

UNIVERSIDAD COMPLUTENSE DE MADRID

FACULTAD DE CIENCIAS FÍSICAS
Departamento de Óptica



**GENERACIÓN DE SEÑALES ÓPTICAS DE REFERENCIA
PARA POSICIONAMIENTO ABSOLUTO DE ALTA
RESOLUCIÓN**

**MEMORIA PARA OPTAR AL GRADO DE DOCTOR
PRESENTADA POR**

José Sáez Landete

Bajo la dirección de los doctores:
Eusebio Bernabeu y José Alonso

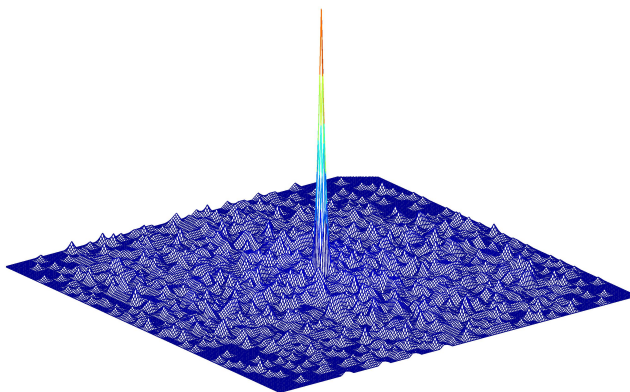
Madrid, 2005

ISBN: 978-84-669-2927-1



UNIVERSIDAD COMPLUTENSE DE MADRID
DEPARTAMENTO DE ÓPTICA
FACULTAD DE CIENCIAS FÍSICAS

**GENERACIÓN DE SEÑALES ÓPTICAS DE
REFERENCIA PARA POSICIONAMIENTO
ABSOLUTO DE ALTA RESOLUCIÓN**



Memoria presentada para optar al grado de Doctor
por José Sáez Landete
Madrid, Mayo de 2006

A mis padres, y a mi hermana

¿Por qué esta magnífica tecnología, que ahorra trabajo y nos hace la vida mas fácil, nos aporta tan poca felicidad? La repuesta es, simplemente: porque aún no hemos aprendido a usarla con acierto.

(Albert Einstein)

Agradecimientos

Deseo expresar mi agradecimiento a los directores de este trabajo, el Prof. Eusebio Bernabeu y el Dr. José Alonso. Quiero agradecer el constante apoyo tanto en lo profesional como en lo personal durante estos últimos años. En primer lugar, agradezco al Prof. Bernabeu la confianza depositada en mi al integrarme como miembro de su equipo de investigación. Debo agradecer a José Alonso buena parte de lo que he aprendido, su profundo conocimiento en muchas ramas de la física ha sido la referencia principal de este trabajo.

Así mismo, quiero agradecer a Tomás Morlanes y Jose Luis de la Peña la supervisión y orientación en el trabajo, gracias a sus acertadas sugerencias hemos conseguido obtener la mayor parte de los resultados de este trabajo. En general, quiero agradecer a Fagor Automation S. Coop., y en particular a Pedro Ruiz de Aguirre la buena acogida que siempre han tenido a las nuevas ideas. Su planteamiento valiente y acertado en I+D+i han sido claves para obtener los resultados de este trabajo.

También quiero agradecer la calurosa acogida y el aliento que he recibido en el Departamento de Teoría de la Señal y Comunicaciones de la Universidad de Alcalá.

Agradezco la hospitalidad dispensada en el Laboratorio de Aplicaciones Ópticas de la Universidad de Buenos Aires, que hizo sentirme como en casa durante una estancia breve. En especial agradezco la colaboración a Fernando Pérez y Ariel Lutenberg.

Por supuesto, quiero agradecer la ayuda y el ánimo de mis compañeros de “la cueva”. A los que estaban cuando llegué y a los que se han ido incorporando. Todos ellos son los responsables de crear un ambiente de trabajo distendido y divertido. Debo agradecer, tanto el apoyo técnico, como las interesantes discusiones, en las comidas, y en el bar... En particular, aprecio la ayuda de José Antonio, Luis Miguel, Natalia, Rubén, Quiroga...

También debo mostrar mi gratitud a mis amigos, compañeros de piso y huéspedes de “el albergue”, con quienes he compartido horas de agradables discusiones. Entre ellos, Isabel M^a, Justo, Sergio, Javier, Rebeca, Enrique, Viqui, Sylvain, Fleur, Susanne... También deseo agradecer a mis amigos del pueblo los buenos ratos compartidos a lo largo de toda mi vida.

Finalmente, tengo que dar las gracias a mi familia. Su apoyo y ayuda incondicional ha sido imprescindible, a pesar de la distancia y de mis dificultades para explicarles a qué me estaba dedicando durante estos últimos años.

Índice general

Introducción	XI
1. Señales ópticas de referencia	1
1.1. Codificadores ópticos y señales de referencia	1
1.1.1. Codificadores ópticos absolutos	2
1.1.2. Codificadores ópticos incrementales	2
1.1.3. Estándares desarrollados en torno a los sistemas de codificación óptica	4
1.2. Fotolitografía y señales de referencia.	7
1.3. Características de las señales de referencia	9
2. Diseño de ZRCs unidimensionales	11
2.1. Introducción	11
2.2. Consideraciones generales	12
2.3. Optimización con el algoritmo DIRECT	16
2.3.1. Descripción de DIRECT	16
2.3.2. Resultados	17
2.4. Optimización mediante un algoritmo genético	20
2.4.1. Descripción de los algoritmos genéticos	21
2.4.2. Resultados	22
2.5. Conclusiones sobre diseño de ZRCs unidimensionales	23
3. Diseño de ZRCs bidimensionales	27
3.1. Introducción	27
3.2. Consideraciones generales	29
3.3. Cota inferior del segundo máximo	31
3.4. Optimización con el algoritmo DIRECT	35
3.5. Optimización mediante un algoritmo genético	37
3.6. Conclusiones sobre diseño de ZRCs bidimensionales	39
4. Estrechamiento de señales de referencia	41
4.1. Introducción	41
4.2. Generación del perfil de irradiancia en un codificador de efecto Lau	43
4.2.1. Cálculo del perfil de irradiancia	44

4.2.2. Máscara de estrechamiento	46
4.3. Diseño de la máscara de detección	47
4.4. Resultados experimentales	50
4.5. Conclusiones sobre el estrechamiento de señales de referencia	55
Conclusiones	57
A. Lista de publicaciones	59
A.1. Publicaciones	59
A.2. Comunicaciones a congresos	59

Introducción

La medida precisa del desplazamiento y la posición es imprescindible en las tecnologías implicadas en la automatización, cuyo cometido es la mejora de la eficiencia en la fabricación y la caracterización de procesos en todas las ramas de la industria y en gran parte de la instrumentación científica avanzada. Existe una enorme variedad de sistemas, dispositivos o sensores que miden el desplazamiento de unos objetos respecto a otros. Entre todos los dispositivos capaces de medir el desplazamiento, un grupo especialmente importante es el de los codificadores o transductores de movimiento. Un codificador es un dispositivo que mide en tiempo real el desplazamiento de un objeto respecto de otro, normalmente entre partes móviles de un sistema, del cual el codificador forma parte. Los codificadores informan a los operarios de las máquinas que los incluyen —o a los sistemas de control que gobiernan dichas máquinas— de la posición relativa o absoluta de las partes móviles. Esta información es necesaria para controlar el trabajo que realiza la máquina de manera automática.

Entre las innumerables aplicaciones en las que los codificadores son imprescindibles se encuentran: máquina-herramienta (tornos, fresadoras, . . .), orientación de sistemas de apuntamiento (antenas, radar, telescopios, . . .), sistemas robotizados (cadenas de producción, sistemas autónomos, . . .), sistemas informáticos (impresoras, scanners, . . .); en definitiva, cualquier sistema de movimiento controlado y control numérico (“motion control” y “numerical control”). Según su aplicación, los codificadores están basados en diferentes fenómenos y métodos de medida. En las aplicaciones de alta resolución son de notable importancia los codificadores basados en fenómenos ópticos, denominados codificadores ópticos. Son este tipo de codificadores los que forman la base de la ingeniería de precisión y micro-nano-tecnología. El desarrollo de la codificación óptica en términos de resolución, precisión, fiabilidad y rango de medida se ha convertido en un área de gran importancia estratégica. Es, por tanto, un campo en el que se realiza un gran esfuerzo en I+D+i, el cual genera continuos avances según la creciente demanda de la industria.

Por otro lado, la fotolitografía es el proceso por el cual un diagrama grabado en una máscara se transfiere a la superficie de un sustrato por medio de una radiación luminosa. Se pueden utilizar diferentes sustrato, entre los cuales, el más típico es una oblea de silicio. En la oblea se deposita una capa uniforme de resina fotosensible, que es expuesta selectivamente mediante el diagrama grabado en la máscara. Para la exposición se usa una fuente de luz con longitud

de onda –o anchura espectral– adecuada a la sensibilidad espectral de la resina. Como en fotografía, la resina expuesta es revelada, quedando sobre la superficie de la oblea una capa en la que se ha grabado el diagrama deseado. Cada uno de estos pasos descritos requieren complejos procesos químicos y físicos, como pueden ser: la preparación del sustrato, baños de “recocido”, aplicación de adherentes, deposición de la resina (coating), etc... La fotolitografía es la base del desarrollo de la industria microelectrónica, ya que se utiliza en cada fase de la fabricación de un dispositivo semiconductor. Tanto es así, que la fotolitografía supone aproximadamente un tercio del coste total de cada dispositivo. Hay un fuerte incentivo económico en mejorar la resolución de la litografía, ya que esta mejora se traduce directamente en la miniaturización de los circuitos y en un aumento de las velocidades de procesamiento. De hecho, las diferentes generaciones de litografía coinciden con las diferentes generaciones de módulos de memoria (DRAM).

Tanto en codificación óptica como en fotolitografía existen analogías y problemas comunes. En este trabajo hemos analizado un aspecto fundamental en estos dos campos, la detección de la posición absoluta. En los dos ámbitos de aplicación pueden utilizarse sistemas muy similares para realizar esta tarea. Los sistemas de detección de la posición absoluta suelen utilizar una señal de control, comúnmente denominada señal de referencia en codificación y señal de alineamiento en fotolitografía. La diferencia fundamental entre estas dos señales es que en codificación óptica es unidimensional mientras que en fotolitografía es bidimensional. En la detección absoluta de la posición podemos distinguir tres etapas: en primer lugar la generación de la señal de referencia –o alineamiento–, después la etapa de detección de la señal, y finalmente una etapa de procesamiento y control electrónico.

El objetivo principal del presente trabajo consiste en la exploración, estudio y mejora de los subsistemas de generación y detección de señales de referencia y alineamiento. Para ello, se analizan los problemas existentes en la tecnología actual, tanto en codificadores ópticos como en alineadores de máscaras para fotolitografía. La memoria está organizada de la siguiente manera: en el capítulo 1 se presenta una breve descripción de los fundamentos de la codificación óptica, así como de alineamiento de máscaras. Se definen los conceptos y magnitudes que caracterizan este tipo de señales, señalando las analogías en los dos ámbitos de aplicación. En el capítulo 2 se analiza en profundidad las señales de referencia en codificadores ópticos. El objetivo de este capítulo es superar las dificultades existentes en el subsistema de generación de este tipo de señales. En el capítulo 3 se explora el estado de la técnica en señales de alineamiento utilizadas en alineadores de máscaras. El objetivo es proponer nuevos diseños de señales de alineamiento aplicando una generalización de las técnicas desarrolladas en el capítulo 2. Finalmente, en el capítulo 4 se aborda el problema que ocasiona la difracción en señales de referencia en codificadores ópticos. Se establece un método para evitar este problema, desarrollando un subsistema novedoso de detección de la señal de referencia.

El trabajo presentado en esta Memoria se enmarca dentro de la colaboración entre el Departamento de Óptica de la Universidad Complutense de Madrid y

Fagor Automation S. Coop., la única empresa española que diseña y fabrica codificadores ópticos y uno de los líderes mundiales en el campo de la codificación óptica. En el seno de esta colaboración, a lo largo de los últimos años, se han ido definiendo las prestaciones que deben ofrecer los codificadores ópticos para responder a las necesidades de la industria. Así mismo, se han ido determinando los factores limitantes de los sistemas ópticos que se emplean convencionalmente. Este trabajo surge como la continuación de una labor de muchos años, aprovechando la experiencia acumulada anteriormente para buscar soluciones novedosas que satisfagan las necesidades actuales y futuras de la industria. Este trabajo ha sido realizado con el soporte económico del Ministerio de Educación y Ciencia, de los proyectos DPI2001-1238 y DPI2005-02860 y la colaboración de Fagor Automation S. Coop.

Capítulo 1

Señales ópticas de referencia

Ante la enorme variedad de codificadores ópticos, existen varios tipos de clasificaciones atendiendo a las diferentes características de funcionamiento o tipos de medida realizadas. En este primer capítulo presentamos un breve análisis sobre los codificadores ópticos, mostrando una clasificación de codificadores ópticos que permite centrarnos en el tipo que interesa en esta memoria. Analizamos la característica que motiva esta clasificación y que es el objetivo fundamental de esta memoria, la señal de referencia. Exponemos algunos conceptos básicos y la terminología relativa a los codificadores ópticos que utilizaremos en los capítulos siguientes. Por otro lado, analizamos brevemente las técnicas de fotolitografía, así como las analogías entre la alineación de máscaras y la detección de la posición absoluta en codificadores ópticos. Introducimos los parámetros que caracterizan las señales de referencia y que son comunes en la codificación óptica y en el alineamiento de máscaras en fotolitografía.

1.1. Codificadores ópticos y señales de referencia

En primer lugar, es conveniente aclarar un aspecto de la terminología. Con el término de codificadores ópticos se suelen denominar a aquellos sistemas que se emplean para medir tanto rotaciones como desplazamientos lineales. Dado que los mismos principios pueden ser utilizados tanto para la medida de desplazamientos lineales como angulares, en la memoria no haremos distinción sobre el tipo de codificador. Aunque las figuras y algunas suposiciones muestren o hagan referencia al codificador lineal, los resultados de esta memoria son también aplicables a codificadores angulares.

Los codificadores ópticos son sistemas que se basan en la exploración optoelectrónica de un patrón de medida. Estos sistemas registran electrónicamente las

variaciones de irradiancia de luz que se producen al desplazar un cierto sistema óptico a lo largo del patrón de medida. Hay muchos tipos de patrones de medida: redes de difracción (pueden ser de amplitud o de fase), códigos pseudoaleatorios (en codificadores absolutos), longitud de onda de la luz (en codificadores interferométricos), etc. Al soporte físico que contiene el patrón de medida –cuando existe– se le suele llamar regla, y suele consistir de un soporte de vidrio, metálico, o en los casos de baja precisión, de plástico. La exploración optoelectrónica de la regla suele hacerse mediante un sistema óptico que suele ir incluido en lo que denominamos cabeza lectora. La cabeza lectora incluye una fuente de luz con la que se ilumina el patrón situado en la regla. La luz que proviene de la regla vuelve a pasar a la cabeza por una ventana, denominada ventana de lectura. La regla puede transmitir o reflejar la luz que proviene de la cabeza lectora, dando lugar a codificadores ópticos de reflexión o transmisión. Si la regla funciona por transmisión, la ventana de lectura se sitúa en el lado opuesto a la fuente de iluminación, mientras que si la regla actúa por reflexión, la fuente y la ventana están en el mismo lado de la regla. El tipo de clasificación que interesa en esta memoria depende del tipo de medida que realiza el codificador; según ésta, los codificadores ópticos pueden ser absolutos o incrementales. Un codificador absoluto es capaz de detectar la posición absoluta de la regla respecto de la cabeza de lectura, mientras que el incremental mide el desplazamiento que se ha producido entre la cabeza de lectura y la regla, sin determinar la posición de ésta.

1.1.1. Codificadores ópticos absolutos

Como su nombre indica, son sistemas que miden la posición de forma absoluta. Estos codificadores pueden utilizar varias señales periódicas de diferente periodo para determinar la posición. Las señales provienen de una regla con diferentes marcas periódicas, o bien, pueden utilizar un patrón pseudoaleatorio en el cual están codificadas las posiciones. Estos sistemas realizan la medida simultánea de las señales periódicas o de una sección del patrón pseudoaleatorio. A partir de esta medida son capaces de deducir de manera biunívoca la posición de la regla sobre la que se encuentra la cabeza lectora. Los principios físicos que regulan la generación de las señales en un codificador absoluto son similares a los de un codificador incremental. La diferencia fundamental entre ambos es que el número de pistas y/o detectores es mucho mayor en el absoluto, generando cada uno de ellos una señal diferente. La combinación de todas las señales genera la palabra digital que codifica la posición absoluta. En la figura 1.1 mostramos un codificador absoluto con varias pistas que contienen marcas periódicas (redes de difracción). En esta memoria no trataremos este tipo de sistemas, nos centraremos exclusivamente en los codificadores ópticos incrementales.

1.1.2. Codificadores ópticos incrementales

Entre los codificadores ópticos que realizan una medida incremental, en esta memoria nos centramos en los que tienen como patrón de medida una red de

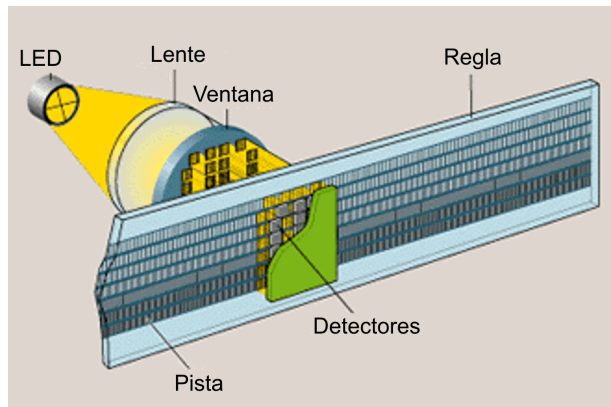


Figura 1.1: Esquema de codificador lineal absoluto. Emplea varias pistas con marcas periódicas.

difracción. Este tipo de codificadores son los más utilizados por su robustez, versatilidad y facilidad de montaje.

Las variaciones de la intensidad de luz que se producen al desplazar la cabeza lectora sobre la regla son periódicas. Normalmente con el mismo periodo de la red de difracción situada en la regla. La ventana situada en la cabeza lectora, a su vez suele contener otra red de difracción de periodo igual o similar al de la regla. Las variaciones de intensidad de la luz provienen de un batido moiré entre la regla y la ventana. También pueden producirse por un fenómeno de formación de imagen similar, como puede ser efecto Lau o autoimagen generalizada. Al desplazar la cabeza lectora, las variaciones periódicas de la intensidad de luz son registradas por un sistema de fotodetectores situados en ella. De esta manera, las variaciones de la intensidad de luz se traducen en variaciones periódicas de voltaje o corriente. Normalmente, estas señales se convierten electrónicamente en un tren de pulsos periódicos cuya frecuencia es proporcional a la velocidad de movimiento. Dicho de otra manera, el codificador discretiza el movimiento generando un pulso eléctrico cada vez que la cabeza lectora se desplaza una cantidad fija, Δx , que es la resolución del codificador. De esta manera, es posible medir el desplazamiento de la cabeza lectora contando el número de pulsos de las señales de lectura. Esta forma de trabajar permite un control electrónico del movimiento muy robusto mediante PIDs digitales. Sin embargo, con este tipo de señales solo es posible medir desplazamientos relativos –o incrementales– a partir de la posición inicial de la regla. En algunas aplicaciones es fundamental el conocimiento de la posición absoluta, y para ello es necesario dotar a estos sistemas con una señal adicional, denominada señal de referencia (Zero Reference Signal, ZRS). La señal de referencia consiste en un pulso eléctrico que sirve para identificar una posición conocida de la regla. Esta posición es el origen a partir del cual se suma o resta el desplazamiento incremental. Para obtener la

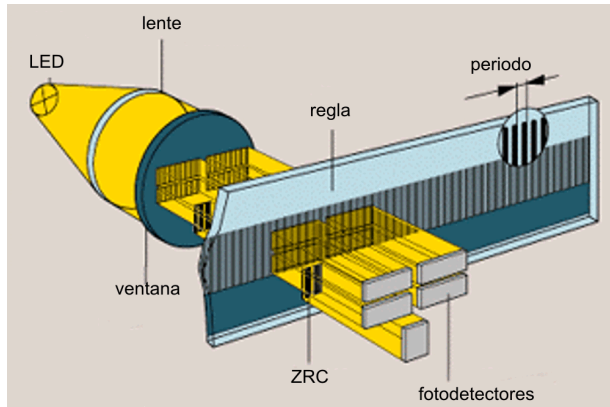


Figura 1.2: Esquema de codificador basado en efecto moiré. En una segunda pista se pueden observar una marca de referencia (ZRC).

señal de referencia es necesario incluir unas marcas de referencia al lado de la red de difracción, comúnmente denominadas ZRCs (Zero Reference Codes). En la figura 1.2 se muestra un codificador basado en efecto moiré. La regla contiene la pista incremental –red de difracción– y una segunda pista que contiene un ZRC. Un ZRC consiste en un grupo de rendijas transparentes y opacas que forman una especie de código de barras pseudo-aleatorio. Los ZRCs son detectados por un sistema óptico que genera la señal de referencia e identifica su posición. Este sistema óptico también está situado en la cabeza lectora y suele compartir elementos con el sistema encargado de la detección incremental. En sistemas que miden grandes distancias, es difícil encontrar el ZRC que genera la señal de referencia y define el origen de la medida. En estos sistemas se suelen situar varios ZRCs a una distancia codificada a lo largo de la regla. Cuando la cabeza lectora pasa por dos ZRCs consecutivos, mide la distancia que hay entre ellos deduciendo la posición absoluta en la que se encuentra. En la figura 1.3 mostramos una regla que contiene la pista incremental y varios ZRC a distancia codificada. Este diseño capacita a este tipo de codificadores, a pesar de ser incrementales, para la medida absoluta de la posición. A diferencia de los codificadores absolutos, estos codificadores tienen que realizar un desplazamiento inicial antes de obtener la medida absoluta.

1.1.3. Estándares desarrollados en torno a los sistemas de codificación óptica

Existen ciertos estándares desarrollados en codificadores ópticos basados en redes de difracción. Algunos de estos –como la definición de precisión y resolución– son tan generales que se aplican a todos los tipos de codificadores de la posición. Estos estándares determinan cómo deben ser las señales electrónicas

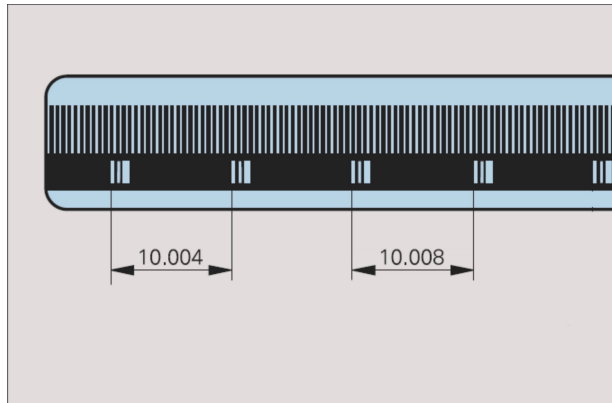


Figura 1.3: Regla típica con dos pistas, la pista superior es incremental (red de difracción) y la pista inferior contiene varios ZRC con distancia codificada.

suministradas por el codificador y establecen la calidad de las medidas obtenidas.

- Señales incrementales.

Los codificadores se suelen diseñar de manera que suministran dos señales sinusoidales, que comúnmente se denominan A y B, y tienen un desfase relativo de 90° . Las señales sinusoidales se convierten electrónicamente en señales cuadradas, bien del mismo periodo que la señal original, o bien de un periodo menor. Las señales eléctricas con un periodo menor que las originales se obtienen mediante una interpolación digital de la señal original. La medida del desplazamiento se suele hacer detectando los flancos de subida y bajada de las señales cuadradas. Cada vez que se detecta un flanco, bien de la señal A, o bien de la señal B, se incrementa la medida en $1/4$ del periodo de la señal. Según el flanco detectado, sea de la señal A o B, se puede deducir el sentido del movimiento. En la parte superior de la figura 1.4 se muestran las señales de lectura de la etapa incremental del codificador. En la izquierda se muestran las señales adquiridas por los fotodetectores y en la derecha se muestran las señales cuadradas generadas electrónicamente sin interpolación.

- Señal de referencia.

La señal de referencia debe ser un pulso de intensidad de luz aislado, que proviene de la autocorrelación entre el ZRC situado en la regla y otro ZRC idéntico situado en la ventana de la cabeza lectora. Los ZRCs se diseñan para que el pulso tenga una anchura a media altura igual al periodo de las señales incrementales. Al igual que las señales incrementales, esta señal se convierte electrónicamente en una señal cuadrada con anchura igual a $1/4$ de la señal incremental. Esta señal también puede ser interpolada según los

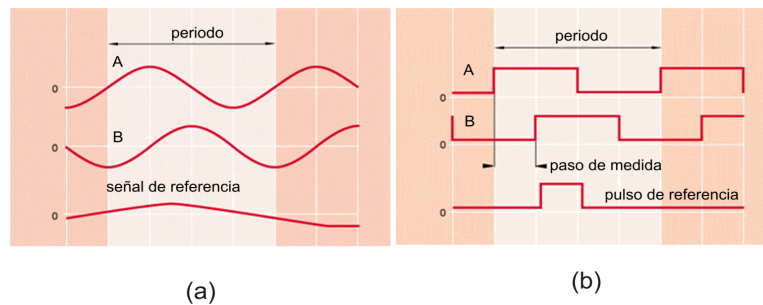


Figura 1.4: Señales incrementales A y B y señal de referencia. Las señales (a) son registradas por los fotodetectores, mientras que en (b) se han convertido a señales cuadradas.

requerimientos de la medida. En la parte inferior de la figura 1.4, se muestra la señal de referencia. En la izquierda se representa la señal registrada por los fotodetectores y en la derecha el pulso generado electrónicamente. Al igual que con las señales incrementales, se ha realizado una conversión electrónica sin interpolación.

- Definición de precisión.

La precisión de un codificador se suele definir como el valor extremo de la desviación de la medida de la posición respecto al valor medio de ésta, en cualquier punto a lo largo de 1m de recorrido de la regla. La precisión en un codificador óptico viene determinada por:

- La calidad del grabado de la red de difracción de la regla.
- Los errores introducidos al interpolar y procesar las señales suministradas por el codificador.
- Los errores de orientación y alineamiento entre regla y cabeza lectora. La sensibilidad a estos errores suele denominarse tolerancia del codificador.
- La sincronía entre la señal incremental y la señal de referencia para detectar la posición absoluta.

- Definición de resolución.

Es el mínimo desplazamiento que el codificador es capaz de detectar, depende del periodo de la señal incremental y de la interpolación realizada. Es un error confundir la resolución con la precisión del codificador.

1.2. Fotolitografía y señales de referencia.

La fotolitografía es el proceso más sofisticado implicado en la fabricación de circuitos integrados (IC). El diagrama que se graba representa las complejas estructuras que componen y conectan los millones de transistores de un IC. La velocidad de procesado de un circuito integrado está directamente relacionada con la compacidad de los circuitos que lo integran, y a su vez, con el mínimo detalle de grabación. Por esta razón, los circuitos que componen un IC se disponen en una compleja estructura de capas. Cada capa exige un proceso litográfico de grabación de una máscara, junto con un proceso de ataque químico y su correspondiente difusión de dopantes, deposición de metales, aislantes, etc. El proceso de fotolitografía es tan costoso que el precio de un IC puede estimarse por los niveles de fotolitografía necesarios (masking levels). Así, por ejemplo, los procesadores de Intel suelen tener entre 20 y 30 procesos de fotolitografía. El sistema de posicionado de la oblea es el sistema mecánico más preciso que existe en la tecnología actual. Típicamente, en un área correspondiente a un diámetro de oblea de 150-200mm, puede haber más de 100 dispositivos semiconductores denominados "chips". La exposición de los chips suele ser secuencial, por lo tanto, su desplazamiento debe realizarse en condiciones de vacío y sobre una plataforma móvil ultraprecisa. El movimiento de la plataforma es controlado por codificadores ópticos interferométricos, con precisión mínima inferior a 20nm. La posición exacta de la oblea sobre la plataforma móvil se desconoce. Por esta razón, antes de cada exposición es necesario detectar la posición exacta de la oblea, realizando un alineamiento entre ésta y la máscara. El alineamiento se realiza mediante un subsistema denominado alineador de máscaras. Para detectar la posición de la oblea, el alineador de máscaras necesita unas marcas de referencia estandarizadas. Estas marcas de referencia son impresas en la oblea en cada nivel previo de fotolitografía, y pueden ser detectadas mediante una gran variedad de técnicas ópticas, algunas automáticas y otras semi-automáticas. El alineamiento entre la oblea y la máscara que contiene el diagrama de grabación, debe realizarse con una precisión entre 1/5 y 1/10 del mínimo detalle de grabación, además, debe mantenerse durante la exposición.

Dependiendo de la configuración óptica utilizada, existen tres tipos de fotolitografía: de contacto, proximidad y de proyección. Como sus nombres indican, la máscara está en contacto con la resina de la oblea en la fotolitografía de contacto, está muy próxima en la de proximidad y es proyectada por un sistema óptico en la de proyección. En la figura 1.5 se muestra un esquema de los tres tipos de fotolitografía. La fotolitografía de contacto y de proximidad son sencillas de implementar, mientras que la de proyección es más complicada, tanto desde el punto de vista óptico, como mecánico.

En la fotolitografía de contacto y proximidad el diagrama de la máscara cubre completamente la oblea, por lo tanto la magnificación del sistema es $1\times$. La fotolitografía de contacto alcanza una gran resolución, por debajo de la longitud de onda utilizada. El inconveniente de esta técnica es el deterioro de la máscara, ésta se desgasta, se raya y se ensucia rápidamente. La fotolitografía de proximidad reduce el riesgo de desgaste, a costa de una disminución de la resolución.

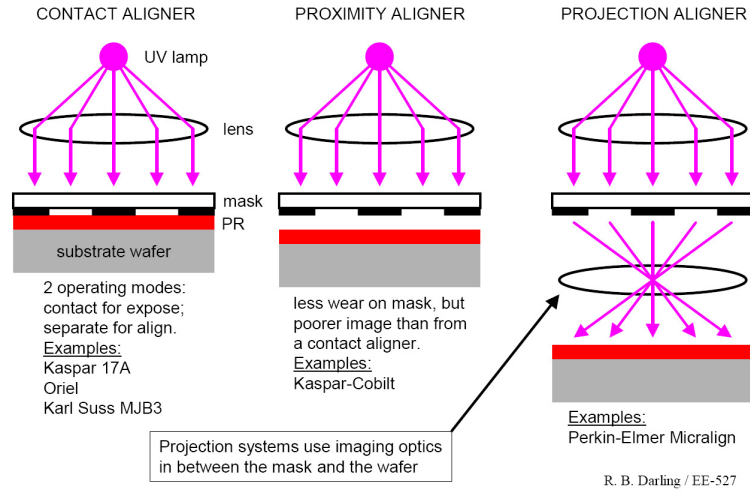


Figura 1.5: Diferentes tipos de fotolitografía, en la figura de la izquierda se muestra un esquema de fotolitografía de contacto, en el centro de proximidad y en la derecha de proyección.

La resolución está limitada por la longitud de onda de la luz utilizada, y por la distancia entre máscara y oblea. Iluminando con ultravioleta cercano, la máxima resolución alcanzada es de unas $2 - 3 \mu\text{m}$. Típicamente, en los sistemas de contacto y proximidad, el alineamiento lo realiza un operario. El operario desplaza la plataforma móvil para superponer una marca de referencia, previamente impresa en la oblea, con otra marca situada en la máscara. La superposición de las marcas se observa en un microscopio con el campo desdoblado. En una parte del campo se observa la marca de la oblea, y en la otra la marca situada en la máscara. También existen sistemas de alineamiento automáticos basados en redes de difracción. Estos sistemas generan una señal de alineamiento por efecto moiré o similar. A lo largo de esta memoria nos centramos en los métodos de alineamiento en sistemas de contacto y proximidad, en particular, en el diseño de nuevas marcas de referencia que permitan mejores señales de alineamiento.

Actualmente, debido a la gran capacidad de diseño de sistemas de proyección libres de aberraciones, la fotolitografía de proyección es la más utilizada en la industria de semiconductores. Se pueden utilizar sistemas de proyección con una reducción de $4 - 10\times$, que alcanzan resoluciones del orden de 90 nm . El campo que permiten exponer estos sistemas es pequeño y necesitan un complejo sistema de control mecánico. Hay diferentes tipos dependiendo de la mecánica: scanners, step-and-repeat o los modernos step-and-scan. El alineamiento en estos sistemas también se realiza mediante el uso de marcas de referencia situadas en la máscara y en la oblea, puede ser: visual, semi-automático o automático. En el caso del alineamiento automático es necesario hacer un prealineamiento y el sistema

utiliza técnicas de reconocimiento automático de las marcas de alineamiento.

1.3. Características de las señales de referencia

En este apartado analizamos una característica común en alineadores de máscaras para fotolitografía y en codificadores ópticos basados en redes de difracción, la señal óptica de referencia. Esta señal es utilizada en codificadores para localizar una posición conocida de la regla, y dotar al sistema de una medida absoluta. El desplazamiento en estos sistemas es unidimensional, y por tanto la señal de referencia también lo es. Por otro lado, en fotolitografía el diagrama que se graba en la oblea debe situarse en una posición determinada, para ello es necesario alinear las marcas de referencia situadas en la máscara y en la oblea. El alineamiento consiste en localizar una posición conocida de la oblea, haciéndola coincidir con una posición conocida de la máscara. En este caso el movimiento es bidimensional y las marcas de referencia utilizadas también lo son. Existen procedimientos de alineamiento que se basan en la generación de una señal de alineamiento, que consiste en un pulso de intensidad luz bidimensional. En lo que sigue, nos centramos en señales ópticas de alineamiento y de referencia, que consisten en un pulso aislado y localizado y que llamaremos genéricamente S . Analizamos las semejanzas y diferencias de las señales de alineamiento bidimensional y las señales de referencia unidimensional. La señal puede tener una forma arbitraria, pero normalmente suele convertirse electrónicamente en un pulso cuadrado. Esta conversión se realiza comparando la señal con un umbral electrónico prefijado. Teniendo en cuenta el resultado de la comparación, la señal es binarizada. Sin pérdida de generalidad, suponemos que el pulso está localizado en el centro de la señal, de manera que el máximo de ésta es $S(0)$. El segundo máximo de la señal (o máximo secundario) está fuera del pulso y lo denominamos σ . A continuación, definimos los parámetros comunes que miden la calidad de estas señales, tanto en una como en dos dimensiones:

- Señal efectiva.
Llamamos señal efectiva a la diferencia entre el primer y el segundo máximo de la señal, $SE = S(0) - \sigma$. Esta es la cantidad de señal útil si situamos el umbral electrónico de comparación a la altura del segundo máximo.
- Factor K
Este factor lo definimos como la relación entre el máximo secundario y el máximo principal de la señal, $K = \sigma/S(0)$. Cuanto más pequeño es este parámetro mejor será la relación entre el primer y el segundo máximo. Si consideramos como ruido la estructura de la señal en la parte externa al pulso central, el factor K sería equivalente a la inversa de la relación señal-ruido.
- Pendiente del máximo central.
La definimos como $SL = (S(0) - S(x_1))/x_1$, donde x_1 es la semianchura

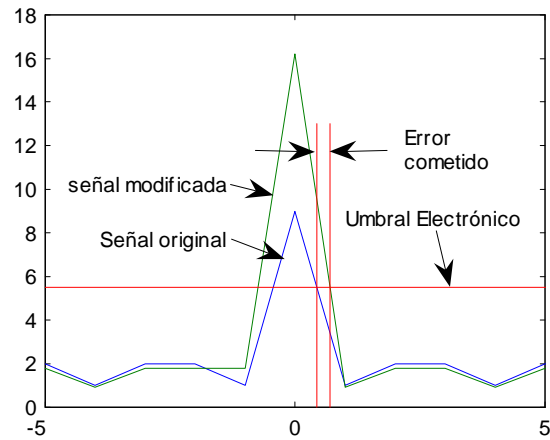


Figura 1.6: Error cometido en la conversión electrónica de una señal que ha sufrido una modificación.

del pico central en su base. Este parámetro está relacionado con la velocidad de crecimiento del pico central de la señal. Es muy importante en la detección de la posición mediante pulsos con forma arbitraria. Dado que la conversión electrónica se realiza por comparación con un umbral de corte con la señal, cuanto mayor sea la velocidad de crecimiento del pulso, más insensible será el punto de corte frente a fluctuaciones de la señal. Este efecto se muestra en la figura 4.15. La señal de referencia ha sufrido una fluctuación, la cual provoca un cambio en la posición del punto de corte con el umbral electrónico.

La detección de una posición de referencia y/o alineamiento debe ser rápida, lo más cercano posible al “tiempo real”. La detección de esta señal suele formar parte de un ciclo de control y posicionamiento en el que el tiempo es una magnitud crítica. Otro aspecto fundamental de las señales de referencia y alineamiento es que su diseño tiene que ser robusto, es decir, la señal deberá ser insensible a pequeños desalineamientos, tolerancias de montaje y de fabricación de los elementos que se utilicen.

Capítulo 2

Diseño de ZRCs unidimensionales

2.1. Introducción

En este capítulo analizamos el sistema de generación de señales de referencia en codificadores ópticos basados en redes de difracción. Típicamente, la generación de la señal de referencia unidimensional se lleva a cabo mediante un sistema óptico como el de la figura 2.1. Este sistema consta de dos ZRCs situados uno en frente del otro. Un haz de luz colimado pasa a través de los ZRCs y después se detecta mediante un fotodetector, el cual está situado en el plano focal de una lente. Uno de los ZRCs está situado en la regla y el otro en la cabeza lectora, solidario al sistema de generación-detección del haz de luz. El desplazamiento relativo entre los dos ZRCs origina un cambio en el área de solapamiento entre ellos, y por lo tanto, en la cantidad de luz transmitida a través de los códigos. Cuando un código pasa a través del otro, la señal registrada por el fotodiodo es la señal de referencia. Si los códigos están suficientemente próximos, la señal de referencia es proporcional a la correlación entre las dos transmitancias de los ZRCs. Para incrementar el máximo de esta señal, típicamente se consideran sistemas con códigos idénticos [1]. En este caso, la señal obtenida será la autocorrelación de la transmitancia asociada al ZRC.

En general, un ZRC viene descrito por una secuencia binaria de datos,

$$c = [c_1, c_2, \dots, c_n], \quad c_j \in \{0, 1\}, \quad j = 1, \dots, n \quad (2.1)$$

donde n es la longitud del ZRC, $c_i = 1$ si una rendija transparente está situada en la posición i y $c_j = 0$ si una rendija opaca está situada en la posición j . El tamaño de las regiones transparentes y opacas del código es un múltiplo entero de la anchura de una rendija, que denominamos b . El número de rendijas transparentes y opacas es n_1 y n_0 respectivamente, de manera que se cumple la relación $n_1 + n_0 = n$. La transmitancia del código se puede expresar como

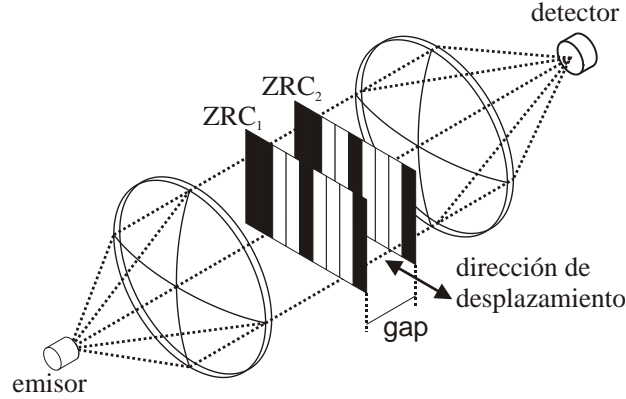


Figura 2.1: Procedimiento de generación de la señal de referencia.

función de la coordenada en la dirección del movimiento, x

$$t(x) = \sum_{j=1}^n c_j \cdot \text{rect} \left(\frac{x-j}{b} \right), \quad (2.2)$$

donde rect es la función rectángulo, definida en la siguiente expresión

$$\text{rect}(x) = \begin{cases} 1 & \text{si } |x| < \frac{1}{2} \\ 0 & \text{si } |x| \geq \frac{1}{2} \end{cases}. \quad (2.3)$$

Los ZRCs fueron empleados por primera vez en [1] y desde entonces su diseño ha sido estudiado por varios autores [1], [2],[3],[4],[5]. En estos trabajos se consideran las propiedades de una señal que proviene de la autocorrelación de la transmitancia del ZRC, ecuación 2.2. Además, se establecen las condiciones necesarias para obtener señales de referencia adecuadas. La principal dificultad en el diseño de códigos que generan señales óptimas es la ausencia de un método de cálculo sistemático para obtener ZRCs de longitud arbitraria. El método más utilizado [3] consiste en formar un código a base de añadir unos y ceros de acuerdo con un criterio predefinido. Este criterio garantiza un código óptimo pero solo es aplicable a códigos de longitud reducida, hasta 30-40 elementos.

El objetivo de este capítulo es el diseño de ZRCs óptimos de longitud arbitraria. Para ello estudiamos algunas características de las señales y los ZRCs que las generan, después proponemos dos herramientas de optimización para el cálculo sistemático de ZRCs.

2.2. Consideraciones generales

Para calcular la señal de referencia registrada en el fotodetector de la figura 2.1, asumimos que el haz de luz incidente en el primer ZRC es plano y los efectos

difractivos son despreciables. Esta aproximación es válida cuando la distancia entre los dos ZRCs es suficientemente pequeña respecto a la anchura de las rendijas del código, además la anchura de las rendijas del código debe ser mucho mayor que la longitud de onda del haz que lo ilumina. Teniendo en cuenta estas aproximaciones, cuando uno de los ZRCs está desplazado una cantidad τ , el flujo de luz que pasa a través de ellos es proporcional a

$$S(\tau) = \int_{-\infty}^{+\infty} t(x) t(x - \tau) dx = \sum_{j=-n+1}^{n-1} a_j \cdot \Delta\left(\frac{\tau - j}{b}\right), \quad (2.4)$$

la función triángulo se define como

$$\Delta(x) \equiv \begin{cases} 1 - |x| & \text{si } |x| \leq 1 \\ 0 & \text{si } |x| > 1 \end{cases} \quad (2.5)$$

y las componentes del vector \mathbf{a} vienen dados por

$$a_k = a_{-k} = \sum_{j=1}^{n-k} c_j \cdot c_{j+k}, \quad k = 0, 1, \dots, n-1. \quad (2.6)$$

Si la señal S es muestreada en los puntos correspondientes a los múltiplos de la anchura de las rendijas, b , la señal coincide con los elementos del vector \mathbf{a} ,

$$S(k \cdot b) = S_k = a_k, \quad k = -n+1, \dots, 0, \dots, n-1. \quad (2.7)$$

Teniendo en cuenta la expresión anterior, el problema de cálculo de la señal de referencia es esencialmente discreto. Para pasar a la versión continua de la señal solo hay que utilizar la expresión 2.4.

Las señales que provienen de una autocorrelación son simétricas respecto del origen ($S_k = S_{-k}$), por lo tanto solo necesitamos conocer los últimos (o los primeros) n elementos para caracterizar completamente la señal. Definimos el vector señal de referencia considerando los índices positivos (últimos n elementos), $\mathbf{S} = [S_0, \dots, S_{n-1}]$. En [1] y [3] se demuestran las siguientes propiedades:

- i. El máximo de la señal de autocorrelación, S_0 , es el número de rendijas transparentes en el ZRC, $S_0 = \sum_{j=1}^n c_j^2 = \sum_{j=1}^n c_j = n_1$. Donde hemos utilizado el carácter binario de las variables.
- ii. Teniendo en cuenta que la señal de referencia es una superposición de triángulos escalados y desplazados, la anchura a mitad de altura del triángulo central de la señal es dos veces la anchura de las rendijas del ZRC.
- iii. En la suma de la ecuación 2.6, cada sumando $c_j \cdot c_{j+k}$ no nulo exige que tanto c_j , como c_{j+k} sean "1", es decir, el código contiene dos "1" separados por una distancia de k elementos. Según la expresión 2.7, el valor de S_k es igual al número de subsecuencias en el ZRC del tipo $[1, \underbrace{X, \dots, X}_{k-1}, 1]$,

donde $X \in \{0, 1\}$. Cuando el desplazamiento relativo entre los dos códigos es $k \cdot b$, uno de los “1” de la subsecuencia desplazada se enfrenta con el otro “1” de la subsecuencia sin desplazar. Esta propiedad es utilizada en el método de diseño de Y. Li en [2]-[5].

En el capítulo anterior hemos definido algunos parámetros que caracterizan las señales de referencia [2]. En lo que concierne a la codificación óptica, el parámetro más importante es

$$K = \frac{\sigma}{S_0}, \quad (2.8)$$

donde σ es el segundo máximo de la señal, $\sigma = \max\{S_1, \dots, S_{n-1}\}$. Una buena ZRS es una señal que contiene un máximo central aislado, localizado y bien definido. La forma del resto de la señal no tiene interés práctico. De acuerdo con esto, el parámetro K cuantifica la relación entre lo que podríamos considerar ruido y la cantidad de señal (valor del máximo central). Cuanto más pequeño es este valor, mayor es la sensibilidad y la robustez de la señal de referencia, lo cual contribuye a la fiabilidad del sistema.

En el método de diseño de ZRCs utilizado por Xiangyang y Chunyong [1] los parámetros S_0 y K (y por tanto, σ) son fijados previamente y el objetivo es encontrar el código con esos parámetros y con el mínimo número de elementos posible. Este proceso de diseño de ZRCs consiste en la adición de subsecuencias (ver la propiedad iii) teniendo en cuenta que el número subsecuencias repetidas no puede ser mayor que σ . Hay muchos posibles códigos, dependiendo del orden en el que se dispongan las subsecuencias. Entre todos estos códigos tenemos que establecer una comparación para seleccionar el óptimo. Este proceso es muy complicado de programar y es muy difícil obtener ZRCs con más de 40 elementos. A partir de la propiedad (iii), Y. Li [3] ha establecido una cota inferior para el segundo máximo de la señal

$$\sigma \geq \frac{\left[(2n+1) - \sqrt{(2n+1)^2 - 4n_1(n_1-1)} \right]}{2}. \quad (2.9)$$

Existen algunos casos muy simples en los que el segundo máximo toma el valor de la cota inferior (normalmente con códigos de muy pocos “1”), pero no hay ninguna evidencia de que se pueda alcanzar el valor de la cota inferior para cualquier valor de n y n_1 . En el procedimiento de diseño de Y. Li [3], los parámetros n y n_1 son fijados y la cota inferior es calculada con la expresión 2.9. El proceso comienza con el código vacío y se van añadiendo “1” intentando conseguir el valor de la cota inferior, teniendo en cuenta la propiedad (iii). Normalmente no es posible alcanzar la cota, en ese caso el valor de la cota se aumenta en una unidad, comenzando el proceso de nuevo. Este proceso se repite hasta encontrar el ZRC con el menor valor de K . Con este método se han encontrado ZRCs con longitud máxima de 20 elementos. En [4], Y. Li propone otro tipo de método de diseño. En este trabajo se propone un programa de cálculo para obtener todas las posibles ZRCs con un valor fijo de n y n_1 , el programa compara las características de todos los códigos para obtener el código óptimo. El

programa tiene que considerar $C(n-1, n_1-2) = n! / (n_1-1)! (n-n_1)!$ ZRCs. En este algoritmo se produce la llamada “explosión combinatoria”. La computadora más potente en este momento (Earth Simulator Center, Japón) tardaría 5×10^{28} años en calcular un caso con $n = 170$ y $n_1 = 85$. Finalmente, en [5] Y. Li propone un método de cálculo de ZRC cuando la señal de autocorrelación es conocida. En general, la señal de autocorrelación óptima no es conocida y este método no tiene utilidad práctica.

Para que una técnica de diseño de ZRCs tenga interés práctico en codificación óptica, debe considerar algunos parámetros que están predeterminados en el codificador óptico. El periodo de la señal incremental debe ser igual a la anchura de la señal de referencia, por esta razón el periodo de la red de difracción suele ser igual a la anchura de las rendijas del ZRC. Además, el diámetro del haz de luz limita la longitud del ZRC. La longitud del ZRC y la anchura de las rendijas fijan el número total de elementos del ZRC. Finalmente, los requerimientos electrónicos del sistema de detección establecen un valor mínimo para el máximo de la señal, n_1 . Teniendo en cuenta estas condiciones de trabajo, los parámetros n y n_1 están predeterminados y la señal óptima será la que tenga más bajo el valor del segundo máximo, σ . Con estos requerimientos, se puede utilizar el método de Y. Li [3] para el diseño de ZRCs de longitud inferior a 20 elementos. La creciente demanda en la resolución de los sistemas de medida nos lleva a trabajar con redes de difracción de unas pocas micras. Por otro lado, para tener una cantidad suficiente de señal necesitamos un haz de varios milímetros de extensión. Estas dos condiciones nos obligan a trabajar con ZRCs de gran longitud. Por ejemplo, un haz de 5mm necesita un código de 625 elementos para obtener una señal de $8\mu\text{m}$ de anchura a mitad de altura. En este ejemplo se hace evidente la necesidad de proponer nuevos métodos de diseño de códigos de gran longitud. Los métodos de diseño deben considerar ZRCs con n y n_1 fijos. El proceso de diseño consiste en disponer los “1” del ZRC de manera que el valor de σ sea el menor posible. Un proceso de diseño con estos requerimientos se puede plantear numéricamente en términos de la optimización de una función objetivo con restricciones. La optimización consiste en la minimización del segundo máximo de la señal como función de la posición de las rendijas en el ZRC. En la siguiente expresión mostramos la función objetivo y las restricciones necesarias,

$$\begin{aligned}
 f(\mathbf{c}) &= \max \{S_1, \dots, S_{n-1}\}, \\
 S_k &= \sum_{j=1}^{n-k} c_j \cdot c_{j+k}, \quad k = 1, \dots, n-1 \\
 \text{s.t.} \quad S_0 &= n_1 = \sum_{j=1}^n c_j \\
 \mathbf{c}_j &\in \text{binario},
 \end{aligned} \tag{2.10}$$

donde $f(\mathbf{c}) = \max \{S_1, \dots, S_{n-1}\} = \sigma$ es el segundo máximo de la señal de autocorrelación, ya que el primer máximo es S_0 . La variable $\mathbf{c} = [c_1, \dots, c_n]$ es un vector cuyos elementos son binarios y su suma es igual a n_1 . Físicamente, n_1

es una cantidad proporcional a la cantidad de energía que llega al fotodetector.

2.3. Optimización con el algoritmo DIRECT

2.3.1. Descripción de DIRECT

DIRECT es un algoritmo desarrollado por D. R. Jones et al. [6] para la obtención del mínimo global de una función de varias variables. Las variables pueden ser enteras y pueden estar sujetas a restricciones. La función objetivo puede ser no lineal, con varios mínimos (multimodal) y puede no ser suave (nonsmooth optimization). DIRECT forma parte de los llamados “algoritmos de muestreo”, dichos algoritmos realizan un muestreo del dominio de la función, evalúan ésta en los puntos seleccionados y utilizan la información para decidir la dirección de búsqueda en el siguiente paso. El algoritmo no necesita información sobre la derivada de la función objetivo. DIRECT es apropiado cuando la función objetivo es complicada, discontinua o no conocemos la relación entrada-salida de la función. El algoritmo realiza un muestreo determinista, es decir, una misma semilla siempre produce la misma solución, por lo tanto, no es necesario realizar múltiples ejecuciones del algoritmo para realizar una búsqueda global. Si la función objetivo y las restricciones son continuas en un entorno del mínimo global la convergencia está asegurada.

El problema genérico que puede resolver DIRECT viene descrito como sigue [7]:

$$\begin{aligned}
 & \underset{\mathbf{x}}{\text{mín}} f(\mathbf{x}) \\
 \text{s.t. } & \mathbf{LB} \leq \mathbf{A} \cdot \mathbf{x}' \leq \mathbf{UB} \\
 & \mathbf{LC} \leq \mathbf{c}(\mathbf{x}) \leq \mathbf{UC} \\
 & \mathbf{Lx} \leq \mathbf{x} \leq \mathbf{Ux} \\
 & x_i \in \mathbb{Z} \quad \forall i \in \mathbf{I},
 \end{aligned} \tag{2.11}$$

donde $f(\mathbf{x}) \in \mathbb{R}$, $\mathbf{x}, \mathbf{Lx}, \mathbf{Ux} \in \mathbb{R}^n$, $\mathbf{LB}, \mathbf{UB} \in \mathbb{R}^m$, $\mathbf{A} \in \mathbb{R}^{m \times n}$, $\mathbf{LC}, \mathbf{c}(\mathbf{x}), \mathbf{UC} \in \mathbb{R}^p$. Donde \mathbf{Lx} y \mathbf{Ux} son las cotas inferior y superior de las variables, \mathbf{LB} y \mathbf{UB} son las cotas inferior y superior para las restricciones lineales y \mathbf{LC} y \mathbf{UC} son las cotas inferior y superior para las restricciones no lineales, respectivamente. \mathbf{A} es una matriz que contiene los coeficientes de la restricción lineal. \mathbf{x}' es la transpuesta de \mathbf{x} y las variables x_i cuyo índice pertenezca a \mathbf{I} son números enteros.

DIRECT toma su nombre de "DIviding RECTangles", siglas que describen la forma en la que el algoritmo avanza hacia el mínimo. El primer paso del algoritmo es la transformación del espacio de búsqueda en un hiper cubo de lados normalizados. La función objetivo es muestreada en el centro del hiper cubo. Después, el algoritmo divide el hiper cubo en hiperrectángulos cuyos centros también serán muestreados. Normalmente, en optimización Lipschitziana, la constante de Lipschitz se utiliza para determinar el próximo rectángulo para dividir y muestrear. La constante de Lipschitz es una cota en la velocidad de

cambio de la función objetivo. En el caso más general, la constante de Lipschitz no es conocida y DIRECT localiza un conjunto de rectángulos “potencialmente óptimos” que corresponden a todas las posibles constantes de Lipschitz para cada rectángulo del mismo tamaño. En cada iteración, todos los rectángulos “potencialmente óptimos” son divididos en rectángulos más pequeños cuyos centros son muestreados. Si no conocemos el óptimo global no hay ninguna forma de asegurar que lo hayamos alcanzado. El algoritmo termina cuando alcanza un número predeterminado de iteraciones o cuando el número de evaluaciones de la función objetivo supere un límite preestablecido. En la práctica, si el mejor valor para la función objetivo no mejora cuando incrementamos el número de evaluaciones de la función y el número de iteraciones, podemos considerar que hemos alcanzado un valor óptimo.

2.3.2. Resultados

Para poder aplicar DIRECT a nuestro problema tenemos que adaptar la definición de nuestro problema 2.10 a la estructura de los problemas que DIRECT puede resolver 2.11. El modelo de nuestro problema se puede expresar:

$$\begin{aligned}
 f(\mathbf{c}) &= \text{máx} \{S_1, \dots, S_{n-1}\}, \\
 S_k &= \sum_{j=1}^{n-k} c_j \cdot c_{j+k} \\
 [0, \dots, 0] &\leq \mathbf{c} \leq [1, \dots, 1] \\
 n_1 &= [1, \dots, 1] \cdot \begin{bmatrix} c_1 \\ \vdots \\ c_n \end{bmatrix} \\
 \mathbf{c} &\in \mathbb{Z}^n.
 \end{aligned} \tag{2.12}$$

El requerimiento de que \mathbf{c} sea un vector binario se cumple si exigimos que las variables sean enteras y que estén acotadas inferior y superiormente por $\mathbf{L}_x = [0, \dots, 0]$ y $\mathbf{U}_x = [1, \dots, 1]$ respectivamente. \mathbf{A} es un vector de unos y $\mathbf{A} \cdot \mathbf{x}'$ es el número de rendijas transparentes en el ZRC, el cual está fijado por n_1 .

Para demostrar el funcionamiento de DIRECT aplicamos esta técnica a problemas de dimensión baja, los cuales pueden ser resueltos con las técnicas descritas en la literatura. Con DIRECT obtenemos los mismos resultados que los obtenidos con los métodos tradicionales, esto demuestra la consistencia de la nueva técnica. Es relativamente sencillo diseñar ZRCs que generan señales con el segundo máximo igual a 1. Este tipo de casos son muy interesantes para el estudio de las capacidades de los algoritmos, ya que se alcanza el valor de la cota inferior teórica (ecuación 2.9), además, el valor óptimo del segundo máximo es un mínimo global. A partir de la propiedad (iii), el número de subsecuencias repetidas no puede ser mayor que uno y el procedimiento de diseño consiste en la adición de

subsecuencias de longitud creciente, sin repetición. Un ejemplo de este tipo de soluciones con 21 elementos es $\mathbf{c} = [1, 1, 0, 1, 0, 0, 0, 1, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 1]$. En [8], Y. Li calcula los ZRCs que generan señales con segundo máximo igual a 1 y establece el número de elementos y de rendijas que deben tener este tipo de códigos. Cuando aplicamos DIRECT a estos casos obtenemos siempre el mínimo global ($f(\mathbf{c}) = 1$) para ZRCs inferiores a 100 elementos. El caso límite es $n = 100$ y $n_1 = 11$. Con estos parámetros el número de puntos compatible con las restricciones del problema (número de puntos en el espacio de búsqueda) es $1,4 \times 10^{14}$. Dado un número de elementos n , el número de rendijas n_1 necesario para obtener una señal con segundo máximo igual a 1 es muy pequeño, esto hace que estos códigos sean muy poco eficientes energéticamente y no tengan interés práctico.

Para demostrar el funcionamiento de DIRECT en un caso general realizamos una comparación entre la cota inferior de σ dada por la ecuación 2.9 y el valor del segundo máximo alcanzado por el algoritmo con diferentes números de rendijas. El resultado se muestra en la figura 2.2. Las optimizaciones se realizan con un número fijo de elementos, $n = 50$ y con un número de rendijas que varía desde 1 a 49. En esta figura puede verse que en pocos casos puede alcanzarse el valor de la cota inferior y esto ocurre únicamente cuando el número de rendijas es pequeño. En un caso general es muy complicado demostrar que una solución óptima es un mínimo global, en nuestro caso hay dos razones que nos permiten pensar que hemos obtenido un mínimo global: en primer lugar, en todos los casos el número de puntos del dominio de la función compatible con las restricciones es menor que $1,4 \times 10^{14}$. En espacios de búsqueda con el mismo número de puntos el algoritmo ha encontrado el mínimo global de la función, en los casos en los que el segundo máximo es 1. En segundo lugar, la solución obtenida no cambia cuando aumentamos el número máximo de evaluaciones de la función y el número de iteraciones del algoritmo. El número máximo de evaluaciones de la función está limitado por la memoria de la computadora. En algunos casos la sensibilidad del detector no es un parámetro restrictivo y el máximo central de la señal de autocorrelación no es fijo. En este caso, la función objetivo a optimizar deberá ser la diferencia entre el primer y segundo máximo de la señal. Como hemos visto en el capítulo 1 este parámetro da cuenta de la cantidad efectiva de señal y está directamente relacionado con la estabilidad del sistema. La optimización de este parámetro está indirectamente calculado en los resultados presentados en la figura 2.2. En esta figura se muestra el valor óptimo del segundo máximo de la señal para cada valor del número de rendijas, n_1 . Para obtener el máximo de la señal efectiva y el número de rendijas que da lugar a ésta, representamos en la figura 2.3 el valor del primer y segundo máximo de la señal, así como la diferencia entre éstos (señal efectiva) respecto de n_1 . El máximo para la señal efectiva es de 15 unidades y se obtiene para $n_1 = 29$. En la figura 2.4 se muestran las señales de autocorrelación óptimas encontradas por el algoritmo para un caso de $n = 50$ y $n_1 = 25$. Este caso es inabordable por las técnicas de diseño de ZRCs tradicionales. El algoritmo encuentra varias soluciones que son equivalentes en cuanto al valor de la función objetivo, en este caso encuentra 136 soluciones con $\sigma = 11$. En la figura 2.5 mostramos los códigos

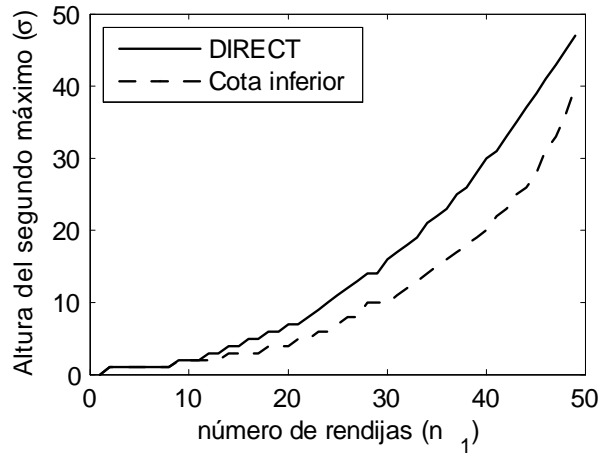


Figura 2.2: Altura del segundo máximo de la señal de autocorrelación respecto del número de rendijas del ZRC. El ZRC tiene 50 elementos. La gráfica continua muestra los valores obtenidos con DIRECT y la discontinua es la cota inferior calculada teóricamente.

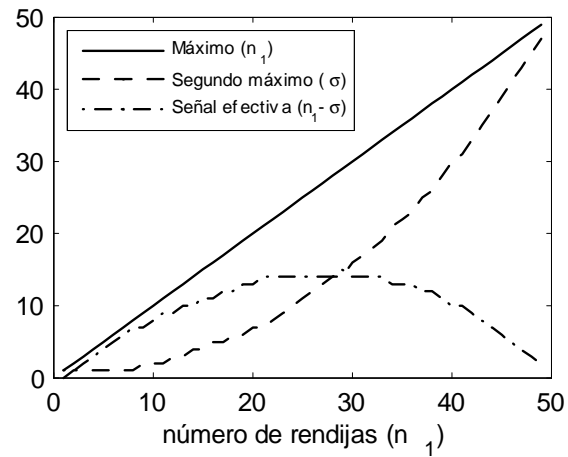


Figura 2.3: Primer máximo, segundo máximo y su diferencia (señal efectiva) respecto del número de rendijas n_1 . El máximo de la señal efectiva se obtiene en $n_1 = 29$.

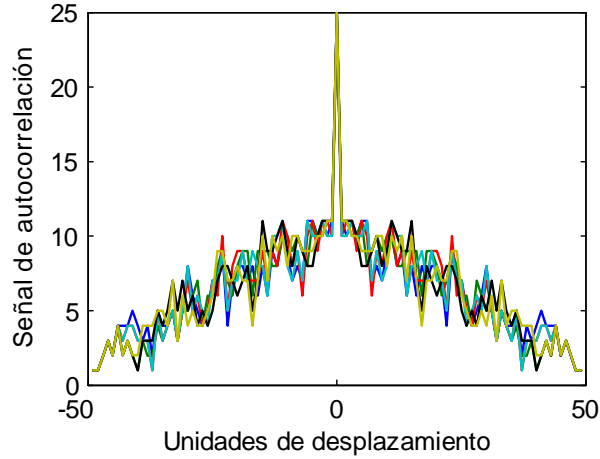


Figura 2.4: Seis de las 136 señales de autocorrelación obtenidas por DIRECT. Todas ellas son equivalentes en cuanto al valor del segundo máximo. El número de elementos $n = 50$ y $n_1 = 25$. El valor del segundo máximo es $\sigma = 11$.

que generan las señales de autocorrelación mostradas en la figura 2.4. Cada uno de los códigos genera una señal diferente pero todas tienen el mismo valor del segundo máximo. Entre las 136 soluciones encontradas, es posible seleccionar los ZRCs teniendo en cuenta otros criterios: sensibilidad a la difracción, sensibilidad respecto de la no uniformidad del haz de luz, etc.

2.4. Optimización mediante un algoritmo genético

En el apartado anterior presentamos una aproximación al problema de diseño de ZRCs basada en técnicas de optimización [9]. Con esta técnica conseguimos aumentar la capacidad de diseño de ZRCs hasta 100 elementos. No obstante, este algoritmo sufre un problema de dimensionalidad, es decir, el tiempo de procesado aumenta muy rápidamente cuando aumenta la dimensión del espacio de búsqueda –número de variables–. Para diseñar códigos de mayor longitud, es necesario disminuir el máximo número de evaluaciones de la función objetivo, obteniendo soluciones subóptimas. Esto nos permite utilizar menos memoria en la computadora, que es el principal factor limitante de este algoritmo. En este apartado proponemos un algoritmo genético para romper definitivamente la barrera en la longitud de los ZRCs, permitiendo el diseño de códigos óptimos de longitud arbitraria. El carácter binario del problema y el procesado paralelo de este tipo de algoritmos hace que sean una herramienta muy eficiente para este tipo de problemas.

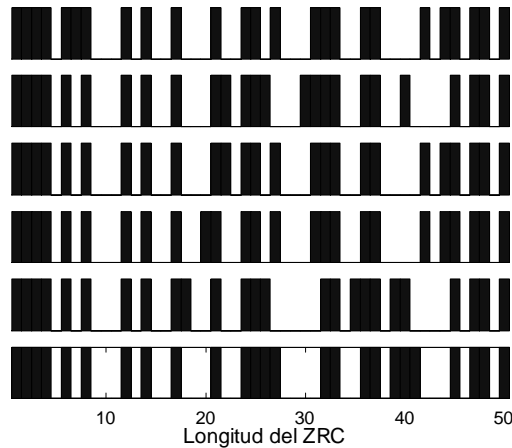


Figura 2.5: Seis de los 136 ZRCs encontrados por DIRECT con $n = 50$ y $n_1 = 25$.

2.4.1. Descripción de los algoritmos genéticos

Los algoritmos genéticos (GAs) son una herramienta robusta de optimización basada en los principios de la evolución y selección natural [10]. Estos algoritmos están basados en poblaciones, en las cuales un conjunto de soluciones potenciales del problema son sometidas a evolución por medio de la aplicación de operadores genéticos. Para aplicar un GA a un problema de optimización, es necesario codificar las variables de las que depende el problema. Normalmente las variables se codifican como cadenas de números binarios aunque también pueden ser enteros o reales. El algoritmo necesita una función objetivo y un conjunto de operadores, los cuales se aplican a la población de soluciones para poder mejorar la función objetivo.

El algoritmo utilizado para resolver nuestro problema comienza con una población de ZRCs binarios –individuos– generados aleatoriamente y con longitud n . Cada ZRC binario representa una posible solución del problema. A cada código se le asocia el valor que genera la función objetivo al ser evaluada con dicho ZRC. Este valor es conocido como “fitness”. La población de ZRCs evoluciona mediante la sucesiva aplicación de los operadores genéticos, selección, cruce y mutación. La selección es el proceso en el cual se eligen a los individuos aleatoriamente con una probabilidad proporcional a su valor de fitness, en nuestro caso el valor de σ . La selección se realiza mediante una estrategia elitista, los ZRCs con mejor fitness tienen mayor probabilidad de pasar a la siguiente generación, manteniendo la mejor solución a través de la evolución. Los individuos con bajo fitness también tienen cierta probabilidad de ser seleccionados, de esta manera se asegura la diversidad en la siguiente generación. El conjunto seleccionado, del mismo tamaño que la población inicial, está sujeto al operador de cruce. En

la figura 2.6 mostramos un ejemplo del funcionamiento del operador de cruce. Este proceso se realiza en dos etapas:

- Los ZRCs binarios se acoplan en parejas de manera aleatoria.
- Para cada pareja de ZRCs se selecciona una posición del código de manera aleatoria. A partir de esta posición se intercambian los elementos de un ZRC con los del otro obteniendo dos ZRCs nuevos. Esta operación se aplica a las parejas con una probabilidad P_c estrictamente menor que uno.

En un GA estándar, el último operador es el de mutación. Por medio de esta operación, cada elemento de cada ZRC de la población puede cambiar de 1 a 0 y viceversa, con una probabilidad muy pequeña, P_m . Este operador previene al GA de una convergencia prematura a soluciones subóptimas. La evolución del GA termina cuando un criterio dado se cumple, normalmente cuando se alcanza un número de generaciones (número de veces que se aplican los operadores genéticos).

La minimización de la función objetivo en el diseño de ZRCs está sujeta a las restricciones de variables binarias y a vectores con un número de predefinido de unos, ecuación 2.10. En un GA estándar la primera restricción se cumple de manera natural ya que el algoritmo normalmente trabaja con una codificación binaria de las variables. Para incluir la segunda restricción en el GA hay que introducir un operador extra, el operador de búsqueda restringida. Este operador ha sido utilizado en la referencia [11]. Después de aplicar los operadores de cruce y mutación, los individuos tendrán un número p de unos que –en general– será diferente a n_1 . Si $p < n_1$, el operador de búsqueda restringida añade aleatoriamente $(n_1 - p)$ unos al individuo. Si $p > n_1$, el operador selecciona aleatoriamente $(p - n_1)$ unos y los sustituye por ceros. El operador de búsqueda restringida reduce el tamaño del espacio de búsqueda de 2^n a $\binom{n}{n_1}$, [11]. La figura 2.6 muestra el pseudo-código de el GA implementado para resolver el problema de diseño de ZRCs.

2.4.2. Resultados

Para demostrar el funcionamiento del GA aplicado al problema descrito en la ecuación 2.10 configuramos el algoritmo con $P_c = 0,6$, $P_m = 0,01$, el número de generaciones lo fijamos a 500 y utilizamos una población de 100 individuos. Los valores de probabilidad de cruce y mutación vienen dados en [10]. En la figura 2.7 mostramos una comparación entre la cota inferior teórica de σ dada por la ecuación 2.9, el valor del segundo máximo alcanzado por DIRECT y el valor alcanzado por el GA. Las optimizaciones se han realizado para ZRCs con un número fijo de elementos, $n = 200$, y con un número de rendijas variable desde 1 a 199. Al igual que en la figura 2.2, en esta figura puede verse que solo es posible alcanzar la cota teórica en muy pocos casos, cuando el número de rendijas es pequeño. Como ya hemos comentado anteriormente, los problemas de más de 100 variables están fuera del régimen de funcionamiento de DIRECT pero podemos obtener soluciones subóptimas reduciendo el número de evaluaciones

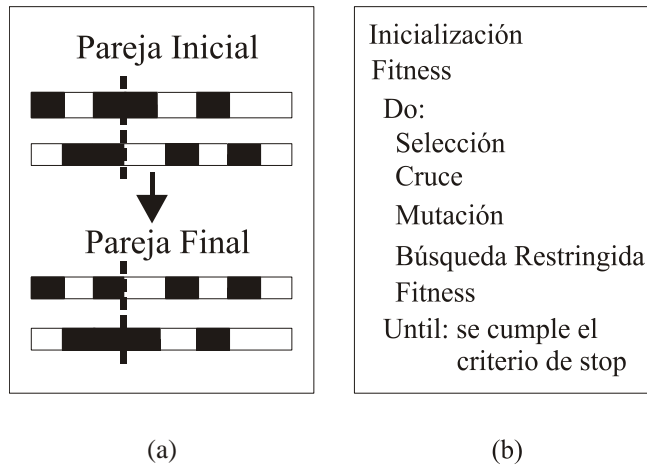


Figura 2.6: (a) Ejemplo de la operación de cruce, la pareja inicial intercambia sus genes a partir de la posición 3. (b) Pseudo-código implementado en el GA.

de la función. Por esta razón, las soluciones de DIRECT en la gráfica 2.7 son peores que las del GA.

Al igual que en la sección anterior, en la figura 2.8 mostramos el valor del primer y segundo máximo de la señal, así como la diferencia entre éstos –señal efectiva– respecto de n_1 . El máximo para la señal efectiva se obtiene para $n_1 = 110$ y la cantidad de señal efectiva es de 54 unidades de flujo. Para demostrar la eficacia del método basado en GAs hemos calculado varios casos con códigos de longitud arbitraria, en particular, un ZRC con $n = 1000$ elementos y $n_1 = 100$. Con estos parámetros el segundo máximo obtenido es 13. En la figura 2.9 mostramos la señal óptima de autocorrelación respecto del desplazamiento relativo entre ZRCs. La dimensión de este problema lo hace inabordable con las técnicas de cálculo tradicionales.

2.5. Conclusiones sobre diseño de ZRCs unidimensionales

A lo largo de este capítulo se han analizado los sistemas de generación de señales de referencia en codificadores ópticos, los ZRCs utilizados y las propiedades las señales generadas. Los cálculos implicados en el procedimiento de diseño de ZRCs son computacionalmente muy complejos y las técnicas tradicionales están limitadas a ZRCs de menos de 20 elementos. En este capítulo planteamos el problema desde el punto de vista de la optimización mediante un modelo muy simple del problema. En la primera parte del capítulo presentamos un algoritmo de optimización global llamado DIRECT, basado en el muestreo determinista de la función objetivo. Con este algoritmo expandimos las capa-

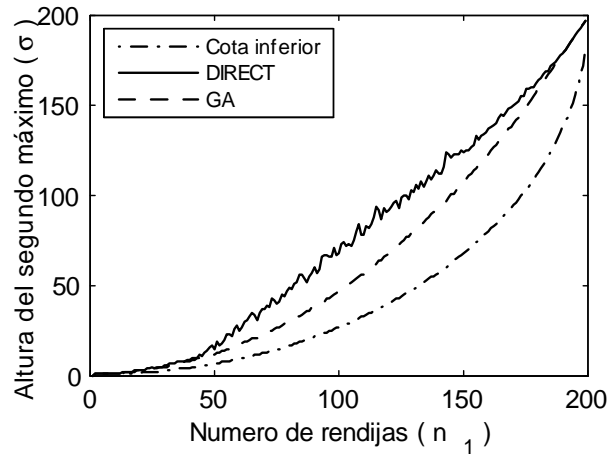


Figura 2.7: Comparación entre el segundo máximo obtenido por DIRECT, GA y el valor de la cota inferior teórica. La longitud del código es $n = 200$ y n_1 varía desde 1 a 199

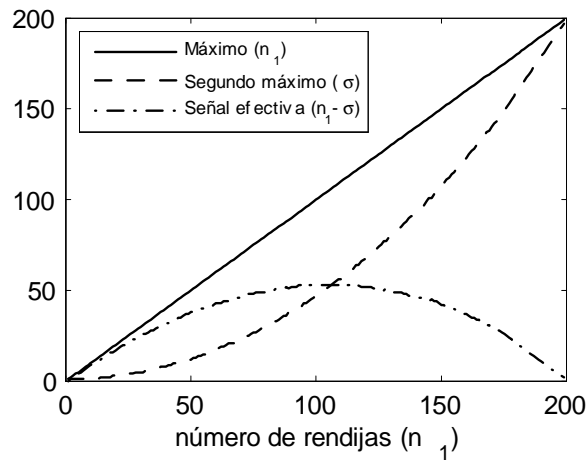


Figura 2.8: Primer máximo, segundo máximo y su diferencia (señal efectiva) respecto del número de rendijas n_1 . El máximo de la señal efectiva se obtiene en $n_1 = 67$.

2.5. CONCLUSIONES SOBRE DISEÑO DE ZRCs UNIDIMENSIONALES 25

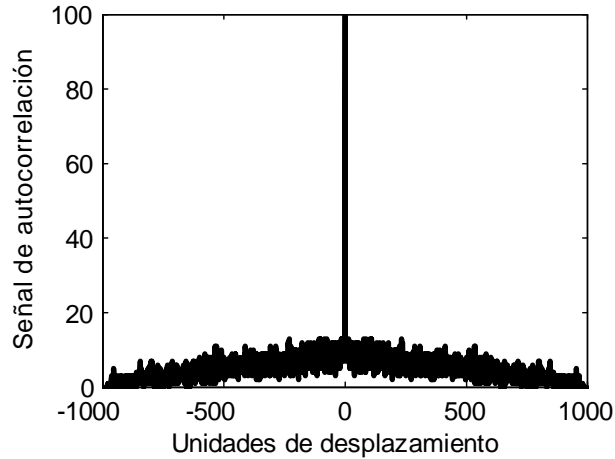


Figura 2.9: Señal de autocorrelación generada con un código óptimo. La longitud del código es 1000 y $n_1 = 100$. El valor del segundo máximo es 13.

ciudades de diseño de ZRCs, obteniendo resultados globalmente óptimos, hasta ZRCs con 100 elementos. El algoritmo obtiene varias soluciones con el mismo valor óptimo de la función objetivo, entre estas soluciones es posible seleccionar ZRCs atendiendo a otras propiedades de interés práctico. Además, basándonos en casos cuya solución es conocida, establecemos los límites de validez de esta técnica. Los requerimientos de sistemas de ultra-alta resolución imponen ZRCs de longitudes mayores. Para poder diseñar ZRCs de longitud arbitraria proponemos un nuevo algoritmo genético con un operador de búsqueda restringida. Este algoritmo se basa en un muestreo estocástico de la función objetivo que depende de variables –cromosomas– binarias. La eficiente gestión de variables binarias y el procesado paralelo de los vectores que describen a los ZRCs hacen que este algoritmo encuentre soluciones óptimas para problemas con miles de variables.

Capítulo 3

Diseño de ZRCs bidimensionales

3.1. Introducción

El proceso más crítico en fotolitografía es el alineamiento entre la máscara (o la imagen de la máscara en la fotolitografía de proyección) y la oblea de silicio. En un proceso CMOS estándar, la oblea contiene una estructura de capas superpuestas correspondientes a la grabación de los esquemas de varias máscaras. Las diferentes capas sobre la oblea deben estar perfectamente superpuestas, para ello es necesario realizar un alineamiento antes de cada exposición. En este capítulo analizamos una de las diversas técnicas de alineamiento usadas en alineadores de máscaras. Normalmente, las técnicas de alineamiento utilizadas en fotolitografía de contacto y proximidad son las mismas, mientras que en fotolitografía de proyección, debido a la distancia entre máscara y oblea, pueden ser diferentes. En 1972, King y Berry [12] desarrollaron un sistema de alineamiento de máscaras para fotolitografía de contacto y proximidad, basándose en una técnica de moire. La técnica consiste en hacer pasar un haz de luz a través de dos redes de difracción radiales; una de las redes está situada en la máscara y la otra en la oblea de silicio. Un operario observa las franjas de moire a través de un microscopio. Cuando las franjas se transforman en círculos concéntricos –en el caso de que los periodos sean ligeramente diferentes– las dos redes están alineadas. En [13], Torii y Mizushima introdujeron un ingenioso sistema de alineamiento para fotolitografía de contacto. En este sistema se utilizan dos redes de difracción lineales con relación de aspecto inferior a 1/2. Si se hace pasar un haz de luz colimado a través de las redes y éstas están en contrafase, se obtiene una red compuesta con la mitad del periodo. En esta nueva red desaparecen los órdenes de difracción impares. Si uno de los órdenes de difracción impar es registrado con un detector, se puede localizar la posición de contrafase de las redes. Más tarde, Makosch desarrolló un sistema interferométrico para la medida de desplazamiento entre capas superpuestas, [14], y para el alineamiento

de ellas en sistemas de fotolitografía de proyección, [15]. Este sistema también utiliza dos redes de difracción, pero en este caso, se basa en la medida de la diferencia de fase entre dos órdenes de difracción a la salida del sistema. La diferencia de fase se calcula mediante la detección del patrón de interferencia, el cual depende del desplazamiento de una de las redes. Este sistema es muy preciso pero es bastante complejo en cuanto al número de elementos utilizados. En 1985, Chitnis et al. [16] utilizó el efecto moire entre dos redes de difracción lineales, de las cuales una de ellas actúa por reflexión. Basándose en este fenómeno, desarrollaron un sistema para fotolitografía de proximidad. La variación de la intensidad de luz en alguno de los órdenes de difracción –típicamente se utiliza el orden 0– depende del desplazamiento relativo entre redes. Esta señal variable es registrada en un fotodiodo y se llama “señal de moire”. En este artículo se estudia teórica y experimentalmente la dependencia de la señal con la distancia entre redes. Se establecen las distancias de máximo contraste de la señal, es decir, las distancias correspondientes a los diferentes planos de Talbot. Además, se propone un sistema con cuatro redes de difracción que generan dos señales de moire desfasadas 180° . La resta entre las dos señales es la señal de alineamiento, esta resta cancela el ruido de las señales que está en fase y el alineador es más robusto. El control del sistema considera que se alcanza el alineamiento en uno de los puntos de corte entre las dos señales de moire, es decir, en uno de los ceros de la señal de alineamiento. En estos puntos la pendiente de la señal es mayor y se mejora la sensibilidad del alineador. Todos los sistemas basados en efecto moire utilizan para el alineamiento una señal periódica, en la cual detectan un máximo, un mínimo, o bien un cero. Estos sistemas son ambiguos para desplazamientos mayores a la mitad del periodo de la señal, y es necesario realizar un prealineamiento en el rango de la mitad del periodo. El prealineamiento suele hacerlo un operario mediante un procedimiento manual. Para evitar la ambigüedad en el alineamiento, Nomura et al. [17] utilizaron un array unidimensional para medir la separación entre los dos pares de redes que generan las señales desfasadas, y de esta manera, realizar un ajuste grueso del alineamiento. Esto hace que el sistema completo de adquisición de señal, tratamiento de la señal y proceso de alineamiento sea muy lento. En [18], Hara et al. da cuenta de los problemas que existen en los sistemas basados en efecto moiré y adapta el método para ser utilizado en fotolitografía de proyección, es decir, formando imagen de la red situada en la oblea sobre la máscara. Entre estos problemas destacamos dos: la múltiple interferencia entre la máscara y la oblea que obliga a mantener la distancia entre ellos con una precisión de $\lambda/2$ (donde λ es la longitud de onda de la luz), y la aberración cromática del sistema de proyección, que está diseñado para la luz de exposición (típicamente láser de excímero, o bien, lámpara de arco o de mercurio-xenón) y no para el láser de He-Ne utilizado en el alineamiento. También Sugimoto et al. [19] describen los problemas de la inhomogeneidad en la reflectividad en la oblea (especialmente en el aluminio), inhomogeneidades en las diferentes capas de resina fotosensible depositada en la oblea y los defectos de fabricación de las redes de difracción. La técnica basada en moire y algunas modificaciones para medir desplazamientos angulares han sido ampliamente estudiados en la literatura.

No es difícil observar las similitudes entre la señal de referencia en un codificador óptico y la señal de alineamiento en un alineador de máscaras de fotolitografía. En el capítulo anterior hemos diseñado ZRCs que generan señales de referencia unidimensionales por autocorrelación de la transmitancia de los ZRCs, [9] y [20]. Huang et al. [21], y más tarde Chen et al. [22] propusieron una nueva clase de ZRCs bidimensionales (2D ZRC) para generar una señal de alineamiento bidimensional en alineadores de máscaras. El funcionamiento del sistema es similar al utilizado en codificadores. Los códigos están enfrentados y cuando el movimiento tiene lugar en el plano XY, la señal obtenida es la autocorrelación bidimensional de los dos 2D ZRCs. De esta manera, se pueden lograr señales de alineamiento bidimensional de alta calidad con un sencillo sistema. En el artículo de Chen et al. se estudian las características de las señales bidimensionales de autocorrelación y se calcula la degradación de la señal por la difracción del haz a través de los códigos. En cuanto al diseño de 2D ZRCs adecuados para generar la señal de alineamiento, no existe un método sistemático de cálculo. De nuevo surge el problema del diseño de 2D ZRCs óptimos, con la diferencia de que este problema es dimensionalmente mucho más complicado que su versión unidimensional. Este capítulo está organizado de la siguiente manera: en primer lugar, revisamos algunas propiedades generales de las señales que provienen de una autocorrelación bidimensional, en segundo lugar calculamos una cota inferior novedosa para este tipo de señales, y finalmente, presentamos la aplicación de las técnicas desarrolladas en el capítulo anterior para el cálculo de 2D ZRCs que generan señales de alineamiento óptimas. Los resultados de este capítulo están publicados en [23] y [24].

3.2. Consideraciones generales

Los ZRCs bidimensionales son una generalización natural de los ZRCs unidimensionales, consisten en un conjunto de elementos –píxeles– opacos o transparentes especialmente distribuidos. Estos píxeles son la versión bidimensional de las rendijas unidimensionales, por lo tanto, también son binarios. Uno de los códigos se sitúa en una esquina de la máscara de grabación y el otro en la misma esquina de la oblea donde se graba la máscara. Los 2D ZRCs pueden fabricarse en una placa de vidrio transparente sobre la que se deposita una capa de cromo. En la figura 3.1 se muestra el sistema de alineamiento utilizado en fotolitografía. Los 2D ZRCs se disponen de forma paralela y uno de ellos está sobre una plataforma móvil en las direcciones X e Y. Un haz colimado pasa perpendicularmente a través de la superposición de los 2D ZRCs y el flujo total transmitido se detecta mediante un fotodiodo. La señal registrada en el fotodiodo depende del desplazamiento relativo entre 2D ZRCs en las direcciones X e Y. Al igual que en el caso unidimensional, consideramos los dos 2D ZRCs idénticos con el objetivo de obtener la mayor cantidad de señal posible.

En general, la estructura de un 2D ZRC puede representarse por la siguiente

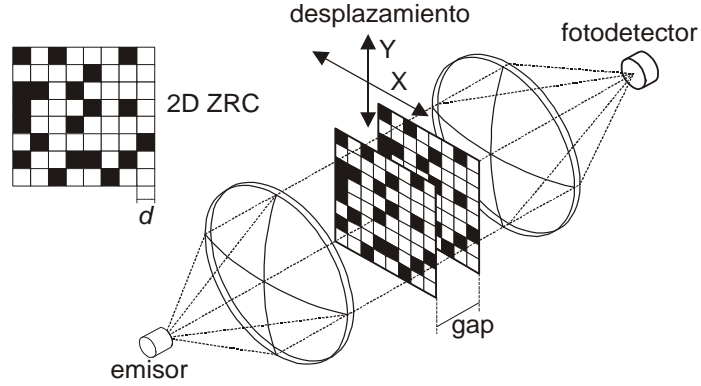


Figura 3.1: Sistema de alineamiento bidimensional basado en ZRCs bidimensionales.

matriz de elementos binarios

$$\mathbf{c} = [c_{ij}] = \begin{bmatrix} c_{11} & \cdots & c_{1n} \\ \vdots & \ddots & \vdots \\ c_{n1} & \cdots & c_{nn} \end{bmatrix}, \quad c_{ij} \in \{0, 1\}, \quad (3.1)$$

donde n^2 es el número total de elementos del 2D ZRC, $c_{ij} = 1$ si en la posición ij hay un pixel transparente y $c_{ij} = 0$ si el pixel es opaco. El número de pixeles transparentes es n_1 . El área de las regiones transparentes u opacas será un múltiplo del área de un pixel. Sin pérdida de generalidad, y para simplificar el problema, consideramos pixeles cuadrados. Al igual que en el capítulo anterior, asumiremos que el haz de iluminación es plano-paralelo y los efectos difractivos son despreciables en las distancias de trabajo.

Si el desplazamiento relativo entre los dos 2D ZRCs es de k y l unidades en las direcciones X e Y respectivamente, la señal registrada en el fotodiodo es proporcional a

$$S_{kl} = \sum_{i=1}^{n-k} \sum_{j=1}^{n-l} c_{ij} c_{i+k, j+l}, \quad (3.2)$$

donde $k, l = -n + 1, \dots, n - 1$, y la señal S_{kl} es la matriz de autocorrelación de los 2D ZRCs definidos en la ecuación 3.1. El factor de proporcionalidad entre S_{kl} y la señal real obtenida en el fotodiodo dependerá de la conversión entre la irradiancia incidente en el área de superposición entre 2D ZRCs y la intensidad eléctrica generada en el fotodiodo. S_{00} es la señal obtenida cuando el desplazamiento relativo entre 2D ZRCs es cero, este valor siempre es el máximo central de la señal y es igual al número de pixeles transparentes, n_1

$$S_{00} = \sum_{i=1}^n \sum_{j=1}^n c_{ij}^2 = \sum_{i=1}^n \sum_{j=1}^n c_{ij} = n_1. \quad (3.3)$$

El máximo secundario de la señal es

$$\sigma = \max_{k^2+l^2 \neq 0} [S_{kl}], \quad (3.4)$$

donde $k^2 + l^2 \neq 0$ significa que k y l no son cero al mismo tiempo.

Al igual que ocurre con la señal de referencia unidimensional, en las señales de alineamiento bidimensionales el parámetro más importante que caracteriza a una señal de referencia es la relación entre el segundo máximo y el máximo central –o primer máximo–, $K = \sigma/S_{00}$. Una señal de alineamiento adecuada debe tener un máximo central aislado y bien definido, de manera que el resto de la señal no tiene ningún interés práctico. Cuanto más pequeño es el parámetro K , más alto será la sensibilidad y robustez de la señal de alineamiento.

En ausencia de difracción, el tamaño de los píxeles de los 2D ZRCs definen la anchura del pico central de la señal de alineamiento y esta anchura es la resolución del sistema de alineamiento. El diámetro del haz de luz limita el número de píxeles en el 2D ZRC y la sensibilidad del fotodiodo determina el valor mínimo para el máximo central de la señal, es decir, el número de píxeles transparentes del 2D ZRC. Teniendo en cuenta estos requerimientos, los valores de n y n_1 están predeterminados y el objetivo es minimizar el segundo máximo de la señal, σ . En la siguiente sección calculamos una nueva cota inferior para el segundo máximo de la señal.

3.3. Cota inferior del segundo máximo

En [3], Y. Li calculó una cota inferior para el segundo máximo de las señales unidimensionales generadas con ZRCs unidimensionales. En alineadores de máscaras, Chen et al. [22], también obtuvieron una cota inferior para la señal bidimensional. En sus cálculos, consideran los n_1 píxeles transparentes concentrados en una esquina del 2D ZRC. La cota inferior calculada de esta manera es muy conservadora, ésta viene dada por la siguiente expresión:

$$\sigma \geq \frac{\sqrt{n_1} (\sqrt{n_1} + 3) (\sqrt{n_1} - 1)}{4(n^2 - 1)} \quad (3.5)$$

Nosotros obtendremos un valor mucho más ajustado para esta cota mediante una generalización bidimensional del cálculo utilizado por Y. Li [3] en el caso de ZRCs unidimensionales. La autocorrelación bidimensional satisface las siguientes propiedades:

- i. Cuando un 2D ZRC se desplaza respecto del otro 2D ZRC a lo largo de toda su área (en las direcciones X e Y), la integral de la intensidad eléctrica

que se registra en el fotodiodo es proporcional a

$$\begin{aligned} \sum_{k=-n+1}^{n-1} \sum_{l=-n+1}^{n-1} S_{kl} &= \sum_{k=-n+1}^{n-1} \sum_{l=-n+1}^{n-1} \sum_{i=1}^n \sum_{j=1}^n c_{ij} c_{i+k, j+l} = \\ &= \sum_{i=1}^n \sum_{j=1}^n c_{ij} \sum_{k=-n+1}^{n-1} \sum_{l=-n+1}^{n-1} c_{i+k, j+l} = n_1^2. \end{aligned} \quad (3.6)$$

- ii. Al igual que su versión unidimensional, la señal de autocorrelación bidimensional es simétrica, $S_{k,l} = S_{-k,-l}$.
- iii. El valor de S_{kl} es igual al número de píxeles transparentes que coinciden (se enfrentan) en los dos 2D ZRCs cuando uno de ellos se ha desplazado k y l unidades en las direcciones X e Y respectivamente. Si dos píxeles coinciden en este caso, existirá un $c_{ij} = 1$ y un $c_{i+k,j+l} = 1$ en el código. En otras palabras, existe en el código una subestructura con dimensiones $(k+1, l+1)$ de la forma,

$$\begin{bmatrix} 1_{i,j} & \cdots & X_{i,j+l} \\ \vdots & \ddots & \vdots \\ X_{i+k,j} & \cdots & 1_{i+k,j+l} \end{bmatrix}, \quad (3.7)$$

donde $X \in \{0, 1\}$. Más aún, el valor de S_{kl} es el número de subestructuras de dimensión $(|k|+1, |l|+1)$ contenidas en el 2D ZRC. En la dimensión de las subestructuras hemos tenido en cuenta que los índices de la autocorrelación pueden ser negativos. Si la dimensión del 2D ZRC es $n \times n$, el máximo número de subestructuras contenidas en él es $(n - |k|)(n - |l|)$, así pues, tenemos una cota superior para el valor de la autocorrelación,

$$S_{kl} \leq (n - |k|)(n - |l|). \quad (3.8)$$

Para valores de $|k|$ y $|l|$ próximos a n , el número de subestructuras contenidas en el 2D ZRC es pequeño (hay poco área de solapamiento) y esta cota superior es también pequeña. En el caso extremo, $S_{kl} \leq 0$ para $k = n$ y $l = n$. Por otro lado, cuando los desplazamientos $|k|$ y $|l|$ son pequeños (próximos a 0), la cota superior es muy grande. En el caso límite $S_{kl} \leq n^2$ cuando $k = 0$ y $l = 0$.

- iv. Considerando el primer máximo de la señal S_{00} en la ecuación 3.3 y el segundo máximo σ en la ecuación 3.4, podemos establecer otro valor para la cota superior de la señal,

$$\frac{S_{kl}}{k^2+l^2 \neq 0} \leq \sigma. \quad (3.9)$$

Esta cota inferior es constante y, obviamente cumple $\sigma = n_1 \leq n^2$.

- v. De las dos cotas superiores, cuando k y l están cercanos a n , la cota más pequeña (la que más se ajusta a la señal) viene dada por la ecuación 3.8, y para k y l cercanos a 0 la mejor cota viene dada por la ecuación 3.9. El límite entre las dos situaciones ocurre cuando las dos cotas son iguales, entonces k y l cumplen,

$$|l_1| = n - \frac{\sigma}{n - |k_1|}. \quad (3.10)$$

Teniendo en cuenta la propiedad (v), cuando $|l| \leq |l_1|$ la cota más pequeña viene dada por la ecuación 3.9, mientras que, si $|l| > |l_1|$ la cota más pequeña viene dada por la ecuación 3.8. Para algunos valores de k_1 y σ , la ecuación 3.10 puede tomar valores negativos y solo tendremos la cota dada por la ecuación 3.8. Estos valores ocurren cuando

$$|k| > n - \frac{\sigma}{n}. \quad (3.11)$$

Resumiendo, si $|k| \leq n - \sigma/n$, S_{kl} tiene dos cotas superiores,

$$S_{kl} \leq \begin{cases} \sigma & 0 \leq |l| = n - \frac{\sigma}{n - |k|} \\ (n - |k|)(n - |l|) & n - \frac{\sigma}{n - |k|} \leq |l| < n \end{cases}. \quad (3.12)$$

Por otro lado, si $|k| > n - \sigma/n$ solo tenemos que considerar una cota superior,

$$S_{kl} = (n - |k|)(n - |l|), \quad 0 \leq |l| \leq n - 1. \quad (3.13)$$

Nuestro objetivo es calcular una cota inferior de S_{kl} . Para ello, sustituimos en la ecuación 3.6 el valor de la señal S_{kl} por sus cotas superiores. Por medio de la propiedad (ii), desdoblamos la suma de la ecuación 3.6 en tres términos:

$$\sum_{k=-n+1}^{n-1} \sum_{l=-n+1}^{n-1} S_{kl} = S_{00} + 2 \sum_{l=1}^{n-1} S_{0l} + 2 \sum_{k=1}^{n-1} \sum_{l=-n+1}^{n-1} S_{kl}. \quad (3.14)$$

Calculamos la primera suma de la parte derecha de la ecuación 3.14, para ello usamos la ecuación 3.12 y volvemos a desdoblar la suma en dos términos, en los cuales sabemos la cota superior de S_{0l}

$$\sum_{l=1}^{n-1} S_{0l} = \sum_{l=1}^{n-\frac{\sigma}{n}} \sigma + \sum_{l=n-\frac{\sigma}{n}+1}^{n-1} n(n-l) = \frac{1}{2} \left(-\frac{\sigma^2}{n} + \sigma(2n-1) \right). \quad (3.15)$$

Ahora calculamos la segunda suma de la ecuación 3.14. Volvemos a desdoblar esta suma teniendo en cuenta el valor de los índices k y l y usando las

desigualdades 3.8 y 3.12 obtenemos,

$$\begin{aligned}
\sum_{k=1}^{n-1} \sum_{l=-n+1}^{n-1} S_{kl} &= \sum_{k=1}^{n-\frac{\sigma}{n}} \sum_{l=-n+1}^{n-1} S_{kl} + \sum_{k=n-\frac{\sigma}{n}+1}^{n-1} \sum_{l=-n+1}^{n-1} S_{kl} \leq \\
&\leq \sum_{k=1}^{n-\frac{\sigma}{n}} \sum_{l=-n+1}^{n-\frac{\sigma}{n}-(n-\frac{\sigma}{n-k})} (n-k)(n+l) + \sum_{k=1}^{n-\frac{\sigma}{n}} \sum_{l=-(n-\frac{\sigma}{n-k})+1}^{n-\frac{\sigma}{n}} \sigma + \\
&+ \sum_{k=1}^{n-\frac{\sigma}{n}} \sum_{l=n-\frac{\sigma}{n-k}+1}^{n-1} (n-k)(n-l) + \sum_{k=n-\frac{\sigma}{n}+1}^{n-1} \sum_{l=-n+1}^0 (n-k)(n+l) + \\
&+ \sum_{k=n-\frac{\sigma}{n}+1}^{n-1} \sum_{l=1}^{n-1} (n-k)(n-l) = \frac{1}{2} \left(-3\sigma^2 + \sigma n(4n-1) - 2\sigma \sum_{k=1}^{n-\frac{\sigma}{n}} \frac{\sigma}{n-k} \right). \tag{3.16}
\end{aligned}$$

Es fácil demostrar que la última suma de la expresión anterior está acotada por $\sum_{k=1}^{n-\frac{\sigma}{n}} \frac{\sigma}{n-k} \geq n \left(n - \frac{\sigma}{n} \right)$. Sustituyendo la ecuación 3.16 y 3.15 en la ecuación 3.14 y usando las ecuaciones 3.3 y 3.6, obtenemos:

$$0 \geq n_1(1-n_1) + \sigma(2n^2+n-1) - \sigma^2 \left(1 + \frac{1}{n} \right). \tag{3.17}$$

El segundo máximo de la señal debe cumplir la desigualdad $\sigma_1 \leq \sigma$, donde σ_1 es la cota inferior del segundo máximo de la señal. El valor de σ_1 es una de las soluciones de la ecuación de segundo grado 3.17. En particular, la solución positiva es

$$\sigma_1 = \frac{-(2n^2+n-1) + \sqrt{(2n^2+n-1)^2 + 4\left(1+\frac{1}{n}\right)n_1(1-n_1)}}{-2\left(1+\frac{1}{n}\right)}. \tag{3.18}$$

Estos resultados son novedosos y han sido publicados en [9]. Existen casos sencillos donde la cota es alcanzada, pero al igual que en el caso unidimensional, dado n y n_1 no se puede garantizar que haya 2D ZRCs con segundo máximo igual a la cota inferior. Así pues, el objetivo es el cálculo sistemático de 2D ZRCs cuya autocorrelación tenga el valor del segundo máximo más pequeño posible. Al igual que en el capítulo anterior, aplicamos el algoritmo DIRECT para resolver este problema y finalmente utilizaremos un algoritmo genético para el diseño de 2D ZRCs con dimensiones arbitrarias.

3.4. Optimización con el algoritmo DIRECT

En esta sección aplicaremos el algoritmo DIRECT al diseño de 2D ZRCs óptimos. En esta aplicación la función objetivo es,

$$\begin{aligned}
 & \min_{\mathbf{c} \in \text{binary}} f(\mathbf{c}) \\
 f(\mathbf{c}) &= \max_{k^2+l^2 \neq 0} \{S_{kl}\} \\
 S_{kl} &= \sum_{i=1}^{n-k} \sum_{j=1}^{n-l} c_{ij} c_{i+k, j+l},
 \end{aligned} \tag{3.19}$$

donde $f(\mathbf{c}) = \sigma$ es el segundo máximo de la señal de autocorrelación y \mathbf{c} es una matriz binaria. Al igual que en el caso unidimensional, una forma de conseguir una matriz binaria es considerando variables enteras y acotadas entre 0 y 1. Además, es necesario fijar el máximo de la señal considerando ZRCs con un número fijo de píxeles transparentes. Teniendo en cuenta estas consideraciones, las restricciones quedan

$$\begin{aligned}
 & \mathbf{c} \in \mathbb{Z}^{n \times n} \\
 & 0 \leq c_{ij} \leq 1 \\
 & \sum_{i=1}^n \sum_{j=1}^n c_{ij} = n_1.
 \end{aligned} \tag{3.20}$$

En la figura 3.2 se muestra una comparación entre la cota inferior del segundo máximo que hemos calculado en la sección anterior –ecuación 3.18–, la cota inferior dada por Chen et al. –ecuación 3.5– y el valor del segundo máximo alcanzado por DIRECT. Las optimizaciones han sido realizadas con 100 elementos, $n = 10$, y un número variable de rendijas desde 1 a 99. En esta figura puede verse que en muy pocos casos se alcanza la cota inferior teórica, y que estos casos tienen lugar cuando el número de píxeles es pequeño. También puede observarse que la cota dada por Chen es muy conservadora mientras que la cota calculada por nosotros se ajusta mejor a las soluciones calculadas. Como en el capítulo anterior, en algunos casos la sensibilidad del detector no es un parámetro restrictivo y el máximo central de la señal de autocorrelación no es fijo. En este caso, la función objetivo a optimizar deberá ser la diferencia entre el primer y segundo máximo de la señal. En la figura 3.3 mostramos el valor del primer y segundo máximo de la señal, así como la diferencia entre éstos (señal efectiva) respecto de n_1 . El máximo para la señal efectiva se obtiene para $n_1 = 62$ y la señal efectiva obtenida es de 33 unidades. En la figura 3.4, se muestra la señal de autocorrelación óptima para un 2D ZRC con 10×10 elementos y 50 píxeles transparentes. Al igual que en el caso unidimensional, además del diseño de 2D ZRCs óptimos, una ventaja del método de calculo basado en DIRECT es la capacidad de encontrar soluciones equivalentes con el mismo valor del segundo máximo. Entre estas soluciones, es posible seleccionar los 2D ZRCs teniendo en cuenta otros criterios: sensibilidad a la difracción, sensibilidad a haces que no son uniformes, etc.

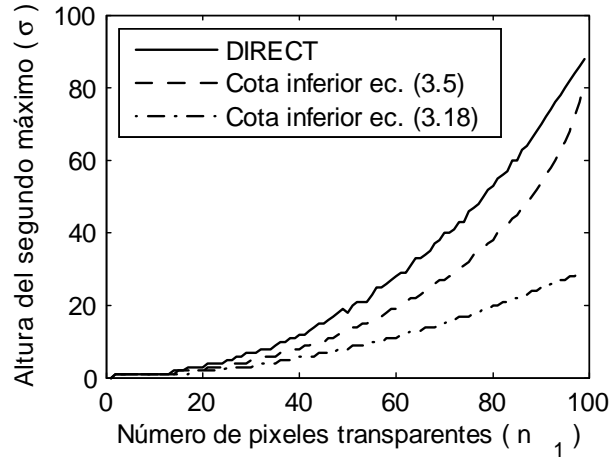


Figura 3.2: Altura del segundo máximo de la señal de autocorrelación con $n = 10$. La gráfica continua es alcanzada con DIRECT, la punteada es la cota inferior calculada en la ecuación 3.18 y la discontinua es la cota calculada por Chen mediante la expresión 3.5.

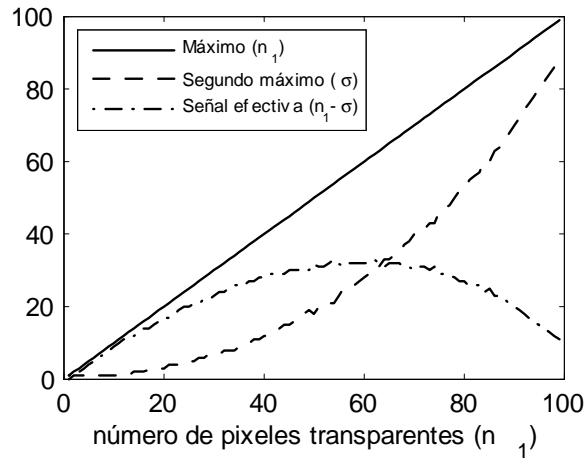


Figura 3.3: Primer máximo, segundo máximo y su diferencia (señal efectiva) respecto del número de rendijas n_1 . El máximo de la señal efectiva es de 33 unidades y se obtiene en $n_1 = 62$.

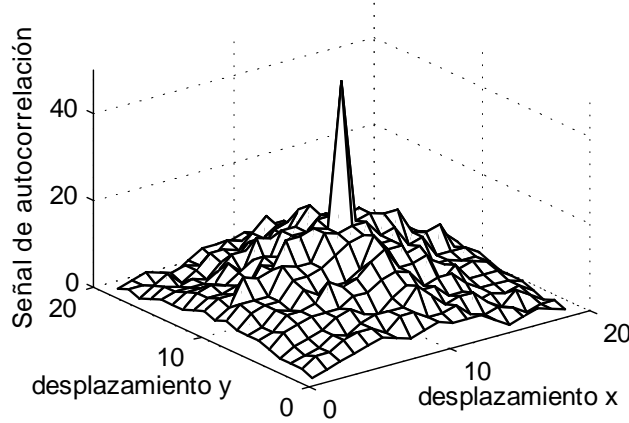


Figura 3.4: Señal de referencia óptima para un caso donde $n = 10$ y $n_1 = 50$.

3.5. Optimización mediante un algoritmo genético

En la sección anterior y en la referencia [23] presentamos un método sistemático de diseño de 2D ZRCs basado en optimización mediante el algoritmo DIRECT. El algoritmo obtiene soluciones óptimas para códigos hasta 10×10 elementos. En 2D ZRCs de mayor tamaño, DIRECT sufre un problema de dimensionalidad. Los requerimientos de memoria del algoritmo hacen que el diseño sea inabordable con una computadora de sobremesa. Como en el caso unidimensional, estos requerimientos de memoria se pueden controlar acotando el número de evaluaciones de la función objetivo, pero esta limitación conlleva la obtención de soluciones subóptimas. En esta sección demostramos que –también en este caso– los algoritmos genéticos superan esta dificultad, obteniendo 2D ZRCs de longitud arbitraria. Para ello, aplicamos el algoritmo genético con el operador de búsqueda restringida desarrollado en el capítulo anterior, [11]. El algoritmo genético se ejecutó con las mismas probabilidades de cruce y mutación que en el capítulo anterior, $P_c = 0,6$ y $P_m = 0,01$, respectivamente. La evolución del GA termina cuando se cumple un criterio de parada preestablecido, en nuestro caso, cuando se alcanza un número de iteraciones dado (en nuestro caso, 400 generaciones). La función objetivo es la misma que la que usamos en la sección previa, ecuación 3.19. A diferencia de DIRECT, el algoritmo genético solo utiliza la restricción del máximo de la señal, ya que este tipo de algoritmos gestiona las variables binarias de manera natural. Así pues, la única restricción es $\sum_{i=1}^n \sum_{j=1}^n c_{ij} = n_1$.

En la figura 3.5 demostramos el funcionamiento del algoritmo genético. Para

ello, calculamos el segundo máximo de la señal generada con códigos de 16×16 píxeles ($n = 16$). El número de píxeles transparentes, n_1 , varía desde 1 a 255. Los cálculos se han realizado con el GA implementado, con el algoritmo DIRECT [25] y evaluando la cota inferior, ecuación 3.18. El algoritmo DIRECT no gestiona esta cantidad de variables y se obtienen soluciones subóptimas con un bajo número de evaluaciones de la función objetivo. En este caso hemos evaluado la función objetivo $1,2 \times 10^5$ veces mientras que normalmente necesitamos evaluarla 3×10^5 veces para garantizar soluciones óptimas. Hay que resaltar que las soluciones obtenidas con el GA están más cercanas a la cota inferior y son significativamente mejores que las obtenidas con DIRECT. En [22], Chen muestra un 2D ZRC de 16×16 elementos con 64 píxeles transparentes ($n_1 = 64$). La autocorrelación de este código tiene un segundo máximo de 16 píxeles. En la figura 3.5, el valor del segundo máximo de la señal alcanzado por el GA es de 13 píxeles, en consecuencia, demostramos que el 2D ZRC usado en [22] no es óptimo. La máxima señal efectiva que podemos conseguir es de 78 unidades y se alcanza con 178 píxeles transparentes, las gráficas del primer, segundo máximo y señal efectiva se muestran en la figura 3.6. La dimensión de los

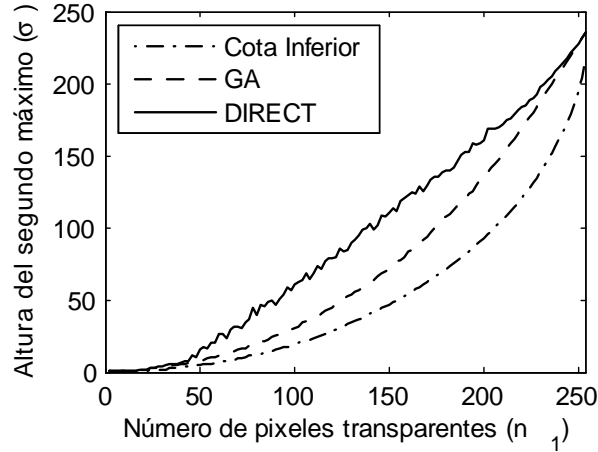


Figura 3.5: Altura del segundo máximo de la señal de autocorrelación respecto al número de píxeles transparentes. Los ZRCs tienen 16×16 píxeles y el número de píxeles transparentes varía desde 1 a 255. La gráfica continua es la cota inferior teórica, la punteada se ha obtenido con GA y la discontinua se ha alcanzado con el algoritmo DIRECT.

códigos analizados anteriormente es relativamente pequeña, aunque es adecuada para comparar el GA con otras técnicas. Para demostrar el potencial del GA con grandes cantidades de variables, hemos calculado un 2D ZRC óptimo con 100×100 píxeles y con 300 píxeles transparentes. La altura del segundo máximo es 14, como se puede observar, se ha conseguido un máximo de autocorrelación

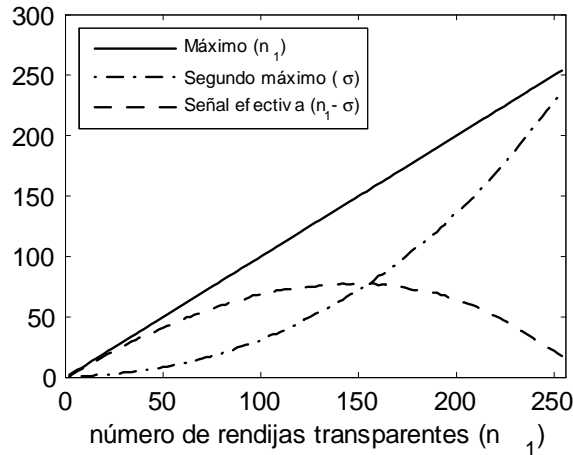


Figura 3.6: Primer máximo, segundo máximo y su diferencia (señal efectiva) respecto del número de rendijas n_1 . El máximo de la señal efectiva es de 78 unidades y se obtiene en $n_1 = 178$.

extremadamente pronunciado. La señal de autocorrelación bidimensional para este caso se muestra en la figura 3.7.

3.6. Conclusiones sobre diseño de ZRCs bidimensionales

Hasta la realización de este trabajo, no existía un método sistemático para el diseño de ZRCs bidimensionales que generan señales de alineamiento óptimas. El diseño de ZRCs bidimensionales, al igual que su versión unidimensional, es una tarea computacionalmente muy costosa. Además, en el caso bidimensional el crecimiento del número de variables respecto del tamaño de los 2D ZRCs es cuadrático. Para desarrollar métodos de diseño de 2D ZRCs, hemos hecho un estudio de las características de las señales que provienen de una autocorrelación bidimensional. En este estudio hemos obtenido una nueva cota inferior del segundo máximo de la señal que mejora a la que existía en la literatura. La aplicación del algoritmo DIRECT para el diseño de 2D ZRCs nos permite el diseño de códigos óptimos hasta 10×10 píxeles y códigos subóptimos hasta 16×16 píxeles, además, este algoritmo nos proporciona diferentes 2D ZRCs que son equivalentes en cuanto al valor del segundo máximo. Es importante obtener varios 2D ZRCs que generan el mismo tipo de señal, ya que podemos hacer una selección atendiendo a otros criterios ópticos. El fuerte crecimiento del número de variables con el tamaño del 2D ZRC hace que DIRECT no sea una herramienta práctica para códigos mayores a 16×16 píxeles. Al igual

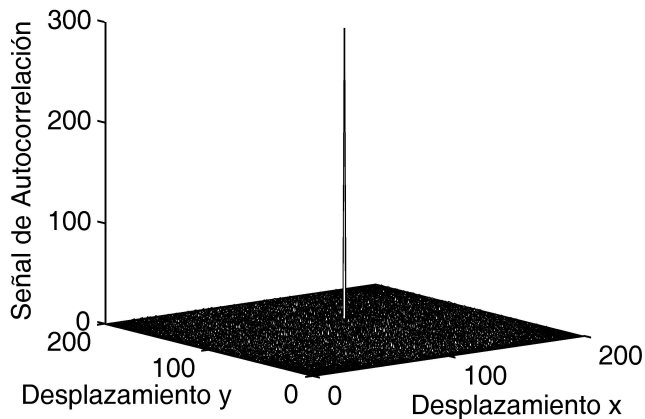


Figura 3.7: Señal de autocorrelación obtenida con un 2D ZRC de 100×100 píxeles y 300 píxeles transparentes (n_1). El segundo máximo conseguido es 14.

que en el capítulo anterior, para resolver este problema proponemos un nuevo algoritmo genético con un operador extra de búsqueda restringida. Con este algoritmo diseñamos un 2D ZRC de 16×16 píxeles que mejora a los existentes en la literatura. Además, hemos diseñado un 2D ZRC de 100×100 píxeles para demostrar que este algoritmo se adapta perfectamente a este tipo de problemas y, definitivamente, superamos las dificultades de diseño de ZRCs de gran dimensión.

Capítulo 4

Estrechamiento de señales de referencia

4.1. Introducción

En el capítulo 2 hemos presentado las técnicas de optimización que definitivamente resuelven el problema de diseño de ZRCs unidimensionales. Los ZRCs diseñados generan señales de referencia óptimas para codificadores ópticos basados en redes de difracción. En todos los trabajos sobre diseño de ZRCs se considera como fuente de iluminación un haz planoparalelo y se ignoran los efectos difractivos del haz en la propagación a través de los ZRCs. Como ya hemos comentado en los capítulos anteriores, esta aproximación solo es válida cuando la distancia entre códigos es mucho menor que la anchura de las rendijas, y a su vez la anchura de las rendijas es mucho mayor que la longitud de onda de la luz utilizada. Teniendo en cuenta estas aproximaciones, la señal de referencia obtenida es la autocorrelación de la transmitancia del ZRC. El pico central de la señal es un triángulo y la anchura de la base del triángulo es el doble de la anchura de las rendijas que componen el ZRC.

Recientemente se ha propuesto el uso de haces de luz espacialmente incoherentes –luz difusa– para incrementar la resolución de los codificadores basados en redes de difracción. La iluminación difusa permite mantener algunas tolerancias de montaje y aumentar otras, [26] [27]. Esta nueva tecnología tiene dos ventajas fundamentales: la primera es que permite una reducción del periodo de las redes de difracción, y la segunda es que permite un aumento en la distancia entre las redes. Normalmente, el periodo de la red de difracción es igual al periodo de la señal incremental, por lo tanto, la reducción del periodo de la red aumenta la resolución del sistema. Para no perder calidad en la medida, el aumento de la resolución de un codificador debe ir acompañado de un aumento en la precisión de éste. Respecto de la segunda ventaja, el incremento de la distancia entre redes de difracción implica mejores tolerancias de montaje, que redundan en codificadores más robustos.

El uso de señales de referencia permite la conversión de la medida incremental a una medida absoluta, para ello solamente hay que desplazar la cabeza de lectura desde la posición actual hasta la referencia. Para que la determinación de la posición absoluta sea válida, la señal de referencia debe tener una anchura a media altura igual al periodo de la señal incremental. Además, y de forma ideal, el flanco de subida de la señal de referencia debe estar en buena sincronía con un flanco de subida de la señal incremental. Según esto, el aumento de la resolución conlleva una reducción en la anchura de la señal de referencia, es decir, una disminución de la anchura de las rendijas de los ZRCs. Por otro lado, la distancia entre redes de difracción típicamente es igual a la distancia entre regla y cabeza, y por tanto, debe ser igual a la distancia entre ZRCs. Por esta razón, si aumenta la distancia entre redes de difracción de la etapa incremental, aumenta en la misma medida la distancia entre ZRCs. La disminución de las rendijas de los ZRCs, junto con un aumento de la distancia entre ellos, contribuyen a un considerable aumento de los efectos difractivos en la propagación del haz. La distancia entre los ZRCs hace que la sombra del primero sobre el segundo se degrade por la difracción de la luz, efecto que no podemos seguir ignorando con rendijas de anchuras inferiores a 20 micras. En esta situación, la señal registrada en el fotodetector no se puede calcular como la autocorrelación entre ZRCs. Como veremos más adelante, la difracción del haz en los ZRCs provoca un suavizado de la señal de referencia respecto de la señal de autocorrelación. Se produce un ensanchamiento del pico central de la señal de referencia y una disminución de la altura de este pico, disminuyendo la cantidad efectiva de señal y el parámetro K de la señal.

En este capítulo tratamos el efecto de la difracción sobre las señales de referencia unidimensionales, estudiamos la degradación que produce este efecto sobre la señal y proponemos un método para compensar el ensanchamiento del pico central de ésta. En primer lugar se genera una señal óptica de referencia por medio de ZRCs. Esta señal consiste en una distribución espacial de luz degradada por la difracción. La distribución espacial de luz –o perfil de irradiancia incidente– se desplaza a lo largo del plano de detección según el movimiento de la cabeza de lectura. Dicho perfil, al desplazarse, se correlaciona con una máscara adecuada situada sobre un sistema de fotodetectores monolíticos. El objetivo es obtener un pulso de señal eléctrica que corresponde al desplazamiento de un periodo entre las redes de difracción y que da lugar a un periodo de la señal incremental.

La organización de este capítulo es como sigue: en primer lugar presentamos una técnica adecuada para la generación de la distribución de irradiancia incidente en la máscara, en la siguiente sección estudiamos el diseño de la máscara de correlación, y finalmente, mostramos la comprobación experimental del método propuesto.

4.2. Generación del perfil de irradiancia en un codificador de efecto Lau

En esta sección analizamos el sistema de generación de la señal de referencia, el cual se muestra en la figura 4.1. Esta configuración corresponde a un codificador basado en efecto Lau. La fuente de iluminación es un LED con un difusor que emite un haz espacialmente incoherente. La luz pasa a través de los ZRCs y es registrada en el plano focal de una lente, donde se sitúa el sistema de detección. La fuente y la lente pueden estar compartidas con la etapa incremental. En nuestra configuración experimental hemos utilizado como sistema de detección un sensor CCD. La razón para esta elección es doble, por un lado podemos registrar el diagrama de irradiancia que llega al plano de detección, y así compararlo con las predicciones teóricas del mismo. Por otro lado, como mostramos más adelante, el desplazamiento del diagrama de irradiancia sobre el CCD puede ser digitalizado, y con esta información podemos simular cualquier geometría bidimensional de detección. La señal de referencia ensanchada por los efectos difractivos, podrá estrecharse de nuevo mediante el uso de una máscara de correlación adecuada en el plano focal de la lente.

La diferencia entre este sistema y el estudiado en el capítulo 2 reside en el tipo de iluminación. En el capítulo 2, para obtener la sombra geométrica de un ZRC sobre otro, se utiliza un haz colimado. Sin embargo, el sistema de la figura 4.1 utiliza un haz espacialmente incoherente.

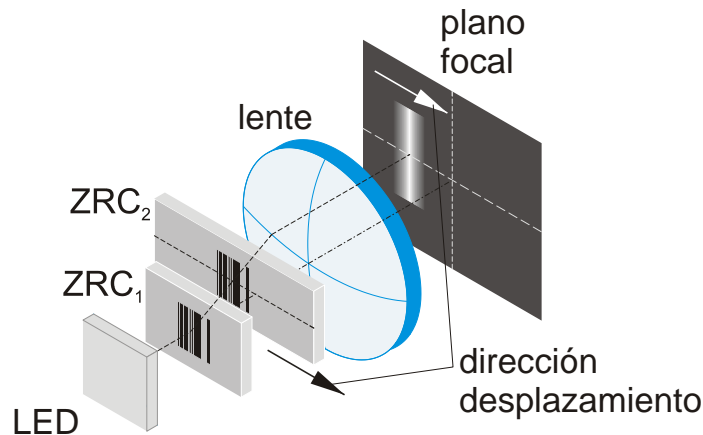


Figura 4.1: Configuración de Lau para la generación de señales de referencia. El plano focal de la lente es donde se sitúa el sistema de detección, que consta de una máscara antepuesta a un fotodiodo.

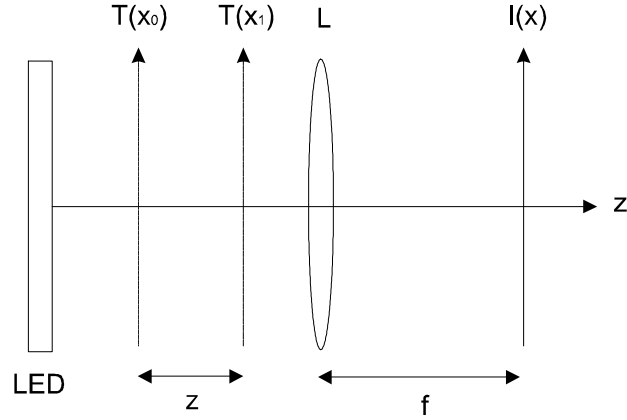


Figura 4.2: Configuración estudiada en el análisis teórico.

4.2.1. Cálculo del perfil de irradiancia

Para calcular el perfil de intensidad incidente en el plano de detección, mostraremos brevemente los cálculos utilizados por Jahns en [28]. La configuración estudiada y la nomenclatura utilizada se muestra en la figura 4.2. Consideramos dos ZRCs con transmitancia T y separados por una distancia z . Los ZRCs son iluminados por un haz monocromático de longitud de onda λ y espacialmente incoherente. El haz pasa primero a través de los ZRCs y después incide en la lente L , propagándose hasta el plano focal de ésta. El objetivo es calcular el perfil de intensidad incidente en el plano focal de la lente. Si la fuente de iluminación es espacialmente incoherente, podemos considerar que está situada sobre la superficie del primer ZRC. De esta manera, suponemos la superficie transparente del ZRC constituida por un conjunto infinito de fuentes puntuales y uniformemente distribuidas. La intensidad en el plano de detección es la suma de las intensidades generadas por cada una de estas fuentes puntuales, matemáticamente:

$$I(x) = \int |T(x_0)|^2 |u(x, x_0)|^2 dx_0, \quad (4.1)$$

donde $u(x, x_0)$ es la amplitud compleja del campo en el plano de detección, x , generada por una fuente puntual situada en la superficie del primer ZRC, x_0 . Para calcular $u(x, x_0)$, consideramos la propagación de una fuente puntual cuya amplitud es una delta de Dirac centrada en x_0 . Utilizando la aproximación de Fresnel [29], obtenemos la amplitud del campo en la superficie del segundo ZRC:

$$u(x_1, x_0) \propto \exp \left[i\pi (x_1 - x_0)^2 / \lambda z \right], \quad (4.2)$$

donde x_1 es la coordenada en la superficie del segundo ZRC. La amplitud en esta superficie se multiplica por la transmitancia del código y se propaga hasta

4.2. GENERACIÓN DEL PERFIL DE IRRADIANCIA EN UN CODIFICADOR DE EFECTO LAU45

el plano de detección. Si despreciamos la apertura de la lente, la amplitud en el plano de detección queda

$$u(x, x_0) \propto TF \left\{ T(x_1) \cdot \exp \left[i\pi (x_1 - x_0)^2 / \lambda z \right] \right\}, \quad (4.3)$$

donde TF es la transformada de Fourier. Reagrupando términos y sustituyendo la amplitud de la ecuación 4.3 en la ecuación 4.1, obtenemos

$$\begin{aligned} I(x) &= \int T(x_0) \cdot \left| \hat{T} \left(x_0 + \frac{z \cdot x}{f} \right) \right|^2 dx_0 = \\ &= \int T(x_0) \cdot \left| \hat{T}(x_0 + u) \right|^2 dx_0 = I \left(\frac{f}{z} u \right), \end{aligned} \quad (4.4)$$

donde hemos utilizado el carácter binario de la transmitancia del ZRC, $u = z \cdot x/f$ y f es la focal de la lente. \hat{T} es la transformada de Fresnel de T , que viene dada por:

$$\hat{T}(x + u) = \int T(x') \exp \left\{ \frac{i\pi}{\lambda z} [x' - (x + u)]^2 \right\} dx'. \quad (4.5)$$

La intensidad en el plano de detección calculada en la expresión 4.4, es proporcional al flujo, que a su vez es proporcional a la irradiancia¹. Según esta expresión, la irradiancia es la correlación entre la transmitancia de un ZRC y un segundo término. Este término depende de la transformada de Fresnel del ZRC a lo largo de la distancia de propagación entre ZRCs. La transformada de Fresnel del ZRC produce una degradación en la transmitancia de éste. Esta degradación transforma la autocorrelación que obtenemos en ausencia de difracción, en una correlación. Como veremos, la correlación es una versión de la autocorrelación ensanchada y degradada. Hay que notar que si la transmitancia de los ZRCs fuera periódica, la transformada de Fresnel de la transmitancia sería igual a la misma transmitancia en los planos de Talbot, recuperando la autocorrelación de los ZRCs.

La verificación experimental de la ecuación 4.4 se ha llevado a cabo con los siguientes parámetros: como fuente de iluminación se ha utilizado un LED verde que incorpora un difusor (LB2512G3, LEDtronics) y una longitud de onda dominante $\lambda = 540$ nm, dos ZRCs idénticos con rendijas de $20\mu m$ de anchura y una lente de 7,7 mm de focal frontal posterior. La distancia entre códigos es de 2 mm. En la figura 4.3 se muestra el ensanchamiento del pico central de la señal debido a la difracción del haz a través de los ZRCs. La gráfica discontinua es el perfil de irradiancia calculado mediante la aproximación de Fresnel (por lo tanto, considerando la difracción), la gráfica discontinua y punteada es la autocorrelación obtenida si se ignora la distancia entre códigos (sin tener en cuenta la difracción) y la continua es el perfil medido en el CCD. Como puede observarse, la ecuación 4.4 predice razonablemente bien el ensanchamiento del pico central y la posición de los picos laterales. Además, se pone de manifiesto la necesidad de considerar los efectos difractivos en la señal.

¹La irradiancia se define como el flujo recibido por unidad de área. El sensor CCD detecta la cantidad total de flujo, es decir, la irradiancia recibida por el área del pixel.

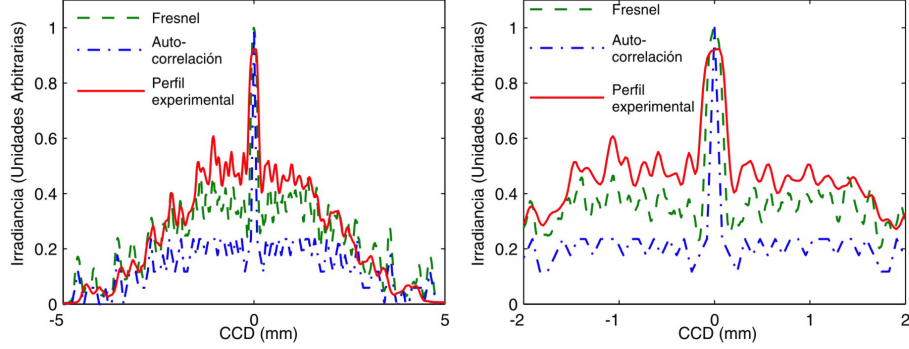


Figura 4.3: Perfiles de intensidad, la gráfica discontinua está calculada propagando el haz mediante la aproximación de Fresnel, la discontinua y punteada es la autocorrelación ignorando la distancia entre ZRCs y la gráfica continua es el perfil medido en el CCD. En la derecha se muestra una ampliación de la zona central.

4.2.2. Máscara de estrechamiento

Cuando uno de los ZRCs se desplaza respecto del otro una cantidad η , el perfil de irradiancia en el plano de detección es:

$$\begin{aligned}
 I'(x) &= \int T(x_0 + \eta) \left| \hat{T} \left(x_0 + \frac{z \cdot x}{f} \right) \right|^2 dx_0 = \\
 &= \int T(x_0) \left| \hat{T}(x_0 + u - \eta) \right|^2 dx_0 = I \left(\frac{f}{z} (u - \eta) \right). \quad (4.6)
 \end{aligned}$$

Según esta expresión, cuando un ZRC se desplaza respecto del otro, el perfil de irradiancia también se desplaza una cantidad proporcional al desplazamiento entre ZRCs. Cuando el perfil de irradiancia incidente se desplaza y pasa frente al fotodetector, la fotocorriente generada en éste es la correlación entre el perfil de irradiancia incidente y la geometría del detector. De modo que tenemos dos correlaciones, una correlación inicial para generar el perfil de irradiancia incidente en el plano de detección y una segunda correlación entre el perfil de irradiancia incidente y la geometría del detector. Esta segunda correlación será la que utilizaremos para estrechar el perfil de irradiancia degradado por la difracción. Si anteponeamos una máscara al fotodetector, la geometría del fotodetector es sustituida por la geometría de la máscara, que llamaremos D , y que puede ser variable respecto de la dirección del movimiento. En estas condiciones, la fotocorriente generada en el fotodetector es:

$$S(\eta) = K \int I(x - \eta) D(x) dx, \quad (4.7)$$

donde K es el factor de conversión entre el flujo de luz y la señal eléctrica. La geometría de la máscara antepuesta al detector equivale a una distribución espacial de detectividad, que puede ser variable a lo largo de la dirección de movimiento. La solución más sencilla consiste en un sistema de rendijas pegadas al fotodetector y cuyas alturas son proporcionales a la detectividad deseada en cada posición. Con este esquema, la transmitancia de la máscara es binaria, ya que las rendijas son transparentes u opacas. Podemos hacer que la función de detectividad tome valores positivos y negativos. Para ello, utilizamos dos máscaras antepuestas a dos fotodiodos y las señales de los fotodiodos se conectan a un restador (conexión push-pull). La resta de las dos fotocorrientes es equivalente a una detectividad que es resta de las detectividades de cada máscara, y que puede tomar valores positivos y negativos.

El problema consiste en el cálculo de la detectividad cuya correlación con el diagrama de irradiancia es una señal de referencia adecuada. Más aún, nos preguntamos en qué condiciones podemos encontrar una máscara cuya correlación es una señal arbitraria. Para responder a esta pregunta, es necesario resolver la ecuación integral dada en 4.7, su resolución numérica implica la resolución de un problema inverso. La resolución de problemas inversos está bien descrito en [30]. Para resolver el problema, es necesario conocer I y S , sin embargo, en nuestro caso la señal S es desconocida. Estamos interesados en señales S con propiedades específicas de las señales de referencia, pero no conocemos exactamente el valor de la señal. En el capítulo 1 se describen los parámetros que caracterizan a una señal de referencia adecuada. En el capítulo 2 se estableció que el parámetro más importante para una señal que proviene de una autocorrelación es $K = \sigma/S(0)$. Ahora la señal no es una autocorrelación y el pico central no es un triángulo. En este caso, es necesario tener en cuenta un parámetro extra que de cuenta de la pendiente del pico central de la señal, a este parámetro lo hemos denominado SL :

$$SL = \frac{S(0) - S(x')}{x'}, \quad (4.8)$$

donde x' es la semianchura del pico central en su base. Hay que notar que hemos denominado $S_0 \equiv S(0)$, ya que en este caso la señal es continua. En el capítulo 2 la señal de referencia utilizada estaba muestreada en cada múltiplo de la anchura de la rendija del ZRC, y por tanto, era una señal discreta. Como ya hemos comentado, la anchura de la señal de referencia debe ser igual al periodo de la señal incremental. Teniendo en cuenta todas estas consideraciones, una señal de referencia apropiada deberá tener un pico central con una anchura adecuada, un valor bajo de K y un valor alto de SL . Aparte de estos condicionantes, la forma de la señal en fuera del pico central es irrelevante.

4.3. Diseño de la máscara de detección

Para enfrentarnos con la resolución del problema inverso es necesario formular la ecuación integral en forma matricial. Llamamos D al vector cuyos elementos son la detectividad del fotodetector en cada punto, S es el vector

señal eléctrica que se registra en los detectores (puede ser fotocorriente o voltaje, ambos difieren en una constante), finalmente, I es el perfil de irradiancia que incide sobre la máscara:

$$\begin{aligned} I &= [I_1, I_2, \dots, I_n] \\ D &= [D_1, D_2, \dots, D_m] \\ S &= [S_1, S_2, \dots, S_{n+m-1}], \end{aligned} \quad (4.9)$$

la matriz correlación es:

$$C = \begin{bmatrix} I_n & 0 & 0 & 0 & 0_{(1,m)} \\ \vdots & I_n & 0 & 0 & 0 \\ I_1 & \vdots & \ddots & 0 & 0 \\ 0 & I_1 & \vdots & \ddots & 0 \\ \vdots & 0 & \vdots & \vdots & I_n \\ 0_{(n+m-1,1)} & \vdots & \vdots & \vdots & \vdots \end{bmatrix}. \quad (4.10)$$

La correlación entre I y D puede obtenerse por medio del producto matricial,

$$S = C \cdot D. \quad (4.11)$$

Este es un sistema sobredeterminado de $n + m - 1$ ecuaciones y m incógnitas. La solución de este tipo de sistemas viene dada por los siguientes teoremas [31]:

Teorema 1 *Existe una solución de $S = C \cdot D$ si, y solo si, existe una matriz denominada “inversa condicional” de C , denominada C' , que cumple $C \cdot C' \cdot S = S$. La inversa condicional de una matriz C es una matriz que cumple: $C \cdot C' \cdot C = C$.*

Teorema 2 *Si existe una solución de $S = C \cdot D$, la solución es única si, y solo si, el rango de la matriz C , de dimensiones $n + m - 1 \times m$, es igual a m .*

Uno de los métodos para obtener la inversa condicional C' es calcular la inversa generalizada. La inversa generalizada es una inversa condicional y, además, cumple otras propiedades [31]. La inversa generalizada puede calcularse siempre y cuando el rango de la matriz es máximo. Cuando el rango de la matriz es m , una forma de calcular C^- es por medio de la siguiente expresión:

$$C^- = (C^T \cdot C)^{-1} \cdot C^T. \quad (4.12)$$

La matriz de correlación C es triangular, por lo tanto siempre tiene determinante no nulo y es de rango m . Si fijamos el valor de la señal eléctrica S , para la cual el sistema de ecuaciones tiene solución, según el Teorema 2 la solución será única y vendrá dada por:

$$D = C^- \cdot S. \quad (4.13)$$

Si el sistema de ecuaciones tiene al menos una solución, decimos que el sistema es consistente. En general, para cualquier señal S el sistema no tiene por qué ser consistente, es decir, la máscara calculada mediante la ecuación $D = C^- \cdot S$, no reproducirá la misma señal al correlacionarla con el perfil de irradiancia. La correlación origina otra señal diferente $S' = C \cdot D$. Para un sistema cualquiera, se puede demostrar el siguiente teorema [31]:

Teorema 3 *Si calculamos la máscara mediante la ecuación $D = C^- \cdot S$, la señal S' obtenida correlacionando esta máscara con el perfil de irradiancia será la más próxima a la señal S en el sentido de mínimos cuadrados. En otras palabras, si el sistema no es consistente, siempre podemos encontrar la señal de correlación S' más próxima a nuestra señal objetivo S .*

Para que el sistema sea consistente, el vector S debe pertenecer al subespacio definido por las columnas de C . Este argumento se ha expresado de una forma más sistemática en el Teorema 1, o en otras palabras, la señal S debe ser un vector propio de $C \cdot C'$ con valor propio igual a uno. El objetivo del problema es encontrar señales que hagan al sistema consistente y que, adicionalmente, cumplan los requerimientos de una señal de referencia. En general, para plantear el problema, debemos considerar el error que cometemos en un sistema no consistente, con una señal arbitraria S , correlacionando el perfil de irradiancia incidente con una máscara también arbitraria:

$$S = C \cdot D + \varepsilon(D), \quad (4.14)$$

donde $\varepsilon(D)$ es la función error, que depende de la máscara usada en la correlación, según el teorema 3. La norma de este vector es mínima² cuando la máscara la calculamos mediante la ecuación $D = C^- \cdot S$ y C^- la calculamos mediante la expresión 4.12. En este caso, si sustituimos la ecuación 4.13 en la ecuación 4.14 y ponemos la máscara como función de S , se obtiene la siguiente expresión para la norma de la función error:

$$\|\varepsilon(S)\|^2 = [\varepsilon(S)]^T \cdot [\varepsilon(S)] = S^T [\mathbf{I} - C \cdot C^-] S, \quad (4.15)$$

donde S^T es la traspuesta de S e \mathbf{I} es la matriz identidad.

Si el sistema de ecuaciones lineales es consistente para una señal S , la ecuación 4.15 será cero. La máscara necesaria para obtener esa señal de referencia la obtenemos mediante la ecuación 4.13. Como la ecuación 4.15 solo depende de la señal y del perfil de irradiancia (que es un vector conocido), el problema se resuelve encontrando los ceros de la ecuación 4.15 como función de la señal S . Ya que el error es positivo, podemos encontrar los ceros buscando los mínimos de la función con restricciones sobre la variable S . Las restricciones consisten en una cota inferior y otra superior que nos aseguran unos parámetros K y SL adecuados. El perfil de irradiancia incidente y la cota superior e inferior de la señal S se muestran en la figura 4.4. En esta figura se muestran las cotas superior e inferior de la señal fuera del origen, el valor de estas cotas permite

²Una demostración completa del Teorema 3 puede encontrarse en [31].

controlar el valor del máximo secundario, que está relacionado con el parámetro K . En el centro de la señal definimos una región central de la señal, en esta región fijamos un valor más alto para la cota inferior y superior. De modo que la anchura del pico central de la señal es, como mínimo la anchura de la región central de la cota inferior, y como máximo la anchura de la región central de la cota superior. Además, la pendiente SL del máximo central estará controlada por la relación entre anchura y altura de las cotas en la región central.

Para demostrar la validez de la técnica, mostramos un ejemplo de estrechamiento. Con un montaje como el de la figura 4.1 obtenemos un perfil de irradiancia como el de la figura 4.3. Este perfil incidente tiene un pico central de $250\mu\text{m}$ en su base. Esta señal óptica se normaliza y se convierte en una señal eléctrica mediante una correlación con una máscara. La señal eléctrica tendrá una anchura de $100\mu\text{m}$ en la base del pico central. En la figura 4.4 mostramos el perfil incidente y las cotas superior e inferior exigidas (restricciones en la variable a optimizar). Las cotas superior e inferior constriñen la señal entre $\pm 0,3$ cuando $|x| > 0,05$ mm, en el centro constriñen la señal entre 1 y 1,5 cuando $|x| < 0,05$ mm. Con todos estos parámetros, el problema de optimización consiste en minimizar la función objetivo que nos proporciona el error cometido en la correlación, ecuación 4.15. Para esta minimización hemos utilizado el algoritmo “fmincon” de Matlab [32]. Con las restricciones más comunes el algoritmo converge en casi todos los casos, es decir, encuentra una señal apropiada con su máscara asociada que la genera. En la figura 4.5 mostramos las siguientes señales: la señal continua y punteada es la señal obtenida con el algoritmo, la señal continua es el perfil de irradiancia incidente y las señales discontinuas son las cotas impuestas a la señal. La figura 4.6 (a) muestra la máscara que reproduce la señal encontrada por correlación con el perfil de intensidad, la región superior es positiva mientras que la región inferior es negativa. En la figura 4.6 (b) se muestra la máscara superpuesta al diagrama de irradiancia con el que se va a correlacionar.

4.4. Resultados experimentales

En esta sección demostraremos la viabilidad de la técnica de estrechamiento de señales por correlación. Para ello realizamos la comprobación experimental de los resultados de la sección anterior. Simulamos la máscara de correlación sobre el sensor CCD y obtenemos la señal eléctrica experimentalmente, que será comparada con la señal teórica que predice el algoritmo. La configuración experimental es la de un sistema de doble difracción basado en efecto Lau, como el que se muestra en la figura 4.1. En la figura 4.7 se muestra el montaje de laboratorio utilizado. Los parámetros experimentales son:

- Las redes de difracción de un sistema de Lau se sustituyen por dos ZRCs, uno fijo y otro sobre una plataforma móvil en la dirección perpendicular a las rendijas. Los ZRCs están separados 3 mm en la dirección de propagación y tienen rendijas de $20\mu\text{m}$.

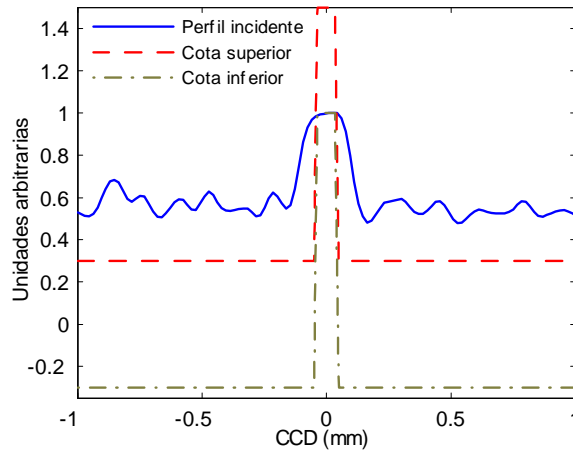


Figura 4.4: Perfil incidente y cotas superior e inferior.

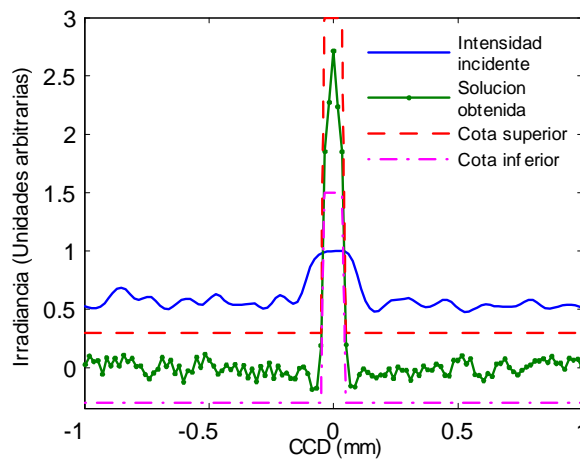


Figura 4.5: Diferentes señales, la señal continua es el perfil incidente medido en el CCD, la señal continua y punteada es la señal obtenida con el algoritmo y las señales discontinuas son las cotas inferiores y superiores impuestas a la señal.

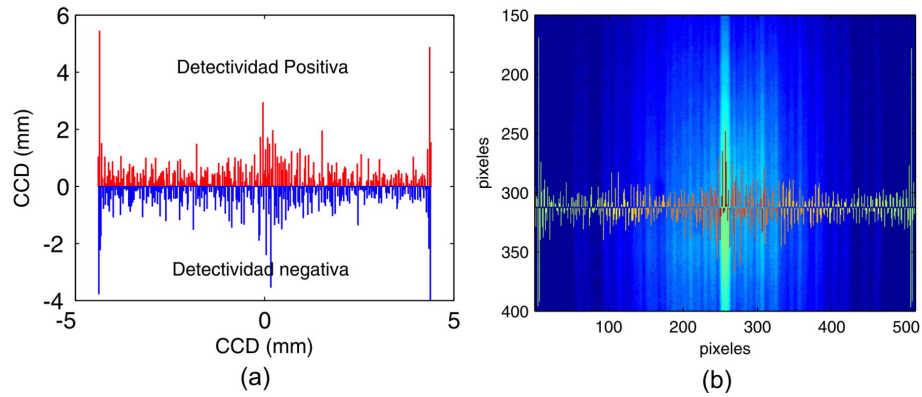


Figura 4.6: En la figura (a) se muestra la máscara que es solución del problema, la región superior es positiva mientras que la región inferior es negativa. La figura (b) muestra la máscara superpuesta al diagrama de irradiancia con el que se va a correlacionar.

- La fuente de iluminación es el LED utilizado en la sección anterior situado a 1 mm del primer ZRC y es alimentado con una corriente de 50 mA.
- Utilizamos una lente de 7,95 mm de focal frontal posterior, situada detrás del segundo ZRC y con el sensor CCD en su plano focal.

La máscara se simula en el CCD multiplicando los pixeles o bien por 1, o bien por 0 según el píxel pertenezca o no a la máscara, respectivamente. La señal registrada por la parte superior de la máscara se resta a la correspondiente de la parte inferior (para simular una conexión en push-pull de fotodetectores). Las medidas han sido realizadas desplazando uno de los ZRCs con un posicionador motorizado mientras se registran las imágenes del perfil incidente. Las imágenes se multiplican por la máscara binaria y el resultado se suma por filas y por columnas, de manera que obtenemos una correlación entre máscara y el perfil incidente.

En la figura 4.8 ilustramos el resultado del experimento: la señal continua es el perfil de irradiancia incidente, la señal discontinua es la señal obtenida teóricamente por el algoritmo y la señal continua y punteada es la señal registrada experimentalmente. El acuerdo entre teoría y experimento es bastante bueno, sobre todo en la región central de la señal. La anchura del pico central de la señal teórica y experimental es el mismo. Fuera de la región central, la diferencia entre la medida experimental y la señal predicha es mayor, sin embargo la señal experimental permanece entre las cotas exigidas. Hay algunos factores que influyen en la discrepancia entre señales fuera del pico central:

- La distorsión de la lente deforma el perfil de irradiancia incidente cuando uno de los ZRCs se desplaza. Esta distorsión del perfil es mayor en las zonas

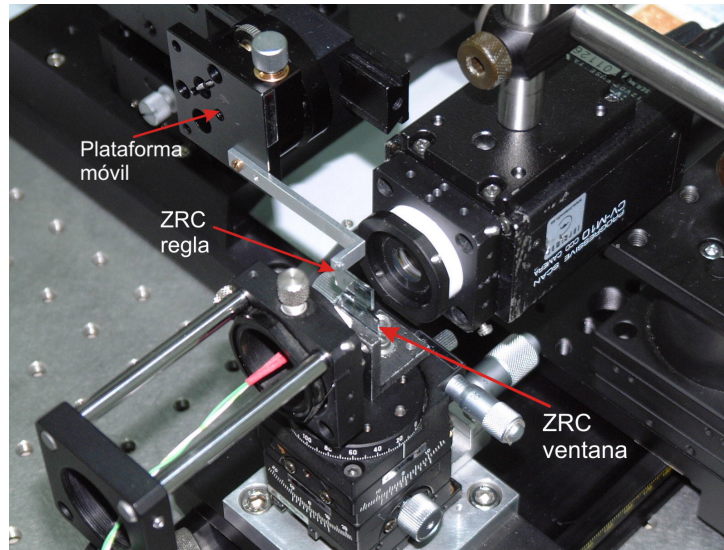


Figura 4.7: Sistema de generación y estrechamiento de señales de referencia en un codificador basado en efecto Lau.

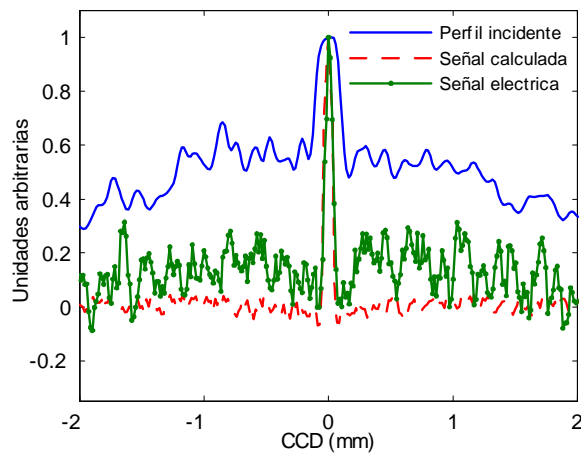


Figura 4.8: La señal continua es el perfil incidente medido sobre el CCD, la discontinua es la señal eléctrica calculada con el algoritmo y la señal continua y punteada es la señal registrada experimentalmente.

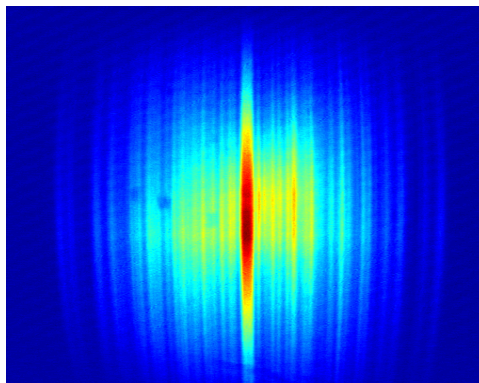


Figura 4.9: Imagen del diagrama de irradiancia registrado en el CCD cuando los ZRCs están enfrentados.

laterales del CCD. En la figura 4.9 se muestra la imagen de la irradiancia registrada en el plano de detección cuando los ZRCs están enfrentados. Puede observarse la distorsión de la lente en los laterales de la imagen.

- La distribución angular del flujo emitido por la combinación LED + difusor, hace que el perfil de irradiancia cuando los ZRCs están desplazados sea inferior al perfil obtenido sin desplazar. Este factor de atenuación angular del perfil de irradiancia afecta a los laterales de la señal de correlación. En la figura 4.10 mostramos los diferentes perfiles de irradiancia registrados en el CCD según se va desplazando uno de los ZRCs. En la figura de la derecha (a) se observa la distribución angular de la fuente de iluminación y en la derecha (b) se observa la deformación de los perfiles desplazados.

El efecto combinado de ambos factores provoca un cambio del perfil de irradiancia según se desplaza por el plano focal. Es en las zonas laterales de dicho plano donde el perfil resulta más deformado por los efectos de la distorsión y la falta de constancia de radiancia de la fuente (el emisor no es lambertiano). En la zona central, la forma del perfil coincide con la predicción teórica, y por tanto la reproducción del máximo central de la señal de referencia es correcta. Afortunadamente, este efecto carece de importancia en el diseño de la máscara de estrechamiento, ya que los detalles de la señal de referencia en su pedestal son, como ya hemos dicho anteriormente, irrelevantes.

4.5. CONCLUSIONES SOBRE EL ESTRECHAMIENTO DE SEÑALES DE REFERENCIA 55

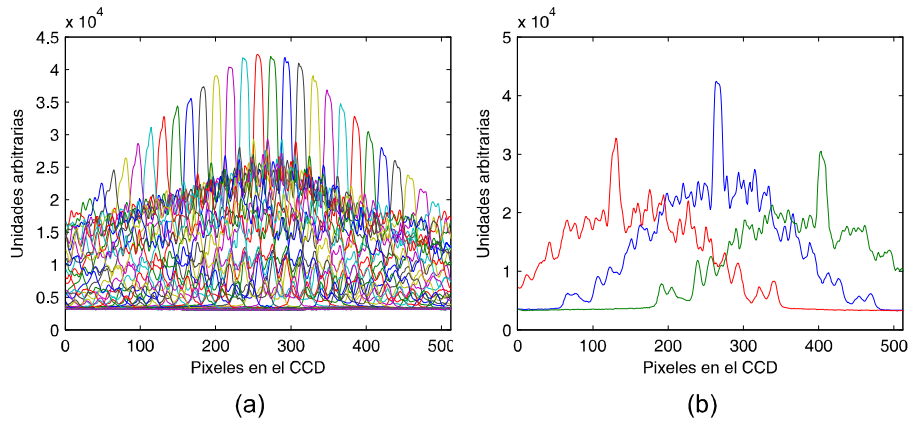


Figura 4.10: Perfiles de irradiancia registrados en el CCD según se desplaza uno de los ZRCs. En la gráfica de la derecha (a) se observa la distribución angular de la fuente y en la izquierda (b) se observa, además, la deformación del perfil.

4.5. Conclusiones sobre el estrechamiento de señales de referencia

En este trabajo se ha desarrollado un algoritmo de diseño de máscaras de detección que corrigen el ensanchamiento de las señales de referencia producido por la difracción en los ZRCs. El algoritmo necesita como parámetro el perfil de irradiancia incidente que es correlacionado con una máscara antepuesta a un sistema de fotodetectores. Se demuestra la posibilidad de diseñar señales de referencia con características preestablecidas, tales como, la relación señal-ruido, rango efectivo de señal o la pendiente del máximo central de la señal. La convergencia del algoritmo está relacionado con la complejidad de la máscara, cuanto más fino es el detalle de la máscara, más posibilidades de convergencia tenemos (mayor número de grados de libertad del problema). Un aspecto de vital importancia en señales de referencia para codificadores ópticos es la estabilidad respecto del desalineamiento entre ZRCs, fotodiodos, etc. Las pruebas hechas en el laboratorio son bastante satisfactorias, superando los requerimientos en cuanto a tolerancias de la tecnología actual.

Conclusiones

A continuación, se exponen las conclusiones generales del trabajo recogido en esta memoria.

1. Se han definido los parámetros que caracterizan las señales de referencia, tanto en codificación óptica como en alineamiento de máscaras. Con estos parámetros, se establecen los criterios de calidad de este tipo de señales.
2. Se ha propuesto una nueva técnica de diseño de códigos de referencia (ZRCs) basada en técnicas de optimización. Para ello, se presenta un algoritmo llamado DIRECT que nos permite diseñar ZRCs de mayor longitud, generando señales de referencia de mejor calidad. Con este algoritmo obtenemos resultados globalmente óptimos con ZRCs de 100 elementos, un orden de magnitud mayores que los métodos tradicionales. Además, el algoritmo obtiene diferentes soluciones equivalentes y nos posibilita la selección de soluciones atendiendo a otros criterios de interés, como puede ser la difracción y estabilidad.
3. Se han utilizado soluciones conocidas del problema para establecer el límite de validez de esta nueva técnica de optimización, comprobándose el carácter global de las soluciones obtenidas en el rango de validez del algoritmo.
4. Se ha propuesto un nuevo algoritmo genético con un operador de búsqueda restringida para diseñar ZRCs de longitud arbitraria. Se demuestra la eficiencia de este algoritmo para diseño de ZRCs de gran longitud, y por tanto con gran cantidad de variables binarias.
5. Se ha calculado una nueva cota inferior para el máximo secundario de las señales de referencia utilizadas en alineadores de máscaras. Esta cota inferior se ajusta mejor a las soluciones reales que la cota existente en la literatura.
6. Se ha propuesto la aplicación de DIRECT al diseño de ZRCs bidimensionales para alineamiento de máscaras, obteniendo soluciones óptimas hasta ZRCs de 10×10 píxeles y soluciones subóptimas hasta 16×16 píxeles.

También en este caso, se han obtenido varias soluciones equivalentes que pueden ser seleccionadas de acuerdo a otros criterios.

7. Se propone el uso de algoritmos genéticos al diseño de ZRCs bidimensionales (2D ZRC). Con este algoritmo se han obtenido 2D ZRCs de 16×16 píxeles que mejoran a los existentes en la literatura. Además, se ha diseñado un 2D ZRC de 100×100 píxeles, con el cual se ha demostrado la eficiencia de este algoritmo para gestionar 2D ZRCs de dimensiones arbitrarias.
8. Se ha calculado una señal de referencia en un codificador óptico basado en efecto Lau. Se ha puesto de manifiesto, con un modelo teórico bien contrastado experimentalmente, el ensanchamiento y la degradación de la señal producidos por la difracción de la señal de referencia en los ZRCs.
9. Se ha propuesto el uso de máscaras de detección de la señal de referencia que corrigen el ensanchamiento de ésta, y se ha propuesto un algoritmo de diseño de este tipo de máscaras. El algoritmo utiliza como parámetro el perfil de irradiancia incidente que es correlacionado con una máscara antepuesta a un sistema de fotodetectores.
10. Se demuestra la posibilidad de diseñar señales de referencia con características preestablecidas, tales como: la relación señal-ruido, rango efectivo de señal o la pendiente del máximo central de la señal.

Apéndice A

Lista de publicaciones

En este apéndice se muestran las publicaciones y comunicaciones que, hasta el momento, ha dado lugar el trabajo recogido en esta memoria.

A.1. Publicaciones

1. J. Saez, J. Alonso, and E. Bernabeu. Design of zero reference codes by means of a global optimization method. *Optics Express*, 13(1):195–201, 2005.
2. J. Sáez, J. Alonso, and E. Bernabeu. Design of two-dimensional zero reference codes by means of a global optimization method. *Optics Express*, 13(11):4230–4236, 2005.
3. J. Sáez, S. Salcedo, M. Rosa, J. Alonso, and E. Bernabeu. Optimal design of optical reference signals using a genetic algorithm. *Optics Letters*, 30(20):2724–2726, 2005.
4. J. Sáez, S. Salcedo, M. Rosa, J. Alonso, and E. Bernabeu. Design of two dimensional zero reference codes with a genetic algorithm. *Optics Letters*, en prensa, 2005.

A.2. Comunicaciones a congresos

1. J. Sáez, J. Alonso, E. Bernabeu, and Tomás Morlanes. Estrechamiento de señales ópticas de referencia por correlación. optoel'03.
2. J. Sáez, J. Alonso, and E. Bernabeu. Design of two-dimensional zero reference marks with a global optimization algorithm. euspen'05.

Índice de figuras

1.1. Esquema de codificador lineal absoluto. Emplea varias pistas con marcas periódicas.	3
1.2. Esquema de codificador basado en efecto moiré. En una segunda pista se pueden observar una marca de referencia (ZRC).	4
1.3. Regla típica con dos pistas, la pista superior es incremental (red de difracción) y la pista inferior contiene varios ZRC con distancia codificada.	5
1.4. Señales incrementales A y B y señal de referencia. Las señales (a) son registradas por los fotodetectores, mientras que en (b) se han convertido a señales cuadradas.	6
1.5. Diferentes tipos de fotolitografía, en la figura de la izquierda se muestra un esquema de fotolitografía de contacto, en el centro de proximidad y en la derecha de proyección.	8
1.6. Error cometido en la conversión electrónica de una señal que ha sufrido una modificación.	10
2.1. Procedimiento de generación de la señal de referencia.	12
2.2. Altura del segundo máximo de la señal de autocorrelación respecto del número de rendijas del ZRC. El ZRC tiene 50 elementos. La gráfica continua es muestras los valores obtenidos con DIRECT y la discontinua es la cota inferior calculada teóricamente.	19
2.3. Primer máximo, segundo máximo y su diferencia (señal efectiva) respecto del número de rendijas n_1 . El máximo de la señal efectiva se obtiene en $n_1 = 29$	19
2.4. Seis de las 136 señales de autocorrelación obtenidas por DIRECT. Todas ellas son equivalentes en cuanto al valor del segundo máximo. El número de elementos $n = 50$ y $n_1 = 25$. El valor del segundo máximo es $\sigma = 11$	20
2.5. Seis de los 136 ZRCs encontrados por DIRECT con $n = 50$ y $n_1 = 25$	21
2.6. (a) Ejemplo de la operación de cruce, la pareja inicial intercambia sus genes a partir de la posición 3. (b) Pseudo-código implementado en el GA.	23

2.7. Comparación entre el segundo máximo obtenido por DIRECT, GA y el valor de la cota inferior teórica. La longitud del código es $n = 200$ y n_1 varía desde 1 a 199	24
2.8. Primer máximo, segundo máximo y su diferencia (señal efectiva) respecto del número de rendijas n_1 . El máximo de la señal efectiva se obtiene en $n_1 = 67$	24
2.9. Señal de autocorrelación generada con un código óptimo. La longitud del código es 1000 y $n_1 = 100$. El valor del segundo máximo es 13.	25
3.1. Sistema de alineamiento bidimensional basado en ZRCs bidimensionales.	30
3.2. Altura del segundo máximo de la señal de autocorrelación con $n = 10$. La gráfica continua es alcanzada con DIRECT, la punteada es la cota inferior calculada en la ecuación 3.18 y la discontinua es la cota calculada por Chen mediante la expresión 3.5.	36
3.3. Primer máximo, segundo máximo y su diferencia (señal efectiva) respecto del número de rendijas n_1 . El máximo de la señal efectiva es de 33 unidades y se obtiene en $n_1 = 62$	36
3.4. Señal de referencia óptima para un caso donde $n = 10$ y $n_1 = 50$	37
3.5. Altura del segundo máximo de la señal de autocorrelación respecto al número de píxeles transparentes. Los ZRCs tienen 16×16 píxeles y el número de píxeles transparentes varía desde 1 a 255. La gráfica continua es la cota inferior teórica, la punteada se ha obtenido con GA y la discontinua se ha alcanzado con el algoritmo DIRECT.	38
3.6. Primer máximo, segundo máximo y su diferencia (señal efectiva) respecto del número de rendijas n_1 . El máximo de la señal efectiva es de 78 unidades y se obtiene en $n_1 = 178$	39
3.7. Señal de autocorrelación obtenida con un 2D ZRC de 100×100 píxeles y 300 píxeles transparentes (n_1). El segundo máximo conseguido es 14.	40
4.1. Configuración de Lau para la generación de señales de referencia. El plano focal de la lente es donde se sitúa el sistema de detección, que consta de una máscara antepuesta a un fotodiodo.	43
4.2. Configuración estudiada en el análisis teórico.	44
4.3. Perfiles de intensidad, la gráfica discontinua está calculada propagando el haz mediante la aproximación de Fresnel, la discontinua y punteada es la autocorrelación ignorando la distancia entre ZRCs y la gráfica continua es el perfil medido en el CCD. En la derecha se muestra una ampliación de la zona central.	46
4.4. Perfil incidente y cotas superior e inferior.	51

4.5. Diferentes señales, la señal continua es el perfil incidente medido en el CCD, la señal continua y punteada es la señal obtenida con el algoritmo y las señales discontinuas son las cotas inferiores y superiores impuestas a la señal.	51
4.6. En la figura (a) se muestra la máscara que es solución del problema, la región superior es positiva mientras que la región inferior es negativa. La figura (b) muestra la máscara superpuesta al diagrama de irradiancia con el que se va a correlacionar.	52
4.7. Sistema de generación y estrechamiento de señales de referencia en un codificador basado en efecto Lau.	53
4.8. La señal continua es el perfil incidente medido sobre el CCD, la discontinua es la señal eléctrica calculada con el algoritmo y la señal continua y punteada es la señal registrada experimentalmente.	53
4.9. Imagen del diagrama de irradiancia registrado en el CCD cuando los ZRCs están enfrentados.	54
4.10. Perfiles de irradiancia registrados en el CCD según se desplaza uno de los ZRCs. En la gráfica de la derecha (a) se observa la distribución angular de la fuente y en la izquierda (b) se observa, además, la deformación del perfil.	55

Bibliografía

- [1] Y. Xiangyang and Y. Chunyong. A new method for the design of zero reference marks for grating measurement systems. *Journal of Physics E Scientific Instruments*, 19(1):34–7, 1986.
- [2] Y. Li. Autocorrelation function of a bar code system. *Journal of Modern Optics*, 34(12):1571–5, 1987.
- [3] Y. Li. Optical valve using bar codes. *Optik*, 79(79):67–74, 1988.
- [4] Y. Li. Characterization and design of bar code systems for accurate alignment. *Applied Optics*, 27(12):2612–20, 1988.
- [5] Y. Li and F. T. S. Yu. Design of bar code systems for accurate alignment: a new method. *Applied Optics*, 29(5):723–5, 1990.
- [6] D. R. Jones, C. D. Perttunen, and B. E. Stuckman. Lipschitzian optimization without the lipschitz constant. *Journal of Optimization Theory and Application*, 79(1):157–181, 1993.
- [7] M. Bjorkman and K. Holmstrom. Global optimization using the direct algorithm in matlab. *Advanced Modeling and Optimization*, 1(2):17–37, 1999.
- [8] Y. Li. Bar codes with special correlations like those of the barker codes. *Optics Communications*, 83(1-2):15–20, 1991.
- [9] J. Saez, J. Alonso, and E. Bernabeu. Design of zero reference codes by means of a global optimization method. *Optics Express*, 13(1):195–201, 2005.
- [10] D. E. Goldberg. *Genetic algorithms in search, optimization and machine learning*. Addison Wesley, 1988.
- [11] S. Salcedo, G. Camps, F. Pérez, J. Sepúlveda, and C. Bousoño-Calzón. Enhancing genetic feature selection through restricted search and walsh análisis. *IEEE Trans. System, Man and Cybern. Part C*, 34(4):749, 2005.
- [12] M. C. King and D. H. Berry. Photolithografic mask alignment using moiré techniques. *Applied Optics*, 11(11):2455–2459, 1972.

- [13] Y. Torii and Y. Mizushima. Optical ultramicrometer technique utilizing a composite diffraction grating. *Optics Communications*, 23(1):135–8, 1977.
- [14] G. Makosch and F. J. Schoenes. Interferometric method for checking the mask alignment precision in the lithographic process. *Applied Optics*, 23(4):628–632, 1984.
- [15] G. Makosch and F. Prein. Phase-locked interferometry for automatic mask alignment in projection printers. *Applied Optics*, 26(14):2828–35, 1987.
- [16] V. T. Chitnis, Y. Uchida, K. Hane, and S. Hattori. Moiré signals in reflection. *Optics Communications*, 54(4):207–211, 1985.
- [17] T. Nomura, S. Kimura, Y. Uchida, and S. Hattori. Moiré alignment technique for the mix and match lithographic system. *The Journal of Vacuum Science and Technology B*, 6(1):394–8, 1988.
- [18] K. Hara, Y. Uchida, T. Nomura, S. Kimura, D. Sugimoto, A. Yoshida, H. Miyake, T. Iida, and S. Hattori. An alignment technique using diffracted moiré signals. *The Journal of Vacuum Science and Technology B*, 7(6):1977–9, 1989.
- [19] D. Sugimoto, S. Kimura, M. Eishima, T. Nomura, Y. Uchida, and S. Hattori. Projection moiré alignment technique for mix and match lithographic system. *Proc. SPIE, Electron-Beam, X-Ray, and Ion-Beam Technology: Submicrometer Lithographies IX, Douglas J. Resnick Ed.*, 1263:306–311, 1990.
- [20] J. Sáez, S. Salcedo, M. Rosa, J. Alonso, and E. Bernabeu. Optimal design of optical reference signals using a genetic algorithm. *Optics Letters*, 30(20):2724–2726, 2005.
- [21] W. Huang, X. Dang, Q. Hu, Y. Xia, M. Tello, M. Calleja, R. García, J. M. Gómez-Rodríguez, and A. Baro. Nanoalignment mask fabricated directly on si by afm. *Surface and interface analysis*, 32:130–132, 2001.
- [22] Y. Chen, W. Huang, and X. Dang. Design and analysis of two-dimensional zero-reference marks for alignment systems. *Review of Scientific Instruments*, 74(7):3549–53, 2003.
- [23] J. Sáez, J. Alonso, and E. Bernabeu. Design of two-dimensional zero reference codes by means of a global optimization method. *Optics Express*, 13(11):4230–4236, 2005.
- [24] J. Sáez, S. Salcedo, M. Rosa, J. Alonso, and E. Bernabeu. Design of two dimensional zero reference codes with a genetic algorithm. *Optics Letters*, en prensa, 2005.
- [25] D. R. Jones. Direct global optimization algorithm. In C.A. Floudas and P.M. Pardalos, editors, *Encyclopedia of Optimization*, pages 431–440. Kluwer Academic Publishers, Dordrecht, 2001.

- [26] D. Crespo. *Nuevas herramientas aplicadas a la codificación óptica*. PhD thesis, Facultad de Ciencias Físicas, Universidad Complutense de Madrid, 2001.
- [27] D. Crespo, J. Alonso, T. Morlanes, and E. Bernabeu. Optical encoder based on the lau effect. *Optical Engineering*, 39(3):817–24, 2000.
- [28] J. Jahns and A. W. Lohmann. The lau effect (a diffraction experiment with incoherent illumination). *Optics Communications*, 28(3):263–267, 1979.
- [29] J. W. Goodman. *Introduction to Fourier Optics*. McGraw-Hill, 1996.
- [30] W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling. *Numerical Recipes in C : The Art of Scientific Computing*. Cambridge University Press, 1992.
- [31] W. K. Pratt. *Digital Image Processing*. John Wiley and Sons, 1991.
- [32] Inc. The MathWorks. Optimization toolbox for use with matlab. Technical report, Natick, Massachusetts, 2003.