = get(sld1, 'Value');
l = sourcesld2.Value;
z = get(sld3, 'Value');

offset = mod(k*X + l*Y + z, 2*pi);
fase_final = mod(fasefoco + pi + offset, 2*pi);

% actualiza imagenes
h2.CData = offset;
h3.CData = fase_final;

actualizatxt;
end

function actualizasld3(sourcesld3, callbackdatasld3)
k = get(sld1, 'Value');
l = get(sld2, 'Value');
z = sourcesld3.Value;

offset = mod(k*X + l*Y + z, 2*pi);
fase_final = mod(fasefoco + pi + offset, 2*pi);

% actualiza imagenes
h2.CData = offset;
h3.CData = fase_final;

actualizatxt;
end

function cierra(sourcebtn, callbackdatbtn)
% actualiza variables en el workspace y cierra
assignin('base', 'fasefinal', h3.CData);
assignin('base', 'kfinal', sld1.Value);
assignin('base', 'lfinal', sld2.Value);
assignin('base', 'zfinal', sld3.Value);

close(f)
end

function actualizatxt(sourcetxt, callbackdatatxt)
% actualiza variables en el workspace y cierra
set(txt, 'String', ['k = ', num2str(sld1.Value), '; l = ', num2str(sld2.Value), '; z = ', num2str(sld3.Value)]);

end
end

```