



(12) **Patentschrift**

(21) Deutsches Aktenzeichen: **11 2016 000 419.2**
 (86) PCT-Aktenzeichen: **PCT/US2016/014167**
 (87) PCT-Veröffentlichungs-Nr.: **WO 2016/118651**
 (86) PCT-Anmeldetag: **20.01.2016**
 (87) PCT-Veröffentlichungstag: **28.07.2016**
 (43) Veröffentlichungstag der PCT Anmeldung
 in deutscher Übersetzung: **23.11.2017**
 (45) Veröffentlichungstag
 der Patenterteilung: **05.10.2023**

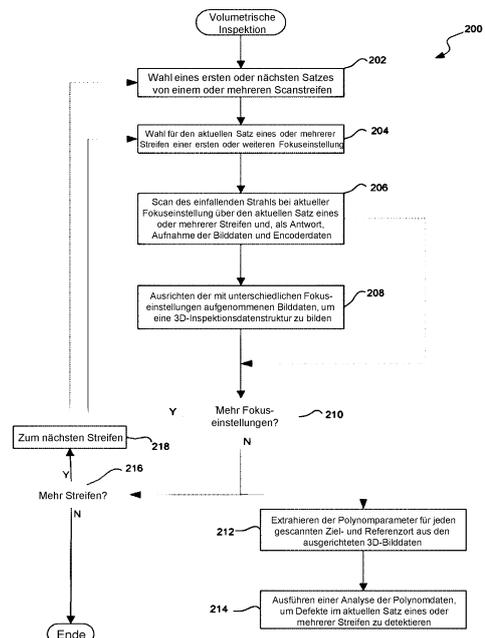
(51) Int Cl.: **H01L 21/66 (2006.01)**

Innerhalb von neun Monaten nach Veröffentlichung der Patenterteilung kann nach § 59 Patentgesetz gegen das Patent Einspruch erhoben werden. Der Einspruch ist schriftlich zu erklären und zu begründen. Innerhalb der Einspruchsfrist ist eine Einspruchsgebühr in Höhe von 200 Euro zu entrichten (§ 6 Patentkostengesetz in Verbindung mit der Anlage zu § 2 Abs. 1 Patentkostengesetz).

<p>(30) Unionspriorität:</p> <table border="0"> <tr> <td>62/105,979</td> <td>21.01.2015</td> <td>US</td> </tr> <tr> <td>15/001,158</td> <td>19.01.2016</td> <td>US</td> </tr> </table> <p>(73) Patentinhaber: KLA-Tencor Corporation, Milpitas, Calif., US</p> <p>(74) Vertreter: Reichert & Lindner Partnerschaft Patentanwälte, 93049 Regensburg, DE</p>	62/105,979	21.01.2015	US	15/001,158	19.01.2016	US	<p>(72) Erfinder: Chen, Grace H., Los Gatos, Calif., US; Wells, Keith, Santa Cruz, Calif., US; Huber, Markus, Oakland, Calif., US; Oh, Sebaek, Millbrae, Calif., US</p> <p>(56) Ermittelte Stand der Technik:</p> <table border="0"> <tr> <td>DE</td> <td>10 2010 025 033</td> <td>A1</td> </tr> <tr> <td>US</td> <td>2014 / 0 212 023</td> <td>A1</td> </tr> </table>	DE	10 2010 025 033	A1	US	2014 / 0 212 023	A1
62/105,979	21.01.2015	US											
15/001,158	19.01.2016	US											
DE	10 2010 025 033	A1											
US	2014 / 0 212 023	A1											

(54) Bezeichnung: **Waferinspektion mit einem Verfahren mit volumetrischen Fokus**

(57) Hauptanspruch: Ein Verfahren zur Detektion von Defekten auf einer Halbleiterprobe (106), wobei das Verfahren umfasst:
 Benutzen eines Inspektionswerkzeugs (102), um Intensitätsdatensätze (152a - 152e) bei einer Vielzahl von Fokuseinstellungen von jeder einer Vielzahl von xy-Positionen der Probe (106) zu sammeln;
 Extrahieren einer Polynomgleichung als Funktion der Fokuseinstellung, die eine Vielzahl von Koeffizienten für jeden der gesammelten Intensitätsdatensätze (152a - 152e) einer xy-Position aufweist;
 Darstellen eines jeden der Sätze von Werten der Koeffizienten für die Vielzahl von xy-Positionen mit einer entsprechenden Koeffizienten-Bildebene, um eine Vielzahl von Koeffizienten-Bildebenen zu bilden; und
 Analysieren einer Vielzahl von Differenz-Koeffizienten-Bildebenen, die auf einem Vergleich eines Zielsatzes der Koeffizienten-Bildebenen und eines Referenzsatzes der Koeffizienten-Bildebenen basieren, um Defekte auf der Probe (106) zu detektieren.



Beschreibung

QUERVERWEIS AUF VERWANDTE ANMELDUNG

[0001] Diese Anmeldung beansprucht die Priorität der provisorischen US-Patentanmeldung Nr. 62/105,979, mit dem Titel: „Wafer Inspection With Focus Volumetric Method“, eingereicht am 21. Januar 2015 von Grace Chen.

TECHNISCHES GEBIET DER ERFUNDUNG

[0002] Die Erfindung bezieht sich allgemein auf das Gebiet der Inspektionssysteme für Wafer und Retikel. Insbesondere betrifft die vorliegende Erfindung die Inspektionstechniken mit einer erhöhten Empfindlichkeit für die Fehlerdetektion.

HINTERGRUND

[0003] Im Allgemeinen beinhaltet die Industrie für die Halbleiterfertigung hochkomplexe Techniken zur Herstellung der integrierten Schaltungen unter Verwendung von Halbleitermaterialien, wie beispielsweise Silizium, die auf einem Substrat geschichtet und gemustert sind. Eine integrierte Schaltung wird in der Regel aus einer Vielzahl von Masken hergestellt. Die Herstellung der Masken und die anschließende optische Inspektion solcher Masken sind Standardschritte bei der Herstellung von Halbleitern geworden. Die Herstellung von Halbleiterbauelementen, wie Logik- und Speicherbauteile, beinhaltet typischerweise die Verarbeitung eines Halbleiterwafers mit einer großen Anzahl von Halbleiterherstellungsprozessen mit mehreren Masken, um verschiedene Funktionen und mehrere Ebenen der Halbleiterbauelemente auszubilden. Mehrere Halbleiterbauelemente können in einer Anordnung auf einem einzigen Halbleiterwafer hergestellt und dann in einzelne Halbleiterbauelemente getrennt werden.

[0004] Wenn Defekte auf der Maske oder Wafer vorhanden sind, funktionieren die resultierenden Halbleiterbauelemente möglicherweise nicht ordnungsgemäß. In einer Chip-zu-Chip-Inspektion werden Bilder eines Test-Chips mit Bildern eines Referenz-Chips verglichen und die Unterschiede analysiert, um Defekte zu erkennen. Herkömmlicherweise wird ein optimaler Fokus für ein einzelnes Ziel des Wafers gefunden und dann wird dieser optimale Fokus für die Gesamtheit der Wafer verwendet. Jedoch kann der optimale Fokus des Test- und Referenzchips für verschiedene Waferpositionen aufgrund einer Reihe von Faktoren variieren, so dass unterschiedliche Positionen auf dem Wafer nicht mit einer optimalen Fokuseinstellung überprüft werden. Eine Chip-zu-Chip-Inspektion kann dabei Bilder über den Wafer sammeln, die nicht optimal fokussiert sind, so dass ein Rauschen in die Inspektionsergebnisse eingeführt wird und Defekte schwieriger zu erkennen sind.

[0005] US 2014 / 0 212 023 A1 offenbart ein Musterinspektionsverfahren und eine entsprechende Vorrichtung. Licht wird in unterschiedlichen optischen Zuständen auf ein Substrat mit einem Muster gerichtet. Für jeden optischen Zustand wird ein Bild erzeugt und eine Differenz zwischen diesem Bild und einem Referenzbild als Grauwert ausgegeben. Die Position eines Defekts in einer Stapelrichtung einer mehrlagigen Struktur wird aus den Grauwerten zu verschiedenen optischen Zuständen erschlossen.

[0006] DE 10 2010 025 033 A1 offenbart ein Verfahren und eine Vorrichtung zur Defekterkennung und Reparatur von EUV-Masken. Dabei wird an der fehlerhaften Stelle der Maske mit einem EUV-Maskeninspektionsmikroskop ein Fokusstapel erzeugt.

ZUSAMMENFASSUNG

[0007] In einer Ausführungsform wird ein Verfahren zum Detektieren von Defekten in einer Halbleiterprobe offenbart. Ein Inspektionswerkzeug wird verwendet, um Intensitätsdatensätze mit einer Vielzahl von Fokuseinstellungen von jeder einer Vielzahl von xy-Positionen der Probe zu sammeln. Eine Polynomgleichung, die eine Vielzahl von Koeffizienten aufweist, wird für jede xy-Position der gesammelten Intensitätsdatensätze als eine Funktion der Fokuseinstellung extrahiert. Jeder Satz der Werte der Koeffizienten für die Vielzahl von xy-Positionen wird von einem entsprechenden Koeffizienten der Bildebene repräsentiert. Ein Zielsatz von Koeffizienten-Bildebenen und ein Referenzsatz von Koeffizienten-Bildebenen werden dann analysiert, um Defekte auf der Probe zu detektieren.

[0008] In einer speziellen Implementierung werden Intensitätsdaten mit mehreren Fokuseinstellungen mit einem ersten Satz von einem oder mehreren Streifen vor einem nächsten Satz von einem oder mehreren Streifen gesammelt, und die Anzahl der Streifen in dem ersten und die nächsten Streifen, von denen aus die Intensitätsdaten gesammelt werden, wird derart ausgewählt, das eine tatsächliche Fokusposition kleiner ist als eine Zeit der thermischen Ausdehnung zur Beeinflussung einer aktuellen Fokusposition. In einem weiteren Aspekt werden xy- und z-Positionsdaten des Encoders in Echtzeit vom Inspektionswerkzeug gesammelt. Vor dem Extrahieren der Polynomgleichung für jeden der an xy-Positionen gesammelten Intensitätsdatensätze, können die Intensitätsdatensätze von jeder Fokuseinstellung auf Basis der gesammelten Echtzeit-xy-Positionsdaten ausgerichtet werden. In einem weiteren Aspekt werden Intensitätsdaten von jedem Paar von Streifen von xy-Positionen bei den mehreren Fokuseinstellungen gesammelt, und zwar bevor der Ausrichtungsvorgang für diese Intensitätsdaten durchgeführt wird.

[0009] In einer weiteren Ausführungsform umfasst das Verfahren (i) ein Initiieren der Aufzeichnung der xy-Positionsdaten des Encoders in einen Puffer des Encoders, bevor die Intensitätsdaten von einem ersten der Streifen an einer ersten der Fokuseinstellungen gesammelt werden, (ii) während der Tisch des Inspektionswerkzeugs umkehrt, um sich für die Sammlung der Intensitätsdaten von einem zweiten der Streifen an der ersten Fokuseinstellung zu konfigurieren, werden die xy-Positionsdaten aus dem Puffer des Encoders in einen Systemspeicher kopiert, um einen Zugriff für die Verwendung im Ausrichtvorgang zu haben und dann wird die Aufzeichnung der xy-Positionsdaten des Encoders in einem Puffer des Encoders initiiert, bevor die Intensitätsdaten eines zweiten Streifens gesammelt werden, (iii) während der Tisch des Inspektionswerkzeugs umkehrt, um sich für die Sammlung der Intensitätsdaten von dem ersten Streifen an der zweiten Fokuseinstellung nach dem Sammeln der Intensitätsdaten des zweiten Streifens zu konfigurieren, werden die xy-Positionsdaten aus dem Puffer des Encoders in einem Systemspeicher kopiert, um einen Zugriff für die Verwendung im Ausrichtvorgang zu haben und dann wird die Aufzeichnung der xy-Positionsdaten des Encoders in einem Puffer des Encoders initiiert, bevor die Intensitätsdaten des ersten Streifens an der zweiten Fokuseinstellung gesammelt werden, und (iv) wiederholen der Operationen des Initiierens der Aufzeichnung und des Kopierens für aufeinanderfolgende Paare von Streifen an jeder der Fokuseinstellungen.

[0010] In einer anderen Implementierung wird die Analyseoperation durch (1) Berechnen einer Vielzahl von Differenzkoeffizienten der Bildebenen, die für jeden Koeffizienten eine Vielzahl von Werten der Differenzkoeffizienten haben, mittels einer Subtraktion eines jeden Targetsatzes (Zielsatzes) von jedem der Referenzsätze durchgeführt, und (2) Analysieren der Bildebenen der Differenzkoeffizienten, um Defekte zu erkennen. In einem weiteren Aspekt werden die Bildebenen der Differenzkoeffizienten dadurch analysiert, dass ein Bildpunkt für die Werte der Differenzkoeffizienten an jeder xy-Position der Bildebenen aus den Differenzkoeffizienten in ein Streudiagramm gedruckt werden, das eine Achse für jeden Koeffizienten besitzt, und dass diese Bildpunkte des Streudiagramms in Clustern von Defekten der Bildpunkte von Interesse, Bildpunkten von Störungen oder Bildpunkten des Hintergrunds angehäuft werden. In noch einem weiteren Aspekt werden die Bildebenen der Differenzkoeffizienten dadurch analysiert, dass ein Bildpunkt für die Werte der Differenzkoeffizienten für jede xy-Position der Bildebenen aus den Differenzkoeffizienten auf eine Einheitskugel projiziert wird, und dass die projizierten Bildpunkte in Clustern von Defekten der Bildpunkte von Interesse, der Bildpunkte von Störungen oder der Bildpunkte des Hintergrunds angehäuft werden.

[0011] In einem weiteren Beispiel umfasst das Verfahren (i) das Erzeugen einer Vielzahl von Differenzbildern aus Intensitätsdatensätzen, die von einem Ziel und einer Referenz an jeder Fokusebene gesammelt wurden, (ii) das Kombinieren der Differenzbilder, um ein fusioniertes Bild über den Fokus zu bilden, und (iii) das Analysieren des fusionierten Bildes für die Fehlererkennung. In einem weiteren Aspekt umfasst das Verfahren (i) das Gruppieren der Koeffizienten mit ähnlichen Werten, um eine Vielzahl von verschiedenen Segmenten zu bilden, wobei jedes Segment einem unterschiedlichen Abschnitt einer tatsächlichen Struktur der Vorrichtung entspricht, und (ii) Analysieren der verschiedenen Segmente mit unterschiedlicher Stringenz, um Defekte zu detektieren, die dem tatsächlichen Typ der verschiedenen Segmente der Vorrichtung entsprechen.

[0012] In einer anderen Implementierung umfasst das Verfahren das Analysieren eines Unterschieds zwischen Intensitätsänderungen als eine Funktion der Änderungen der Fokuseinstellung für jede xy-Position in einem zweiten Ziel-Satz von Intensitätsdatensätzen und einem zweiten Referenz-Satz von Intensitätsdatensätzen. In einem anderen Aspekt werden die Fokuseinstellungen von Paaren von Fokuseinstellungen umfasst, wobei jedes Paar der Fokuseinstellung durch einen Schrittwert getrennt ist, der innerhalb eines Bruchteils der Tiefenschärfe des Fokus des Inspektionssystems ist.

[0013] In einer alternativen Ausführungsform betrifft die Erfindung ein Inspektionssystem, das ein Beleuchtungsoptik-Modul zur Erzeugung und zum Richten eines einfallenden Strahls in Richtung auf eine Halbleiterprobe bei einer Vielzahl von Fokuseinstellungen und ein Sammeloptik-Modul zum Sammeln der Intensitätsdatensätze bei einer Vielzahl von Fokuseinstellungen von jedem einer Vielzahl von xy-Positionen der Probe in Reaktion auf den einfallenden Strahl umfasst. Das System umfasst ferner eine Steuereinheit, die zur Ausführung von einer oder mehrerer der oben aufgeführten Verfahren-Operationen konfiguriert ist.

[0014] Diese und andere Aspekte der Erfindung werden nachfolgend unter Bezugnahme auf die Figuren weiter beschrieben.

KURZE BESCHREIBUNG DER ZEICHNUNGEN

Fig. 1A ist eine schematische Darstellung eines Ansatzes einer volumetrischen Inspektion in Übereinstimmung mit einer Ausführungsform der vorliegenden Erfindung.

Fig. 1 B zeigt eine schematische Darstellung eines Verfahrens zum Extrahieren von einzigartigen elektromagnetischen (EM) 3D volumetrischen (VM) Parametern in Übereinstimmung

mit bestimmten Ausführungsformen der vorliegenden Erfindung.

Fig. 2 zeigt ein Flussdiagramm, das einen volumetrischen Prüfprozess in Übereinstimmung mit einer Ausführungsform der vorliegenden Erfindung veranschaulicht.

Fig. 3A ist eine schematische Darstellung eines Scanvorgang eines Wafers, um die Konstruktion einer 3D-Intensitätsdatenstruktur in Übereinstimmung mit einer spezifischen Implementierung der vorliegenden Erfindung zu ermöglichen.

Fig. 3B zeigt eine Vielzahl von Scanstreifen (Streifen_1, Streifen_2, ... Streifen_n) für den Wafer aus **Fig. 3A**.

Fig. 3C ist ein Ablaufdiagramm, das ein Ausrichtungsverfahren in Übereinstimmung mit einer Ausführungsform der vorliegenden Erfindung veranschaulicht.

Fig. 3D ist eine schematische Darstellung einer 3D-Bildkonstruktion und Extraktion volumetrischer (VM) Parameter in Übereinstimmung mit einer Ausführungsform der vorliegenden Erfindung.

Fig. 4 ist ein 3D-Streudiagramm, das zum Beispiel aus der Differenz von VM-Bildern gemäß einer Ausführungsform der vorliegenden Erfindung konstruiert wurde.

Fig. 5A veranschaulicht eine 3D-Projektion von VM-Differenzbildpixeln auf eine Einheitskugel in Übereinstimmung mit einer Ausführungsform der vorliegenden Erfindung.

Fig. 5B zeigt ein Beispiel einer projizierten Einheitskugel mit der Trennung von DOI-Pixeln und Hintergrund / Störungs-Pixeln gemäß einer Ausführungsform der vorliegenden Erfindung.

Fig. 6 zeigt ein Beispiel eines Segmentierungsbereiches, basierend auf extrahierten VM-Parametern gemäß einer anderen Ausführungsform der vorliegenden Erfindung.

Fig. 7 ist eine schematische Darstellung eines beispielhaften Inspektionssystems, bei dem Techniken der vorliegenden Erfindung implementiert werden können

Fig. 8 zeigt eine schematische Darstellung eines Inspektionswerkzeugs in Übereinstimmung mit bestimmten Ausführungsformen.

DETAILLIERTE BESCHREIBUNG BEISPIELHAFTER AUSFÜHRUNGSFORMEN

[0015] In der folgenden Beschreibung werden zahlreiche spezifische Details dargelegt, um ein gründliches Verständnis der vorliegenden Erfindung zu

schaffen. Die vorliegende Erfindung kann ohne einige oder all dieser spezifischen Details praktiziert werden. In anderen Fällen sind ausreichend bekannte Komponenten oder Prozessoperationen nicht im Detail beschrieben, um die vorliegende Erfindung nicht unnötig zu verschleiern. Während die Erfindung in Verbindung mit den speziellen Ausführungsformen beschrieben wird, versteht es sich, dass die Erfindung nicht auf die Ausführungsformen beschränkt ist.

[0016] Die hierin beschriebenen Inspektionstechniken können in Bezug auf jede geeignete Probe implementiert sein, die identisch ausgelegte Bereiche besitzt. Ein Beispiel ist eine Maske für die Halbleiterherstellung, die mehrere Chip- oder Zellenanordnungen aufweist, die ausgelegt sind, identisch zu sein. Die Proben können auch die Form einer Halbleitermaske, eines Solarpanels, einer Computer-Diskette, usw. annehmen.

[0017] Ein Wafer mit integrierten Schaltkreisen umfasst typischerweise mehrere identische Chips mit Elementen. Bestimmte herkömmliche Inspektionsarten für Halbleiter basieren auf dem Vergleich von Merkmalen auf der Probe, die identisch konstruiert sind. Daraus resultierende Abweichungen der Vergleichsergebnisse können als Fehlerkandidaten identifiziert werden. Zum Beispiel umfasst die Chip-zu-Chip-Inspektion den Vergleich von zwei Chips, die identisch ausgebildet sind. Eine Zelle-zu-Zelle-Inspektion umfasst zwei Zellen, die identisch ausgelegt sind. Jede festgestellte Differenz zwischen zwei Bildern hat das Potenzial, im Vergleich zu einem „Störungs“-Defekt oder Rauschen, ein tatsächlicher Fehler zu sein. Das heißt, dass einige von den detektierten Fehlerkandidaten keine Auswirkung auf den entsprechenden Bereich der integrierten Schaltung haben und können somit als „falsche“ Defekte angesehen werden.

[0018] Ein typisches Ergebnis der Chip-zu-Chip-Inspektion kann oft durch den Grad der relativen Fokus- oder Defokus-Unterschiede zwischen den gleichen Chips oder den gleichen Zellbereichen beeinflusst werden, die untersucht und verglichen werden. Zum Beispiel können zwei korrespondierende, aber ungleich fokussierte Chipbereiche zweier Chips in der Erkennung falscher Fehler oder in dem Versagen der Erkennung realer Defekte in diesen ungleich fokussierten Bereichen resultieren. Obwohl es viele Techniken zur Linderung dieses Fokusproblems in einer Chip-zu-Chip-Inspektion gibt, ist es schwierig, eine optimale Fokussierung über den gesamten Wafer zu halten.

[0019] Bestimmte Ausführungsformen der vorliegenden Erfindung umfassen eine Chip-zu-Chip-Inspektion, bei der Bilder aufgrund eines Satzes von Fokus-Offset-Werten über den gesamten Wafer

erhalten werden. Die Verwendung als dreidimensionale (3D) Fokusdaten verbessert die Differenzierung zwischen Defekt oder Störung bei der Inspektion von gemusterten Wafern. **Fig. 1A** ist eine schematische Darstellung eines volumetrischen Inspektionsprozesses in Übereinstimmung mit einer Ausführungsform der vorliegenden Erfindung. Bei bestimmten Ausführungsformen wird ein Inspektionssystem 102 verwendet, um Bilder einer Probe 106, wie beispielsweise von einem Halbleiterwafer, mittels einer Vielzahl von verwendeten Fokuseinstellungen (z) zu erhalten. Eine 3D-Fokus-Datenstruktur oder eine „volumetrische“ Inspektionsdatenstruktur für jedes Ziel (wie z.B. 104a) und Referenz (104b) wird mittels einer Stapelbildung der erfassten Bilder bei verschiedenen Fokuseinstellungen (z) gebildet, die dann für verschiedene Inspektionsanwendungen verwendet werden kann.

[0020] In bestimmten Anwendungen, wie weiter unten beschrieben, werden die Informationen aus den 3D-Bildern kombiniert, um (1) eine räumliche Segmentierung zu ermöglichen, die eine feinere Auflösung erreicht, als dies durch eine optische Standardauflösung herkömmlicher Inspektionssysteme erreicht werden würde, um so Inspektionsanwendungen im Mikrobereich zu unterstützen, (2) ermöglichen der Extraktion von elektromagnetischen (EM) volumetrischen Parametern (VM-Parameter), um die Separierung zwischen Defekt und Hintergrund oder Störung zu erhöhen; und (3) ermöglichen der Fusion von Bildern, die aus solchen VM-Parametern aufgebaut sind, um ebenso die Sensitivität zwischen Defekt und Hintergrund oder Störung zu erhöhen. Diese Anwendungen sowie andere der unten beschriebenen Anwendungen können die Gesamttempfindlichkeit, sowie andere Aspekte der Wafer-Inspektion verbessern.

[0021] Theoretisch kann die Erfassung von Bildern eines Wafers bei verschiedenen Fokushöhen als Hinzufügung einer bekannten Phase zu dem EM-Feld, das von der Waferoberfläche zurückkommt, bezeichnet werden. Es ist auf dem Gebiet der Mikroskopie bekannt, dass die Phaseninterferenz für die Erhöhung des Defekt / Hintergrund-Kontrasts wirksam sein kann. Das Durchführen der Fehlererkennung durch Bewegen der Probe an verschiedene Fokusebenen, führt letztlich zu einer Phaseninterferenz, die die Extraktion sowohl von Phaseninformation als auch vom Hintergrund und Defekten ermöglicht. Die Nachweisempfindlichkeit wird sowohl von Phasen- als auch von Intensitätsdaten verbessert. Eine Standard-Waferinspektion beruht auf Intensitätsunterschieden zwischen einem Defekt-von-Interesse (DOI) und dem Hintergrund, da die Intensität direkt nachweisbar ist.

[0022] In einem speziellen Beispiel können Fourier-Optiken und Fokusphasenstörungseffekte kombi-

nirt werden, so dass die eindeutige Darstellung der Fokusphaseninterferenz mit dem EM-Wafer-Rückfluss prägnant ausgedrückt werden kann, wie z.B. in einer Polynomgleichung, wie weiter unten beschrieben. **Fig. 1B** zeigt eine schematische Darstellung eines Prozess zur Extraktion einzigartiger EM-3D-volumetrischer Parameter aus 3D-Bilddaten, die von einer Probe in Übereinstimmung mit bestimmten Ausführungsformen der vorliegenden Erfindung erhalten werden. Wie gezeigt, werden die Intensitätsdatensätze (152a - 152e) (wie z.B. ein Bild) von einer Probe an einer Vielzahl von x, y, z-Positionen gesammelt. Anders ausgedrückt, kann jede Probe in eine Vielzahl von x, y-Positionen aufgeteilt werden und jede x, y-Position kann bei mehreren z Fokuseinstellungen abgebildet werden. Die 3D-Intensitätsdaten können als eine Vielzahl von Intensitäten in den z-Ebenen betrachtet werden, um eine 3D-Bilddatenstruktur zu bilden. Zum Beispiel stellt die Ebene 152a die Intensitätsdaten für eine erste Fokuseinstellung und eine Vielzahl von xy-Positionen dar, während die Ebene 152b, die Intensitätsdaten für eine zweite Einstellung und die gleichen xy-Positionen wie bei der Ebene 152a umfasst. Tatsächlich beziehen sich all die Ebenen auf die Intensitätsdaten für den gleichen Satz von xy-Positionen.

[0023] Die 3D-Bilddaten können dann in eine Vielzahl von Polynomgleichungen für jede xy-Position in Abhängigkeit von z transformiert werden. Das Polynom für jede xy-Position kann jede geeignete Anzahl an Termen und entsprechende Koeffizienten umfassen. In dem vorliegenden Beispiel hat ein aus den 3D-Bilddaten erzeugtes Polynom drei Terme und entsprechende Koeffizienten A, B und C. Die Anzahl der extrahierten Terme kann variieren. Beispielsweise, kann die Anzahl der ausgewählten Koeffizienten vom Bereich des verwendeten Fokus für den Erhalt der Volumendaten abhängen. Ungefähr drei Koeffizienten (A ~ C) eignen sich gut für einen Fokusbereich nahe bei 1 oder 2 DOF (Fokustiefe), während etwa fünf Koeffizienten (A ~ E) sich gut für einen Fokusbereich der Nähe von 3 oder 4 DOF eignen. Über sieben Koeffizienten (A ~ G) können für einen Fokusbereich in der Nähe 5 oder 6 DOF verwendet werden, und so weiter.

[0024] Der Fokusschritt zwischen jedem Paar von aufeinanderfolgenden Fokuseinstellungen wird im Allgemeinen so gesteuert, dass die Bilder zwischen den aufeinanderfolgenden Fokusebenen genug Korrelation haben, um eine gute Bildausrichtung zu ermöglichen. Der Fokusschritt zwischen aufeinanderfolgenden Bildern kann in einem Bruchteil des DOF des Systems gewählt werden. Ein Beispiel der Einstellung der Fokusschritte sind ein 1/3 oder 1/2 des DOF.

[0025] Die Koeffizientenwerte für die verschiedenen xy-Positionen können jeweils auch als eine Bild-

ebene oder eine Matrix von Werten dargestellt werden. Wie gezeigt ist, enthält eine erste Matrix 154 für jede xy-Position A Koeffizientenwerte; eine zweite Matrix 156 weist für jede xy-Position B Koeffizientenwerte; und eine dritte Matrix 158 weist C Koeffizientenwerte für jede xy-Position auf. Die folgende Gleichung [1] kann für jeden Satz von Intensitätswerten für jede xy-Position erzeugt werden:

$$I(x, y, z) = A(x, y)\Delta z^2 + B(x, y)\Delta + C(x, y)$$

Gleichung [1]

[0026] Zusammenfassend repräsentiert ein VM-Formalismus das Verhalten von $I(z)$, die Intensität einer bestimmten xy-Position des Wafers und als Funktion der Fokusebene z , durch ein Polynom n -ter Ordnung. Die Polynomkoeffizienten [wie z.B. $A(x, y)$, $B(x, y)$ und $C(x, y)$] enthalten sowohl Intensität als auch Phaseninformation des EM-Feldes, das vom Wafer erzeugt wird. Als Effekt werden elektromagnetische (EM) volumetrische Parameter (VM-Parameter) aus den 3D-Intensitätsdaten extrahiert. Diese VM-Parameter können in zahlreichen Anwendungen eingesetzt werden, wie weiter unten beschrieben wird, um die Inspektion zu verbessern.

[0027] Fig. 2 zeigt ein Flussdiagramm, das einen volumetrischen Inspektionsprozess 200 gemäß einer Ausführungsform der vorliegenden Erfindung darstellt. Zu Beginn wird bei Schritt 202 ein erster Satz von einem oder mehreren Scan-Streifen ausgewählt. Eine erste Fokuseinstellung wird auch für den aktuellen Satz von Streifen im Schritt 204 ausgewählt. Ein einfallender Strahl wird mit der aktuell gewählten Fokuseinstellung über den aktuellen Satz von einem oder mehr Streifen gescannt und als Antwort werden Bilddaten im Schritt 206 aufgenommen. Encoderdaten können ebenfalls im Schritt 206 gesammelt werden.

[0028] Fig. 3A zeigt eine schematische Darstellung eines Wafer-Scanvorgangs, der die Erstellung einer 3D-Intensitätsdatenstruktur in Übereinstimmung mit einer spezifischen Implementierung der vorliegenden Erfindung unterstützt. Fig. 3B zeigt eine Vielzahl von Scan-Streifen 371 (Streifen_1, Streifen_2, ..., Streifen_n) für den in Fig. 3A gezeigten Wafer 350. Wie in Fig. 3 A dargestellt, ist der Abschnitt 354 eine Vergrößerung eines Abschnitts 351 des Wafers 350, der eine Zeile mit von Chips bzw. DIES (wie z.B. 356a ~ 356e) umfasst. In diesem dargestellten Beispiel wird ein erster Fokus zum Abscannen von zwei Streifen einer Reihe der Chips mit einem Serpentinmuster ausgewählt. Zum Beispiel umfasst die Inspektion, dass ein einfallender Strahl mit einer ersten Fokuseinstellung scannt und Bild- oder Intensitätsdaten in der ersten Richtung sammelt (wie z.B. entlang des Streifens_0), der sich von links nach rechts bewegt, und das Scannen mit der gleichen

ersten Fokuseinstellung in einer entgegengesetzten zweiten Richtung (wie z.B. entlang des Streifens_1), wobei sich der Strahl von rechts nach links bewegt. Andere Scanschemata können eingesetzt werden. Beispielsweise kann das Scannen dadurch erreicht werden, dass entweder der Tisch oder der einfallende Strahl bewegt wird oder beide bewegt werden, um so zu einer Relativbewegung zwischen dem Tisch und dem Strahl zu gelangen. Ein Beispiel eines alternativen Scansystems ist ein Rücksprung-Flut-System. Bei einem Flut-System ist der Fußabdruck des Strahls / Beleuchtung stationär, während der Tisch sich als Scanner bewegt. Ebenso ist die Auflösung eines Beleuchtungssystem hierfür im Allgemeinen durch die Abbildungsoptik, anstelle von der Spotgröße der Beleuchtung bestimmt.

[0029] Der nächste Schritt 208 zur Ausrichtung der Bildsätze verschiedener Fokuseinstellungen ist nur dann anwendbar, nachdem mehr als ein Satz von Bilddaten für mehr als eine Fokuseinstellung gesammelt wurde. Somit wird der Schritt 208 für die erste Fokuseinstellung übersprungen. Stattdessen kann dann bei Schritt 210 bestimmt werden, ob mehr Fokuseinstellungen vorhanden sind, um so jeden vordefinierten Satz von Fokuseinstellungen zu durchlaufen. Somit kann, wenn mehr Fokuseinstellungen auszuwählen sind, der Prozess bei Schritt 204 weitergeführt werden, wobei eine nächste Fokuseinstellung ausgewählt wird. Die Bilddaten werden für diesen nächsten Fokus bei Schritt 206 gesammelt.

[0030] Die Inspektion kann ein kontinuierliches Scannen über eine beliebige Anzahl der DIES / Chips des Wafers oder Abschnitte der DIES / Chips bei einer bestimmten Fokuseinstellung umfassen, und das Scanmuster hängt für jede Fokuseinstellung von einer beliebigen Anzahl von Faktoren ab. Ein Faktor ist die thermische Ausdehnung, die die tatsächliche Fokuseinstellung oder die Bildausrichtung quer zum Fokus, nachdem das Inspektionswerkzeug für eine bestimmte Zeitdauer im Betrieb ist, beeinflusst. Zum Beispiel ist es von Vorteil, alle Fokuseinstellungen in einer Zeit zu durchlaufen, die weniger als die durchschnittliche oder minimale Zeit ist, damit die thermische Ausdehnung signifikant die Positionseinstellungen beeinflusst, sowie, wie schnell das Inspektionswerkzeug die Streifen scannen kann. Beispielsweise kann die Anzahl der Streifen in der Regel in weniger als fünf Minuten den Fokus durchlaufen. Das Durchlaufen der Fokuseinstellungen für zwei Streifen scheint gut zu funktionieren, obwohl auch andere Zahlen funktionieren können.

[0031] Nachdem zwei Bilddatensätze für zwei Fokuseinstellungen gesammelt wurden, können die bei den verschiedenen Fokuseinstellungen gesammelten Bilddatensätze ausgerichtet werden, um bei Schritt 208 eine 3D-Inspektionsdatenstruktur zu bil-

den. Alternativ können die Bilddaten, nachdem alle Bilddatensätze bei den verschiedenen Fokuseinstellungen gesammelt wurden, ausgerichtet werden.

[0032] Unabhängig davon, wann der Ausrichtungsvorgang durchgeführt wird, kann eine räumlich präzise x/y Ausrichtung zwischen den durchfokussierten Bildern idealerweise erreicht werden, um somit vollständig die 3D-Bilddaten zu nutzen. In einer Ausführungsform werden x/y-Encoderdaten des Tisches in Echtzeit gesammelt, um eine perfekte Ausrichtung der bei verschiedenen Fokushöhen aufgenommenen Bildern zu ermöglichen. Der Zähler des z-Encoders des Tisches wird auch in Echtzeit erfasst, so dass die VM-Parameter, wie weiter unten beschrieben, ohne Fehler extrahiert werden können.

[0033] Jede geeignete Prozedur kann verwendet werden, um die 3D-Bilddaten aus verschiedenen Fokuseinstellungen auszurichten. **Fig. 3C** ist ein Flussdiagramm, das eine Prozedur 300 der Bildausrichtung gemäß einer Ausführungsform der vorliegenden Erfindung zeigt. Anfangs kann bei Schritt 302 bestimmt werden, ob ein neuer Scan eines Streifens mit einer neuen Fokuseinstellung gestartet wurde. Zum Beispiel ist das Inspektionswerkzeug zum Scannen des Streifens 358a aus **Fig. 3A** bei einer ersten Fokuseinstellung bereit.

[0034] Um Ortsdaten des Encoders zu erhalten, kann eine Instruktion an das Encodersystem des Tisches gesendet werden, um xyz-Positionsdaten in einen oder mehreren Puffern des Encoders aufzuzeichnen. Anschließend wird dann eine andere Anweisung an das Encodersystem des Tisches gesendet, um die xyz-Positionsdaten aus den Puffern des Encoders in den Systemspeicher hochzuladen oder „zu schreiben“.

[0035] Während das xyz-Encodersystem angewiesen werden kann, in Echtzeit xyz-Positionsdaten des Encoders zu erfassen und diese Daten in den Systemspeicher während des Scans eines Streifens zu schreiben, kann eine spezielle Firmware konfiguriert werden, um in Echtzeit eine bidirektionale Kommunikation zu ermöglichen. Alternativ würde eine sequentielle Lese-/Schreibtechnik auch funktionieren, wobei das System das Wenden für den Scan verwendet, um das Schreiben und die Aufnahme der xyz-Positionsdaten des Encoders zu ermöglichen. Ein Beispiel zur Handhabung des Encoders ist in **Fig. 3C** veranschaulicht. Im Schritt 306 wird das Aufzeichnen der xyz-Encoderdaten in Echtzeit in den Puffer des Encoders initiiert. Zum Beispiel, wird der Recorder des Encoders am Punkt 360a vor dem Scan des Streifens 358a (**Fig. 3A**) eingeschaltet. Somit werden die Bilddaten während des Scans des Streifens 358a gesammelt und die xyz-Positionsdaten des Encoders werden im Puffer des Encoders aufgezeichnet.

[0036] Bei Schritt 308 kann dann bestimmt werden, ob der aktuelle Scan des Streifens abgeschlossen ist. In dem Beispiel aus **Fig. 3A** wartet der Prozess 300 bis der Scan 358a des Streifens beendet ist. Der Recorder des Encoders kann ausgeschaltet werden und die xyz-Positionsdaten des Encoders können dann von dem Puffer / den Puffern bei Schritt 310 in den Systemspeicher geschrieben werden. Beispielsweise werden die Datensätze vom Encoder in Echtzeit mit ihren zugeordneten Bilddatensätzen einem Bildverarbeitungssystem bereitgestellt (wie z.B. das System 700 der **Fig. 7**). Wie in **Fig. 3A** gezeigt, wird der Recorder des Encoders bei Punkt 360b nach dem Scan 358a des Streifens abgeschaltet und die Datensätze vom Encoder können dann bei Punkt 360c an das System hochgeladen werden. Dieser Schreibvorgang findet statt, während der Tisch einen Wendevorgang für einen Streifen durchführt, welcher länger dauert als die Zeit zum Hochladen der xyz-Positionsdaten des Encoders in den Systemspeicher. Somit kann die Zeit zum Wenden eines Streifens parallel zum Sammeln der xyz-Positionsdaten des Encoders aus dem Puffer / den Puffern des Encoders genutzt werden. Zum Beispiel wird der Tisch zu einer Position bewegt, um das Abscannen der nächsten Streifen 358b von rechts nach links zu beginnen.

[0037] Bei Schritt 312 kann dann bestimmt werden, ob die Inspektion abgeschlossen ist. Zum Beispiel kann bestimmt werden, ob das Inspektionswerkzeug alle Streifen gescannt hat. Wenn die Inspektion nicht beendet ist, kann dann bei Schritt 314 festgestellt werden, ob der nächste Scan eines Streifens in der gleichen Fokuseinstellung erfolgen soll. Bei dem dargestellten Beispiel wird der nächste Scan eines Streifens 358b mit der gleichen Fokuseinstellung durchgeführt, wie der vorherige Scan eines Streifens 358a. Dementsprechend läuft der Ausrichtungsprozess bei Schritt 306, wobei die Aufzeichnung der Echtzeit xyz-Encoderdaten eingeleitet wird. Zum Beispiel wird der Recorder des Encoders am Punkt 360d der **Fig. 3A** angeschaltet. Am Ende des Scans des Streifens 358b mit der gleichen Fokuseinstellung, wird der Recorder des Encoders wieder am Punkt 360e gestoppt, und die Encoderdaten am Punkt 360f hochgeladen. Für eine neue Fokuseinstellung ist der Prozess zum Erlangen der Encoderdaten ähnlich mit der Hinzufügung der Erlangung der Echtzeit z-Encoderdaten bei Schritt 304. Der Ausrichtungsvorgang wird wiederholt, bis der aktuelle Satz von einem oder mehreren Streifen (wie z.B. 2 Streifen) bei allen Fokuseinstellungen gescannt ist. Die 3D-Bilddaten aus den verschiedenen Fokuseinstellungen können dann genau auf der Basis der Echtzeit xyz-Encoderdaten bei Schritt 316 ausgerichtet werden.

[0038] Unter erneute Bezugnahme auf **Fig. 2** kann dann bestimmt werden, ob alle Streifen bei Schritt 216 gescannt wurden. Falls mehr Streifen übrig blei-

ben, kann dann der Prozess zu dem nächsten Satz von Streifen bei Schritt 218 verschoben werden, und die Bildsammlung wird für solche Streifen durch die Datenerfassung der Fokussierung sowie der Encoderposition wiederholt.

[0039] Parallel zu der Sammlung der Encoderdaten und der Bilddaten von jedem Satz aus einem oder mehreren Streifen bei verschiedenen Fokuseinstellungen, können Polynomparameter für jeden Ziel- und Referenzort aus den ausgerichteten 3D-Bilddaten bei Schritt 212 extrahiert werden. Verschiedene Analyseverfahren können dann auf die extrahierten Polynom-Daten angewendet werden, um Defekte in dem aktuellen Satz von einem oder mehreren Streifen bei Schritt 214 zu erfassen. Dies bedeutet, dass jeder Satz von 3D-Bilddaten für spezifische Streifen bei Durchfokussieren analysiert wird, da diese 3D-Bilddaten für jeden Satz von einem oder mehreren Streifen konstruiert wird. Eine Alternative, aber weniger effizientes Beispiel, ist, dass die 3D-Bilddaten nach dem Durchfokussieren aller Streifen des Wafers abgetastet wurden.

[0040] Fig. 3D ist eine schematische Darstellung von einer 3D-Bild-Konstruktion und der VM-Parameterextraktion in Übereinstimmung mit einer Ausführungsform der vorliegenden Erfindung. Dieses Diagramm veranschaulicht die Bildung eines Polynoms mit fünf Termen und mit Koeffizienten A ~ E.

[0041] Die 3D-Bilddaten, die in Form von Bildebenen 382a ~ 382e dargestellt sind, können zunächst in eine 2D-Matrix 388 angeordnet werden. Zum Beispiel wird Ebene 382a in einer ersten Spalte, wie z.B. 384a, der 2D-Matrix 388 angeordnet. Die anderen Ebenen 382b ~ 382e werden jeweils in einzelne andere der Spalten der 2D-Matrix 388 angeordnet. Die folgende Gleichung [1] stellt die gesammelten Intensitätsdaten in einer 2D-Matrix dar:

$$I_{kl} = \sum_{m=0}^4 T_m Z_{ml} \quad \text{Gleichung [1]}$$

[0042] Die obige Gleichung [1] repräsentiert eine 2D - Matrix, I_{kl} , die aus m Ebenen von Bilddaten an den verschiedenen Fokuseinstellungen z gebildet ist, und besitzt k Zeilen (stellt Pixelintensität der räumlichen Abbildung dar) und l Spalten (die unterschiedliche m -Ebenen repräsentieren). Jedes der Bilder wird neu geordnet, so dass ein 2D-Bild als ein Vektor dargestellt wird. Jedes (i, j) im Originalbild wird in den k-Index abgebildet, wobei $k = i * n_x + j$ gilt. Es ist zu beachten, dass, i der Zeilenindex und j der Spaltenindex ist. In anderen Worten, i stellt die Position in y und j stellt die Position in x dar. Die Matrix T_{km} stellt die Koeffizientenwerte für verschiedene xy-Positionen dar, die verschiedenen Zeilen entsprechen, und

kann ausgedrückt werden durch:

$$T = \begin{bmatrix} A_1 & B_1 & C_1 & D_1 & E_1 \\ A_2 & B_2 & C_2 & D_2 & E_2 \\ \dots & \dots & \dots & \dots & \dots \\ A_k & B_k & C_k & D_k & E_k \end{bmatrix}$$

[0043] Die Matrix Z_{ml} stellt die Änderungen in z für verschiedene Ebenen dar, an denen die 3D volumetrischen Bilder erfasst werden, und kann ausgedrückt werden als:

$$Z = \begin{bmatrix} \Delta z_1^4 & \Delta z_2^4 & \dots & \Delta z_{N_{Ebene-1}}^4 & \Delta z_{N_{Ebene}}^4 \\ \Delta z_1^3 & \Delta z_2^3 & \dots & \Delta z_{N_{Ebene-1}}^3 & \Delta z_{N_{Ebene}}^3 \\ \Delta z_1^2 & \Delta z_2^2 & \dots & \Delta z_{N_{Ebene-1}}^2 & \Delta z_{N_{Ebene}}^2 \\ \Delta z_1^1 & \Delta z_2^1 & \dots & \Delta z_{N_{Ebene-1}}^1 & \Delta z_{N_{Ebene}}^1 \\ \Delta z_1^0 & \Delta z_2^0 & \dots & \Delta z_{N_{Ebene-1}}^0 & \Delta z_{N_{Ebene}}^0 \end{bmatrix}$$

[0044] Die Koeffizienten-Matrix kann dann bestimmt werden durch:

$$T_{km} = \sum_{m=1}^{N_{Ebene}} I_{kl} Z_m^{-1}$$

[0045] Das Polynom für die 3D-Bilddaten kann dann dargestellt werden durch:

$$I_{ij}(\Delta z) = A_{ij} \Delta z^4 + B_{ij} \Delta z^3 + C_{ij} \Delta z^2 + D_{ij} \Delta z + E_{ij} \quad (\Delta z = 0)$$

[0046] Die extrahierten, volumetrischen (VM) Koeffizienten können für jeden geeigneten Zweck verwendet werden, wie beispielsweise, zur Erhöhung des Defekt/Hintergrund-Kontrasts und/oder zum Extrahieren einer höheren räumlichen Segmentierung als die optische Auflösung, um eine Definition von feinen Strukturen zu ermöglichen, wie beispielsweise die Definition eines SRAM - Mikrobereichs zur Begutachtung.

[0047] In einer Anwendung können die VM-Koeffizienten als VM Bilder dargestellt werden, die EM-Informationen von dem Wafermuster und den DOI enthalten, und diese V-Bilder können verwendet werden, um die Sichtbarkeit von DOIs gegenüber den Hintergrund/Störungen zu verbessern. Während der Detektion werden die VM-Bilder sowohl von Zielpositionen als auch von Referenzpositionen erhalten. In einem Beispiel mit drei Koeffizienten können VM-Differenzbilder zwischen dem Ziel und Referenz (bezeichnet als $A_{dif}(x, y)$, $B_{dif}(x, y)$ und $C_{dif}(x, y)$) bestimmt werden.

[0048] Theoretisch hat ein DOI entweder eine andere physikalische Erscheinung (wie z.B. physikalische Defekte) vor dem Hintergrund und/oder ist aus anderem Material als der Hintergrund zusammengesetzt (wie z.B. Verunreinigungen). Daher wird eine EM-Antwort vom DOI erwartet, die anders ist als die EM-Antwort des Hintergrunds. Dementsprechend sollte die Prüfung der Unterschiede in den VM-Parametern zu einer besseren Differenzierung des DOI gegenüber der Störung / des Hintergrunds führen. Diese Theorie wurde auf strukturierten Wafern getestet. **Fig. 4** ist eine 3D-Punktwolke, die aus beispielhaften Differenz-VM-Bildern gemäß einer Ausführungsform der vorliegenden Erfindung konstruiert wurde. Aus dieser beispielhaften 3D-Punktwolke, die aus $A_{\text{dif}}(x, y)$, $B_{\text{dif}}(x, y)$ und $C_{\text{dif}}(x, y)$ konstruiert wurde, ist es klar, dass die DOI-Pixel 404 im Wesentlichen unterschiedlich sind von den Pixeln 402 des Hintergrunds. Cluster-Prozesse können implementiert werden, um DOI Pixel von der Störung / den Pixeln des Hintergrunds zu unterscheiden. Jeder geeignete Cluster-Prozess kann verwendet werden. Beispiele hierfür sind das Clustern nächster Nachbarn, das Clustern um einen Schwerpunkt, auf Verteilung basiertes Clustern, auf Dichte basierendes Clustern usw.

[0049] Um die Unterscheidung zwischen DOI und Hintergrund / Störung zu erleichtern, kann jeder Punkt in der 3D-Punktwolke in eine Einheitskugel projiziert werden. Die Projektion ermöglicht es, dass jeder Punkt eine ausgezeichnete Winkelposition (Θ , Φ) hat. Pixel außerhalb von $(\Theta_{\text{DOI}}, \Phi_{\text{DOI}})$ können innerhalb eines festgelegten Bereichs der Toleranzbereiche der Winkel gefiltert werden, um die Sichtbarkeit des DOI / der Störung zu verbessern. Das heißt, dass Ausreißer in der 3D-Projektion (oder Streudiagramm) als DOIs bezeichnet werden. Ein Winkelfilter kann auf die projizierten Punkte angewendet werden, so dass die Sichtbarkeit von Defekten wesentlich verbessert wird.

[0050] Die Differenzpixel können für die Fehlererkennung mit einem Schwellwert versehen werden. Zum Beispiel, wenn ein gefilterter Differenzwert (Absolutwert) zwischen der Referenz und Ziel größer ist als ein vorgegebener Schwellwert, kann das Pixel als ein Defekt von Interesse (DOI) definiert und als Kandidat für einen Defekt überprüft werden. Das Streudiagramm oder die projizierten Pixel können mit einem Schwellwert versehen werden. Zusätzlich können bestimmten Bereichen (oder Cluster von Punkten) des Streudiagramms oder Bereichen bezüglich des Winkels (geclustert) der 3D-Projektion unterschiedliche Schwellwerte zugewiesen werden. Zum Beispiel können den Regionen von Ausreißern ein empfindlicherer Schwellwert zugewiesen werden, so dass Defekte weit häufiger gefunden werden, während den Nicht-Ausreißern oder den gruppierten

Regionen eine weniger sensitiver Schwellwert oder kein Schwellwert zugeordnet werden kann.

[0051] **Fig. 5A** veranschaulicht eine 3D-Projektion von Pixeln des VM-Differenzbilds auf eine Einheitskugel 500 in Übereinstimmung mit einer Ausführungsform der vorliegenden Erfindung. **Fig. 5B** zeigt ein Beispiel einer projizierten Einheitskugel mit DOI Pixeln 552 und Pixel (nicht bezeichnet) des Hintergrunds / der Störung. Die Einheitskugel 500 (**Fig. 5A**) hat drei Dimensionen: C_{dif} , B_{dif} und C_{dif} .

[0052] Auf der Grundlage von experimentellen Daten, neigen verschiedene Bereiche einer Kugel, auf der die Pixelwerte projiziert werden, dazu, DOIs zu enthalten. Zum Beispiel neigen die Regionen 502a, welche dem Plot 504a entsprechen, dazu, die DOIs zu enthalten, die Positionen mit optimalen Fokus haben, die weg von einer Null-Fokusebene sind und DOIs mit einer relativ schwachen Signalstärke haben. Ebenso hat Region 502b (entsprechend dem Plot 504b) DOIs mit einer relativ schwachen Signalstärke eines von Null verschiedenen, optimalen Fokus. Umgekehrt neigen die Region 502c (entsprechend dem Plot 504c) und die Region 502d (entsprechend dem Plot 504d) dazu, dass sie DOIs bei einem optimalen Fokus von Null besitzen. Jedoch neigt die einzige Region 504d dazu, DOIs mit einer relativ starken Signalstärke zu haben, während die Regionen 504c DOIs mit einer relativen schwachen Signalstärke haben. Als Ergebnis können die DOIs aus bestimmten gruppierten Regionen für die weitere Analyse (wie z.B. Anwendung eines Schwellwerts) ausgewählt werden oder können mit mehr Empfindlichkeit (wie z.B. Schwellwert mit erhöhter Sensitivität) als andere Regionen analysiert werden. Der Ort eines starken/schwachen Objekts, das auf die Kugel projiziert wurde, kann durch einen Blick auf das Verhalten der Polynomfunktion verstanden werden. $C_{\text{dif}}(x,y)$ ist der konstante Term in der Polynomfunktion. Dieser konstante Term beschreibt die Größe des globalen Maximums (absteigendes Polynom zweiter Ordnung) oder globales Minimum (aufsteigendes Polynom zweiter Ordnung). Ein größeres C_{dif} entspricht einer größeren Größe des globalen Maximums oder globalen Minimums und somit einem stärkeren Defekt. Die Kugel mit großem C_{dif} ist auf der oberen oder unteren Halbkugel aufgetragen (wie z.B. 502d aus **Fig. 5A**).

[0053] In einer anderen Anwendung der Verwendung der 3D-Bilddaten, kann eine Verschmelzung des Fokus umgesetzt werden. Wenn die Art des DOI und des Hintergrunds / der Störung unterschiedlich sind, wird erwartet, dass das Verhalten des Durchfokussierens des DOI und der Störung / des Hintergrunds unterschiedlich sein kann. Die Verschmelzung des Fokus ist eine Technik zum weiteren unterscheiden des DOI und der Störung. Die Verschmelzung des Fokus kann durch die Kombination

der Intensität über die Werte der Fokusebene mit jedem Pixel in jeder geeigneten Weise durchgeführt werden. In einer Implementierung wird die Intensität über die Fokusebene miteinander multipliziert, um ein zusammengesetztes Bild zu bilden, das DOIs deutlicher zeigt. Zum Beispiel wird ein verschmolzenes Bild durch Multiplizieren gebildet, wie z.B. Bild $(x, y, \text{deltaZ}_1) * \text{Bild}(x, y, \text{DeltaZ}_2)$ usw. Ein zusammengesetztes Bild, das durch dieses Verfahren zum Verschmelzen des Fokus gebildet wird, neigt dazu, ein relativ höheres Signal-zu-Rausch-Verhältnis (SNR) zu haben als die ursprünglichen Bilder mit festen Fokalebene haben. Das Verschmelzen des Fokus kann mit der VM-Winkelprojektion und der Winkelfiltertechnik kombiniert werden, um ein gefiltertes zusammengesetztes Bild zu erhalten, die ein höheres SNR als die zusammengesetzten Bilder und die Originalbilder aufweisen. Die sich ergebenden zusammengesetzten Bilder der Ziel- und Referenzbereiche können dann mit der oben beschriebenen Schwellwertbildung behandelt werden.

[0054] Die extrahierten VM-Parameter können auch zum Extrahieren von Strukturen des Bauteils verwendet werden, die durch die optische Auflösung nicht auflösbar ist. Wie oben erwähnt, sind VM-Parameter eng mit dem elektromagnetischen Feld verknüpft, das von dem Muster auf dem Wafer generiert wird. Die Informationen können verwendet werden, um die optischen Bilder in verschiedene Mikroregionen, basierend auf dem vom Wafer emittierten EM-Feld, zu unterteilen. **Fig. 6** zeigt ein Beispiel der Bereichssegmentierung, die auf extrahierte VM-Parameter gemäß einer anderen Ausführungsform der vorliegenden Erfindung basiert. Zwei vergrößerte Bildabschnitte 602a und 602b eines zergliederten Bilds sind gezeigt. Das gesamte optische Bild wird zum Zwecke der Klarheit nicht gezeigt. Basierend auf den Eigenschaften der VM-Parameter kann jedes Bild in verschiedene Bereiche unterteilt werden, und jeder Bereich kann durch eine andere Farbe (in **Fig. 6** als schwarz und weiß umrissene Bereiche gezeigt) oder eine andere Kennung dargestellt werden, um die verschiedenen Bereiche zu unterscheiden. Zum Beispiel enthält der Bildabschnitt 602a die Bereiche R1, R2 und R3, während der Bildabschnitt 602b die Bereiche R4, R5 und R6 enthält. Es ist zu erkennen, dass die unterschiedlichen Bereiche tatsächlichen Bauteilstrukturen entsprechen. Die einzelnen Segmente können mit unterschiedlichen Stufen der Gewichtung analysiert werden. Zum Beispiel können die Segmente, die einer bestimmten sensitiven SRAM-Struktur entsprechen, mit empfindlicheren Schwellenwerten als andere Strukturen analysiert werden.

[0055] In einer alternativen Ausführungsform kann die Intensitätsänderung, die beim Durchfokussieren ($\Delta I / \Delta z$) gesammelt wird, verwendet werden, um Unterschiede zwischen den Referenz- und den Tar-

getbildern zu vergleichen. Indem eine $\Delta I / \Delta z$ Messung durchgeführt wird, kann die Intensitätsmessung bezüglich der Phase empfindlich werden, um eine höhere Empfindlichkeit zur Verfügung zu stellen. Dies gilt insbesondere für die durch Defekte induzierte Phasenänderung. Zum Beispiel können neu erzeugte Differenzbilder zwischen den Referenz- und Zielbildern, die durch die Intensitätsableitung über die Defokussierung berechnet werden, verwendet werden, um Defekte mittels der Techniken der adaptiven Schwellwertbildung zu detektieren.

[0056] In einem Beispiel kann zuerst die Differenz der Intensität für zwei unterschiedliche Fokuseinstellungen an der gleichen x-y Position bestimmt werden. Diese Berechnung des Unterschieds kann in einem Bereich und nicht nur an einem einzelnen Punkt durchgeführt werden. Dabei ist von Interesse wie sich $(dI)/(dz)$ in Abhängigkeit von x und y ändern. Falls Bilder von mehr als zwei Fokussen vorhanden sind, schätzt ein Ansatz $(dI)/(dz)$ durch $(I1 - I2)/(\text{delta } z)$ weitaus genauer ab, wobei Terme höherer Ordnung vernachlässigt werden. Da $(dI)/(dz)$ eine mathematisch glatte Funktion ist, die mehrere Messungen von mehreren Brennebenen aufweist, kann diese eine bessere Schätzung für $(dI)/(dz)$ liefern.

[0057] Beispielsweise wird ein Satz von Gewichten für das Bild so gewählt, dass alle bis auf der 1. Ordnung gewählt werden. Bei einem anderen Ansatz liefert eine einfache Methode zum Kurvenfit eine bessere Leistung, auf Kosten von mehr Rechenzeit. Bei dieser Technik wird die Intensität gegen die z-Daten für jedes Pixel an ein Polynommodell gefittet, und dann wird die gewünschte Komponente erster Ordnung zum Berechnen der Phase extrahiert. Durch Fitten an Polynome höherer Ordnung kann eine genauere Schätzung der Ableitung erster Ordnung erhalten werden. Ein Fit der kleinsten Quadrate an Polynome, der alle Bilder gleich gewichtet, kann verwendet werden. Die Ordnung der Fitfunktion der Polynome ist in der Regel kleiner als die Anzahl der verwendeten Bilder, und mehrere Bilder führen zu einem besseren Rauschverhalten, ohne dass die Genauigkeit geopfert wird. Rechnerisch kann jedes Pixel unabhängig behandelt werden, und an ein Polynom durch Standardfittetechniken (Kurvenfit durch kleinste Quadrate) gefittet werden. Die pixelweise Behandlung eignet sich gut für eine parallele Verarbeitung durch den Computer, wie die Berechnung auf einer graphischen Prozessoreinheit (Graphics Processing Unit) (GPL).

[0058] Techniken der vorliegenden Erfindung können in jeder geeigneten Kombination von Hardware und/oder Software implementiert werden. **Fig. 7** ist eine schematische Darstellung eines beispielhaften Inspektionssystems 700, bei dem Techniken der vorliegenden Erfindung implementiert werden können. Das Inspektionssystem 700 kann eine Eingabe 702

von einem Inspektionswerkzeug oder Scanner (nicht gezeigt) empfangen. Das Inspektionssystem kann auch ein Datenverteilungssystem (wie z.B. 704a und 704b) zum Verteilen des empfangenen Eingangs 702, ein Verarbeitungssystem (oder Patch) für das Intensitätssignal (wie z.B. Patch-Prozessoren und Speicher 706a und 706b) für die Verarbeitung von spezifischen Teilen/Patches des empfangenen Eingangs 702, ein VM-Generator-System (wie z. B. ein VM-Generator-Prozessor und Speicher 712) zur Erzeugung von 3D-Bilddaten bei mehreren Fokussen usw., ein Netzwerk (wie z.B. geschaltetes Netzwerk 708), um die Kommunikation zwischen den Komponenten des Inspektionssystems zu ermöglichen, einen optionalen Massenspeicher 716 und eine oder mehrere Inspektionssteuerungen und/oder Review-Stationen (wie z.B. 710), um die Inspektionsergebnisse zu überprüfen, umfassen. Jeder Prozessor des Inspektionssystems 700 kann typischerweise eine oder mehrere Mikroprozessoren mit integrierten Schaltungen umfassen und kann auch eine Schnittstelle und/oder einen Speicher aus integrierten Schaltungen enthalten, die zusätzlich an eine oder mehrere gemeinsame und/oder globale Speichervorrichtungen gekoppelt sein können.

[0059] Der Scanner oder das Datenerfassungssystem (nicht gezeigt) kann zum Erzeugen von Eingangsdaten 702 die Form eines beliebigen und geeigneten Instrumentes annehmen (wie z.B. hier weiter beschrieben ist), um die Intensitätssignale oder Bilder von einem Wafer zu erhalten. Zum Beispiel kann der Scanner ein optisches Bild aufbauen oder Intensitätswerte eines Teil des Wafers erzeugen, die auf einen Teil des detektierten Lichts basieren, das zu einem oder mehreren Lichtsensoren reflektiert, übertragen oder auf andere Weise gerichtet wird. Der Scanner kann dann die Intensitätswerte oder Bilder sowie Positionsdaten des Encoders an das Inspektionssystem ausgeben.

[0060] Der Wafer wird im Allgemeinen in eine Vielzahl von Patch-Portionen aufgeteilt, aus denen mehrere Intensitätswerte von mehreren Punkten erhalten werden. Die Patch-Portionen des Wafers können gescannt werden, um diese Intensitätsdaten zu erhalten. Die Patch-Portionen können eine beliebige Größe und Form haben, die von den speziellen Systemen und den Anforderungen der Anwendung abhängig sind. Im Allgemeinen können mehrere Intensitätswerte für jede Patch-Portion durch Scannen des Wafers in jeder beliebigen und geeigneten Weise erhalten werden. Beispielsweise können mehrere Intensitätswerte für jede Patch-Portion durch Raster-scannen des Wafers erhalten werden. Alternativ können die Bilder durch Scannen des Wafers mit einem beliebigen und geeigneten Muster, wie z.B. einem kreisförmigen oder spiralförmigen Muster, erhalten werden. Selbstverständlich können die Sensoren unterschiedlich angeordnet werden (wie z.B. in

einem kreisförmigen Muster) und/oder der Wafer kann während des Scannens anders bewegt (wie z.B. gedreht) werden, um eine kreis- oder spiralförmige Form von dem Wafer zu scannen.

[0061] Im unten veranschaulichten Beispiel wird Licht, während der Wafer sich an den Sensoren vorbei bewegt, von einem rechteckigen Bereich (hier als „Streifen“ des Wafers bezeichnet) detektiert, und dieses detektierte Licht wird in eine Vielzahl von Intensitätswerten an mehreren Punkten in jedem Patch umgewandelt. In dieser Ausführungsform sind die Sensoren des Scanners in einem rechteckigen Muster angeordnet, um Licht zu empfangen, das von dem Wafer reflektiert und/oder übertragen wird. Die Sensoren erzeugen hieraus einen Satz von Intensitätsdaten, der einem Streifen von Patches des Wafers entspricht. In einem speziellen Beispiel kann jeder Streifen etwa 1 Million Pixel breit und etwa 1000 bis 2000 Pixel hoch sein, während jedes Patch etwa 2000 Pixel breit und 1000 Pixel hoch sein kann.

[0062] Die Intensitätswerte für jedes Patch können mit einem optischen Inspektionswerkzeug erhalten werden, das in jeder geeigneten Weise aufgebaut sein kann. Das optische Werkzeug wird in der Regel mit einem Satz von Betriebsparametern oder ein „Rezept“ eingerichtet, das für die verschiedenen Inspektionsdurchläufe zum Gewinnen der Intensitätswert im Wesentlichen das gleiche ist. Die Rezept-einstellungen können eine oder mehrere der folgenden Einstellungen umfassen: eine Einstellung für das Scannen des Wafers mit einem bestimmten Muster, die Pixelgröße, eine Einstellung zum Gruppieren benachbarter Signale von einzelnen Signalen, die Fokuseinstellungen, eine Einstellung für die Beleuchtungs- oder Detektionsapertur, eine Einstellung des Einfallswinkels und der Wellenlängen, eine Einstellung des Detektors, eine Einstellung für die Menge des reflektierten oder übertragenen Lichts, der Parameter zur Modellierung eines Luftbildes, usw.

[0063] Die Intensität oder Bilddaten, einschließlich Positionsdaten des Encoders, können als Eingang 702 vom Datenverteilungssystem über das Netzwerk 708 begutachtet werden. Das Datenverteilungssystem kann einer oder mehreren Speichervorrichtungen, wie z.B. RAM-Puffer, zugeordnet sein, um zumindest einen Teil der empfangenen Daten (Eingang 702) zu halten. Vorzugsweise ist der Gesamtspeicher groß genug, um einen ganzen Streifen von Daten zu halten. Zum Beispiel funktioniert ein Gigabyte an Speicher gut für einen Streifen aus 1 Million Pixel auf 1000 Pixel oder Punkten gut.

[0064] Das Datenverteilungssystem (wie z.B. 704a und 704b) kann auch die Verteilung von Teilen der empfangenen Eingangsdaten 702 an die Prozess-

oren (wie z.B. 706a und 706b) steuern. Zum Beispiel kann das Datenverteilungssystem für einen ersten Patch (bei mehreren Fokuseinstellungen) die Daten zu einem ersten Patch-Prozessor 706a und Daten für einen zweiten Patch (bei mehreren Fokuseinstellungen) zu dem Patch-Prozessor 706b leiten. Mehrere Sätze von Daten für mehrere Patches können auch zu jedem Patch-Prozessor geleitet werden.

[0065] Die Patch-Prozessoren können Intensitätswerte oder ein Bild erhalten, das zumindest einem Abschnitt oder Patch des Wafers entspricht. Die Patch-Prozessoren können jeweils auch mit einer oder mehreren Speichervorrichtungen (nicht gezeigt) gekoppelt oder in eine oder mehrere Speichervorrichtungen (nicht gezeigt) integriert sein, wie z.B. DRAM-Bauteile, die lokale Speicherfunktionen bieten, wie z.B. das Halten der empfangenen Datenabschnitte. Vorzugsweise ist der Speicher groß genug, um Daten zu halten, die einem Patch des Wafers entsprechen. Beispielsweise kann ein Patch 512 mal 1024 Pixel betragen. Die Patch-Prozessoren können sich auch einen gemeinsamen Speicher teilen.

[0066] Jeder Satz von Eingangsdaten 702 kann einem Streifen des Wafers entsprechen. Ein oder mehrere Sätze von Daten können in dem Speicher des Datenverteilungssystems gespeichert werden. Dieser Speicher kann in dem Datenverteilungssystem durch einen oder mehrere Prozessoren gesteuert werden, und der Speicher kann in eine Vielzahl von Partitionen unterteilt werden. Zum Beispiel kann das Datenverteilungssystem eine Datenmenge, die einem Teil eines Streifens entspricht, in einer ersten Speicherpartition (nicht gezeigt) empfangen, und das Datenverteilungssystem kann einen anderen Datensatz, der einem anderen Streifen entspricht, in einer zweiten Speicherpartition (nicht gezeigt) empfangen. Vorzugsweise hält jede der Speicherpartitionen des Datenverteilungssystems nur die Teile der Daten, die zu einem Prozessor weitergeleitet werden sollen, der einer solchen Speicherpartition zugeordnet ist. Zum Beispiel kann die erste Speicherpartition des Datenverteilungssystems erste Daten halten und sie zu dem Patch-Prozessor 706a weiterleiten, und die zweite Speicherpartition kann zweite Daten halten und sie zu dem Patch-Prozessor 706b weiterleiten.

[0067] Das einfallende Licht oder das detektierte Licht können durch eine beliebige geeignete räumliche Apertur geleitet werden, um jedes einfallende oder detektierte Lichtprofil bei irgendwelchen geeigneten Einfallswinkeln zu erzeugen. Beispielsweise können programmierbare Beleuchtungs-Detektionsaperturen verwendet werden, um ein bestimmtes Strahlprofil, wie z.B. einen Dipol, einen Quadrupol, einen Quasar, einen Ring usw. zu erzeugen. In einem bestimmten Beispiel können die Ausgangsmaskenoptimierung (Source Mask Optimization

(SMO)) oder die gepixelte Beleuchtungstechnik verwendet werden. In anderen Ausführungsformen werden ein Polarisator und Analysator verwendet, um polarisierte Intensitäts-/Bilddaten zu erhalten.

[0068] Das Datenverteilungssystem kann jeden Satz von Daten, die auf Daten jeder geeigneten Parameter der Daten basieren, definieren und verteilen. Beispielsweise können die Daten definiert und verteilt werden, die einer entsprechenden Position des Patches auf dem Wafer entsprechen. In einer Ausführungsform ist jeder Streifen einer Reihe von Spaltenpositionen zugeordnet, die horizontalen Positionen der Pixel innerhalb des Streifens entsprechen. Zum Beispiel entsprechen die Spalten 0 bis 256 des Streifens einem ersten Patch, und die Pixel innerhalb dieser Spalten werden das erste Bild oder einen Satz von Intensitätswerten umfassen, die zu einem oder mehreren Patch-Prozessoren geleitet werden. Ebenso können die Spalten 257 bis 512 des Streifens einem zweiten Patch entsprechen, und die Pixel in den Spalten werden das zweite Bild oder einen Satz von Intensitätswerten umfassen, die zu einem anderen Patch-Prozessor oder zu anderen Patch-Prozessoren geleitet werden. Die Spalten können durch das System auf der Basis der Positionsdaten des Encoders oder von dem Scanner ausgerichtet werden.

[0069] Die hierin beschriebenen Scantechniken für die Inspektion können auf oder in Bezug zu Bildern und Positionsdaten des Encoders implementiert werden, die von verschiedenen und speziell konfigurierten Inspektionssystemen, wie das in **Fig. 8** schematisch dargestellte Inspektionssystem, erhalten werden. Das dargestellte System 850 enthält eine Beleuchtungsquelle 860, die mindestens einen Lichtstrahl bereitstellt, der durch eine Beleuchtungsoptik 851a auf eine Probe S, wie z.B. eine Photomaske oder einen Wafer, in die Ebene 852 gerichtet wird. Das Inspektionssystem 850 kann eine numerische Apertur 851 b in der Ebene 852 haben. Die Beleuchtungsoptik 851 a kann auch verschiedene Linsen und Module für das Ausbilden eines oder mehrerer einfallender Strahlen mit unterschiedlichen Eigenschaften umfassen. Die zu inspizierende / zu vermessende Probe S ist auf einem Tisch-Mechanismus 804 in der ausgesetzten Ebene 852 platziert. Der Tisch-Mechanismus umfasst einen Encoder, der derart betreibbar ist, dass Positionsinformationen des Encoders in seinem Puffer aufgezeichnet werden und dass diese Daten geschrieben werden, indem das System der **Fig. 7** diese beispielhaft empfängt.

[0070] Das übertragene Bild von einer Probe S (wie z.B. einer Maske) kann durch eine Reihe von optischen Elementen 853a gerichtet werden, die das gemusterte Bild auf einen Sensor 854a projizieren. Optische Elemente (wie z.B. Strahlteiler 876 und Detektionslinse 878) sind derart angeordnet, dass

das Licht von der Probe S, welches reflektiert und/oder gestreut wurde, erfasst und auf den Sensor 854b gerichtet wird. Geeignete Sensoren umfassen ladungsgekoppelte Bauteile (CCD), CCD-Arrays, Time-Delay-Integration (TDI) Sensoren, TDI-Sensoranordnungen, Photovervielfacherröhren (PMT) und andere Sensoren. Bestimmte Inspektionswerkzeuge können nur eine reflektierende Sammeloptik 853b (oder eine andere Streuoptik) und einen Sensor 854b umfassen, wobei die Optik 853a und der Sensor 854a ausgeschlossen sind.

[0071] Die Säule der Beleuchtungsoptik kann in Bezug auf den Maskentisch und/oder den Tisch durch jeden geeigneten Mechanismus relativ zum Detektor oder der Kamera bewegt werden, um die Patches des Retikels zu scannen. Beispielsweise kann ein Motormechanismus verwendet werden, um den Tisch zu bewegen. Der Motormechanismus kann zum Beispiel aus einem Spindeltrieb und Schrittmotor, einem Lineartrieb mit Rückinformation zur Position oder Bandtrieb mit Schrittmotor-Stellantrieb gebildet werden.

[0072] Die Bilderfassung mittel Durchfokussierung kann entweder durch Bewegen des Tisches (804) nach oben und unten oder durch Bewegen der Detektoren (854b) entlang der optischen Achse oder durch Bewegen des Tisches und der Detektoren erreicht werden.

[0073] Die von jedem Sensor (wie z.B. 854a und/oder 854b) erfassten Signale können durch ein Steuersystem 873, wie z.B. eine oder mehr Signalverarbeitungseinrichtungen, verarbeitet werden, das jeweils einen Analog-zu-Digital-Wandler umfassen kann, der konfiguriert ist, um analoge Signale von jedem Sensor für die Verarbeitung in digitale Signale zu wandeln. Die Systemsteuerung 873 kann einen oder mehr mit Eingangs-/Ausgangsports gekoppelte Prozessoren und einen oder mehrere über geeignete Busse oder andere Kommunikationsmechanismen gekoppelte Speicher umfassen.

[0074] Die Systemsteuerung 873 kann auch eine oder mehrere Eingabelemente (wie z.B. eine Tastatur, eine Maus, einen Joystick) umfassen, um mit einer Benutzereingabe die Fokusparameter und andere Parameter des Inspektionsrezepts zu ändern. Das Steuerungssystem 873 kann auch mit dem Positionierungsmechanismus des Tisches verbunden werden, um z.B. den Tisch bei der Positionierung (Fokussierung und Scannen) der Probe zu steuern, und kann mit anderen Inspektions-/ Metrologiesystemkomponenten zur Kontrolle der Inspektionsparameter und zur Konfiguration solcher Komponenten verbunden sein.

[0075] Das Steuerungssystem 873 kann (zum z.B. mit Programmieranweisungen) konfiguriert werden,

um eine Benutzerschnittstelle (wie z.B. einem Computer-Bildschirm) für die Darstellung der sich ergebenden Intensitätswerte, Bilder und anderer Inspektions-/Metrologieergebnisse zu konfigurieren. Das Steuerungssystem 873 kann zur Erzeugung von 3D-Bildstrukturen, extrahierten Koeffizienten, Fusionsergebnissen, Segmentierung von Bildern, projizierten Daten, und/oder anderen Umwandlungen von reflektierten und/oder transmittierten und dem erfassten Lichtstrahl konfiguriert werden. Das Steuerungssystem 873 kann (wie z. B. mit Programmieranweisungen) konfiguriert werden, um eine Benutzerschnittstelle (wie. z.B. auf einem Computer-Bildschirm) zur Verfügung zu stellen, über die resultierende Intensitätswerte, Bilder, Diagramme, Projektionen und andere Inspektions-/Metrologieigenschaften dargestellt werden. Bei bestimmten Ausführungsformen ist das Steuersystem 873 konfiguriert, um die oben ausführlich erläuterten Prüftechniken auszuführen.

[0076] Da solche Informationen und Programmieranweisungen auf einem speziell konfigurierten Computersystem implementiert werden können, umfasst ein solches System Programmieranweisungen/Computercodes, die auf einem nicht-flüchtigen computerlesbaren Medien gespeichert werden können, um die hierin beschriebenen verschiedenen Operationen (Schritte) durchzuführen. Beispiele für maschinenlesbare Medien umfassen, sind aber nicht darauf beschränkt, magnetische Medien, wie z.B. Festplatten, Disketten und Magnetband; optische Medien, wie z.B. CD-ROM-Disks; magneto-optische Medien, wie z.B. optische Platten, und Hardwareelemente, die speziell zum Ausführen von Programmieranweisungen konfiguriert sind, wie z.B. Nur-Lese-Speicherelemente (ROM) und Direktzugriffsspeicher (RAM). Beispiele von Programmieranweisungen umfassen sowohl Maschinencode, wie z.B. von einem Compiler erzeugt, als auch Dateien mit einem Code höherer Ebene, der von dem Computer unter Verwendung eines Interpreters ausgeführt werden kann.

[0077] In bestimmten Ausführungsformen kann ein System zur Inspektion eines Halbleiterwafers mindestens einen Speicher und mindestens einen Prozessor umfassen, die die hierin beschriebenen Techniken ausführen können. Beispiele für ein Inspektionssystem umfassen speziell konfigurierte 29xx, 8xxx oder 3xxx Familien von Inspektionssystemen, die über KLA-Tencor in Milpitas, Kalifornien erhältlich sind.

Patentansprüche

1. Ein Verfahren zur Detektion von Defekten auf einer Halbleiterprobe (106), wobei das Verfahren umfasst:
Benutzen eines Inspektionswerkzeugs (102), um Intensitätsdatensätze (152a - 152e) bei einer Viel-

zahl von Fokuseinstellungen von jeder einer Vielzahl von xy-Positionen der Probe (106) zu sammeln; Extrahieren einer Polynomgleichung als Funktion der Fokuseinstellung, die eine Vielzahl von Koeffizienten für jeden der gesammelten Intensitätsdatensätze (152a - 152e) einer xy-Position aufweist; Darstellen eines jeden der Sätze von Werten der Koeffizienten für die Vielzahl von xy-Positionen mit einer entsprechenden Koeffizienten-Bildebene, um eine Vielzahl von Koeffizienten-Bildebenen zu bilden; und Analysieren einer Vielzahl von Differenz-Koeffizienten-Bildebenen, die auf einem Vergleich eines Zielsatzes der Koeffizienten-Bildebenen und eines Referenzsatzes der Koeffizienten-Bildebenen basieren, um Defekte auf der Probe (106) zu detektieren.

2. Verfahren nach Anspruch 1, wobei die Intensitätsdaten (152a - 152e) bei den mehreren Fokuseinstellungen von einem ersten Satz von einem oder mehreren Streifen vor einem nächsten Satz von einem oder mehreren Streifen gesammelt werden, wobei eine Anzahl der Streifen in dem ersten und dem nächsten Satz, von welchem die Intensitätsdaten (152a - 152e) gesammelt werden, derart ausgewählt werden, dass sie kleiner ist als eine Zeit der thermischen Ausdehnung, um eine tatsächliche Fokusposition zu beeinflussen.

3. Verfahren nach Anspruch 2, wobei die Anzahl der Streifen einen ersten und zweiten Streifen umfasst.

4. Verfahren nach Anspruch 3, ferner umfassend:
Sammeln von xy- und z- Positionsdaten eines Encoders des Inspektionswerkzeugs (102) in Echtzeit; und
vor einem Extrahieren der Polynomgleichung für jeden der an xy-Positionen gesammelten Intensitätsdatensätze (152a - 152e), erfolgt ein Ausrichten der gesammelten Intensitätsdaten jeder Fokuseinstellung, das auf den in Echtzeit gesammelten xy-Positionsdaten basiert.

5. Verfahren nach Anspruch 4, wobei die Intensitätsdaten von dem ersten Streifen und dem zweiten Streifen der xy-Positionen bei der Vielzahl von Fokuseinstellungen gesammelt werden, bevor das Ausrichten durchgeführt wird.

6. Verfahren nach Anspruch 2, ferner umfassend:
ein Initiieren der Aufzeichnung der xy- Positionsdaten eines Encoders in einem Puffer des Encoders, bevor die Intensitätsdaten von einem ersten der Streifen bei einer ersten der Fokuseinstellungen gesammelt werden,
während ein Tisch des Inspektionswerkzeugs (102) umkehrt, um sich für die Sammlung der Intensitäts-

daten von einem zweiten der Streifen bei der ersten Fokuseinstellung zu konfigurieren, werden die xy-Positionsdaten aus dem Puffer des Encoders in einen Systemspeicher kopiert, um einen Zugriff für die Verwendung im Ausrichtvorgang zu haben und dann wird die Aufzeichnung der xy-Positionsdaten des Encoders in den Puffer des Encoders initiiert, bevor die Intensitätsdaten des zweiten Streifens gesammelt werden;
während der Tisch des Inspektionswerkzeugs (102) umkehrt, um sich für die Sammlung der Intensitätsdaten von dem ersten Streifen bei einer zweiten der Fokuseinstellungen nach dem Sammeln der Intensitätsdaten von dem zweiten Streifen zu konfigurieren, werden die xy-Positionsdaten aus dem Puffer des Encoders in einen Systemspeicher kopiert, um einen Zugriff für die Verwendung im Ausrichtvorgang zu haben und dann wird die Aufzeichnung der xy-Positionsdaten des Encoders in den Puffer des Encoders initiiert, bevor die Intensitätsdaten des ersten Streifens bei der zweiten Fokuseinstellung gesammelt werden; und
Wiederholen der Operationen des Initiierens der Aufzeichnung und des Kopierens für aufeinanderfolgende Paare von Streifen bei jeder der Fokuseinstellungen.

7. Verfahren nach Anspruch 1, wobei das Analysieren durchgeführt wird mittels:
einer Berechnung der Vielzahl von Differenz-Koeffizienten-Bildebenen, die für jeden Koeffizienten eine Vielzahl von Werten der Differenzkoeffizienten haben, mittels einer Subtraktion eines jeden Zielsatzes von jedem der Referenzsätze.

8. Verfahren nach Anspruch 7, wobei die Differenz-Koeffizienten-Bildebenen dadurch analysiert werden, dass ein Bildpunkt für die Werte der Differenzkoeffizienten an jeder xy-Position der Differenz-Koeffizienten-Bildebenen in ein Streudiagramm geplottet wird, das eine Achse für jeden Koeffizienten besitzt, und dass diese Bildpunkte des Streudiagramms in Clustern von Bildpunkten interessierender Defekte, Bildpunkten von Störungen oder Bildpunkten des Hintergrunds angehäuft werden.

9. Verfahren nach Anspruch 7, wobei die Differenz-Koeffizienten-Bildebenen dadurch analysiert werden, dass ein Bildpunkt für die Werte der Differenzkoeffizienten für jede xy-Position der Differenz-Koeffizienten-Bildebenen auf eine Einheitskugel projiziert wird, und dass die projizierten Bildpunkte in Clustern von Bildpunkten von interessierenden Defekten, Bildpunkten von Störungen oder Bildpunkten des Hintergrunds angehäuft werden.

10. Verfahren nach Anspruch 7, wobei die Differenz-Koeffizienten-Bildebenen analysiert werden durch:
Erzeugen einer Vielzahl von Differenzbildern aus

Intensitätsdatensätzen, die von einem Ziel und einer Referenz an jeder Fokusebene gesammelt wurden; Kombinieren der Differenzbilder, um ein fusioniertes Bild über den Fokus zu bilden; und Analysieren des fusionierten Bildes für die Defekterkennung.

11. Verfahren nach Anspruch 7, weiterhin umfassend:

Gruppieren der Koeffizienten mit ähnlichen Werten, um eine Vielzahl von verschiedenen Segmenten zu bilden, wobei jedes Segment einem unterschiedlichen Abschnitt einer tatsächlichen Bauelementstruktur entspricht; Analysieren der verschiedenen Segmente mit unterschiedlicher Stringenz, um Defekte zu detektieren, basierend darauf, welcher Typ eines tatsächlichen Bauelements den verschiedenen Segmenten entspricht.

12. Verfahren nach Anspruch 1, weiterhin umfassend, dass eine Differenz zwischen Intensitätsänderungen als eine Funktion der Änderungen der Fokuseinstellung für jede xy-Position in einem zweiten Ziel-Satz von Intensitätsdatensätzen und einem zweiten Referenz-Satz von Intensitätsdatensätzen analysiert wird.

13. Verfahren nach Anspruch 1, wobei die Fokuseinstellungen Paare von Fokuseinstellungen umfassen, wobei jedes Paar der Fokuseinstellung durch einen Schrittwert getrennt ist, der innerhalb eines Bruchteils der Tiefenschärfe des Fokus des Inspektionssystems ist.

14. Ein Inspektionssystem (850) zur Inspektion einer Halbleiterprobe (S), umfasst:

Eine Beleuchtungsquelle (860) zur Erzeugung eines einfallenden Strahles;

eine Beleuchtungsoptik (851 a) zum Richten des einfallenden Strahls in Richtung auf eine Halbleiterprobe (S) bei einer Vielzahl von Fokuseinstellungen; einen Sensor (854a, 854b) zum Sammeln von Intensitätsdatensätzen von jeder einer Vielzahl von xy-Positionen der Probe in Reaktion auf den einfallenden Strahl; und

eine Steuereinheit (873), die zur Ausführung der folgenden Operationen konfiguriert ist:

Extrahieren einer Polynomgleichung als Funktion der Fokuseinstellung mit einer Vielzahl von Koeffizienten für jeden der an der xy-Position gesammelten Intensitätsdatensätze;

Darstellen eines jeden Satzes von Werten der Koeffizienten für die Vielzahl von xy-Positionen mit einer entsprechenden Koeffizienten-Bildebene, um eine Vielzahl von Koeffizienten-Bildebenen zu erzeugen; und

Analysieren einer Vielzahl von Differenz-Koeffizienten-Bildebenen, die auf einem Vergleich eines Ziel-satzes von Koeffizienten-Bildebenen und eines

Referenzsatzes von Koeffizienten-Bildebenen basieren, um Defekte auf der Probe (S) zu detektieren.

15. Das Inspektionssystem (850) nach Anspruch 14, wobei die Intensitätsdaten bei der Vielzahl von Fokuseinstellungen von einem ersten Satz von einem oder mehreren Streifen sammelbar sind, bevor ein nächster Satz von einem oder mehreren Streifen sammelbar ist, wobei eine Anzahl der Streifen in dem ersten und dem nächsten Satz, von welchem die Intensitätsdaten gesammelt werden, derart ausgewählt wird, dass sie kleiner ist als eine Zeit der thermischen Ausdehnung, um eine tatsächliche Fokusposition zu beeinflussen.

16. Das Inspektionssystem (850) nach Anspruch 15, wobei die Anzahl der Streifen einen ersten und einen zweiten Streifen umfasst.

17. Das Inspektionssystem (850) nach Anspruch 16, wobei die Steuereinheit (873) ferner konfiguriert ist zum:

Sammeln von xy- und z- Positionsdaten eines Encoders des Inspektionswerkzeugs in Echtzeit; und vor einem Extrahieren der Polynomgleichung für jeden der an xy-Positionen gesammelten Intensitätsdatensätze, erfolgt ein Ausrichten der Intensitätsdaten jeder Fokuseinstellung, basierend auf den in Echtzeit gesammelten xy-Positionsdaten.

18. Das Inspektionssystem (850) nach Anspruch 15, ferner umfassend:

einen Tisch (804) zum Aufnehmen der Probe (S); einen Encoder-Puffer zum Speichern von Encoder-Positionsdaten für die Probe (S); und einen Systemspeicher;

wobei die Steuereinheit (873) ferner konfiguriert ist zum:

Initiieren der Aufzeichnung der xy-Encoder-Positionsdaten in den Encoder-Puffer, bevor Intensitätsdaten von einem ersten der Streifen bei einer ersten der Fokuseinstellungen gesammelt werden, während ein Tisch (804) des Inspektionswerkzeugs (850) umkehrt, um sich für die Sammlung der Intensitätsdaten von einem zweiten der Streifen bei der ersten Fokuseinstellung zu konfigurieren, werden die xy-Positionsdaten aus dem Encoder-Puffer in den Systemspeicher kopiert, um einen Zugriff für die Verwendung im Ausrichtvorgang zu haben und dann wird die Aufzeichnung der xy-Positionsdaten des Encoders in den Encoder-Puffer initiiert, bevor die Intensitätsdaten von dem zweiten Streifen gesammelt werden;

während der Tisch (804) des Inspektionswerkzeugs (850) umkehrt, um sich für die Sammlung der Intensitätsdaten von dem ersten Streifen bei einer zweiten der Fokuseinstellungen nach dem Sammeln der Intensitätsdaten des zweiten Streifens zu konfigurieren, werden die xy-Positionsdaten aus dem Enco-

der-Puffer in den Systemspeicher kopiert, um einen Zugriff für die Verwendung im Ausrichtvorgang zu haben und dann wird die Aufzeichnung der xy-Positionsdaten des Encoders in den Encoder-Puffer initiiert, bevor die Intensitätsdaten des ersten Streifens bei der zweiten Fokuseinstellung gesammelt werden; und

Wiederholen der Operationen des Initiierens der Aufzeichnung und des Kopierens für aufeinanderfolgende Paare von Streifen bei jeder der Fokuseinstellungen.

19. Das Inspektionssystem (850) nach Anspruch 14, wobei das Analysieren durchgeführt wird mittels: einer Berechnung einer Vielzahl von Differenz-Koeffizienten-Bildebenen, die für jeden Koeffizienten eine Vielzahl von Werten der Differenz-Koeffizienten haben, mittels einer Subtraktion eines jeden Zielsatzes von jedem der Referenzsätze.

20. Das Inspektionssystem (850) nach Anspruch 19, wobei das Analysieren der Differenz-Koeffizienten-Bildebenen erfolgt durch:

Erzeugen einer Vielzahl von Differenzbildern aus Intensitätsdatensätzen, die an jeder Fokusebene gesammelt wurden;

Kombinieren der Differenzbilder, um ein fusioniertes Bild über den Fokus zu bilden; und

Analysieren des fusionierten Bildes für die Defekterkennung.

21. Das Inspektionssystem (850) nach Anspruch 19, wobei die Differenz-Koeffizienten-Bildebenen dadurch analysiert werden, dass ein Bildpunkt für die Werte der Differenzkoeffizienten an jeder xy-Position der Differenz-Koeffizienten-Bildebenen in ein Streudiagramm geplottet wird, das eine Achse für jeden Koeffizienten besitzt, und dass diese Bildpunkte des Streudiagramms in Clustern von Bildpunkten interessierender Defekte, Bildpunkten von Störungen oder Bildpunkten des Hintergrunds angehäuft werden.

22. Das Inspektionssystem (850) nach Anspruch 19, wobei die Differenz-Koeffizienten-Bildebenen dadurch analysiert werden, dass ein Bildpunkt für die Werte der Differenz-Koeffizienten für jede xy-Position der Differenz-Koeffizienten-Bildebenen auf eine Einheitskugel projiziert wird, und dass die projizierten Bildpunkte in Clustern von Bildpunkten interessierender Defekte, Bildpunkten von Störungen oder Bildpunkten des Hintergrunds angehäuft werden.

23. Das Inspektionssystem (850) nach Anspruch 19, wobei die Differenz-Koeffizienten-Bildebenen durch die Kombination der Differenz-Koeffizienten-Bildebenen in ein einziges Fusionsbild analysiert werden und das Fusionsbild zur Defekterkennung ausgewertet wird.

24. Das Inspektionssystem (850) nach Anspruch 19, wobei die Steuereinheit (873) ferner konfiguriert ist zum:

Gruppieren der Koeffizienten mit ähnlichen Werten, um eine Vielzahl von verschiedenen Segmenten zu bilden, wobei jedes Segment einem unterschiedlichen Abschnitt einer tatsächlichen Bauelementstruktur entspricht;

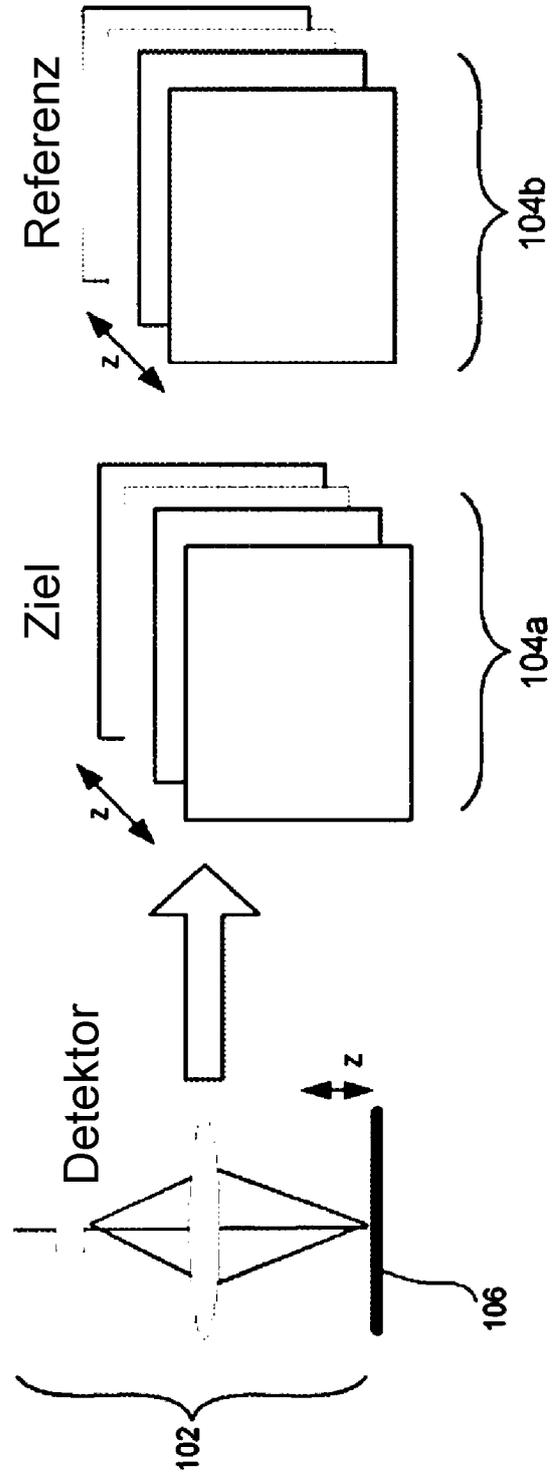
Analysieren der verschiedenen Segmente mit unterschiedlicher Stringenz, um Defekte zu detektieren, basierend darauf, welcher Typ eines tatsächlichen Bauelements den verschiedenen Segmenten entspricht.

25. Das Inspektionssystem (850) nach Anspruch 14, wobei die Steuereinheit (873) ferner konfiguriert ist zum:

Analysieren einer Differenz zwischen Intensitätsänderungen als einer Funktion der Änderungen der Fokuseinstellung für jede xy-Position in einem zweiten Ziel-Satz von Intensitätsdatensätzen und einem zweiten Referenz-Satz von Intensitätsdatensätzen.

Es folgen 12 Seiten Zeichnungen

Anhängende Zeichnungen



Figur 1A

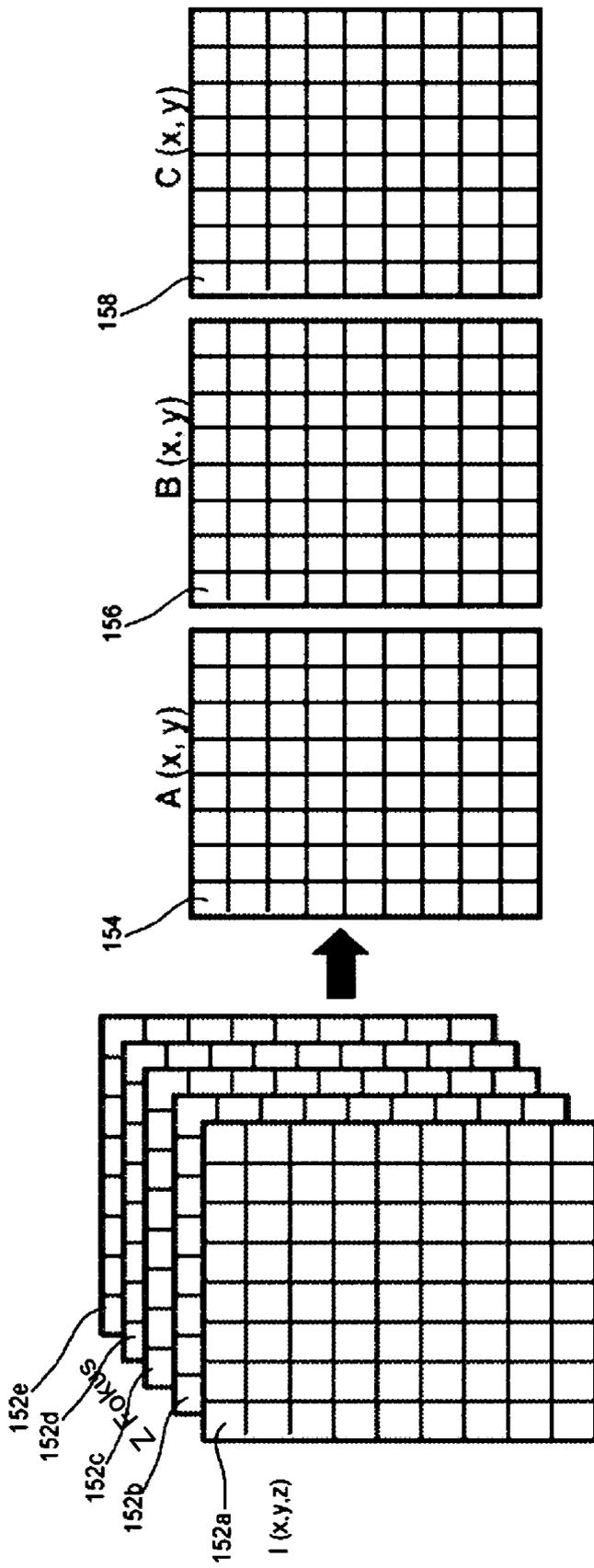
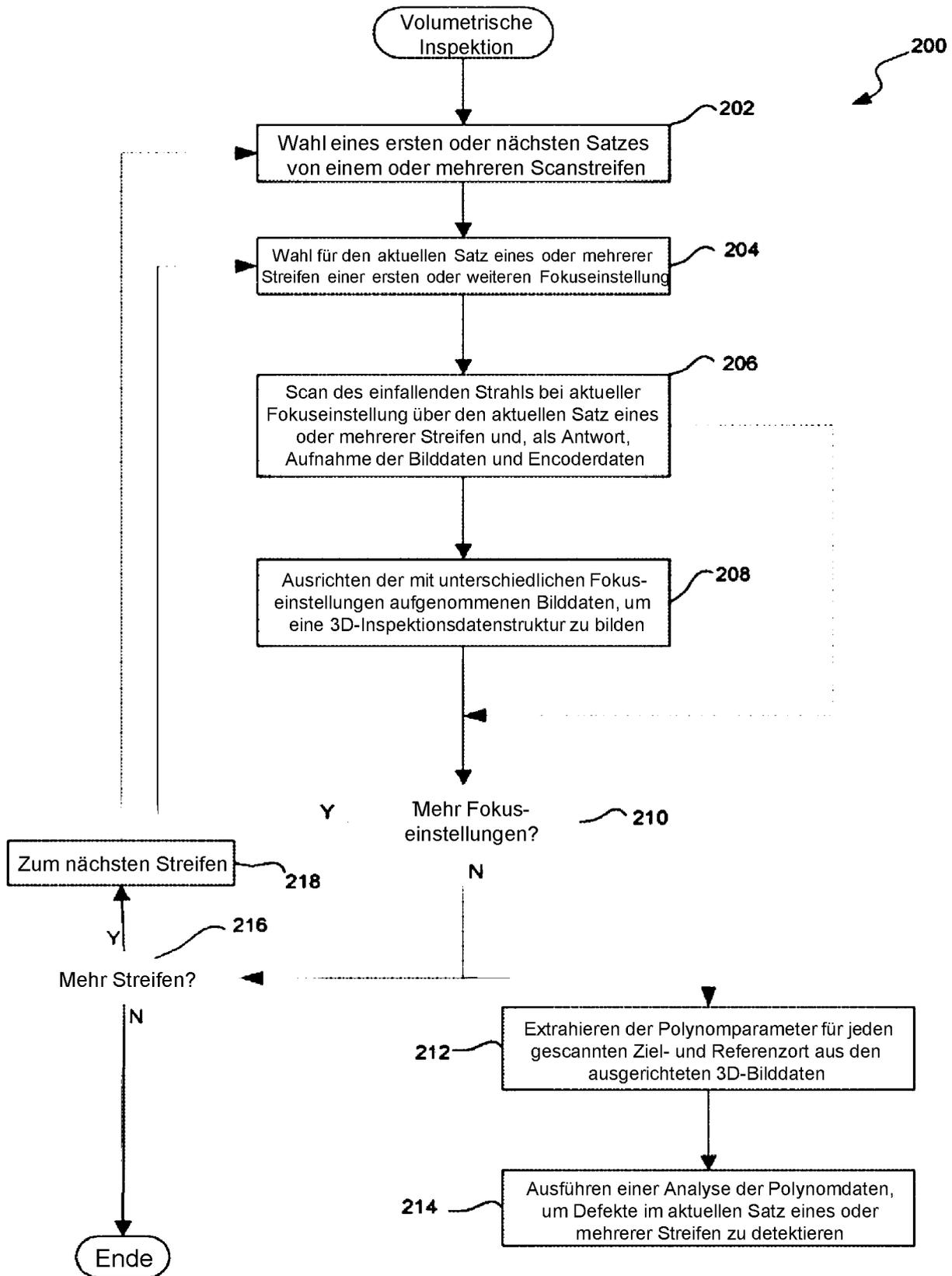
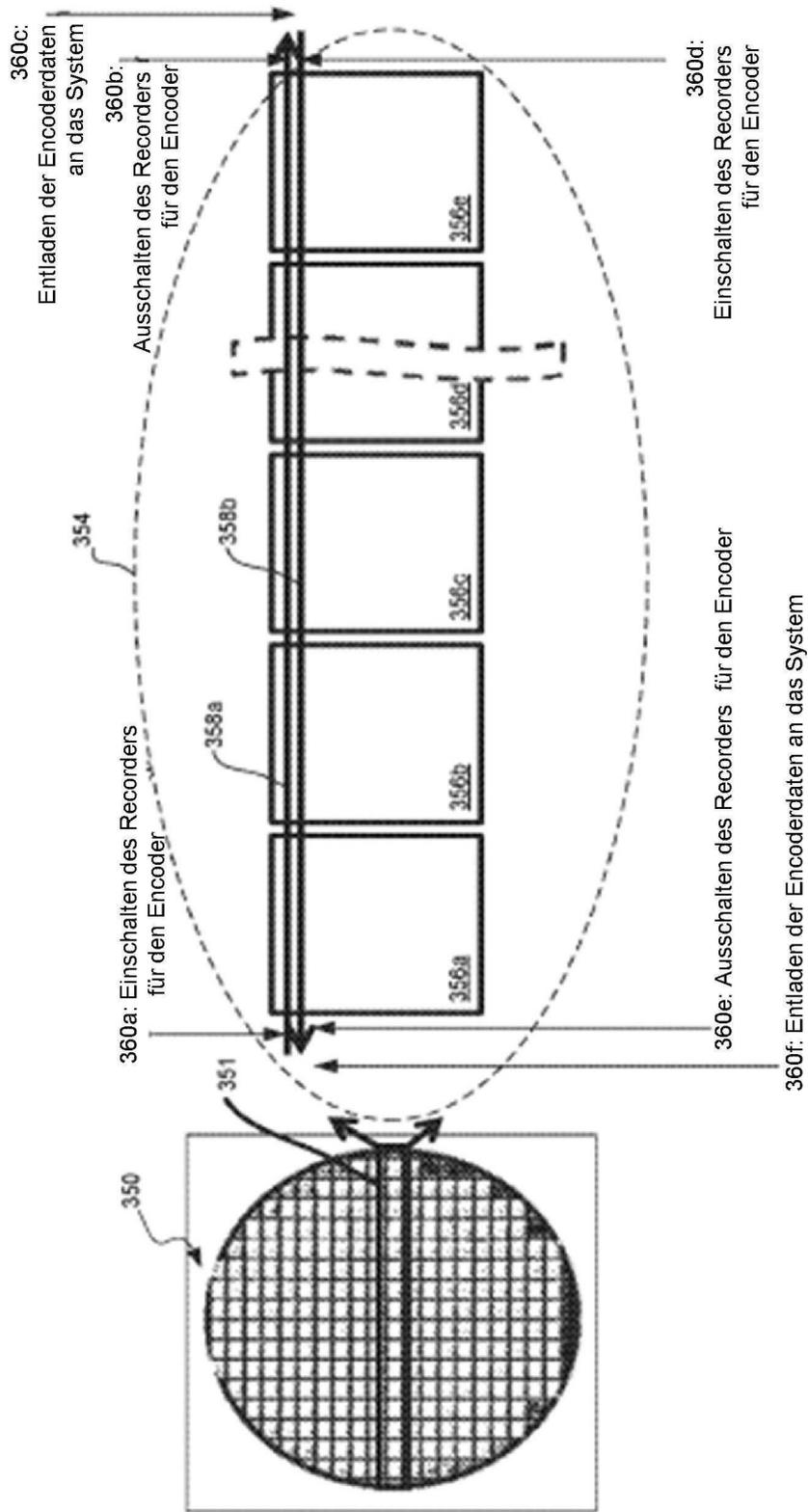


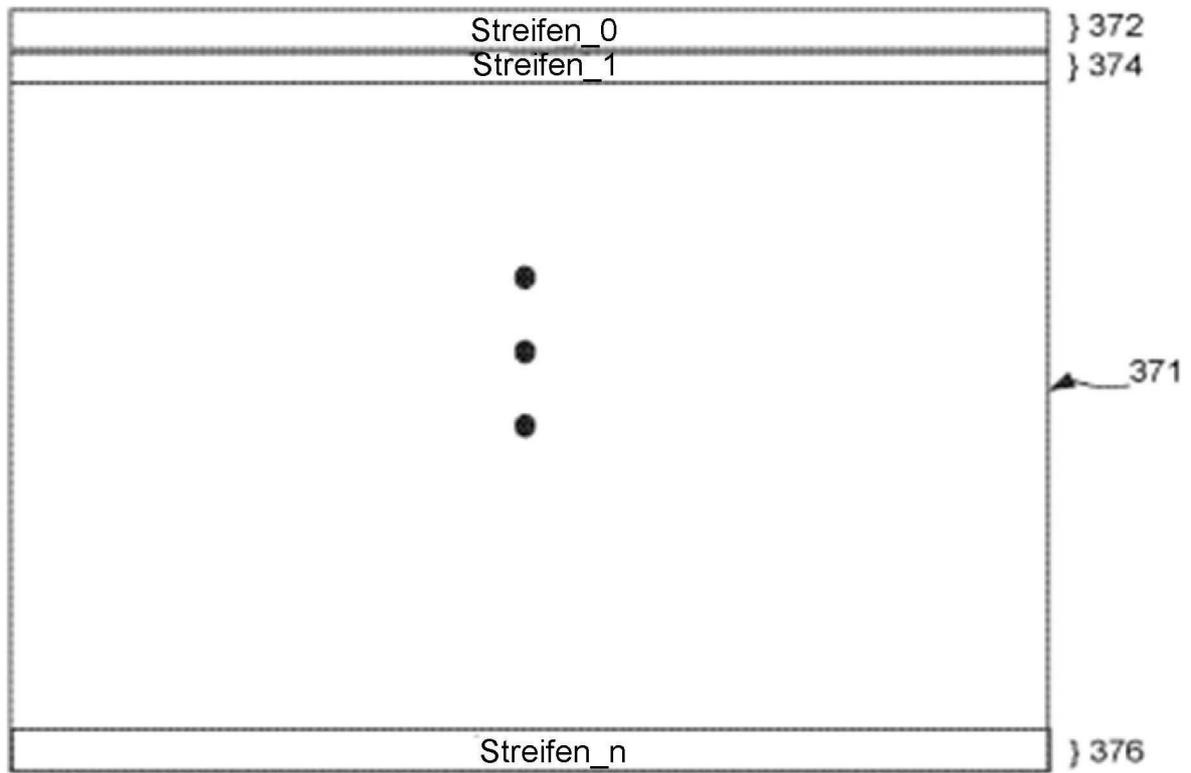
Figure 1B



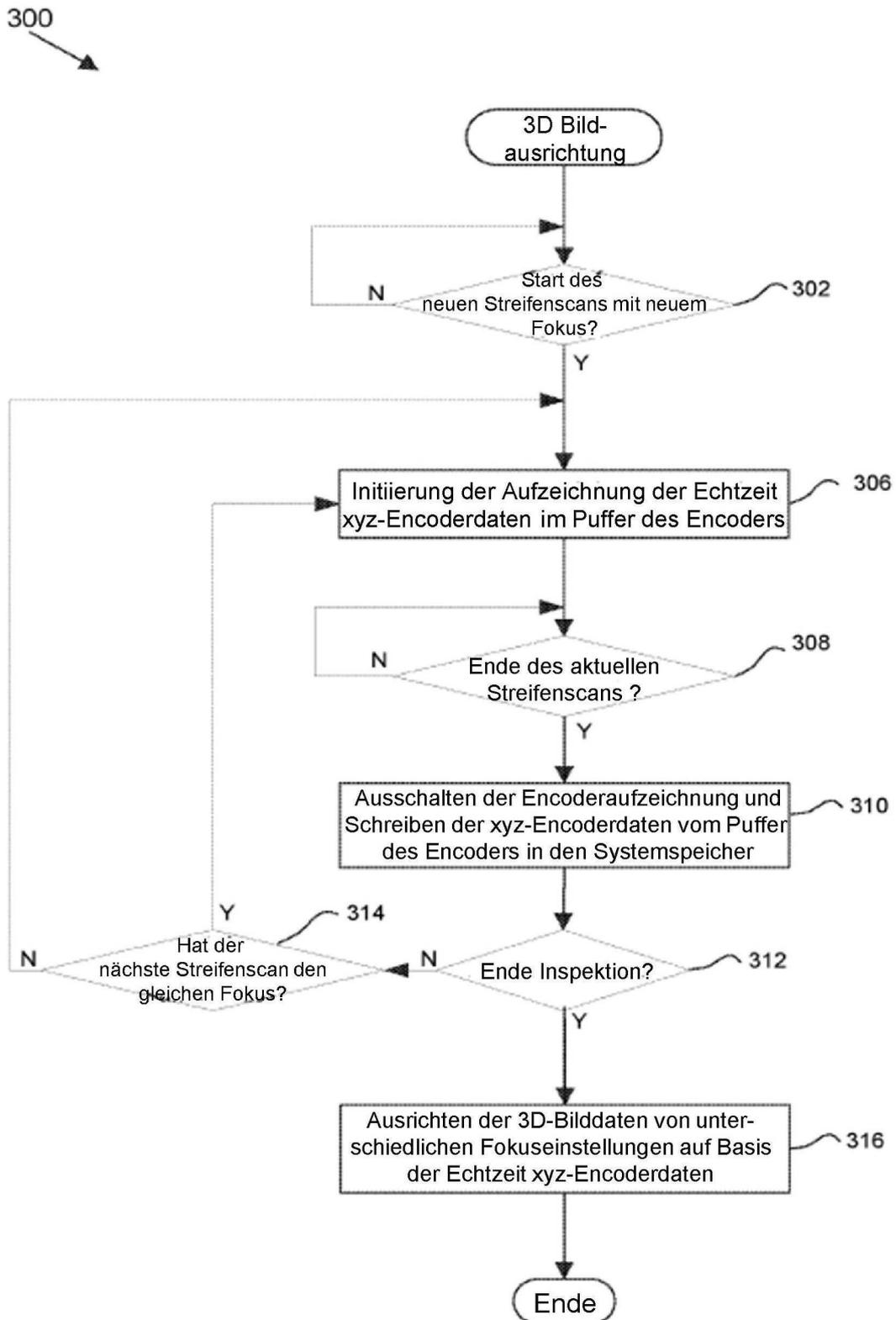
Figur 2



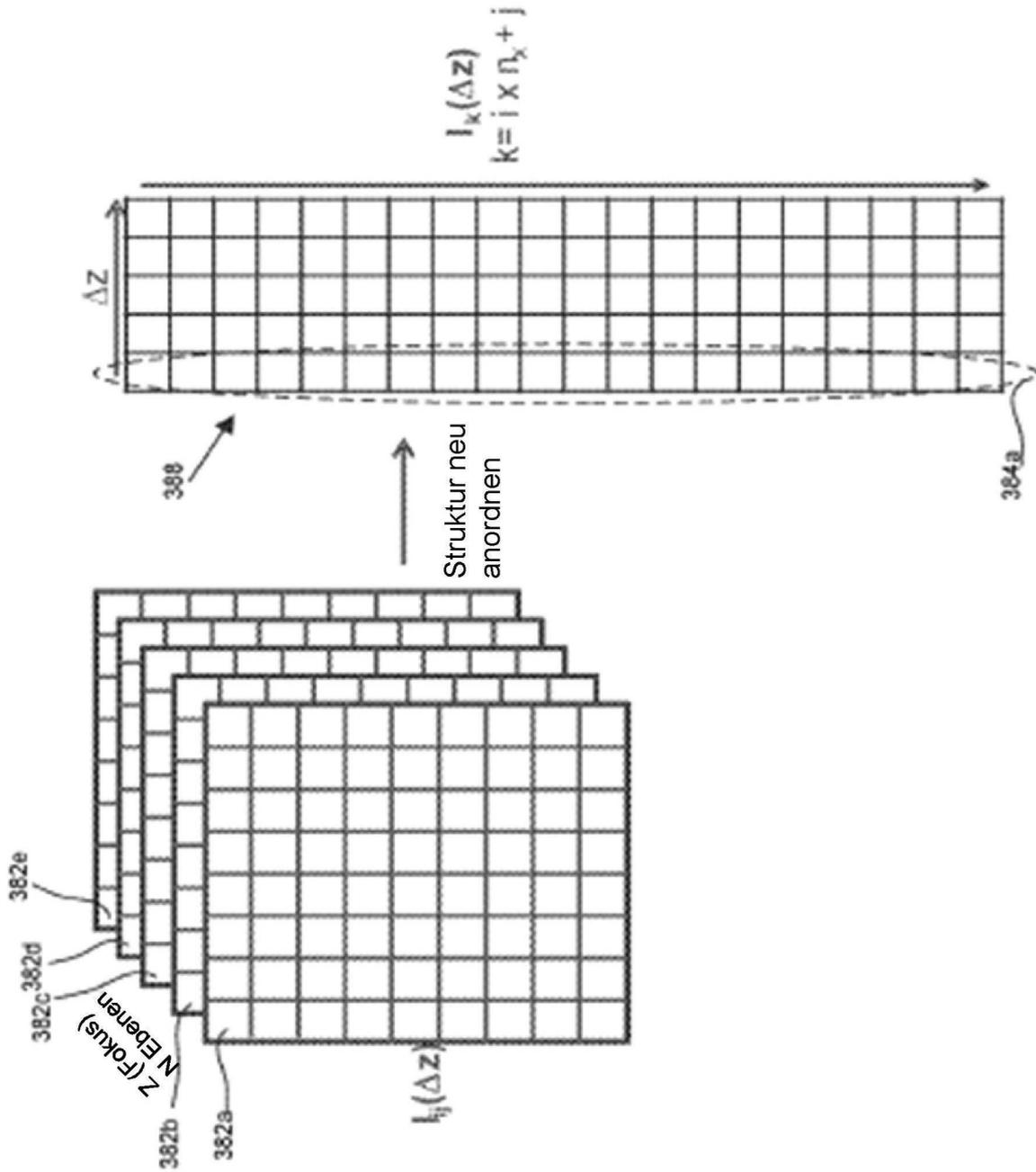
Figur 3A



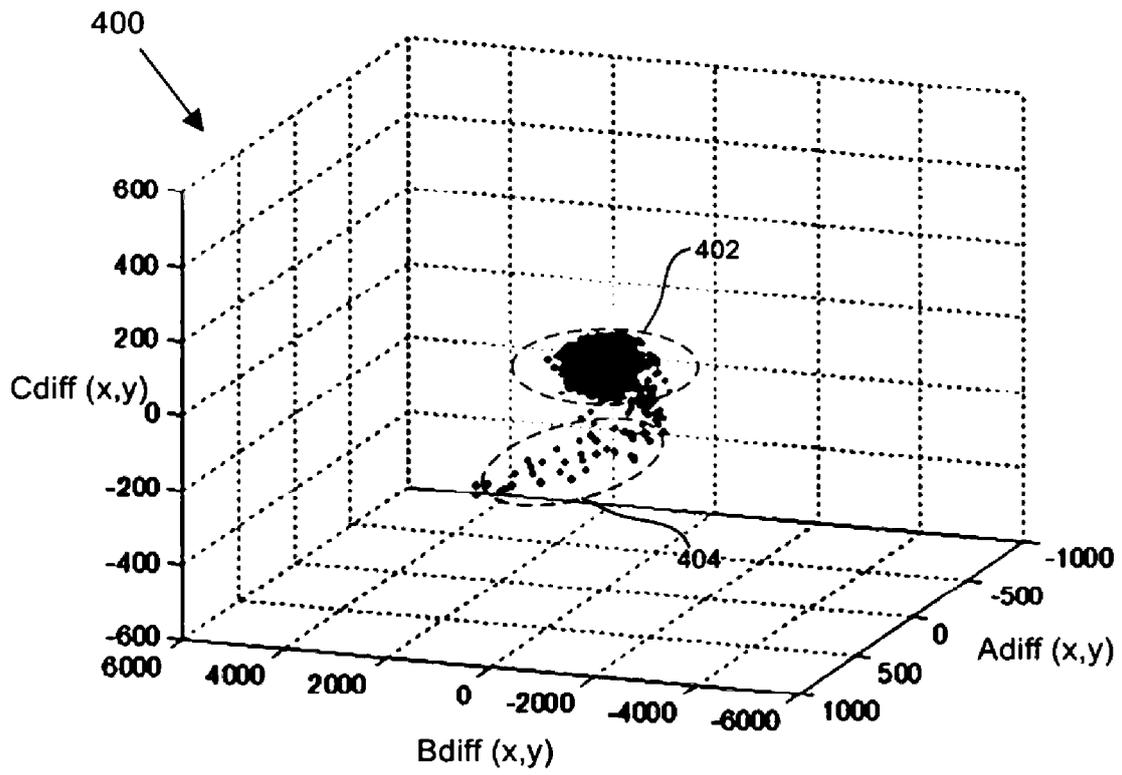
Figur 3B



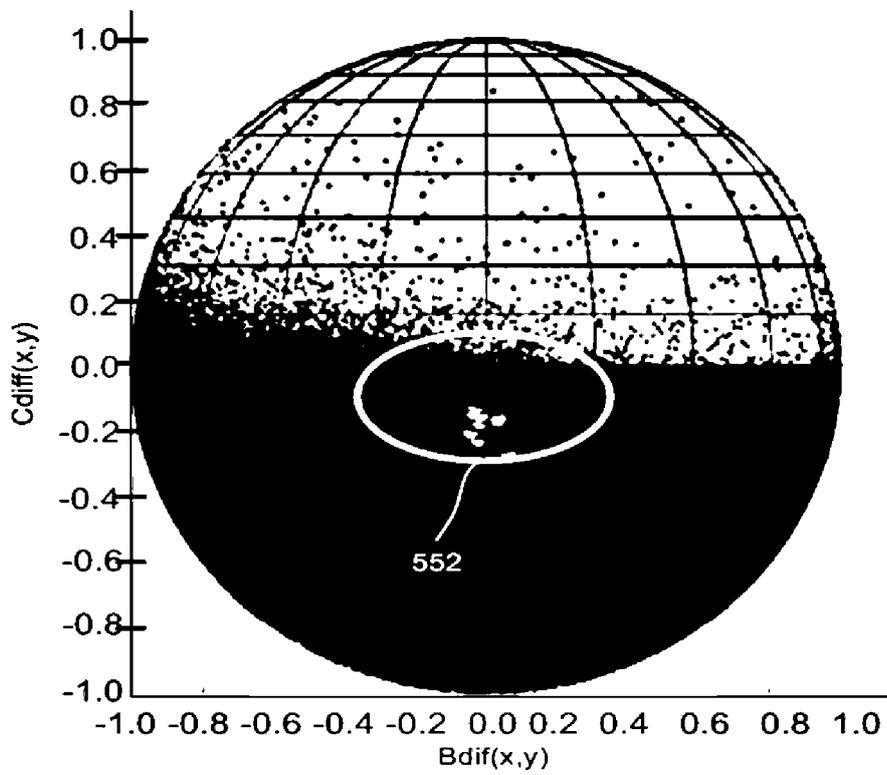
Figur 3C



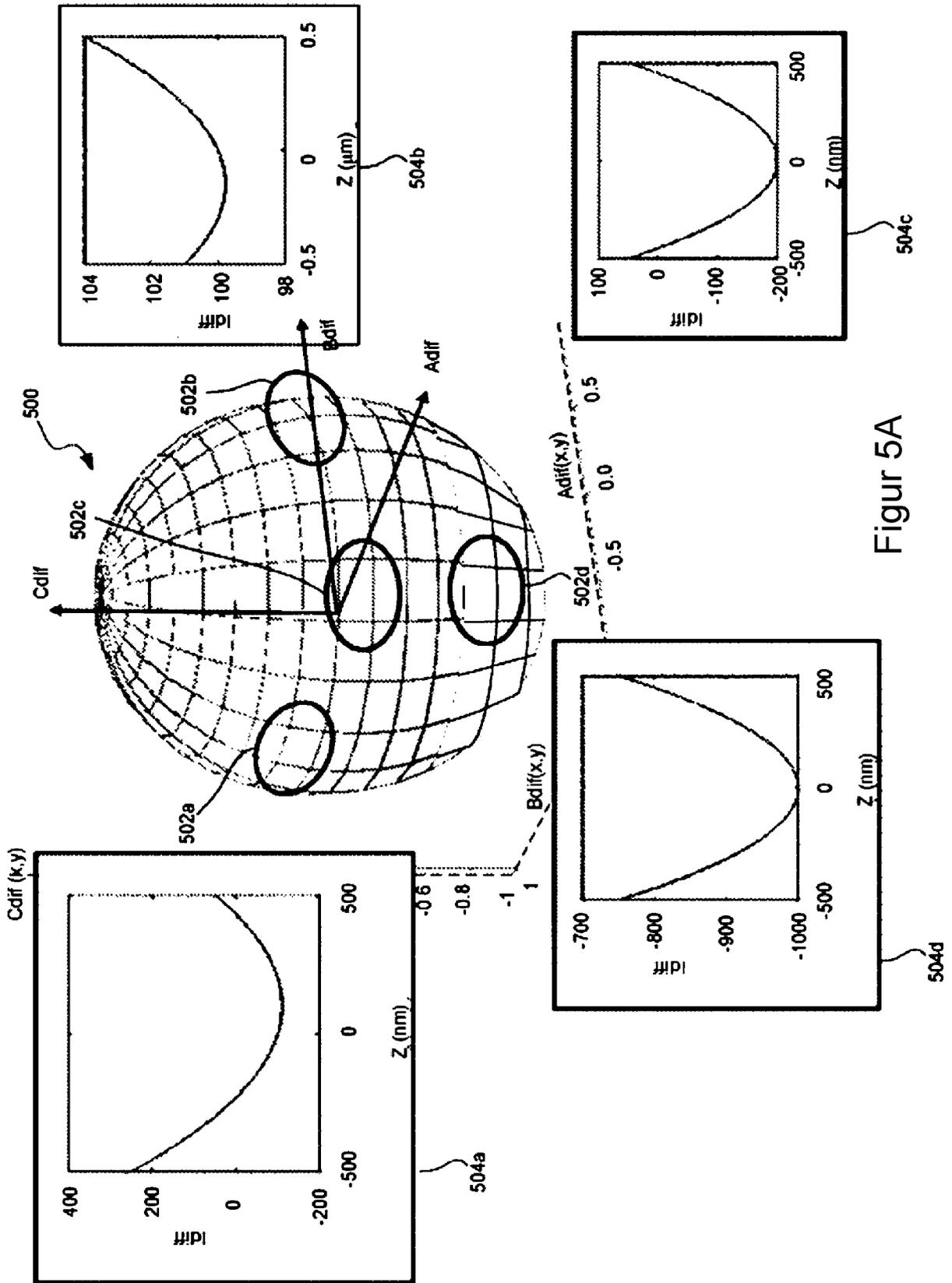
Figur 3D



Figur 4



Figur 5B



Figur 5A

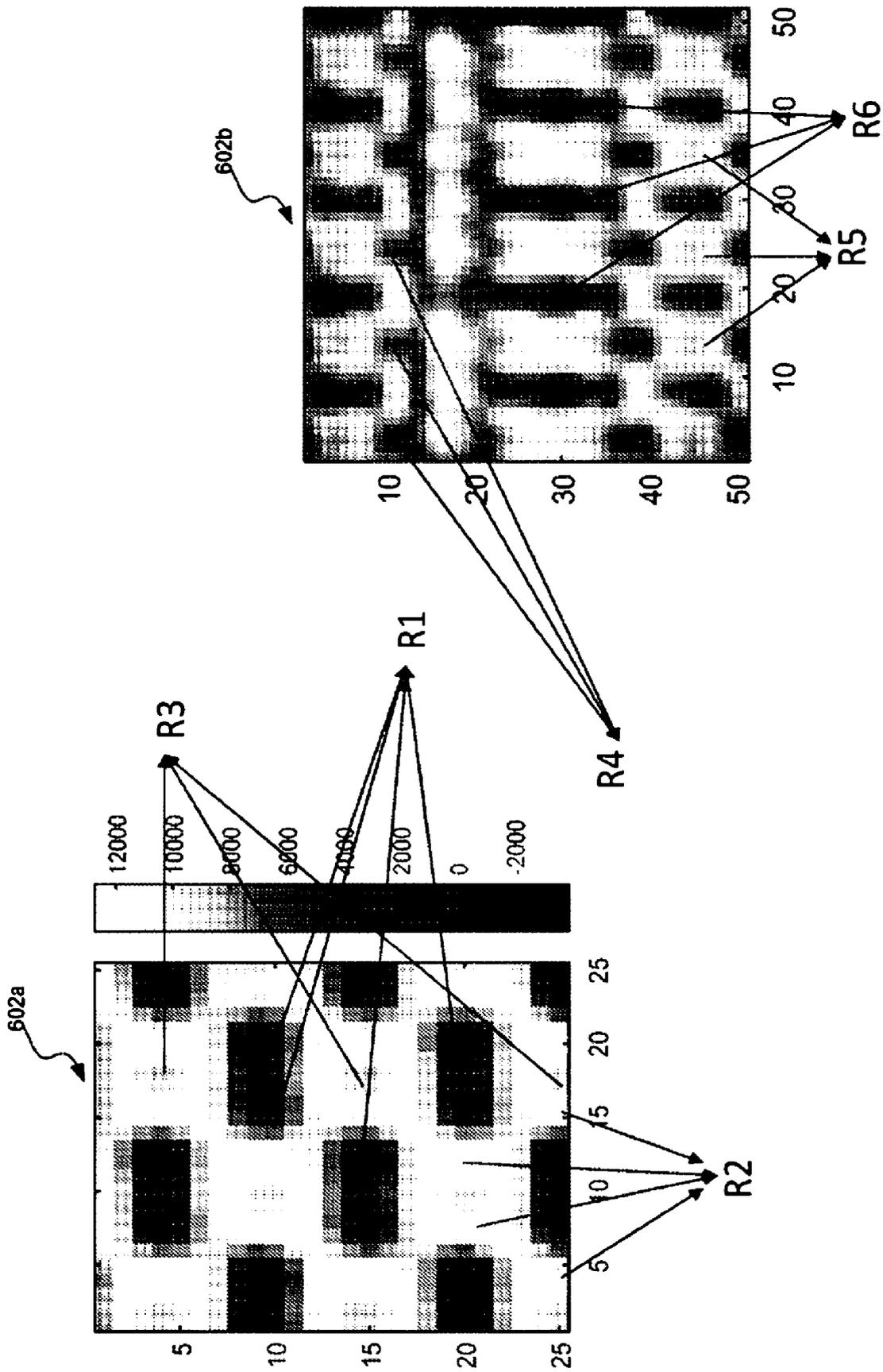
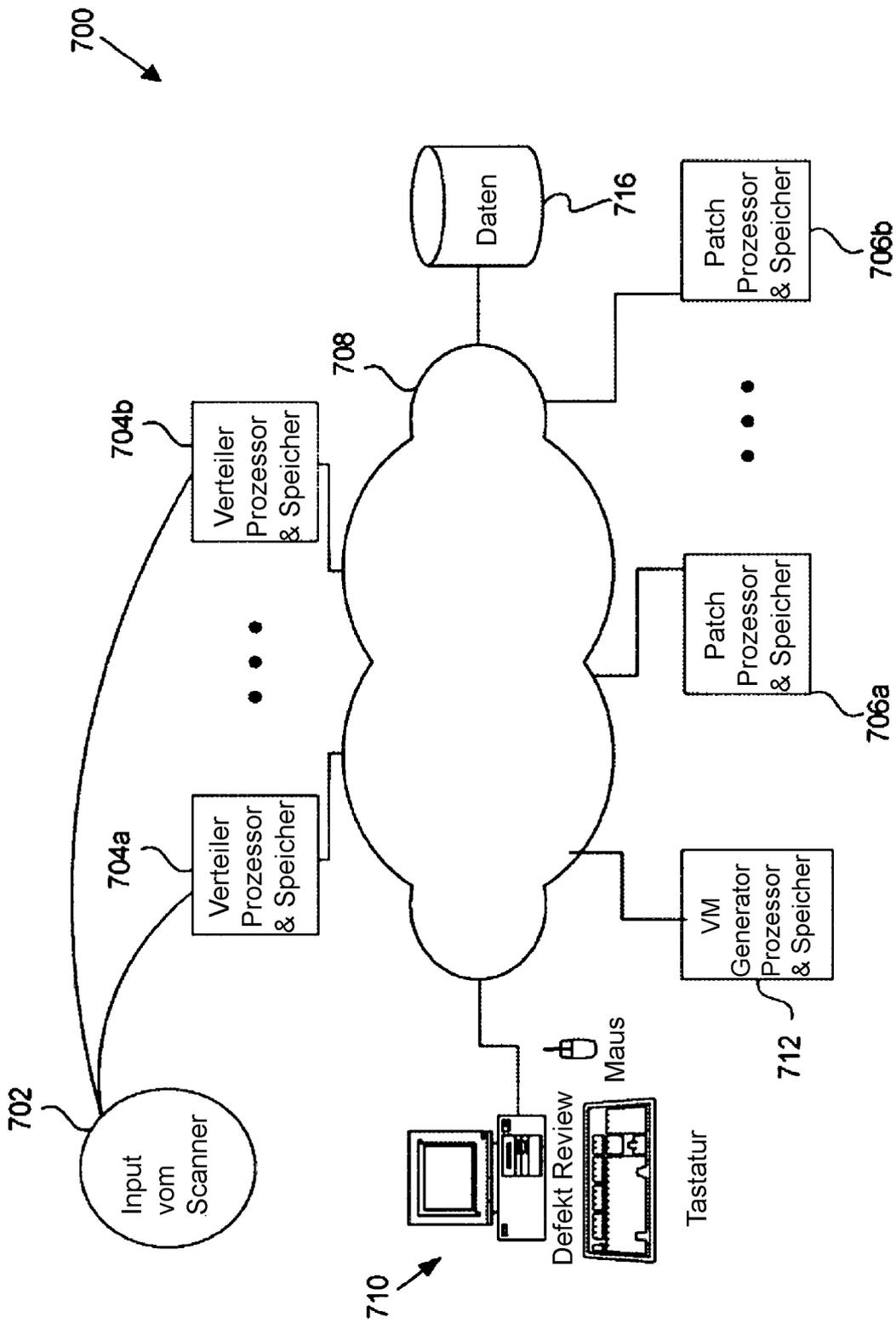
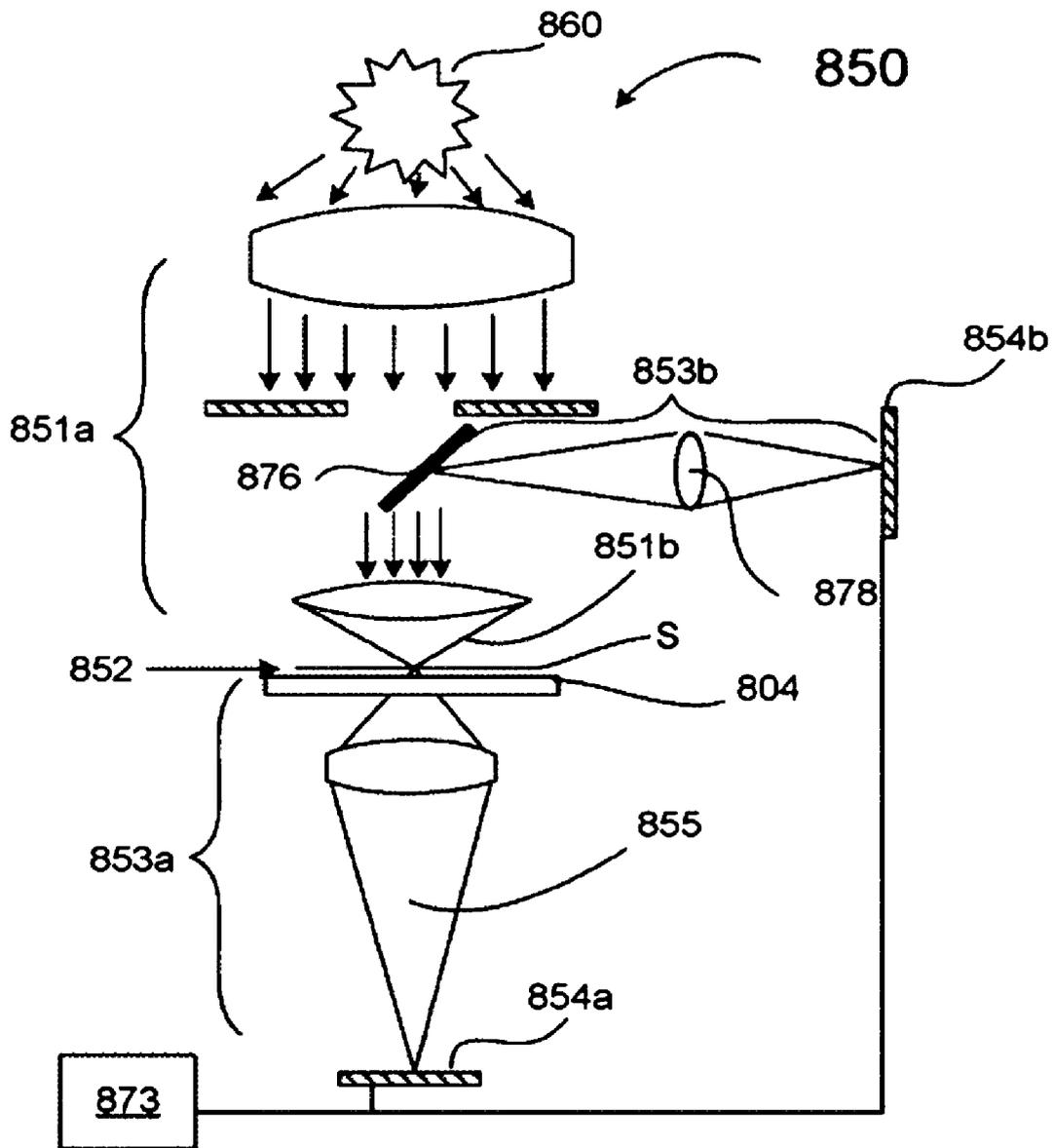


Figure 6



Figur 7



Figur 8