



(19)
Bundesrepublik Deutschland
Deutsches Patent- und Markenamt

(10) **DE 696 32 490 T2** 2005.05.12

(12)

Übersetzung der europäischen Patentschrift

(97) **EP 1 108 501 B1**

(21) Deutsches Aktenzeichen: **696 32 490.3**

(96) Europäisches Aktenzeichen: **01 100 501.4**

(96) Europäischer Anmeldetag: **28.03.1996**

(97) Erstveröffentlichung durch das EPA: **20.06.2001**

(97) Veröffentlichungstag

der Patenterteilung beim EPA: **12.05.2004**

(47) Veröffentlichungstag im Patentblatt: **12.05.2005**

(51) Int Cl.⁷: **B24B 37/04**

B24B 47/12, B24D 7/12, G01B 11/06

(30) Unionspriorität:

413982	28.03.1995	US
605769	22.02.1996	US

(84) Benannte Vertragsstaaten:

DE, FR, GB, IT, NL

(73) Patentinhaber:

Applied Materials, Inc., Santa Clara, Calif., US

(72) Erfinder:

Birang, Manoocher, Los Gatos, California 95030, US; Johansson, Nils, Los Gatos, California 95032, US; Gleason, Allan, San Jose, California 95112, US; Pyatigorsky, Grigory, Santa Clara, California 95051, US

(74) Vertreter:

v. Fünér Ebbinghaus Finck Hano, 81541 München

(54) Bezeichnung: **Verfahren und Vorrichtung zur In-Situ-Kontrolle und Bestimmung des Endes von chemisch-mechanischen Planiervorgängen**

Anmerkung: Innerhalb von neun Monaten nach der Bekanntmachung des Hinweises auf die Erteilung des europäischen Patents kann jedermann beim Europäischen Patentamt gegen das erteilte europäische Patent Einspruch einlegen. Der Einspruch ist schriftlich einzureichen und zu begründen. Er gilt erst als eingelegt, wenn die Einspruchsgebühr entrichtet worden ist (Art. 99 (1) Europäisches Patentübereinkommen).

Die Übersetzung ist gemäß Artikel II § 3 Abs. 1 IntPatÜG 1991 vom Patentinhaber eingereicht worden. Sie wurde vom Deutschen Patent- und Markenamt inhaltlich nicht geprüft.

Beschreibung

[0001] Die Erfindung bezieht sich auf die Halbleiterherstellung, und insbesondere auf eine Vorrichtung und ein Verfahren zum chemischen mechanischen Polieren (CMP) und eine In-situ-Endpunkterfassung während des CMP-Prozesses.

[0002] In dem Prozess zur Fertigung moderner integrierter Halbleiterschaltungen (ICs) ist es erforderlich, verschiedene Materialschichten und Strukturen über vorher ausgebildeten Schichten und Strukturen zu bilden. Die bisherigen Formationen belassen jedoch häufig die Topographie der oberen Oberfläche eines in Behandlung befindlichen Wafers stark unregelmäßig mit Höckern, Bereichen ungleicher Höhe, Mulden, Rinnen und/oder anderen Oberflächenunregelmäßigkeiten. Diese Unregelmäßigkeiten können Probleme verursachen, wenn die nächste Schicht gebildet wird. Wenn beispielsweise ein photolithographisches Muster mit kleinen Geometrien über vorher ausgebildeten Schichten gedruckt wird, ist eine sehr flache Fokustiefe erforderlich. Deshalb ist es wichtig, dass man eine ebene und planare Oberfläche hat, da sonst einige Teile des Musters im Brennpunkt liegen und andere Teile nicht. Tatsächlich wären Oberflächenvariationen in der Größenordnung von weniger als 1000 Å über eine Form von 25 × 25 mm bevorzugt. Wenn die Unregelmäßigkeiten bei jedem Hauptbehandlungsschritt nicht nivelliert werden, kann zusätzlich die Oberflächentopographie des Wafers noch unregelmäßiger werden, was zu weiteren Problemen führt, wenn sich die Lagen während der weiteren Behandlung aufeinander stapeln. Abhängig von der Art der Form und der Größe der zugehörigen Geometrien können die Oberflächenunregelmäßigkeiten zu einer schlechten Ausbeute und einer ungenügenden Vorrichtungsleistung führen. Es ist deshalb erwünscht, eine Art von Planarisierung oder Nivellierung der IC-Strukturen zu bewirken. Die Fertigungstechniken für ICs mit höchster Dichte machen tatsächlich Gebrauch von einem Verfahren zur Bildung einer planarisierten Waferoberfläche an kritischen Punkten bei dem Herstellungsprozess.

[0003] Ein Verfahren zur Erzielung einer Halbleiterwaferplanarisierung oder Topographiebeseitigung ist das chemische mechanische Polieren (CMP). Zu dem chemischen mechanischen Polierprozess (CMP) gehören das Halten und/oder Drehen des Wafers gegenüber einer drehenden Polierplatte bei einem gesteuerten Druck. Wie in [Fig. 1](#) gezeigt ist, hat eine typische CMP-Vorrichtung **10** einen Polierkopf **12** zum Halten des Halbleiterwafers **14** gegen die Polierplatte **16**. Die Polierplatte **16** ist mit einem Kissen **18** abgedeckt. Dieses Kissen **18** hat gewöhnlich eine Rückseitenschicht **20**, die der Oberfläche der Platte zugewandt ist, und eine Abdeckschicht **22**, die in Verbindung mit einer chemischen Polierschlämme zum Polieren des Wafers **14** verwendet wird. Einige Kissen

haben jedoch nur eine Abdeckschicht und keine Rückseitenschicht. Die Abdeckschicht **22** ist gewöhnlich entweder offenzelliges geschäumtes Polyurethan (beispielsweise Rodel IC1000) oder eine Polyurethanbahn mit einer genuteten Oberfläche (beispielsweise Rodel EX2000). Das Kissenmaterial ist mit der chemischen Polierschlämme benetzt, die sowohl ein Schleifmittel als auch Chemikalien enthält. Eine typische chemische Schlämme enthält KOH (Kaliumhydroxid) und Teilchen aus Quarzstaub. Die Platte wird gewöhnlich um ihre zentrale Achse **24** gedreht. Zusätzlich wird gewöhnlich der Polierkopf um seine zentrale Achse **26** gedreht und quer über die Oberfläche der Platte **16** über einen Translationsarm **28** bewegt. Obwohl in [Fig. 1](#) nur ein Polierkopf gezeigt ist, haben CMP-Vorrichtungen gewöhnlich mehr als einen dieser Köpfe, die im Abstand am Umfang um die Polierplatte herum angeordnet sind.

[0004] Ein spezielles Problem, das während eines CMP-Prozesses auftritt, ist die Bestimmung, dass ein Teil auf eine gewünschte Ebenheit oder relative Dicke planarisiert worden ist. Insgesamt muss gemessen werden, wann die gewünschten Oberflächeneigenschaften oder wann der gewünschte planare Zustand erreicht ist. Dies hat man auf einer Vielzahl von Wegen erreicht. Früher war es nicht möglich, die Eigenschaften des Wafers während des CMP-Prozesses zu überwachen. Gewöhnlich wurde der Wafer aus der CMP-Vorrichtung entfernt und an einer anderen Stelle geprüft. Wenn der Wafer den gewünschten Spezifizierungen nicht genügte, musste er erneut in die CMP-Vorrichtung geladen und erneut behandelt werden. Dies war ein zeitraubender und arbeitsintensiver Prozess. Alternativ könnte die Prüfung gezeigt haben, dass eine zu große Materialmenge entfernt worden ist, wodurch das Teil unbrauchbar wäre. Man wollte deshalb eine Vorrichtung haben, die in situ während des CMP-Prozesses erfassen kann, wann die gewünschten Oberflächeneigenschaften oder die Dicke erreicht worden ist.

[0005] Man hat mehrere Vorrichtungen und Verfahren für die In-situ-Erfassung von Endpunkten während des CMP-Prozesses entwickelt. Beispielsweise hat man Vorrichtungen und Verfahren unter Zuordnung der Verwendung von Ultraschallwellen, der Erfassung von Änderungen des mechanischen Widerstands, der elektrischen Impedanz oder der Waferoberflächentemperatur eingesetzt. Diese Vorrichtungen und Verfahren beruhen auf der Bestimmung der Dicke des Wafers oder einer Schicht von ihm und der Erstellung eines Prozessendpunkts durch Überwachen der Änderung in der Dicke. In dem Fall, in welchem die Oberflächenschicht des Wafers dünner gemacht werden soll, wird die Änderung der Dicke zur Bestimmung verwendet, wann die Oberflächenschicht die gewünschte Tiefe hat. Im Falle einer Planarisierung eines strukturierten Wafers mit einer unregelmäßigen Oberfläche wird der Endpunkt dadurch

bestimmt, dass die Änderung der Dicke überwacht wird und die annähernde Tiefe der Oberflächenunregelmäßigkeiten bekannt ist. Wenn die Änderung der Dicke gleich der Tiefe der Unregelmäßigkeiten ist, ist der CMP-Prozess beendet. Obwohl diese Vorrichtungen und Verfahren bei den für sie vorgesehenen Einsätzen vernünftig arbeiten, besteht ein Bedürfnis für Systeme, die eine genauere Bestimmung des Endpunktes bereitstellen. Das Dokument EP-A-0 663 265 A1, veröffentlicht am 19.07.95 und somit einen Stand der Technik unter Art. 54(3) EPÜ darstellend, offenbart ein In-situ-Endpunkt-Erfassungsverfahren, bei dem ein Fenster aus einem transparenten Material innerhalb der Platte vorgesehen wird, die das Kissen hält.

[0006] Die vorliegende Erfindung ist auf einen Endpunktdetektor und auf Verfahren für seine Verwendung gerichtet, um eine verbesserte Genauigkeit und weitere nützliche Informationen über den Polierprozess zu erhalten. Die Vorrichtung und das Verfahren nach der vorliegenden Erfindung verwenden Interferometertechniken zur In-situ-Bestimmung der Dicke des entfernten Materials oder der Ebenheit einer Wafereoberfläche während des CMP-Prozesses.

[0007] Erfindungsgemäß wird ein Polierkissen bereitgestellt, wie es im Anspruch 1 definiert ist. Anspruch 11 definiert die Verwendung des Kissens. Das Polierkissen hat vorzugsweise eine rückseitige Schicht, die mit der Platte verknüpft ist, und eine abdeckende Schicht, die mit einer chemischen Schläme bedeckt ist und mit dem Wafer zusammenwirkt. Der Wafer ist aus einem Halbleitersubstrat aufgebaut, dem eine Oxidschicht zugrunde liegt. Der Endpunktdetektor hat ein Laserinterferometer, das in der Lage ist, einen Laserstrahl zu erzeugen, der zu dem Wafer gerichtet wird, und das davon reflektierte Licht zu erfassen, sowie ein Fenster, das angrenzend an ein Loch angeordnet ist, das durch die Platte gehend ausgebildet ist. Dieses Fenster bildet einen Durchlass für den Laserstrahl, um auf dem Wafer aufzutreffen, wenigstens während der Zeit, während der der Wafer über dem Fenster liegt.

[0008] Das Fenster kann mehrere Formen aufweisen. So kann das Fenster die Form eines Teils des Polierkissens haben, von dem die benachbarte rückseitige Schicht entfernt worden ist. Dies ist möglich, weil die abdeckende Schicht aus Polyurethan für den Laserstrahl wenigstens teilweise durchlässig ist. Schließlich kann das Fenster die Form eines Stopfens aufweisen, der in der abdeckenden Schicht des Kissens ausgebildet ist und keine rückseitige Schicht hat. Dieser Stopfen ist vorzugsweise aus einem Polyurethanmaterial hergestellt, das für den Laserstrahl in hohem Maße durchlässig ist.

[0009] Bei einer Ausführungsform der vorliegenden Erfindung haben das durch die Platte gehende Loch

und das Fenster eine Kreisform. Bei einer anderen Ausgestaltung sind das Loch und das Fenster bogenförmig. Das bogenförmige Fenster hat einen Radius mit einem Ursprung, der mit der Drehmitte der Platte zusammenfällt. Einige Ausgestaltungen der Erfindung haben auch einen Laserstrahl, dessen Strahldurchmesser am Punkt seines Auftreffens auf den Wafer beträchtlich größer als der kleinstmögliche Durchmesser für die verwendete Wellenlänge ist.

[0010] Die CMP-Vorrichtung kann auch einen Lagesensor aufweisen, der feststellt, wann sich das Fenster in der Nähe des Wafers befindet. Dies gewährleistet, dass der von dem Laserinterferometer erzeugte Laserstrahl ungestört durch das Fenster hindurchgehen und auf den Wafer auftreffen kann. Bei einer bevorzugten Ausführungsform der Erfindung hat der Sensor einen Abdeckschirm, der längs eines Teils des Umfangs der Platte festgelegt ist, der sich radial davon nach außen erstreckt. Zusätzlich gibt es einen optischen Unterbrechungssensor, der an dem Rahmen am Umfang der Platte angebracht ist. Dieser Sensor kann einen optischen Strahl produzieren, der die Erzeugung eines Signals verursacht, solange der optische Strahl von dem Abdeckschirm unterbrochen ist. Der Abdeckschirm ist deshalb an dem Umfang der Platte in einer solchen Position befestigt, dass der optische Strahl an dem Abdeckschirm immer dann unterbrochen wird, wenn der Laserstrahl ungehindert durch das Fenster hindurchgehen und auf den Wafer treffen kann.

[0011] Das Laserinterferometer hat eine Vorrichtung zur Erzeugung eines Messsignals immer dann, wenn von dem Wafer reflektiertes Licht erfasst wird. Der Positionssensor hat ein Element zur Abgabe eines Erfassungssignals immer dann, wenn sich das Fenster angrenzend an den Wafer befindet. Dies ermöglicht es einer Messwerterfassungsvorrichtung, das Messsignal aus dem Laserinterferometer während der Dauer des Erfassungssignals aus dem Positionssensor abzutasten. Die Messwerterfassungsvorrichtung benutzt dann ein Element zur Abgabe eines Datensignals, welches das abgetastete Messsignal darstellt. Diese Messwerterfassungsvorrichtung kann auch ein Element zum Integrieren des abgetasteten Messsignals von dem Laserinterferometer über einen vorgegebenen Zeitraum aufweisen, so dass das Ausgangssignal ein Datensignal ist, welches die integrierten Abtastwerte des Messsignals darstellt. Wenn die vorstehend erwähnte Abtastwertperiode während nur einer Drehung der Platte nicht erreicht werden kann, kann ein anderes Verfahren einer stückweisen Messwerterfassung verwendet werden. Insbesondere kann die Messwerterfassungsvorrichtung Elemente zur Durchführung des Verfahrens der Abtastung des Messsignalausgangs von dem Laserinterferometer während jeder vollständigen Umdrehung der Platte für eine Abtastzeit aufweisen, wobei jeder Abtastwert des Messsignals über der Abtastzeit

integriert wird, um einen dem Abtastwert entsprechenden integrierten Wert zu erzeugen, und jeder integrierte Wert gespeichert wird. Die Messwerterfassungsvorrichtung verwendet dann andere Elemente zum Berechnen einer kumulativen Abtastzeit nach jeder vollständigen Umdrehung der Platte (wobei die kumulative Abtastzeit die Summe der Abtastzeiten ist, die jedem Abtastwert des Messsignals zugeordnet ist), zum Vergleichen der kumulativen Abtastzeit mit der gewünschten minimalen Abtastzeit und zum Überführen der gespeicherten integrierten Werte von dem Speicherelement zum Element für die Summenberechnung daraus immer dann, wenn die kumulative Abtastzeit zur vorgegebenen minimalen Abtastzeit gleich ist oder diese überschreitet. Dementsprechend ist das erwähnte Ausgangssignal ein Datensignal, welches eine Reihe von von dem Summierungselement gebildeten Summen des integrierten Werts darstellt.

[0012] Der Datensignalausgang aus der Messwerterfassungsvorrichtung ist aufgrund der Interferenz zwischen dem Teil des Laserstrahls, der von der Oberfläche des Oxidschicht des Wafers reflektiert wird, und dem Teil, der von der Oberfläche des darunter liegenden Wafersubstrats reflektiert wird, zyklisch, da die Oxidschicht während des CMP-Prozesses dünner wird. Deshalb kann der Endpunkt eines CMP-Prozesses zum Dünnmachen der Oxidschicht eines blanken Oxidwafers unter Verwendung zusätzlicher Vorrichtungselemente zum Zählen einer Anzahl von Zyklen bestimmt werden, die von dem Datensignal aufgezeigt werden, wobei eine Materialdicke berechnet wird, die während eines Zyklus des Ausgangssignals von der Wellenlänge des Laserstrahls und dem Brechungsindex der Oxidschicht des Wafers entfernt wird, eine gewünschte, von der Oxidschicht zu entfernende Materialdicke mit einer entfernten Dicke verglichen wird, die das Produkt der Anzahl von Zyklen aufgezeigt durch das Datensignal und der während eines Zyklus entfernten Materialdicke aufweist, und das CMP immer dann beendet wird, wenn die entfernte Dicke der gewünschten zu entfernenden Materialdicke gleich ist oder diese überschreitet. Anstelle des Zählens von vollständigen Zyklen kann alternativ ein Teil eines Zyklus gezählt werden. Das Verfahren ist nahezu identisch mit der Ausnahme, dass die Dicke des entfernten Materials für den Teil des Zyklus anstelle für einen gesamten Zyklus bestimmt wird.

[0013] Eine andere Art zur Bestimmung des Endpunkts bei der CMP-Behandlung eines blanken Oxidwafers verwendet Vorrichtungselemente, die die Zeit messen, die für das Datensignal erforderlich ist, um entweder eine vorgeschriebene Anzahl von Zyklen oder einen vorgeschriebenen Teil eines Zyklus zu vollenden, die die Dicke des entfernten Materials während der gemessenen Zeit berechnen, die eine Entfernungsrage bestimmen, indem die Dicke des

entfernten Materials durch die gemessene Zeit geteilt wird, die eine bleibende zu entfernende Dicke feststellen, indem die Dicke des entfernten Materials von der gewünschten Dicke des von der Oxidschicht zu entfernenden Materials subtrahiert wird, die eine CMP-Restzeit bestimmen, indem die verbliebene Entfernungsdicke durch die Entfernungsrage geteilt wird, und die den CMP-Prozess nach dem Ablauf der CMP-Restzeit beenden.

[0014] Zusätzlich kann diese CMP-Restzeit nach jedem Auftreten der zuvor erwähnten Anzahl von Zyklen oder von Teilen davon aktualisiert werden, um Änderungen der Materialentfernungsrage zu kompensieren. In diesem Fall ist das Verfahren nahezu identisch mit der Ausnahme, dass die Bestimmung der Dicke des Materials zuerst das Summieren aller Dicken, die bei früheren Iterationen entfernt wurden, und das Subtrahieren dieser kumulativen Dicke von der gewünschten Dicke umfasst, um die Größe der Restentfernungsdicke zu bestimmen.

[0015] Wenn jedoch der Wafer am Anfang eine unregelmäßige Oberflächentopographie hat und während des CMP-Prozesses eben gemacht werden muss, ist das Datensignal nur zyklisch, nachdem die Waferoberfläche glatt geworden ist. In diesem Fall erhält man einen Endpunkt für den CMP-Prozess, der einer Bestimmung entspricht, dass der Wafer eben gemacht worden ist, indem zusätzliche Vorrichtungselemente zum Erfassen einer zyklischen Änderung des Datensignals verwendet und der CMP immer dann beendet wird, wenn das Messelement die zyklische Änderung erfasst. Das Messelement ist vorzugsweise in der Lage, eine zyklische Änderung des Datensignals innerhalb höchstens eines Zyklus des Beginns dieser Änderung zu erfassen.

[0016] In manchen Fällen möchte man die Filmdicke steuern, die über einer Form auf einem strukturierten Wafer liegt. Diese Filmdicke kann nicht immer durch die vorher erwähnte Planarisierung erreicht werden. Diese Steuerung kann jedoch dadurch erhalten werden, dass das Datensignal gefiltert wird, um alle Frequenzen außer der auszuschließen, die der speziellen Struktur oder einer Gruppe von Strukturen ähnlicher Größe zugeordnet sind, über denen eine spezielle Filmdicke erwünscht ist. Wenn das Signal einmal gefiltert worden ist, kann im Wesentlichen jede der vorher aufgezählten Arten zur Bestimmung des CMP-Endpunkts für einen blanken Oxidwafer auf dem strukturierten Wafer verwendet werden.

[0017] Der Datensignalausgang aus der Messwerterfassungsvorrichtung kann vorteilhafterweise auch für andere Dinge als die Bestimmung des Endpunkts eines fortlaufenden CMP-Prozesses als für ein In-situ-Verfahren zum Messen der Gleichförmigkeit einer Schicht auf einem Substrat während des Polierens

der Schicht verwendet werden. Zu dem Verfahren gehören die Schritte, einen Lichtstrahl auf die Schicht während des Polierens zu richten, ein Interferenzsignal zu überwachen, das von dem Lichtstrahl erzeugt wird, der von dem Substrat wegreflektiert, und das Berechnen eines Gleichförmigkeitsmaßes aus dem Interferenzsignal.

[0018] Zu dem Berechnungsschritt kann die Entnahme von Eigenschaftsinformationen aus dem Interferenzsignal und das Berechnen des Gleichförmigkeitsmaßes aus den entnommenen Eigenschaftsinformationen gehören. Das Verfahren kann auch den Schritt aufweisen, das Gleichförmigkeitsmaß mit einer Referenz zu vergleichen und einen Alarm zu erzeugen, wenn das Gleichförmigkeitsmaß von der Referenz um mehr als eine vorgegebene Größe abweicht. Das Interferenzsignal hat eine Niederfrequenzkomponente, und zu dem Entnahmeschritt gehört das Messen einer ersten Charakteristik der Niederfrequenzkomponente sowie das Ableiten der entnommenen Informationen aus der ersten Charakteristik. Tatsächlich hat das Interferometersignal auch eine Hochfrequenzkomponente, und zu dem Entnahmeschritt gehört ferner das Messen einer zweiten Charakteristik des Hochfrequenzsignals sowie das Ableiten der entnommenen Informationen aus der ersten und zweiten Charakteristik. Insbesondere sind die erste und die zweite Charakteristik Amplituden der Hochfrequenz- bzw. Niederfrequenzsignale, und zu dem Ableitschritt gehört das Berechnen eines Verhältnisses der Amplituden der Hochfrequenz- und Niederfrequenzsignale.

[0019] Zusätzlich zu den gerade beschriebenen Vorteilen werden Ziele und Vorteile der vorliegenden Erfindung aus der folgenden, ins Einzelne gehenden Beschreibung in Verbindung mit den beiliegenden Zeichnungsfiguren ersichtlich.

[0020] Die speziellen Merkmale, Aspekte und Vorteile der vorliegenden Erfindung lassen sich besser unter Bezug auf die folgende Beschreibung, die beiliegenden Ansprüche und die zugehörigen Zeichnungen verstehen.

[0021] [Fig. 1](#) ist eine Seitenansicht einer für den Stand der Technik typischen Vorrichtung zum chemischen mechanischen Polieren (CMP).

[0022] [Fig. 2](#) ist eine Seitenansicht einer Vorrichtung zum chemischen mechanischen Polieren mit einer Endpunkterfassung, die mit einigen Aspekten nach der vorliegenden Erfindung, jedoch ohne massives Fensterelement gebaut ist.

[0023] [Fig. 3A](#) ist nicht Teil der Erfindung, sondern wird zur Erläuterung gezeigt.

[0024] [Fig. 3B](#) und [Fig. 3C](#) sind vereinfachte

Schnittansichten von entsprechenden Ausgestaltungen des Fensterteils der Vorrichtung von [Fig. 2](#).

[0025] [Fig. 4](#) ist eine vereinfachte Schnittansicht eines Fensterteils der Vorrichtung von [Fig. 2](#) und zeigt Komponenten eines Laserinterferometers zur Erzeugung eines Laserstrahls und der Erfassung eines reflektierten Interferenzstrahls, wobei diese Ausgestaltung nicht Teil der Erfindung ist.

[0026] [Fig. 5](#) ist eine vereinfachte Schnittansicht eines blanken Oxidwafers, der mit der Vorrichtung von [Fig. 2](#) zu behandeln ist, und zeigt schematisch den auf den Wafer auftreffenden Laserstrahl und Reflexionsstrahlen, die einen sich ergebenden Interferenzstrahl bilden.

[0027] [Fig. 6](#) ist eine vereinfachte Draufsicht auf die Platte der Vorrichtung von [Fig. 2](#) und zeigt eine mögliche relative Anordnung zwischen dem Fenster und der Sensorabschirmung und dem Sensor und dem Laserinterferometer.

[0028] [Fig. 7](#) ist eine Draufsicht auf die Platte der Vorrichtung von [Fig. 2](#) und zeigt eine relative Anordnung zwischen dem Fenster und der Sensorabschirmung sowie dem Sensor und dem Laser, wobei das Fenster die Form eines Bogens hat.

[0029] [Fig. 8](#) ist ein Ablaufdiagramm für ein Verfahren einer stückweisen Messwerterfassung nach der vorliegenden Erfindung.

[0030] [Fig. 9A](#) und [Fig. 9B](#) sind Diagramme, die die zyklische Änderung des Datensignals aus dem Laserinterferometer über der Zeit während des Dünnermachens eines blanken Oxidwafers zeigen, wobei das Diagramm von [Fig. 9A](#) die integrierten Werte des Datensignals integriert über einer gewünschten Abtastzeit und das Diagramm von [Fig. 9B](#) eine gefilterte Version der integrierten Werte zeigen.

[0031] [Fig. 10A](#) ist ein Ablaufdiagramm eines rückwärts schauenden Verfahrens zur Bestimmung des Endpunkts eines CMP-Prozesses zum Dünnermachen der Oxidschicht eines blanken Oxidwafers nach der vorliegenden Erfindung.

[0032] [Fig. 10B](#) ist ein Ablaufdiagramm eines vorwärts schauenden Verfahrens zur Bestimmung des Endpunkts eines CMP-Prozesses zum Dünnermachen der Oxidschicht eines blanken Oxidwafers nach der vorliegenden Erfindung.

[0033] [Fig. 11A](#) bis [Fig. 11C](#) sind vereinfachte Schnittansichten eines strukturierten Wafers mit einer unregelmäßigen, mit der Vorrichtung von [Fig. 2](#) zu bearbeitenden Oberfläche, wobei

[0034] [Fig. 11A](#) den Wafer am Anfang des

CMP-Prozesses, [Fig. 11B](#) den Wafer etwa in der Mitte des Prozesses und [Fig. 11C](#) den Wafer nahe am Ende seiner Planarisierung zeigen.

[0035] [Fig. 12](#) ist ein Ablaufdiagramm eines Verfahrens zur Bestimmung des Endpunkts eines CMP-Prozesses zum Planarisieren eines strukturierten Wafers mit einer unregelmäßigen Oberfläche nach der vorliegenden Erfindung.

[0036] [Fig. 13](#) ist ein Diagramm, das die Änderung des Datensignals von dem Laserinterferometer über der Zeit während der Planarisierung eines strukturierten Wafers zeigt.

[0037] [Fig. 14](#) ist ein Ablaufdiagramm eines erfindungsgemäßen Verfahrens zur Bestimmung des Endpunktes eines CMP-Prozesses, um die Filmdicke einzustellen, die über einem speziell bemessenen Aufbau oder einer Gruppe von ähnlich bemessenen Aufbauten liegt.

[0038] [Fig. 15A](#) ist eine vereinfachte Schnittansicht eines Wafers mit einer Oberflächenungenauigkeit, die mit einem Laserstrahl mit schmalen Durchmesser beleuchtet wird.

[0039] [Fig. 15B](#) ist eine vereinfachte Schnittansicht eines Wafers mit einer Oberflächenungenauigkeit, die mit einem Laserstrahl mit breitem Durchmesser beleuchtet wird.

[0040] [Fig. 16](#) ist ein Diagramm, das die zyklische Änderung des Datensignals aus dem Laserinterferometer über der Zeit während des Dünnmachens eines blanken Oxidwafers zeigt und das Hochfrequenzsignal aufweist, das einer nicht gleichförmigen Waferoberfläche zugeordnet ist.

[0041] [Fig. 17](#) ist eine schematische Darstellung eines CMP-Systems mit einem Interferometer und einem Rechner, der zum Analysieren und Ansprechen auf das Ausgangssignal der Interferometerwellenform programmiert ist.

[0042] [Fig. 18](#) ist ein Ablaufdiagramm, das die Funktionalität zeigt, die dem Rechner zur Ausführung der In-situ-Überwachung der Gleichförmigkeit eingegeben ist.

[0043] [Fig. 19A](#) bis [Fig. 19C](#) zeigen Beispiele eines Interferometersignals, nachdem es von einem Niederfrequenz-Bandpassfilter bzw. von einem Hochfrequenz-Bandpassfilter gefiltert worden ist.

[0044] [Fig. 20A](#) und [Fig. 20B](#) sind Ablaufdiagramme, die den Vorgang der Erzeugung und der anschließenden Verwendung einer Signalfolge eines CMP-Systems zeigen, um sie für den Produktionseinsatz zu qualifizieren.

[0045] [Fig. 21A](#) ist eine vereinfachte Schnittansicht einer Ausführungsform des Fensterabschnitts der Vorrichtung von [Fig. 2](#), wobei das Polierkissen als das Fenster verwendet wird und eine Reflexion von der Rückseite des Kissens gezeigt ist.

[0046] [Fig. 21B](#) ist ein Diagramm, das die zyklische Änderung des Datensignals aus dem Laserinterferometer über der Zeit mit einer großen Gleichspannungskomponente zeigt, die durch die Reflexion von der Rückseite des Kissens der Ausführungsform von [Fig. 21A](#) verursacht wird.

[0047] [Fig. 21C](#) ist eine vereinfachte Schnittansicht einer Ausführungsform des Fensterabschnitts der Vorrichtung von [Fig. 2](#), wobei das Polierkissen als Fenster mit einer diffus gemachten rückseitigen Oberfläche verwendet wird, um Reflexionen zu unterdrücken.

[0048] [Fig. 21D](#) ist ein Diagramm, das die zyklische Änderung des Datensignals aus dem Laserinterferometer über der Zeit ohne die große Gleichspannungskomponente zeigt, die durch die Reflexion von der Rückseite des Kissens infolge der diffusen rückseitigen Oberfläche der Ausführungsform von [Fig. 21C](#) verursacht wird.

[0049] [Fig. 2](#) zeigt einen Teil einer CMP-Vorrichtung, die entsprechend einer Ausgestaltung der vorliegenden Erfindung teilweise modifiziert ist. In der Platte **16** und dem darüber liegenden Plattenkissen **18** ist ein Loch **30** ausgebildet. Dieses Loch **30** ist so positioniert, dass es einen von dem Polierkopf **12** gehaltenen Wafer **14** während eines Teils der Plattendrehung unabhängig von der Translationsposition des Kopfes **12** sehen kann. Unter der Platte **16** ist ein Laserinterferometer **32** in einer Position festgelegt, die es ermöglicht, dass ein von dem Laserinterferometer **32** projizierter Laserstrahl **34** durch das Loch **30** in der Platte **16** hindurchgeht und auf die Oberfläche des darüber liegenden Wafers **14** während einer Zeit trifft, wenn sich das Loch **30** nahe an dem Wafer **14** befindet.

[0050] In [Fig. 3A](#) bis [Fig. 3C](#) ist eine ins Einzelne gehende Ansicht des Plattenlochs **30** und des Wafers **14** (zu einer Zeit, wenn er über dem Plattenloch **30** liegt) gezeigt. Wie in [Fig. 3A](#), die nicht Teil der Erfindung ist und nur zu Veranschaulichungszwecken dient, zu sehen ist, hat das Plattenloch **30** einen abgestuften Durchmesser, wodurch eine Schulter **36** gebildet wird. Die Schulter **36** wird benutzt, um einen Quarzeinsatz **38** aufzunehmen und zu halten, der als Fenster für den Laserstrahl **34** dient. Die Zwischenfläche zwischen der Platte **16** und dem Einsatz **38** ist abgedichtet, so dass der Teil der chemischen Schlämme **40**, der seinen Weg zwischen den Wafer **14** und den Einsatz **38** findet, nicht als Leckstrom zum Boden der Platte **16** hindurchgelangen kann.

Der Quarzeinsatz **38** steht über die obere Oberfläche der Platte **16** und teilweise in das Plattenkissen **18** vor. Dieser Vorsprung des Einsatzes **38** soll den Spalt zwischen der oberen Fläche des Einsatzes **38** und der Oberfläche des Wafers **14** minimieren. Durch Minimieren dieses Spaltes wird die Menge der Schlämme **40** auf ein Minimum reduziert, die in dem Spalt eingeschlossen wird. Dies ist vorteilhaft, weil die Schlämme **40** dazu neigt, durch sie hindurchgehen- des Licht zu streuen und somit den Laserstrahl zu dämpfen, der von dem Laserinterferometer **32** emittiert wird. Je dünner die Schicht der Schlämme **40** zwischen dem Einsatz **38** und dem Wafer **14** ist, desto weniger werden der Laserstrahl **34** und das von dem Wafer reflektierte Licht gedämpft. Man glaubt, dass ein Spalt von etwa 1 mm zu akzeptablen Dämpfungswerten während des CMP-Prozesses führen würde. Bevorzugt wird jedoch, den Spalt noch kleiner zu machen. Der Spalt sollte so klein wie möglich sein, während noch gewährleistet ist, dass der Einsatz **38** den Wafer **14** zu irgendeiner Zeit während des CMP-Prozesses berührt. In einer geprüften Version wurde der Spalt zwischen dem Einsatz **38** und dem Wafer **14** auf 10 mils (250 µm) mit zufrieden stellenden Ergebnissen eingestellt.

[0051] Dieser Aufbau entspricht nicht der Erfindung.

[0052] [Fig. 3B](#) zeigt eine andere Ausgestaltung der Platte **16** und des Kissens **18** gemäß der vorliegenden Erfindung. Bei dieser Ausgestaltung ist der Quarzeinsatz weggelassen, und es gibt in dem Kissen **18** kein durchgehendes Loch. Stattdessen ist die rückseitige Schicht **20** (wenn sie vorhanden ist) des Kissens **18** in dem Bereich entfernt worden, der über dem Loch **30** in der Platte **16** liegt. Dies belässt nur die abdeckende Schicht **22** des Kissens **18** aus Polyurethan zwischen dem Wafer **14** und dem Boden der Platte **16**. Es hat sich gezeigt, dass das bei der abdeckenden Schicht **22** verwendete Polyurethanmaterial den Laserstrahl **34** von dem Laserinterferometer **32** im Wesentlichen durchlässt. Somit wirkt der Teil der abdeckenden Schicht **22**, der über dem Plattenloch **30** liegt, als Fenster für den Laserstrahl **34**. Diese Anordnung hat einen beträchtlichen Vorteil. Da das Kissen **18** selbst als Fenster verwendet wird, gibt es zunächst keinen merkbaren Spalt. Deshalb ist nur sehr wenig Schlämme **40** vorhanden, die das schädliche Streuen des Laserstrahls verursacht. Ein weiterer Vorteil dieser Ausgestaltung besteht darin, dass der Kissenverschleiß unbedeutend wird. Bei der baulichen Gestaltung von [Fig. 3A](#) wurde der Spalt zwischen dem Quarzeinsatz **38** und dem Wafer **14** so klein wie möglich gemacht. Wenn jedoch das Kissen **18** verschleißt, neigt dieser Spalt dazu, noch kleiner zu werden. Schließlich kann der Verschleiß so groß werden, dass die obere Oberfläche des Einsatzes **38** den Wafer **14** berührt und beschädigt. Da das Kissen **18** als Fenster bei der Ausgestaltung von [Fig. 3B](#) verwendet wird und so ausgelegt ist, dass es in Kontakt

mit dem Wafer **14** steht, gibt es keine schädlichen Effekte aufgrund des Verschleißes des Kissens **18**. Anzumerken ist, dass Versuche, bei denen sowohl offenzellige als auch genutete Oberflächenarten von Kissens verwendet wurden, gezeigt haben, dass der Laserstrahl bei einem Kissen mit genuteter Oberfläche weniger gedämpft wird. Dementsprechend wird bevorzugt, dass diese Art von Kissens verwendet wird.

[0053] Obwohl das Polyurethanmaterial, das bei der abdeckenden Schicht des Kissens verwendet wird, im Wesentlichen für den Laserstrahl durchlässig ist, enthält es bestimmte Zusatzstoffe, die seine Durchlässigkeit verhindern. Dieses Problem wird bei der Ausgestaltung der Erfindung beseitigt, die in [Fig. 3C](#) gezeigt ist. Bei dieser Ausführung ist das typische Kissensmaterial in dem über dem Plattenloch **30** liegenden Bereich durch einen massiven Polyurethanstopfen **42** ausgetauscht. Dieser Stopfen, der als Fenster für den Laserstrahl wirkt, ist aus einem Polyurethanmaterial hergestellt, das keine Nuten (oder eine offenzellige Struktur) des umgebenden Kissensmaterials hat und frei von Zusatzstoffen ist, die die Durchlässigkeit unterbinden. Dementsprechend ist die Dämpfung des Laserstrahls **34** durch den Stopfen **42** hindurch auf ein Minimum reduziert. Vorzugsweise ist der Stopfen **42** integral in das Kissen **18** eingeformt.

[0054] In Betrieb verwendet eine CMP-Vorrichtung nach der vorliegenden Erfindung den Laserstrahl aus dem Laserinterferometer, um die Materialmenge zu bestimmen, die von der Oberfläche des Wafers entfernt ist, oder um zu bestimmen, wann die Oberfläche eben geworden ist. Der Anfang dieses Prozesses wird unter Bezugnahme auf [Fig. 4](#) erläutert. Als Elemente des Laserinterferometers **42** sind ein Laser und Kollimator **44**, ein Strahlteiler **46** und ein Detektor **48** gezeigt. Dadurch wird die vorstehend erwähnte Erläuterung der Funktion der CMP-Vorrichtung erleichtert. Zusätzlich ist aus Zweckmäßigkeitsgründen der beim Aufbau von [Fig. 3A](#) verwendete Quarzeinsatz **38** als Fenster gezeigt. Natürlich ist die gezeigte Ausgestaltung nur eine mögliche Anordnung, wobei auch andere verwendet werden können. Beispielsweise kann jede der vorstehend erwähnten Fensteranordnungen zum Einsatz kommen, und es sind alternative Ausgestaltungen des Laserinterferometers **32** möglich. Eine andere Interferometeranordnung verwendet einen Laser zur Erzeugung eines Strahls, der auf die Oberfläche des Wafers in einem Winkel einfällt. Bei dieser Ausgestaltung wird ein Detektor an einer Stelle positioniert, wo das von dem Wafer reflektierende Licht auf ihn trifft. Bei dieser anderen Ausgestaltung ist ein Strahlenteiler nicht erforderlich.

[0055] Wie in [Fig. 4](#) gezeigt ist, erzeugen der Laser und Kollimator **44** einen Laserparallelstrahl **34**, der auf den unteren Teil des Strahlenteilers **46** einfällt. Ein Teil des Strahls **34** pflanzt sich durch den Strah-

lenteiler **46** und die Quarzeinlage **38** fort. Wenn dieser Teil des Strahls **34** das obere Ende des Einsatzes **38** verlässt, pflanzt er sich durch die Aufschlammung **40** fort und trifft auf die Oberfläche des Wafers **14**. Der Wafer, der im Einzelnen in [Fig. 5](#) gezeigt ist, hat ein Substrat **50** aus Silicium und eine darüber liegende Oxidschicht **52** (beispielsweise SiO_2).

[0056] Der Teil des Strahls **34**, der auf den Wafer **14** trifft, wird an der Oberfläche der Oxidschicht **52** teilweise reflektiert und bildet einen ersten reflektierten Strahl **54**. Ein Teil des Lichts wird jedoch durch die Oxidschicht **52** durchgelassen und bildet einen Durchgangsstrahl **56**, der auf das darunter liegende Substrat **50** trifft. Wenigstens etwas von dem Licht aus dem durchgelassenen Strahl **56** erreicht das Substrat **50** und wird durch die Oxidschicht **52** zurückreflektiert, um einen zweiten reflektierten Strahl **58** zu bilden. Der erste und der zweite reflektierte Strahl **54**, **58** bilden miteinander eine aufbauende oder zerstörende Interferenz abhängig von ihrer Phasenbeziehung unter Bildung eines daraus resultierenden Strahls **60**, bei dem die Phasenbeziehung hauptsächlich eine Funktion der Dicke der Oxidschicht **52** ist.

[0057] Obwohl der vorstehend beschriebene Aufbau ein Siliciumsubstrat mit einer einzigen Oxidschicht verwendet, ist dem Fachmann bekannt, dass der Interferenzvorgang auch bei anderen Substraten und anderen Oxidschichten eintritt. Der Schlüssel besteht darin, dass die Oxidschicht teilweise reflektiert und teilweise durchlässt und dass das Substrat wenigstens teilweise den auftreffenden Strahl reflektiert. Zusätzlich kann der Interferenzvorgang auf Wafer mit Mehrschichten anwendbar sein, die über dem Substrat liegen. Wenn jede Schicht teilweise reflektierend und teilweise durchlässig ist, wird wieder ein daraus resultierender Interferenzstrahl erzeugt, obwohl er eine Kombination der von allen Schichten und dem Substrat reflektierten Strahlen ist.

[0058] Aus [Fig. 4](#) ist zu sehen, dass der sich ergebende Strahl **60** die Kombination des ersten und zweiten reflektierenden Strahls **54**, **58** ([Fig. 5](#)) darstellt und zurück durch die Schlämme **40** und den Einsatz **38** zum oberen Teil des Strahlenteilers **46** fortschreitet. Der Strahlenteiler **46** fängt einen Teil des resultierenden Strahls **60** zu dem Detektor **48** ab.

[0059] Gewöhnlich dreht sich die Platte **16** während des CMP-Prozesses. Deshalb hat das Plattenloch **30** nur einen Blick auf den Wafer **14** während eines Teils der Drehung. Dementsprechend kann das Messsignal von dem Laserinterferometer **32** nur erhalten werden, wenn der Wafer **14** von dem Laserstrahl **34** getroffen wird. Wichtig ist, dass das Messsignal nicht genommen wird, wenn der Laserstrahl **34** teilweise durch das Loch **34** hindurchgeht, wenn ein Teil durch den Boden der Platte **16** an dem Lochrand blockiert

ist, da dies bei dem Signal zu einem beträchtlichen Rauschen führt. Damit sich dies nicht ereignet, hat man eine Positionssensorvorrichtung vorgesehen. Es kann jeder bekannte Annäherungssensor benutzt werden, beispielsweise ein Hall-Effekt-Wirbelstrom-optischer-Unterbrecher- oder akustischer Sensor, obwohl ein Sensor in der Bauweise eines optischen Unterbrechers bei den untersuchten Ausgestaltungen der Erfindung eingesetzt wurde und bei den folgenden Figuren gezeigt wird. Eine Vorrichtung nach der vorliegenden Erfindung zum Synchronisieren des Laserinterferometers **32** ist in [Fig. 6](#) gezeigt, wobei ein Sensor **62** in der Bauweise eines optischen Unterbrechers (beispielsweise ein LED-/Photodiodenpaar) an einem festen Punkt an dem Rahmen der CMP-Vorrichtung so angebracht ist, dass er eine Sicht auf den Umfangsrand der Platte **16** hat. Diese Art von Sensor **62** wird aktiviert, wenn ein optischer Strahl, den er erzeugt, unterbrochen wird. An dem Umfang der Platte **16** ist eine Lagesensorabschirmung **64** befestigt. Der Befestigungspunkt und die Länge der Abschirmung **64** erfolgen so, dass sie das optische Signal des Sensors nur unterbrechen, wenn der Laserstrahl **34** von dem Laserinterferometer **32** vollständig durch den vorstehend beschriebenen Fensteraufbau **66** hindurchgeht. Wie beispielsweise in [Fig. 6](#) gezeigt ist, kann der Sensor **62** diametral gegenüber dem Laserinterferometer **32** bezogen auf die Mitte der Platte **16** angebracht werden. Die Abschirmung **64** wird an der Platte **16** in einer Position diametral gegenüber dem Fensteraufbau **66** befestigt. Die Länge der Abschirmung **64** wird annähernd von den gestrichelten Linien **68** bestimmt, obwohl die genaue Länge der Abschirmung **64** fein abgestimmt wird, um zu gewährleisten, dass der Laserstrahl von der Platte **16** während der gesamten Zeit vollständig ungestört bleibt, während der die Abschirmung **64** von dem Sensor **62** erfasst wird. Diese Feinabstimmung kompensiert jedes Positionssensorrauschen oder jede Ungenauigkeit, das Ansprechvermögen des Laserinterferometers, usw.. Wenn der Sensor **62** einmal aktiviert ist, wird ein Signal erzeugt, das dazu verwendet wird, zu bestimmen, wann das Detektorsignal von dem Interferometer **32** genommen werden kann.

[0060] Messwerterfassungssysteme, die in der Lage sind, das Positionssensorsignal zu verwenden, um das Laserinterferometersignal während solcher Zeiten zu nehmen, zu denen der Wafer für den Laserstrahl sichtbar ist, sind bekannt und bilden keinen neuen Teil der vorliegenden Erfindung. Deshalb erfolgt hier auch keine ins Einzelne gehende Beschreibung. Bei der Wahl eines geeigneten Systems sind jedoch einige Überlegungen zu berücksichtigen. Beispielsweise ist es bevorzugt, dass das Signal aus dem Interferometer über einen Zeitraum integriert wird. Diese Integration verbessert das Signal-Rausch-Verhältnis durch Mitteln des Hochfrequenzrauschens über dem Integrationszeitraum.

Dieses Rauschen hat verschiedene Ursachen, beispielsweise eine Vibration aus der Rotation der Platte und des Wafers, und aufgrund von Änderungen in der Oberfläche des Wafers in einer ungleichen Planarisierung. Bei der oben beschriebenen Vorrichtung bestimmen der Durchmesser des Quarzfensters und die Drehgeschwindigkeit der Platte, wie lange ein Zeitraum während jeder Drehung der Platte verfügbar ist, um das Signal zu integrieren. Unter bestimmten Umständen kann jedoch diese verfügbare Zeit unzureichend sein. Beispielsweise kann ein akzeptables Signal-Rausch-Verhältnis eine längere Integrationszeit erfordern, oder die Zwischenflächenschaltung, die bei einem gewählten Messwerterfassungssystem verwendet wird, kann eine minimale Integrationszeit erfordern, die diejenige überschreitet, die bei einem Durchgang zur Verfügung steht.

[0061] Eine Lösung für dieses Problem besteht darin, das Plattenloch längs der Drehrichtung der Platte auszudehnen. Mit anderen Worten, der Fensteraufbau **66'** (d. h. der Einsatz, das Kissen oder der Stopfen) würde die Form eines Bogens annehmen, wie es in [Fig. 7](#) gezeigt ist. Natürlich wird die Abschirmung **64'** zur Anpassung an den längeren Fensteraufbau **66'** ausgedehnt. Alternativ kann das Fenster das gleiche bleiben, jedoch wird das Laserinterferometer an der drehenden Platte direkt unter dem Fenster angebracht. In diesem Fall muss die CMP-Vorrichtung für die Aufnahme des Interferometers unter der Platte modifiziert werden, und es muss dafür gesorgt werden, das Detektorsignal von dem Interferometer aus zu leiten. Das Nettoergebnis einer jeden Methode wäre jedoch eine Verlängerung der Messwerterfassungszeit für jede Umdrehung der Platte.

[0062] Obwohl das Verlängern des Plattenlochs und des Fensters von Vorteil ist, wird dadurch der Oberflächenbereich des Plattenkissens etwas verringert. Deshalb wird die Planarisierungsrate in den Bereichen der Scheibe, die über dem Fenster liegen, während eines Teils der Plattendrehung verringert. Außerdem müssen sich die Länge des Plattenlochs und das Fenster nicht über die Ränder des Wafers hinaus erstrecken, und die Datenentnahme muss nicht erfolgen, wenn sich das Fenster jenseits des Randes des Wafers unabhängig von der Wafertranslationsposition befindet. Deshalb ist die Länge des erweiterten Plattenlochs und des Fensters oder die Zeit, während der das an der Platte montierte Interferometer abtasten kann, durch jede Translationsbewegung des Polierkopfes begrenzt.

[0063] Deshalb ist ein stärker bevorzugtes Verfahren zum Erhalten einer ausreichenden Datenerfassungsintegrationszeit das Sammeln der Daten über mehr als eine Umdrehung der Platte. Während des Schritts **102** von [Fig. 8](#) wird das Laserinterferometersignal während der verfügbaren Datenerfassungszeit bei jeder Drehung der Platte abgefragt. Als Nächstes

wird bei den Schritten **104** und **106** jedes abgefragte Signal über der vorstehend erwähnten Datenerfassungszeit integriert, und die integrierten Werte werden gespeichert. Dann wird bei den Schritten **108** und **110** eine kumulative Abfragezeit nach jeder vollständigen Umdrehung der Platte berechnet und mit einer gewünschten minimalen Abfragezeit verglichen. Natürlich ergibt dies nur eine Abfragezeit, wenn nur eine Abfrage erfolgt ist. Wenn die kumulative Abfragezeit zu der gewünschten minimalen Abfragezeit gleich ist oder sie überschreitet, werden die gespeicherten integrierten Werte überführt und summiert, wie im Schritt **112** gezeigt ist. Wenn dies nicht der Fall ist, gehen der Prozess des Abfragens, des Integrierens, des Speicherns, des Berechnens der kumulativen Abfragezeit und des Vergleichens mit der gewünschten minimalen Abfragezeit weiter. In einem abschließenden Schritt **114** werden die summierten integrierten Werte, die jedes Mal erzeugt werden, wenn die gespeicherten integrierten Werte übertragen und summiert werden, als ein Datensignal abgegeben. Das gerade beschriebene Datensammelverfahren kann auf eine Anzahl bekannter Weisen durchgeführt werden, wobei Logikschaltungen oder Software-Algorithmen verwendet werden. Da diese Verfahren bekannt sind, ist jede ins Einzelne gehende Beschreibung überflüssig und wird deshalb weggelassen. Das Verfahren des stückweisen Datensammelns bildet eine Lösung für das Problem, eine gewünschte minimale Abfragezeit unabhängig vom Durchmesser des Fensters oder der Geschwindigkeit der Plattendrehung zu bekommen. Wenn der Prozess mit der Positionssensorvorrichtung verknüpft wird, kann die Plattendrehgeschwindigkeit variiert werden, und man erhält trotzdem zuverlässige Daten. Es würde sich nur die Anzahl der Plattenumdrehungen ändern, die erforderlich ist, um die notwendigen Daten zu erhalten.

[0064] Der vorstehend erwähnte erste und zweite reflektierte Strahl, die den resultierenden Strahl **60** bilden, wie es in [Fig. 4](#) und [Fig. 5](#) gezeigt ist, verursachen eine Interferenz, die am Detektor **48** zu sehen ist. Wenn der erste und der zweite Strahl zueinander in Phase sind, verursachen sie am Detektor **48** ein Maximum. Wenn die Strahlen um 180° außer Phase sind, verursachen sie am Detektor **48** ein Minimum. Jede andere Phasenbeziehung zwischen den reflektierten Strahlen führt zu einem Interferenzsignal zwischen den Maxima und den Minima, wie sie am Detektor **48** zu sehen sind. Das Ergebnis ist ein Signalausgang aus dem Detektor **48**, der sich zyklisch mit der Dicke der Oxidschicht **52** ändert, wenn sie während des CMP-Prozesses verringert wird. Man hat beobachtet, dass der Signalausgang aus dem Detektor **48** sich auf eine sinusförmige Weise ändert, wie es in den Diagrammen **9A** und **9B** gezeigt ist. Das Diagramm von [Fig. 9A](#) zeigt die integrierte Amplitude des Detektorsignals (y-Achse) über der Abtastperiode gegenüber der Zeit (x-Achse). Diese Daten erhält

man durch Überwachen des Laserinterferometer-Ausgangssignals der Vorrichtung von [Fig. 4](#), während der CMP-Vorgang an einem Wafer ausgeführt wird, der eine glatte Oxidschicht hat, die über einem Siliciumsubstrat (d. h. einem blanken Oxidwafer) liegt. Das Diagramm von [Fig. 9B](#) zeigt eine gefilterte Version der Daten aus dem Diagramm von [Fig. 9A](#). Diese gefilterte Version zeigt die zyklische Änderung in dem Interferometer-Ausgangssignal ganz deutlich. Die Periode des Interferenzsignals wird durch die Rate gesteuert, mit der Material von der Oxidschicht während des CMP-Prozesses entfernt wird. Somit bestimmen Faktoren, wie die Abwärtskraft, die auf den Wafer gegen das Plattenkissen aufgebracht wird, und die Relativgeschwindigkeit zwischen der Platte und dem Wafer die Periode. Während jeder Periode des Ausgangssignals, das in den [Fig. 9A](#) und [Fig. 9B](#) aufgetragen ist, wird eine bestimmte Dicke der Oxidschicht entfernt. Die entfernte Dicke ist proportional zur Wellenlänge des Waferstrahls und dem Brechungsindex der Oxidschicht. Insbesondere beträgt die Größe der pro Periode entfernten Dicke etwa $\lambda/2n$, wenn λ die Freiraumwellenlänge des Laserstrahls und n der Brechungsindex der Oxidschicht sind. Dadurch ist es möglich, in situ während des CMP-Prozesses unter Verwendung des in [Fig. 10A](#) gezeigten Verfahrens zu bestimmen, wie viel von der Oxidschicht entfernt ist. Zunächst wird im Schritt **202** die Anzahl der Zyklen gezählt, die von dem Datensignal vorgewiesen werden. Als Nächstes wird im Schritt **204** die Dicke des Materials, das während eines Zyklus des Ausgangssignals entfernt wird, aus der Wellenlänge des Laserstrahls und dem Brechungsindex der Oxidschicht des Wafers berechnet. Dann wird die gewünschte Dicke des von der Oxidschicht zu entfernenden Materials mit der tatsächlichen entfernten Dicke im Schritt **206** verglichen. Die tatsächliche entfernte Dicke entspricht dem Produkt der Anzahl der Zyklen, die von dem Datensignal ausgewiesen werden, und der Dicke des während eines Zyklus entfernten Materials. In dem abschließenden Schritt **208** wird der CMP-Prozess immer dann beendet, wenn die entfernte Dicke der gewünschten Dicke des zu entfernenden Materials gleich ist oder sie überschreitet.

[0065] Alternativ kann weniger als ein ganzer Zyklus zur Bestimmung der Menge des entfernten Materials verwendet werden. Auf diese Weise kann jedes überschüssige Material, das über die gewünschte Größe hinaus entfernt wird, minimiert werden. Wie in den in Klammern gesetzten Teilen des Schritts **202** in [Fig. 10A](#) gezeigt ist, wird die Anzahl des Auftretens eines vorgeschriebenen Teils eines Zyklus jeder Iteration gezählt. Beispielsweise bildet jedes Auftreten eines Maximums (d. h. einer Spitze) und eines Minimums (d. h. einer Vertiefung) oder umgekehrt den vorgeschriebenen Teil des Zyklus. Dieser spezielle Teil des Zyklus ist passend, da Maxima und Minima leicht über bekannte Signalverarbeitungsverfahren

erfassbar sind. Als Nächstes wird im Schritt **204** nach dem Feststellen, wie viel Material während eines Zyklus entfernt worden ist, diese Dicke mit dem Bruchteil eines Zyklus multipliziert, der den vorstehend erwähnten vorgeschriebenen Teil darstellt. Im Falle des Zählens des Auftretens eines Maximums oder Minimums, die eine Hälfte eines Zyklus darstellen, würde beispielsweise die berechnete Dicke eines Zyklus mit einer Hälfte multipliziert, um die Dicke der während des vorgeschriebenen Teils des Zyklus entfernten Oxidschicht zu erhalten. Die restlichen Schritte des Verfahrens bleiben unverändert. Das Nettoergebnis dieser alternativen Methode besteht darin, dass der CMP-Prozess nach dem Auftreten eines Teils des Zyklus beendet werden kann. Dementsprechend ist jegliches entfernte Überschussmaterial in den meisten Fällen kleiner als es es wäre, wenn ein voller Zyklus als Basis zum Bestimmen des entfernten Materials verwendet würde.

[0066] Die gerade beschriebenen Verfahren blicken von dem Ende eines Zyklus oder eines Teils von ihm zurück, um zu bestimmen, ob die gewünschte Materialmenge entfernt worden ist. Wie jedoch vorstehend gefolgert wurde, kann die entfernte Materialmenge die gewünschte Menge überschreiten. In manchen Fällen kann dieses zu große Entfernen von Material nicht akzeptabel sein. In diesen Fällen kann ein alternatives Verfahren verwendet werden, welches nach vorne schaut und vorhersieht, wie viel Material über einen anstehenden Zeitraum zu entfernen sein wird, und das den Vorgang beendet, wenn die vorhergesehene gewünschte Dicke entfernt worden ist. Eine bevorzugte Ausgestaltung dieses alternativen Verfahrens ist in [Fig. 10B](#) gezeigt. Wie zu sehen ist, gehört zu dem ersten Schritt **302** das Messen der Zeit zwischen dem ersten Auftreten eines Maximums und eines Minimums oder umgekehrt bei dem Detektorsignal (obwohl ein gesamter Zyklus oder ein Teil davon verwendet werden könnte). Als Nächstes wird im Schritt **304** die Materialmenge über die vorher beschriebenen Verfahren bestimmt, die während dieses Teils des Zyklus entfernt wird. Dann wird eine Entfernungsrate dadurch berechnet, dass die entfernte Materialmenge durch die gemessene Zeit geteilt wird, wie dies im Schritt **306** gezeigt ist. Dies ergibt die Rate, mit der Material bei dem vorhergehenden Teil des Zyklus entfernt wurde. Im nächsten Schritt **308** wird die Dicke des entfernten Materials, wie sie im Schritt **304** berechnet wurde, von der gewünschten zu entfernenden Dicke abgezogen, um eine zu entfernende Restdicke zu bestimmen. Im Schritt **310** wird dann diese zu entfernende Restdicke durch die Entfernungsrate geteilt, um zu bestimmen, wie viel länger der CMP-Prozess fortgesetzt werden muss, bevor er abgeschlossen wird.

[0067] Der Zeitraum des Detektorsignals, und somit die Entfernungsrate, ändert sich jedoch gewöhnlich mit dem Fortschreiten des CMP-Prozesses. Deshalb

wird das oben beschriebene Verfahren wiederholt, um dies zu kompensieren. D. h. mit anderen Worten, wenn einmal eine Restzeit berechnet worden ist, wird der Prozess für jedes Auftreten eines Maximums und eines Minimums oder umgekehrt wiederholt. Dementsprechend wird die Zeit zwischen dem nächsten auftretenden Maximum und Minimum gemessen, die Dicke des Materials, das während des Teils des Zyklus entfernt wird, der durch dieses Auftreten des Maximums und Minimums (d. h. ein halb) dargestellt wird, durch die gemessene Zeit geteilt und die Entfernungsrates berechnet, wie bei der ersten Iteration des Verfahrens. Bei dem nächsten Schritt **308** jedoch wird, wie in Klammern gezeigt ist, die Gesamtmenge des Materials bestimmt, das während allen vorherigen Iterationen entfernt wurde, ehe es von der gewünschten Dicke abgezogen wird. Der Rest des Verfahrens bleibt der gleiche, nämlich dass die restliche zu entfernende Dicke durch die neu berechnete Entfernungsrates geteilt wird, um die restliche CMP-Prozesszeit zu bestimmen. Auf diese Weise wird die restliche Prozesszeit nach jedem Auftreten des vorgeschriebenen Teils eines Zyklus des Detektorsignals wieder berechnet. Dieser Prozess setzt sich fort, bis die restliche CMP-Prozesszeit abläuft, ehe die nächste Iteration beginnen kann. Wie im Schritt **312** zu sehen ist, ist an diesem Punkt der CMP-Prozess beendet. Gewöhnlich wird die zu entfernende Dicke in dem ersten Halbzyklus des Detektorsignals nicht erreicht, wobei jede Änderung der Entfernungsrates nach der Berechnung für den vorhergehenden Halbzyklus klein ist. Deshalb geht man davon aus, dass dieses vorwärts blickende Verfahren ein sehr genauer Weg ist, um genau die gewünschte Dicke von dem Wafer zu entfernen.

[0068] Während das gerade beschriebene Überwachungsverfahren gut für das Dünnermachen von blanken Oxidwafern mit glatter Oberfläche funktioniert, hat sich gezeigt, dass der Vorgang zum Planarisieren der meisten strukturierten Wafer, bei denen die Oberflächentopographie in hohem Maße unregelmäßig ist, nicht erfolgreich eingesetzt werden kann. Der Grund dafür besteht darin, dass ein typischer strukturierter Wafer Formen enthält, die eine breite Vielfalt von unterschiedlich bemessenen Oberflächenelementen haben. Diese Oberflächenelemente mit unterschiedlicher Größe neigen zu einem Polieren mit unterschiedlichen Raten. Beispielsweise neigt ein kleineres Oberflächenelement, das sich relativ weit von den anderen Elementen befindet, dazu, dass es schneller als andere größere Elemente verkleinert werden kann. [Fig. 11A](#) bis [Fig. 11C](#) zeigen einen Satz von Oberflächenelementen **72**, **74**, **76** der Oxidschicht **52**, die zu darunter liegenden Strukturen **78**, **80**, **82** gehören, wie sie bei einem typisch strukturierten Wafer **14** zu finden sind, sowie die Änderungen, denen sie während des CMP-Prozesses unterliegen. Das Element **72** ist ein relativ kleines Element, das Element **74** ist ein mittelgroßes Element, und das

Element **76** ist ein relativ großes Element. [Fig. 11A](#) zeigt die Elemente **72**, **74**, **76** vor dem Polieren, [Fig. 11B](#) zeigt die Elemente **72**, **74**, **76** etwa bei der Hälfte des Polierprozesses, während [Fig. 11C](#) die Elemente **72**, **74**, **76** zum Ende des Polierprozesses hin zeigt. In [Fig. 11A](#) wird das kleinere Element **72** mit einer größeren Rate als das mittlere Element **74** oder das große Element **76** reduziert. Außerdem wird das mittlere Element **74** mit einer größeren Rate als das große Element **76** reduziert. Die Rate, mit der die Elemente **72**, **74**, **76** reduziert werden, nimmt auch ab, wenn der Polierprozess fortschreitet. Beispielsweise hat das kleinere Element **72** anfänglich eine hohe Reduktionsrate. Diese Rate fällt jedoch während des Polierprozesses ab. Dementsprechend zeigt [Fig. 11B](#) die Höhe der Elemente **72**, **74**, **76**, die sich auszugleichen beginnt, während [Fig. 11C](#) die Höhe der Elemente **72**, **74**, **76** im Wesentlichen gleich zeigt. Da die Elemente mit unterschiedlicher Größe mit unterschiedlichen Raten reduziert werden und sich diese Raten ändern, hat das Interferenzsignal, das von jedem Element erzeugt wird, eine andere Phase und Frequenz. Dementsprechend schwankt das resultierende Interferenzsignal, welches partial aus allen einzelnen Reflexionen von jedem der Elemente **72**, **74**, **76** erstellt wird, in einer anscheinend willkürlichen Weise im Gegensatz zu dem vorher beschriebenen periodischen sinusförmigen Signal.

[0069] Wie jedoch oben erwähnt wurde, neigen die Polieraten der Elemente **72**, **74**, **76** zu einer engeren Konvergenz zum Planarisierungspunkt hin. Deshalb nähert sich die Differenz in der Phase und Frequenz zwischen den Interferenzstrahlen, die von den Elementen **72**, **74**, **76** erzeugt werden, null. Dies führt dazu, dass das resultierende Interferenzsignal als eine periodische sinusförmige Wellenform erkennbar wird. Dadurch ist es möglich zu bestimmen, wann die Oberfläche des strukturierten Wafers planarisiert worden ist, indem festgestellt wird, wann ein sinusförmiges Interferenzsignal beginnt. Dieses Verfahren ist in [Fig. 12](#) dargestellt. Zuerst wird im Schritt **402** eine Suche nach der vorstehend erwähnten sinusförmigen Änderung des Interferometersignals ausgeführt. Wenn die sinusförmige Änderung entdeckt wird, wird der CMP-Vorgang beendet, wie es im Schritt **404** gezeigt wird.

[0070] In dem Diagramm von [Fig. 13](#) ist die Amplitude des Detektorsignals über der Zeit für einen strukturierten Wafer aufgetragen, der einem CMP-Vorgang unterliegt. Der abgefragte Datensatz, der zur Erstellung dieses Diagramms verwendet wird, wird auf seinem vorherigen integrierten Wert gehalten, bis der nächste Wert berichtet wird, was die rechteckig abgeschnittenen gezeigten Spitzenwerte erklärt. Eine nähere Untersuchung zeigt, dass ein annehmbarer sinusförmiger Zyklus bei etwa 250 s aufzutauchen beginnt. Das fällt mit dem Punkt zusam-

men, an dem der strukturierte Wafer zuerst planarisiert worden ist. Natürlich wäre es bei einer Realzeitüberwachung des Ausgangssignals des Interferometers nicht möglich, genau zu erfassen, wann der Zyklus beginnt. Stattdessen muss wenigstens ein Teil des Zyklus erfolgt sein, ehe man sicher sein kann, dass der Zyklus begonnen hat. Vorzugsweise darf nicht mehr als ein Zyklus vorübergehen, ehe der CMP-Vorgang beendet wird. Eine Ein-Zyklus-Grenze ist eine praktische Wahl, da dies ein hohes Vertrauen gibt, dass der Zyklus tatsächlich begonnen hat, anstatt dass das Signal lediglich Änderungen des Rauschens wiedergibt, das durch das Polieren der unterschiedlich bemessenen Elemente auf der Oberfläche des Wafers erzeugt wird. Die Ein-Zyklus-Grenze gewährleistet zusätzlich, dass nur eine kleine Materialmenge von der Oberfläche des Wafers entfernt wird, nachdem er planarisiert worden ist. Es hat sich gezeigt, dass das Ausmaß der Planarisierung im Wesentlichen das gleiche nach zwei Zyklen ist, wie es nach einem war. Deshalb würde eine Fortsetzung des CMP-Vorgangs nur dazu dienen, mehr Material von der Oberfläche des Wafers zu entfernen. Obwohl ein Zyklus in dem Fall bevorzugt wird, in welchem der CMP-Prozess zu beenden ist, wenn der strukturierte Wafer planarisiert worden ist, soll die vorliegende Erfindung nicht auf diesen zeitlichen Rahmen begrenzt sein. Wenn das Signal besonders stark ist, kann es möglich sein, den gleichen Vertrauenspegel nach nur einem Teil eines Zyklus zu erhalten. Wenn das Signal besonders schwach ist, kann es alternativ mehr als einen Zyklus brauchen, um das erforderliche Vertrauen zu erhalten. Die Wahl hängt von den Eigenschaften des verwendeten Systems ab. Beispielsweise hat die Größe des Spaltes zwischen dem Quarzfenster und der Oberfläche des Wafers eine Wirkung auf die Signalstärke und somit auf die Entscheidung, wie viele Zyklen abzuwarten sind, bevor der CMP-Prozess beendet wird.

[0071] Die tatsächliche Bestimmung, wann das Ausgangssignal aus dem Laserinterferometer tatsächlich einem Zyklus unterliegt und somit anzeigt, dass die Oberfläche des Wafers planarisiert worden ist, kann in vielfältiger Weise erfolgen. Beispielsweise kann das Signal digital verarbeitet und ein Algorithmus verwendet werden, um die Bestimmung auszuführen. Ein solches Verfahren ist in dem US-Patent 5,097,430 offenbart, wo die Neigung des Signals zur Bestimmung verwendet wird. Zusätzlich stehen verschiedene bekannte Kurvenanpassalgorithmen zur Verfügung. Diese Verfahren werden im Wesentlichen dazu verwendet, das Interferometersignal mit einer sinusförmigen Kurve zu vergleichen. Wenn eine Anpassung innerhalb einer bestimmten vorgegebenen Toleranz erfolgt, wird bestimmt, dass der Zyklus begonnen hat. Manche Halbleiteranwendungen erfordern, dass die Dicke des Materials, das über einer Struktur liegt, die auf einem Element eines strukturierten Wafers ausgebildet ist (d. h. die Filmdicke),

eine bestimmte Tiefe hat und dass diese Filmdicke von Element zu Element und von Wafer zu Wafer wiederholbar ist. Die vorher beschriebenen Verfahren zum Planarisieren eines typischen strukturierten Wafers ergeben nicht notwendigerweise diese gewünschte wiederholbare Filmdicke. Der Zweck der Planarisierungsverfahren besteht darin, eine glatte und ebene Oberfläche, nicht jedoch eine spezielle Filmdicke zu erzeugen. Deshalb muss, wenn man die Filmdicke über einer speziellen Struktur oder einer Gruppe von ähnlich bemessenen Strukturen regulieren will, ein alternatives Verfahren verwendet werden. Dieses alternative Verfahren wird nachstehend beschrieben.

[0072] Wie vorstehend erläutert, neigt jede in der Größe unterschiedliche Oberflächenform, die sich aus einer Oxidschicht ergibt, die über einem strukturierten Aufbau auf einem Element gebildet wird, zur Erzeugung eines reflektierten Interferenzsignals mit einmaliger Frequenz und Phase. Nur in der Nähe des Planarisierungspunktes konvergieren Frequenz und Phase einer in der Größe unterschiedlichen Form. Vor dieser Konvergenz lassen sich die einmalige Frequenz und Phase der Interferenzsignale, die von den verschiedenen Formen unterschiedlicher Größe verursacht werden, zur Erzeugung eines Detektorsignals kombinieren, das willkürlich zu variieren scheint. Es ist jedoch möglich, dieses Signal so zu bearbeiten, dass die Interferenzsignalbeiträge von allen Formen, die mit unterschiedlichen Raten poliert werden, ausgenommen eine Form mit spezieller Größe oder eine Gruppe von Formen mit ähnlich bemessener Größe, beseitigt werden. Wenn einmal das Interferenzsignal, das der Form mit speziell bemessener Größe oder der Gruppe von Formen zugeordnet ist, isoliert worden ist, kommen die Verfahren zum Einsatz, die in Zuordnung zu dem Entfernen von Material von einer blanken Oxidscheibe diskutiert wurden, um genau die Materialmenge zu entfernen, die erforderlich ist, um die gewünschte Filmdicke zu erreichen.

[0073] Natürlich muss die Frequenz der Interferenzsignalkomponente, die durch die interessierende Form verursacht wird, vor der Signalverarbeitung bestimmt werden. Man nimmt an, dass diese Frequenz leicht durch Ausführen eines CMP-Prozesses an einer Testprobe bestimmt werden kann, die Elemente enthält, die ausschließlich mit Aufbauten strukturiert sind, die dem Aufbau entsprechen, der eine spezielle, darüber liegende Filmdicke haben soll. Das während des CMP-Prozesses erzeugte Detektorsignal wird nach bekannten Verfahren analysiert, um die einmalige Frequenz des Interferenzsignals zu bestimmen, die durch die Oberflächenformen verursacht wird, die den vorstehend erwähnten Strukturen zugeordnet sind.

[0074] Die spezifischen Schritte, die erforderlich sind, um das oben beschriebene Verfahren der Steu-

erung der Filmdicke über einem speziellen Aufbau oder einer Gruppe von Aufbauten ähnlicher Größe auf einem Element in situ während der CMP-Behandlung eines Wafers auszuführen, werden nun unter Bezug auf [Fig. 14](#) beschrieben. Bei dem Schritt **502** wird das Detektorsignal gefiltert, so dass nur die Komponente des Signals hindurchgeht, die die vorher festgelegte Frequenz hat, die dem interessierenden Aufbau zugeordnet ist. Dieser Schritt wird unter Verwendung von bekannten Bandpassfiltertechniken ausgeführt. Danach wird im Schritt **504** eine Messung der Zeit zwischen dem ersten Auftreten eines Maximums und eines Minimums oder umgekehrt in dem Detektorsignal ausgeführt (obwohl ein gesamter Zyklus oder ein Teil davon verwendet werden könnte). Die Materialmenge, die während des Teils des Zyklus (d. h. ein halber Zyklus) entfernt wird, wird im Schritt **506** über vorher beschriebene Verfahren bestimmt. Dann wird eine Entfernungsrates berechnet, indem die entfernte Materialmenge durch die gemessene Zeit geteilt wird, was im Schritt **508** gezeigt ist. Dies ergibt die Rate, mit der das Material bei dem vorhergehenden Teil des Zyklus entfernt wurde. In dem nächsten Schritt **510** wird die Dicke des Materials, das, wie im Schritt **506** berechnet, entfernt wurde, von der gewünschten zu entfernenden Dicke subtrahiert (d. h. die Dicke, die, wenn sie entfernt ist, die gewünschte Filmdicke ergibt, die über dem interessierenden Aufbau liegt), um eine zu entfernende Restdicke zu bestimmen. Diese zu entfernende Restdicke wird dann durch die vorstehend erwähnte Entfernungsrates geteilt, um zu bestimmen, wie lang der CMP-Prozess vor seiner Beendigung im Schritt **512** fortgesetzt werden muss. Wenn die Restzeit einmal berechnet worden ist, wird der Prozess für jedes Auftreten eines Maximums und Minimums oder umgekehrt wiederholt. Dementsprechend wird die Zeit zwischen dem nächsten auftretenden Maximum und Minimum gemessen, wobei die Dicke des Materials, das während des Teils des Zyklus entfernt wurde, der durch dieses Auftreten des Maximums und Minimums (d. h. eine Hälfte) dargestellt wird, durch die gemessene Zeit geteilt, und es wird die Entfernungsrates genauso wie bei der ersten Iteration des Verfahrens berechnet. Im nächsten Schritt **510** jedoch wird, wie in den Klammern gezeigt ist, die Gesamtmenge des Materials, das während der vorherigen Iterationen entfernt wurde, bestimmt, bevor sie von der gewünschten Dicke abgezogen wird. Der Rest des Verfahrens ist der gleiche, indem die zu entfernende Restdicke durch die neu berechnete Entfernungsrates geteilt wird, um die restliche CMP-Prozesszeit zu berechnen. Dieser Prozess wird wiederholt, bis die Restzeit abläuft, bevor die nächste Iteration beginnen kann. An diesem Punkt wird der CMP-Prozess beendet, was im Schritt **514** zu sehen ist.

[0075] Obwohl das oben beschriebene Verfahren zur Bereitstellung der Filmdicke das Verfahren zur Bestimmung des CMP-Prozessendpunkts benutzt,

das in [Fig. 10B](#) dargestellt ist, kann auch jedes der anderen Endpunktbestimmungsverfahren, die hier beschrieben sind, verwendet werden, falls dies erwünscht ist.

[0076] Der Strahldurchmesser (d. h. der Fleck) und die Wellenlänge des von dem Laserinterferometer erzeugten Laserstrahls können vorteilhaft manipuliert werden. Wie in [Fig. 15A](#) und [Fig. 15B](#) gezeigt ist, bedeckt ein schmaler Strahl **84**, beispielsweise ein bei der verwendeten Wellenlänge auf den kleinstmöglichen Fleck fokussierter, einen kleineren Bereich der Oberfläche des Wafers **14** als ein breiterer, weniger fokussierter Strahl **86**. Dieser schmale Strahl **84** ist für das Abtasten (d. h. Strahl **88**) aufgrund von Oberflächenunregelmäßigkeiten **90** empfindlicher als der weitere Strahl **86**, da der weitere Strahl **86** sich über einen größeren Teil der Oberfläche des Wafers **14** ausbreitet und mehr von Oberflächenunregelmäßigkeiten **90** umfasst. Deshalb hat ein breiterer Strahl **86** einen integrierten Effekt und ist weniger empfindlich für extreme Änderungen des reflektierten Interferenzsignals, wenn er über die Oberfläche des Wafers **14** läuft. Deshalb wird aus diesem Grund ein breiterer Strahl **86** bevorzugt. Die Laserstrahlbreite kann unter Verwendung bekannter optischer Vorrichtungen vergrößert werden.

[0077] Zu erwähnen ist, dass der breitere Strahl die verfügbare Datenerfassungszeit pro Plattenumdrehung verringert, da die Zeit, in welcher der Strahl vollständig in den Grenzen des Fensters enthalten ist, kleiner ist als bei einem schmaleren Strahl. Mit den vorher beschriebenen Verfahren der Datenerfassung stellt dies jedoch kein ernsthaftes Problem dar. Da der breitere Strahl auch die Lichtenergie über eine größere Fläche als ein schmalere Strahl ausbreitet, nimmt zusätzlich die Stärke der Reflexionen etwas ab. Dieser Nachteil kann dadurch geheilt werden, dass die Leistung des Laserstrahls aus dem Laserinterferometer so gesteigert wird, dass der Intensitätsverlust der reflektierten Strahlen bei der Messung keine Rolle spielt.

[0078] Was die Wellenlänge des Laserstrahls angeht, so ist es durchführbar, irgendeine Wellenlänge von weit infrarot bis ultraviolett zu verwenden. Bevorzugt wird jedoch, dass ein Strahl im Rotlichtbereich verwendet wird. Der Grund für diese Bevorzugung ist ein doppelter. Zuerst ergeben kürzere Wellenlängen eine Steigerung der Streugröße, die durch die chemische Schlämme verursacht wird, da diese Streuung proportional zur vierten Potenz der Frequenz des Laserstrahls ist. Deshalb ist die Streuung umso kleiner, je länger die Wellenlänge ist. Längere Wellenlängen führen jedoch auch dazu, dass pro Periode des Interferenzsignals mehr Oxidschicht entfernt wird, da die Materialmenge, die pro Periode entfernt wird, etwa $\lambda/2n$ entspricht. Deshalb wird in einer Periode umso weniger Material entfernt, je kürzer die Wellenlänge

ist. Man möchte so wenig wie möglich Material während jeder Periode entfernen, so dass die Möglichkeit des Entfernens von zu viel Material minimiert wird. Beispielsweise würde in einem System, welches das vorstehend beschriebene Verfahren verwendet, bei dem die Anzahl der Zyklen oder eines Teils davon gezählt werden, um die Dicke der entfernten Oxidschicht zu bestimmen, jedes überschüssige entfernte Material, das über die gewünschte Menge entfernt wird, minimiert, wenn die Materialmenge, die während jedes Zyklus oder eines Teils davon entfernt wird, so klein wie möglich ist.

[0079] Man nimmt an, dass diese beiden entgegengesetzten Faktoren bei der Wahl der Wellenlänge optimal ausgeglichen sind, wenn ein Rotlicht-Laserstrahl gewählt wird. Rotlicht bietet ein akzeptables Maß an Streuung und führt nicht zu einer nicht handhabbaren, pro Zyklus zu entfernenden Materialmenge.

Weitere Ausführungsformen

[0080] Die erzeugte Interferenzwellenform gibt beträchtliche Zusatzinformationen über den Polierprozess. Diese zusätzlichen Informationen können dazu verwendet werden, eine In-situ-Messung der Gleichförmigkeit der polierten Schicht zu liefern. Sie können auch dazu verwendet werden, festzustellen, wann das CMP-System nicht innerhalb des Spektrums arbeitet (d. h. nicht arbeitet wie gewünscht). Diese beiden Anwendungen werden nun beschrieben.

Gleichförmigkeitsmessung

[0081] Die Polier- und/oder Planarisierungsvorgänge, die bei dem CMP-System ausgeführt werden, erfordern insgesamt die Erzeugung einer Oberflächenschicht, die über der Oberfläche des Wafers/Substrats gleichförmig ist. Mit anderen Worten, die Mitte des Wafers muss mit der gleichen Rate wie der Rand des Wafers poliert werden. Gewöhnlich darf sich die Dicke der polierten Schicht nicht um mehr als etwa 5 bis 10% ändern. Wenn dieser Gleichförmigkeitspegel nicht erreicht wird, ist es wahrscheinlich, dass der Wafer nicht verwendbar ist, da die Vorrichtungsaussbeuten unannehmbar niedrig sind. In der Praxis ist es häufig sehr schwierig, über dem Wafer eine gleichförmige Polierrate zu erreichen. Gewöhnlich ist die Optimierung vieler unterschiedlicher Variablen erforderlich, um ihre Ausführung innerhalb der Spezifizierungen zu halten. Der oben beschriebene Endpunktdetektor bildet ein äußerst nützliches Mittel zur Überwachung der Gleichförmigkeit der zu polierenden Schicht, wobei dieses Überwachen sowohl durch In-situ-Datenerfassung als auch durch Verarbeitung ausgeführt werden kann.

[0082] Wir haben gefunden, dass die Interferenzwellenform, die von dem Interferometer während

des Polierens erzeugt wird, Informationen über die Gleichförmigkeit der Schicht, die poliert wird, gibt. Wie oben erwähnt, erscheint das Ausgangssignal des Interferometers als sinusförmiges Signal, wenn die Oberflächenschicht (beispielsweise die Oxidschicht) poliert wird. Der Abstand zwischen den Spitzen dieses Signals zeigt an, wie viel Material entfernt worden ist. Auf diesem sinusförmigen Signal ist auch ein weiteres höherfrequentes sinusförmiges Signal vorhanden. Die Amplitude dieses höherfrequenten Signals zeigt an, um wie viel sich die Dicke der polierten Schicht über der Oberfläche des Wafers ändert.

[0083] Der Grund für das Auftreten des Hochfrequenzsignals ist folgender. Wenn das Polieren ausgeführt wird, tastet das Interferometer gewöhnlich verschiedene Stellen über der Oberfläche des Wafers ab (oder sieht auf diese). Der Grund dafür besteht darin, dass während des Polierens sowohl die Platte als auch der Wafer drehen und zusätzlich der Wafer auch axial bezüglich der Platte bewegt wird. Deshalb gehen während des Polierens unterschiedliche Bereiche der Waferoberfläche über das Loch in der Platte, durch welches das Interferometer die zu polierende Schicht sieht. Wenn die polierte Schicht vollständig gleichförmig ist, bleibt die erhaltene Interferenzwellenform von der Abtastung der unterschiedlichen Stellen quer über der Waferoberfläche uneinträchtigt. D. h., sie hat im Wesentlichen die gleiche Amplitude. Wenn andererseits die polierte Schicht nicht gleichförmig ist, führt die Abtastung unterschiedlicher Stellen zu einer weiteren Änderung an dem sinusförmigen Basissignal. Diese weitere Änderung hat eine Frequenz, die von der Rotations- und der Schwenkrate, die verwendet werden, abhängt, und hat eine Amplitude, die proportional zum Ausmaß der Ungleichförmigkeit der polierten Schicht ist. Ein Beispiel für eine solche Wellenform ist in [Fig. 16](#) gezeigt. Bei diesem speziellen Beispiel war die Ungleichförmigkeit relativ groß, um das Hochfrequenzsignal deutlich darzustellen.

[0084] Ein Maß der Gleichförmigkeit ist das Verhältnis aus Spitze-zu-Spitze-Amplitude A_{hf} des Hochfrequenzsignals und Spitze-zu-Spitze-Amplitude A_{lf} des Niederfrequenzsignals. Je kleiner dieses Verhältnis ist, desto gleichförmiger ist die polierte Schicht, und umgekehrt ist dieses Verhältnis umso größer, je ungleichförmiger die Schicht ist.

[0085] Ein CMP-System, welches ein Maß für die Gleichförmigkeit erzeugt, ist in [Fig. 17](#) gezeigt. Zusätzlich zu den in der vorstehend beschriebenen [Fig. 2](#) gezeigten Komponenten ist auch ein Rechner **150** vorgesehen, der so programmiert ist, dass er den Betrieb des Interferometers steuert und die Signalanalyse durchführt, die erforderlich ist, um aus dem Interferenzsignal ein Maß für die Gleichförmigkeit zu erzeugen. Ferner ist eine Anzeigeeinheit **160** vorgesehen, über die für eine Bedienungsperson verschie-

dene Informationen und Ergebnisse angezeigt werden. Der Rechner **150** kann irgendeine Vorrichtung sein, die in der Lage ist, die Steuer- und Signalverarbeitungsfunktionen auszuführen, beispielsweise ein Standard-PC, der geeignet programmiert ist, und eine zweckorientierte, speziell ausgelegte digitale Verarbeitungseinheit. Die Anzeigeeinheit **140** kann ein Bildschirm, ein Drucker oder irgendeine andere geeignete Vorrichtung oder eine Kombination von Vorrichtungen sein, um der Bedienungsperson des CMP-Systems Informationen zu übermitteln.

[0086] Zur Erzeugung eines Gleichförmigkeitsmaßes wird der Rechner **150** so programmiert, dass er die Signalverarbeitung und andere in [Fig. 18](#) gezeigte Funktionen ausführt. Diesbezüglich hat der Rechner **180** zwei programmierbare Bandpassfilter, nämlich ein Hochfrequenzfilter **152** und ein Niederfrequenzfilter **154**. Das Hochfrequenzfilter **152** hat einen Bandpass, der auf die Frequenz des Hochfrequenzsignals zentriert ist, das die Gleichförmigkeitsinformationen enthält, während das Niederfrequenzfilter **154** einen Bandpass hat, der auf die Frequenz des Niederfrequenzsignals zentriert ist, das die Polierrateninformationen enthält. Die Breite der beiden Bandpässe ist in der Größenordnung weniger Millihertz in dem Fall, wenn die Periode in der Größenordnung von Sekundenzehnern liegt. Die Breite des Bandpasses wird so programmiert, dass sie sich proportional zu der Mittenfrequenz ändert, oder anders gesagt, dass sie sich umgekehrt zur Periode des zu prüfenden Signals ändert. D. h., dass, wenn die Periode des relevanten Signals zunimmt, die Bandbreite des Bandpassfilters abnimmt und umgekehrt.

[0087] [Fig. 19A](#) zeigt ein Beispiel eines Interferometers, das man aus einem aktuellen System erhält. Anfänglich zeigt das Signal an, dass die Schicht durchgehend gleichförmig ist, d. h. oben auf dem Niederfrequenzsignal sitzt kein erkennbares Hochfrequenzsignal. Nach einem Polieren über einem kurzen Zeitraum beginnt ein Hochfrequenzsignal in Erscheinung zu treten, das einen bestimmten Wert für eine Ungleichförmigkeit anzeigt. Das Niederfrequenzfilter **154** wählt die Niederfrequenzkomponente aus und filtert die anderen Frequenzen heraus, um ein Ausgangssignal in der in [Fig. 19B](#) gezeigten Form zu erzeugen. Gleichermaßen wählt das Hochfrequenzfilter **152** die Hochfrequenzkomponente aus und filtert die anderen Frequenz heraus, um ein Ausgangssignal der in [Fig. 19C](#) gezeigten Form zu erzeugen.

[0088] Der Rechner **150** führt zwei Amplitudenmessfunktionen **156** und **158** aus, welche die Spitze-zu-Spitze-Amplituden der Ausgangssignale der Filter **152** bzw. **154** messen. Wenn die Amplituden der beiden gefilterten Signale bestimmt worden sind, berechnet der Rechner **140** ein Verhältnis aus der Sp-Sp-Amplitude des Hochfrequenzsignals und aus

der Sp-Sp-Amplitude des Niederfrequenzsignals (d. h. $A_{\text{HF}}/A_{\text{LF}}$) (siehe Funktionsblock **162**). Nach Berechnung des Verhältnisses vergleicht der Rechner **150** (siehe Block **166**) das berechnete Verhältnis mit einem Schwellen- oder Referenzwert **164**, der vorher in einem lokalen Speicher gespeichert wurde. Wenn das berechnete Verhältnis den gespeicherten Schwellenwert überschreitet, alarmiert der Rechner **150** die Bedienungsperson dahingehend, dass die Ungleichförmigkeit der polierten Schicht eine akzeptable Größe überschreitet. Als Antwort kann die Bedienungsperson die Prozessparameter einstellen, um den Prozess zurück in die Spezifizierungen zu bringen.

[0089] Da das Hochfrequenzsignal nur dann erscheint, wenn etwas poliert worden ist, ist es zweckmäßig abzuwarten, ehe versucht wird, die Ungleichförmigkeit zu messen. Es kann sein, dass man das Verhältnis automatisch und periodisch berechnen möchte, um die Gleichförmigkeit der polierten Schicht während des ganzen Poliervorgangs zu überwachen. In diesem Fall kann es auch erwünscht sein, dass der Rechner **150** die berechneten Verhältnisse während des Prozesses ausgibt, so dass die Bedienungsperson Änderungen und/oder Tendenzen feststellen kann, die bei dem Polierprozess in Erscheinung treten. Dies ist besonders praktisch, wenn die In-situ-Überwachung während der aktuellen Produktion von Wafern während des Polierens ausgeführt wird.

[0090] Die gerade beschriebenen Funktionen können mittels einer Software ausgeführt werden, die auf dem Rechner läuft, oder sie können durch zweckorientierte Schaltungen erfolgen, die für diesen speziellen Zweck gebaut sind.

[0091] Die Bandpassfilter können unter Verwendung von Techniken hergestellt werden, die dem Fachmann bekannt sind. Bei der beschriebenen Ausgestaltung sind dies FIR-Filter (Filter mit Ansprechen auf finite Impulse), die entweder im Frequenz- oder im Zeitbereich ausgeführt sind. Um jedoch das Filtern in Realzeit durchzuführen, wenn das Interferometer-signal verfügbar wird, führt man das Filtern im Zeitbereich aus, indem man die geeignete Funktion mit der Wellenform, wie sie erzeugt wird, zusammenbringt. Die geeignete Funktion ist natürlich einfach die Zeitbereichsdarstellung eines Bandpassfilters, welches die gewünschten Eigenschaften hat (d. h. Mittenfrequenz und Bandbreite).

[0092] Zur Spezifizierung der geeigneten Filterparameter ist es erforderlich, die Frequenz des Signals zu kennen, das von dem Filter ausgewählt werden soll. Diese Information kann leicht aus der Interferometer-Signalwellenform (s) erhalten werden. Beispielsweise kann man die Mittenfrequenz des Niederfrequenzfilters erhalten, indem man eine Charge (bei-

spielsweise 25) von Wafern hernimmt (beispielsweise blanke Wafer nur mit einer Oxidbeschichtung), um eine genaue Messung der Poliertrate zu erhalten. Alternativ kann die Poliertrate zu Beginn eines Poliervorgangs bestimmt werden, indem man die Entfernung zwischen den Spitzen des Niederfrequenzsignals misst. Natürlich erzeugt die Verwendung dieses alternativen Wegs Ergebnisse, die nicht so genau sind wie mittelnnde Messungen über einer größeren Anzahl von Wafern. Auf jeden Fall bestimmt die Poliertrate die Mittenfrequenz des Bandpassfilters, und dadurch, dass man die Mittenfrequenz zusammen mit der gewünschten Bandbreite des Filters kennt, kann man die genaue Form der Zeitbereichs-Filterfunktion und/oder die Koeffizienten des FIR-Filters leicht bestimmen.

[0093] Die Frequenz des Hochfrequenzsignals kann man auf ähnliche Weise erhalten, d. h. direkt aus der Spur, die von dem Interferometer erzeugt wird, wenn das CMP-System den Wafer poliert. D. h. mit anderen Worten, dass die Bedienungsperson einfach die Entfernung zwischen den Spitzen des Hochfrequenzsignals misst. Dieser Vorgang kann leicht automatisiert werden, so dass die Bedienungsperson mit Hilfe einer Hinweisvorrichtung (beispielsweise einer Maus) zwei Punkte an der Wellenform markieren kann, die auf einer Bildschirmanzeige erscheinen, und der Rechner kann so programmiert werden, dass er die Frequenz automatisch berechnet und dann die geeigneten Filterkoeffizienten erzeugt. Die Filterkoeffizienten und/oder die Zeitbereichsdarstellung der Filterfunktionen werden dann in dem lokalen Speicher für eine spätere Verwendung während der Poliervorgänge gespeichert, um die Filtrivorgänge auszuführen.

Prozesssignatur

[0094] Die Interferometerwellenform stellt auch eine Signatur des Systems dar (d. h. sie kennzeichnet es), für das man sie erhält. Deswegen stellt sie Informationen bereit, die für die Bewertung eines Systems für einen Produktionsbetrieb zweckmäßig sind. Wenn man eine Signatur für ein System erhält, von dem man weiß, dass es wie gewünscht arbeitet, kann diese Signaturwellenform (oder die aus der Wellenform extrahierten Merkmale) als Referenz verwendet werden, gegen die darauf folgend erzeugte Signaturen verglichen werden können, um zu bestimmen, ob das System oder die Systeme, aus denen die Signaturen darauf folgend erhalten wurden, innerhalb der Vorgaben arbeiten. Wenn beispielsweise die Polierkissen ausgewechselt sind oder eine neue Charge von Schlämme in dem CMP-System verwendet wird, muss die Bedienungsperson wissen, ob die Änderung die Qualität des Polierens, welches das System ausführt, schädlich beeinflusst. Wir haben festgestellt, dass eine Änderung der Leistung des CMP-Systems zu einer Änderung in der Signatur

führt. D. h., dass bestimmte Merkmale in der Wellenform erscheinen, die vorher nicht vorhanden sind, oder dass vorher existierende Merkmale sich ändern. Durch Erfassen dieser Änderungen ist es möglich, festzustellen, wann ein System nicht wie gewünscht arbeitet.

[0095] Bei der beschriebenen Ausführungsform sind die aus der Interferometerwellenform extrahierten Größen die Poliertrate und die Messung der Gleichförmigkeit. Diese beiden Eigenschaften sind aus der Interferometerwellenform leicht erhältlich, die während des Polierens erzeugt wird, indem die vorher beschriebenen Methoden verwendet werden. Ein richtig arbeitendes System erzeugt eine spezielle Poliertrate und ein spezielles Maß an Gleichförmigkeit. Ein Abweichen von diesen Bezugswerten bildet eine Anzeige dafür, dass das System sich von seinem gewünschten Arbeitspunkt wegbewegt, und alarmiert die Bedienungsperson bezüglich der Notwendigkeit eines Korrekturingriffs, um eine Zerstörung des Produkts zu vermeiden.

[0096] Ein Verfahren, welches eine CMP-Systemsignatur verwendet, ist in [Fig. 20A](#) gezeigt und wird nun beschrieben. Am Anfang wird eine Interferometerwellenform (d. h. eine Signatur) für ein CMP-System erzeugt, von dem man weiß, dass es optimal arbeitet (Schritt 250). Die Entscheidung, ob das System optimal arbeitet, kann empirisch erfolgen, indem ein Satz von Versuchswafern behandelt wird und die Ergebnisse analysiert werden. Wenn die Ergebnisse, die erzeugt werden, innerhalb der Vorgaben liegen, kann die Signatur für diese Ausgestaltung erzeugt werden und die Betriebsbedingungen können eingestellt werden. Bevor ein Teil der Interferometerwellenform genommen wird, ist es wünschenswert, den Wafer zwischen 50 bis 100% des Wegs durch das Oxid zu polieren, so dass die Wellenform eine echte Signatur der Polieranordnung ist.

[0097] Wenn man die Wellenform erhalten hat, extrahiert man aus der erzeugten Wellenform (Schritt 252) bestimmte relevante Merkmale und speichert sie für einen späteren Gebrauch als Referenz, gegenüber der die Systemleistung zu einer späteren Zeit oder späteren Zeitpunkten (Schritt 254) bewertet wird. Alternativ kann die Wellenform selbst gespeichert und als Referenz verwendet werden. Bei der beschriebenen Ausgestaltung sind die extrahierten Merkmale die Poliertrate und die Messung der Gleichförmigkeit, die beide aus der Wellenform, wie oben beschrieben, bestimmt werden können.

[0098] Gemäß [Fig. 20B](#) kann zu einer etwas späteren Zeit die gespeicherte Signatur (oder die extrahierten Merkmale) dazu verwendet werden, das System oder ein anderes System für Fertigungszwecke zu bewerten. Zur Bewertung eines Systems für die Produktion erhält man eine neue Signatur für das Sys-

tem (Schritt **258**), und die relevanten Merkmale werden aus der neuen Signatur (Schritt **260**) extrahiert. Die entnommenen Merkmale werden dann mit dem gespeicherten Referenzsatz von Merkmalen (Schritt **264**) verglichen. Wenn der Betriebspunkt, wie er durch den Satz von extrahierten Merkmalen gekennzeichnet ist, in einen vorgegebenen Bereich um den Referenzpunkt fällt, wie er durch den gespeicherten Referenzsatz von Merkmalen definiert ist, wird geschlossen, dass das System richtig arbeitet und dass es für die Behandlung von Produktwafern (Schritt **266**) online gebracht werden kann. Wenn dieser Prozess automatisiert ist, kann der Rechner an diesem Punkt die Bedienungsperson darauf hinweisen, dass der Prozess innerhalb der Vorgaben abläuft. Wenn andererseits der Betriebspunkt außerhalb des vorgegebenen Bereichs fällt, was eine Anzeige dafür ist, dass das System nicht innerhalb der Vorgaben arbeitet, wird die Bedienungsperson auf dieses Problem aufmerksam gemacht, so dass sie korrigierend eingreifen kann (Schritt **268**). Zu dem Korrekturingriff kann das geeignete Einstellen einiger Prozessparameter gehören, um den Prozess in den Vorgabenbereich zu bringen. Wenn beispielsweise die Poliertrate zu groß ist oder wenn die Oxidungleichförmigkeit größer als erlaubt ist, kann die Bedienungsperson erkennen, dass es zweckmäßig ist, eine neue Charge der Schlämme zu versuchen oder den Druck auf das Kissen einzustellen oder auch das Kissen auszutauschen. Der spezielle Verlauf des Korrekturingriffs, der gewählt wird, hängt natürlich von den Einzelheiten ab, wie das System von dem gewünschten Betriebspunkt abgewichen ist, von der Ausgestaltung und den Betriebsparametern des speziellen Systems und davon, was die Bedienungsperson ihre Erfahrung lehrt.

[0099] Um der Bedienungsperson weitere nützliche Informationen zu geben, gibt der Rechner optional auch über seine Anzeigevorrichtungen) Informationen über die extrahierten Merkmale (Schritt **262**) aus. Die angezeigten Informationen können als die extrahierten Merkmale, die Wellenform, wie nahe die verschiedenen extrahierten Merkmale zu den unterschiedlichen Merkmalen des gespeicherten Referenzsatzes liegen, oder was sich als besonders nützlich für die Bedienungsperson erweist, angegeben werden.

[0100] Natürlich kann der oben beschriebene In-situ-Realzeit-Überwachungsvorgang periodisch während der Behandlung von Produktionswafern oder immer dann verwendet werden, wenn irgendein Prozessparameter in dem CMP-System geändert wird (beispielsweise wenn ein neues Polierkissen verwendet wird, der Polierdruck eingestellt wird oder wenn eine neue Charge von Schlämme verwendet wird). Dabei ist es erforderlich zu wissen, dass sich der CMP-Prozess noch innerhalb der Vorgaben befindet. Anstelle des tatsächlichen Produkts kann der Pro-

zess außerdem bei blanken Wafern verwendet werden, um das CMP-System zu bewerten, bevor es am aktuellen Produkt eingesetzt wird.

[0101] Obwohl wir einen geradlinigen und einfachen Weg beschrieben haben, um Informationen aus der Signaturwellenform zu extrahieren, d. h. durch Verwenden der Poliertrate und des Maßes der Gleichförmigkeit, kann die Signatur oder die Interferometerwellenform dadurch analysiert werden, dass kompliziertere Techniken verwendet werden (beispielsweise Muster oder Merkmalerkennung oder andere Bildanalysealgorithmen oder neurale Netzwerke, um nur einige Alternativen zu nennen). Die Informationen, die die verschiedenen extrahierten Merkmale hinsichtlich des Betriebs des Systems tragen, können durch Versuche bestimmt werden, wobei diejenigen verwendet werden können, die die Informationen tragen, die als äußerst wichtig angesehen werden.

[0102] Anzumerken ist auch, dass ein einfaches Anzeigen der Interferometerwellenform (d. h. der Prozesssignatur) für die Bedienungsperson eine wertvolle Rückkoppelung darüber geben kann, wie gut sich das System verhält. Gewöhnlich ist das menschliche Auge äußerst empfindlich hinsichtlich der Erfassung auch von feinen Änderungen in einem Bild gegenüber dem, was man zu sehen erwartet. Deshalb ist die Bedienungsperson, nachdem sie einige Erfahrung gesammelt hat, in der Lage, Änderungen und immanente Probleme bei der CMP-System-Gesamtleistung einfach durch Blicken auf die Wellenform zu erfassen. Bei der beschriebenen Ausführungsform zeigt somit der Rechner auch die Signaturwellenform für die Bedienungsperson während der Behandlung an, so dass die Bedienungsperson sie auch zur Überwachung der Vorrichtungsleistung verwenden kann.

[0103] Durch Verwenden von dem Fachmann bekannten Techniken kann man leicht Software-Algorithmen entwickeln, die Änderungen automatisch erkennen oder erfassen, nach denen die Bedienungsperson Ausschau hält, und die der Bedienungsperson Hinweise auf bestimmte Probleme geben.

Modifizierung zur Erzielung einer verbesserten Leistung

[0104] Zu einer anderen Ausgestaltung gehört eine Modifizierung des Fensters in dem Kissen zwischen dem Interferometer und dem Wafer. Obwohl das Kissen einen wesentlichen Teil des Interferometer-Laserstrahls durchlässt, hat sich gezeigt, dass es auch eine beträchtliche reflektive Komponente von der unteren Fläche des Kissens gibt. Diese Situation ist in [Fig. 21A](#) gezeigt, wo ein Teil des Laserstrahls **34**, der aus dem Laserinterferometer **32** ausgeht, durch das Kissen **22** hindurchgeht und einen durchgegangenen Strahl **702** bildet, während ein Teil des Laserstrahls **34** von der Rückseitenfläche **704** des Kissens **22** re-

flektiert wird und einen reflektierten Strahl **706** bildet. Der reflektierte Strahl **706** erzeugt eine beträchtliche Gleichstrom-(DC-)Verschiebung in dem Datensignal. **Fig. 21B** zeigt diese Verschiebung (jedoch zum Zwecke der Klarheit übertrieben). Bei diesem Beispiel addiert die DC-Verschiebung, die sich aus dem reflektierten Laserlicht ergibt, etwa 8,0 V zu dem Gesamtsignal hinzu. Die DC-Verschiebung erzeugt Probleme beim Analysieren des nützlichen Teils des Datensignals. Wenn beispielsweise die Datenanalysevorrichtung in einem Bereich von 0 bis 10 V arbeitet, ist eine Verstärkung des DC-verschobenen Signals zur Steigerung des interessierenden Anteils ohne Reduzierung oder Beseitigung der DC-Komponente des Signals nahezu unmöglich. Wenn die DC-Komponente nicht beseitigt wird, würde die Anordnung durch das verstärkte Signal gesättigt. Eine Reduzierung oder Beseitigung der DC-Komponente erfordert elektronisch hinzugefügte Signalverarbeitungselektroniken und kann zu einer Verschlechterung des nützlichen Anteils des Signals führen. Auch wenn die DC-Verschiebung nicht so groß wie hier beschrieben ist, ist wahrscheinlich eine bestimmte Signalverarbeitung für ihre Beseitigung erforderlich. Deshalb möchte man ein nichtelektronisches Verfahren zur Reduzierung oder Beseitigung dieser unerwünschten DC-Komponente haben.

[0105] Es hat sich gezeigt, dass durch Erzeugung einer diffusen Oberfläche **704'** auf der Rückseite des Kissens **22** in dem das Fenster bildenden Bereich, wie es in **Fig. 21C** gezeigt ist, das reflektierte Licht von dieser Oberfläche gedämpft wird. Dadurch wird die unerwünschte DC-Komponente des Datensignals reduziert. Die diffuse Oberfläche **704'** streut tatsächlich das nicht hindurchgegangene Laserlicht **708**, anstatt es zum größten Teil zurück zu dem Interferometer **32** zu reflektieren. Das reflektierte Signal aus dem Wafer muss auch durch die diffuse Oberfläche **704'** gehen, wobei es ebenfalls gestreut wird. Es hat sich jedoch gezeigt, dass dies die Leistung des Interferometers nicht ernsthaft verschlechtert.

[0106] **Fig. 21D** zeigt das Datensignal, das man erhält, wenn die diffuse Oberfläche **704'** verwendet wird. Man sieht, dass bei Beseitigung der DC-Komponente das Signal leicht verstärkt und verarbeitet werden kann, ohne dass es erforderlich ist, irgendeinen DC-Anteil elektronisch zu beseitigen.

[0107] Wie die diffuse Oberfläche erzeugt wird, ist nicht von zentraler Bedeutung. Sie kann durch Sandstrahlen der Rückseite des Polierkissens in der Nähe des Fensters oder durch Aufbringen einer Materialbeschichtung erzeugt werden, die diffus ist (beispielsweise ein Klebeband) oder auf irgendeine andere Weise, die die gewünschten Ergebnisse erzielt.

Patentansprüche

1. Polierkissen zur Verwendung bei einem chemischen mechanischen Poliersystem, wobei das Polierkissen (**22**)
 - (a) eine Polierfläche und
 - (b) eine Bodenfläche sowie ein Fenster mit einem massiven Element aufweist, das für Licht aus einem Interferometer (**32**) durchlässig und das in dem Kissen vorgesehen ist.
2. Polierkissen nach Anspruch 1, bei welchem die obere Fläche des Fensters bündig zur Polierfläche ist.
3. Polierkissen nach Anspruch 1 oder Anspruch 2, bei welchem das Fenster eine diffuse untere Fläche hat.
4. Polierkissen nach einem der vorhergehenden Ansprüche, bei welchem das Fenster ein gesonderter, in dem Kissen (**22**) befestigter Stopfen (**42**) ist.
5. Polierkissen nach Anspruch 2, bei welchem das Fenster ein integraler Teil des Kissens (**22**) ist.
6. Polierkissen nach Anspruch 2, bei welchem das Fenster (**42**) massives Polyurethan ist.
7. Polierkissen nach Anspruch 4 und Anspruch 6, bei welchem das Fenster (**42**) furchenfrei ist.
8. Polierkissen nach Anspruch 2 mit einer Abdeckschicht (**22**) und einer Unterlageschicht (**20**), wobei das Fenster (**42**) nur in der Abdeckschicht ausgebildet ist und die Unterlageschicht in einem Bereich entfernt ist, über dem das Fenster liegt.
9. Polierkissen nach Anspruch 8, bei welchem die Abdeckschicht (**22**) Polyurethan mit Zusatzstoffen aufweist, welche die Durchlässigkeit unterbinden.
10. System zum chemischen mechanischen Polieren eines Wafers (**14**)
 - mit einer Platte (**16**), die das Polierkissen (**20, 22**) nach Anspruch 1 hält,
 - mit einem Polierkopf (**12**) zum Halten des Wafers, und
 - mit einem Endpunktdetektor (**32**), der Licht durch das Fenster in dem Polierkissen richtet und Licht misst, das von dem Wafer durch das Fenster reflektiert wird.
11. Verfahren zur Endpunktmessung während des chemischen mechanischen Polierens eines Wafers (**14**), wobei bei dem Verfahren
 - der Wafer mit einem Polierkissen (**20, 22**) nach Anspruch 1 poliert wird,
 - Licht durch das Fenster (**42**) in dem Polierkissen geschickt wird und

– eine Reflexion des Lichts empfangen wird, das von dem Wafer reflektiert wird und durch das Fenster in dem Polierkissen hindurchgeht.

Es folgen 18 Blatt Zeichnungen

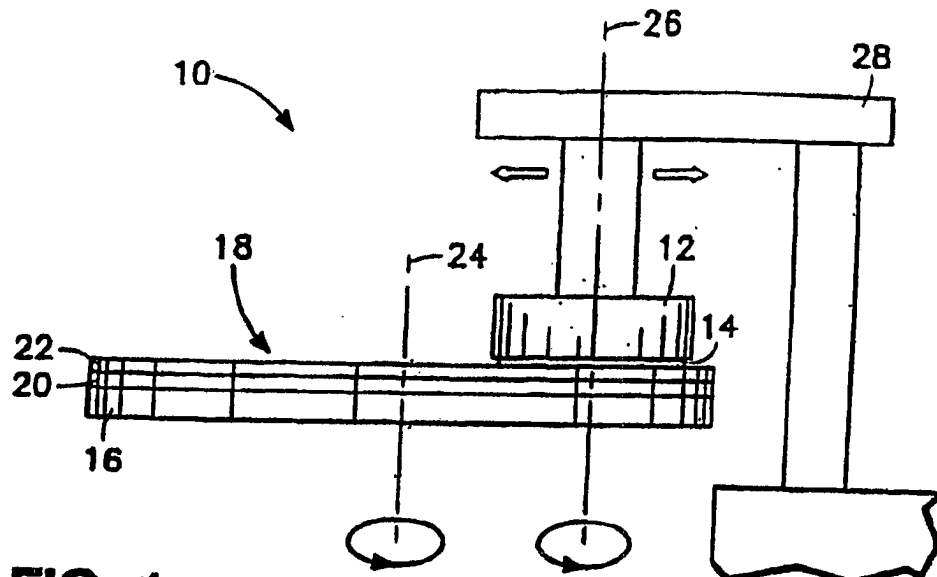


FIG. 1

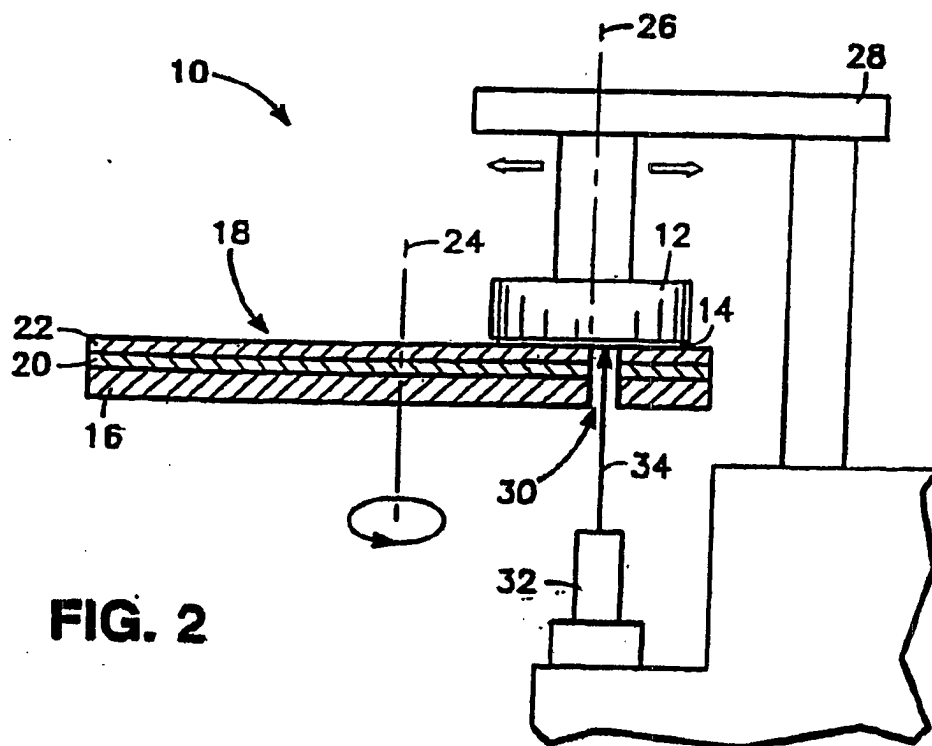
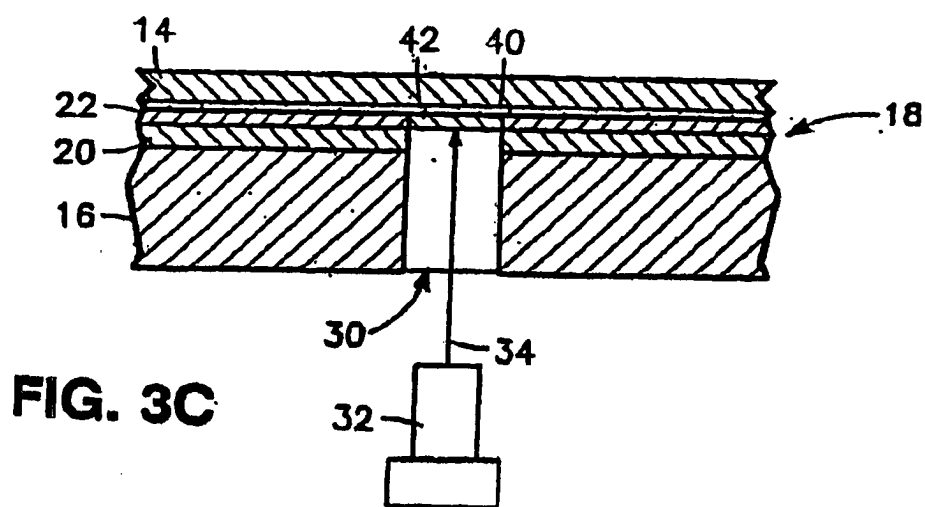
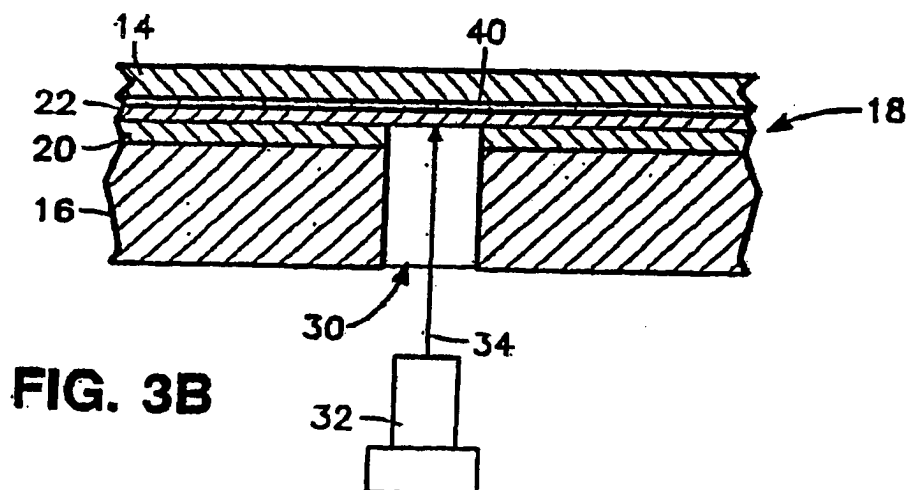
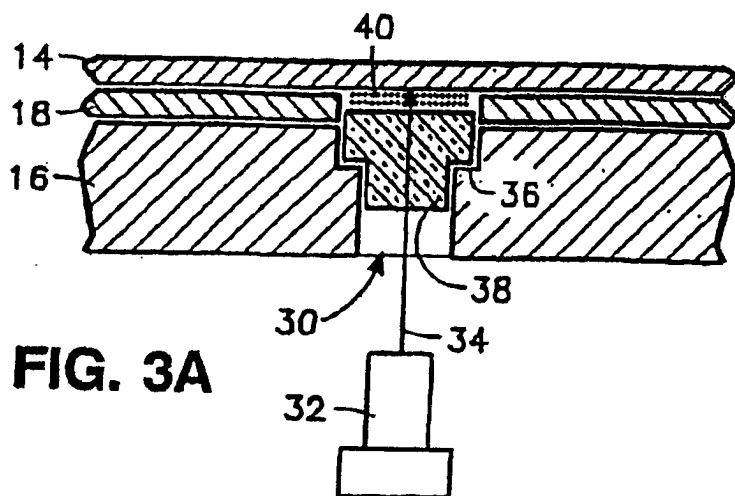
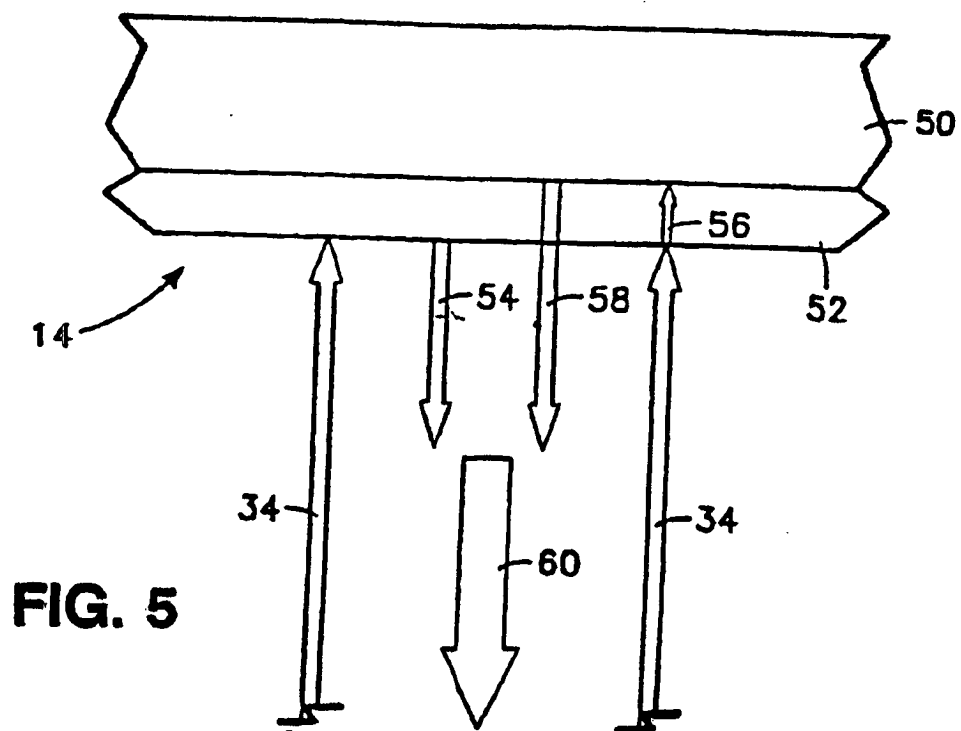
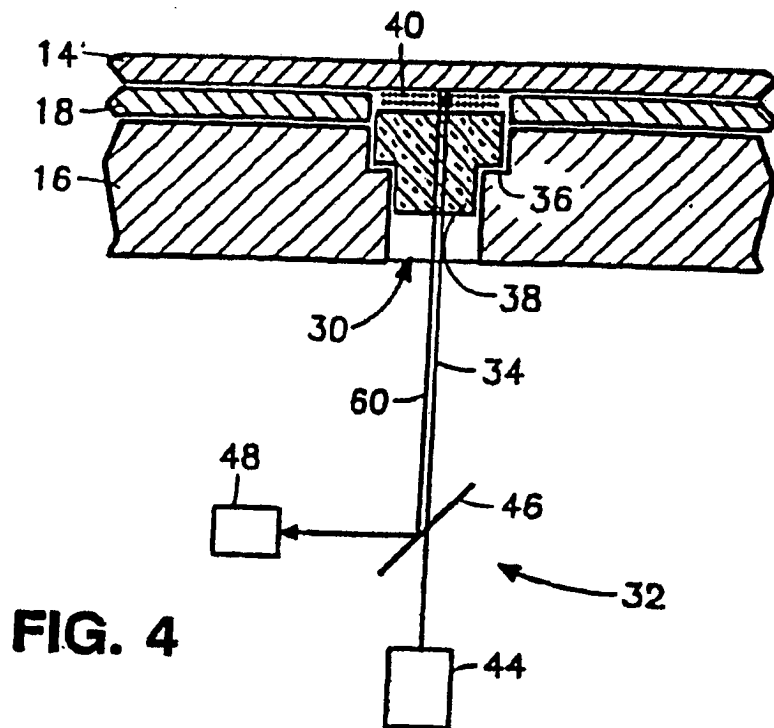


FIG. 2





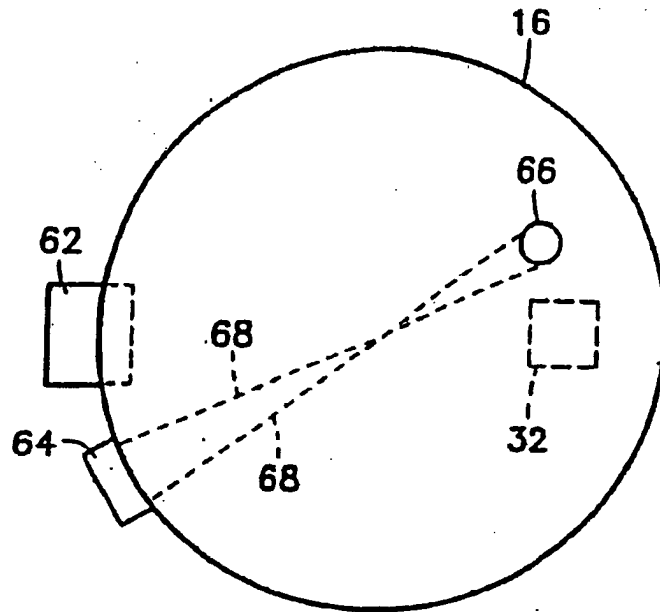


FIG. 6

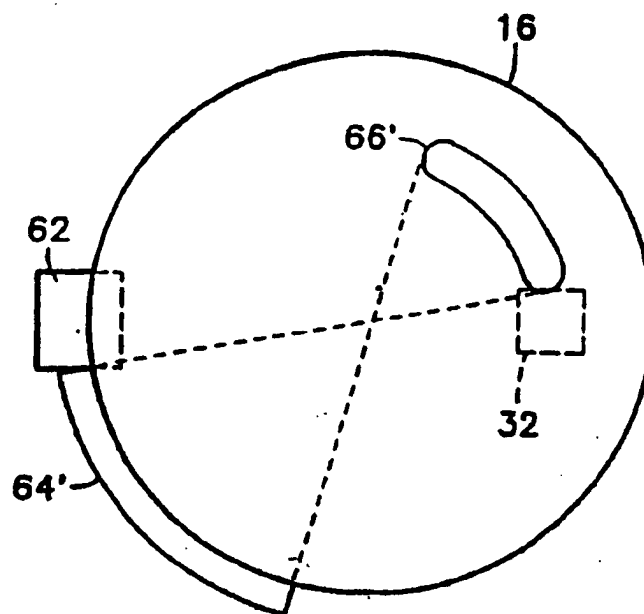


FIG. 7

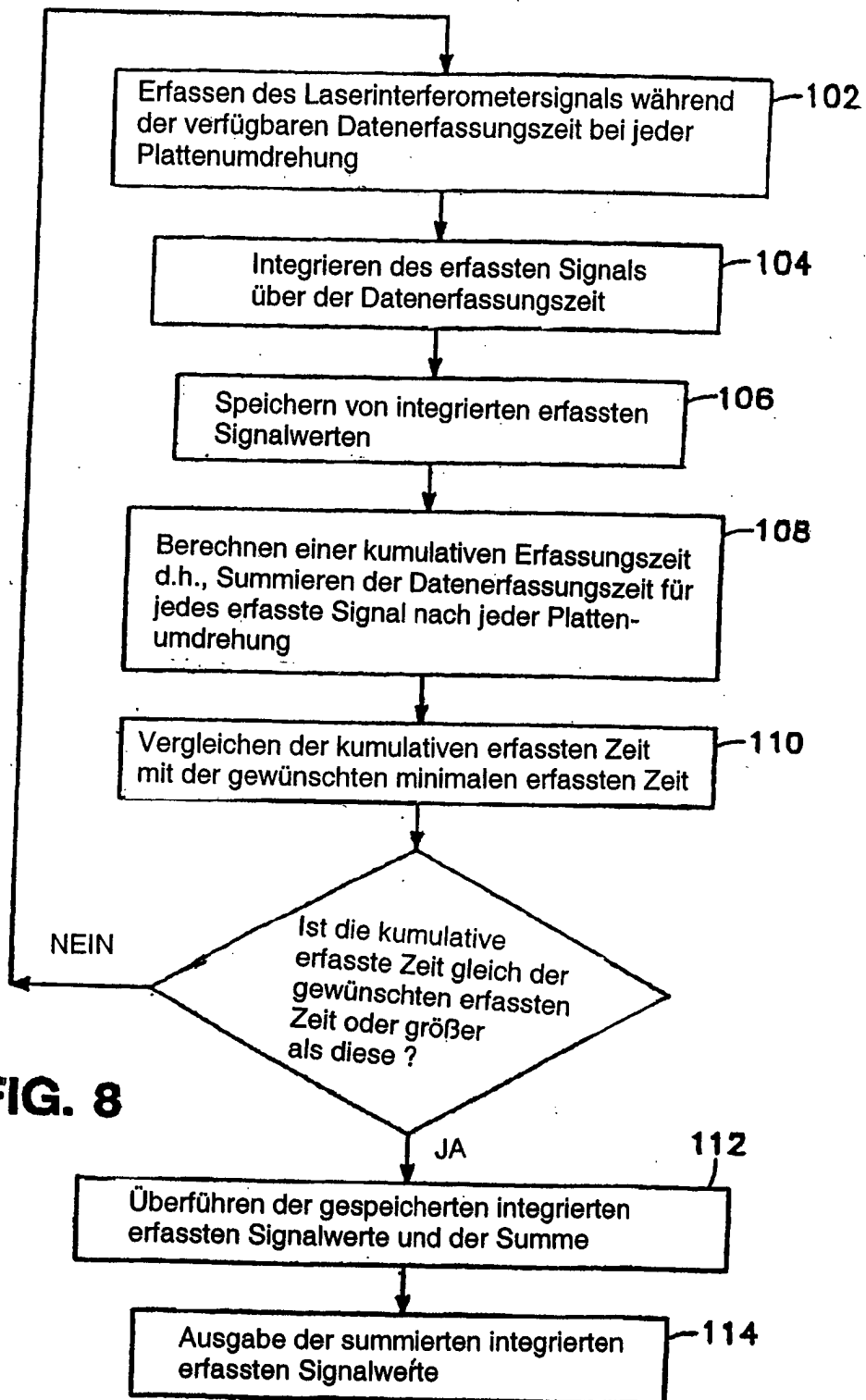


FIG. 8

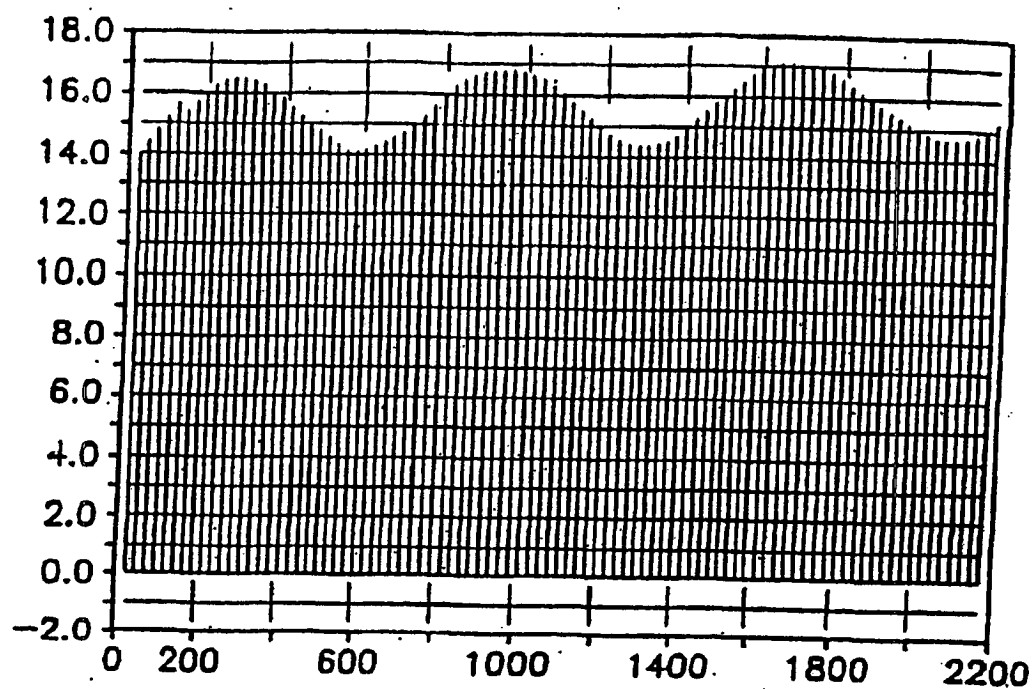


FIG. 9A

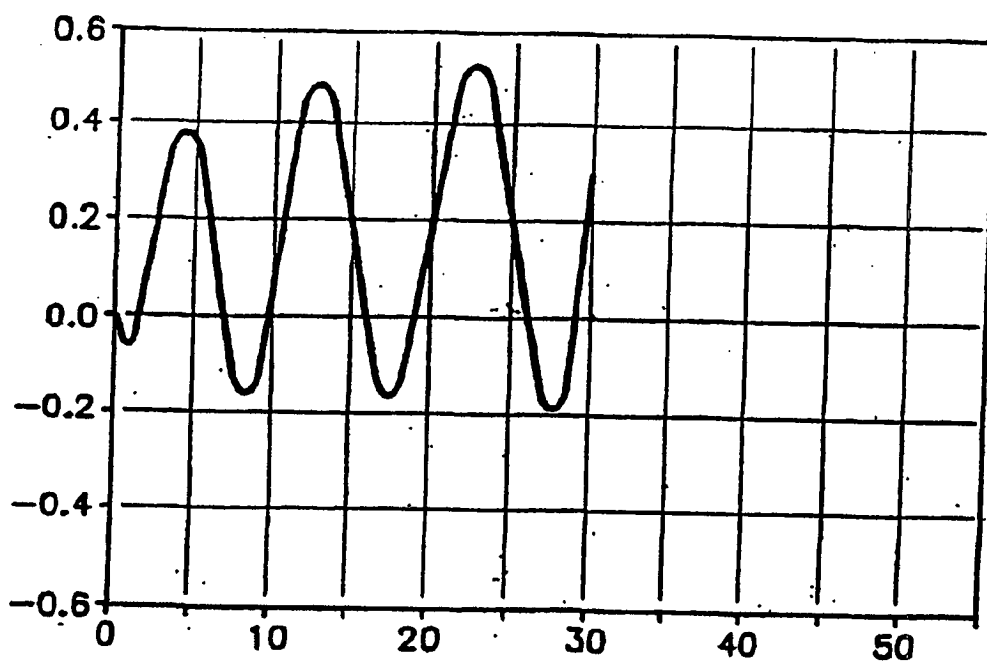
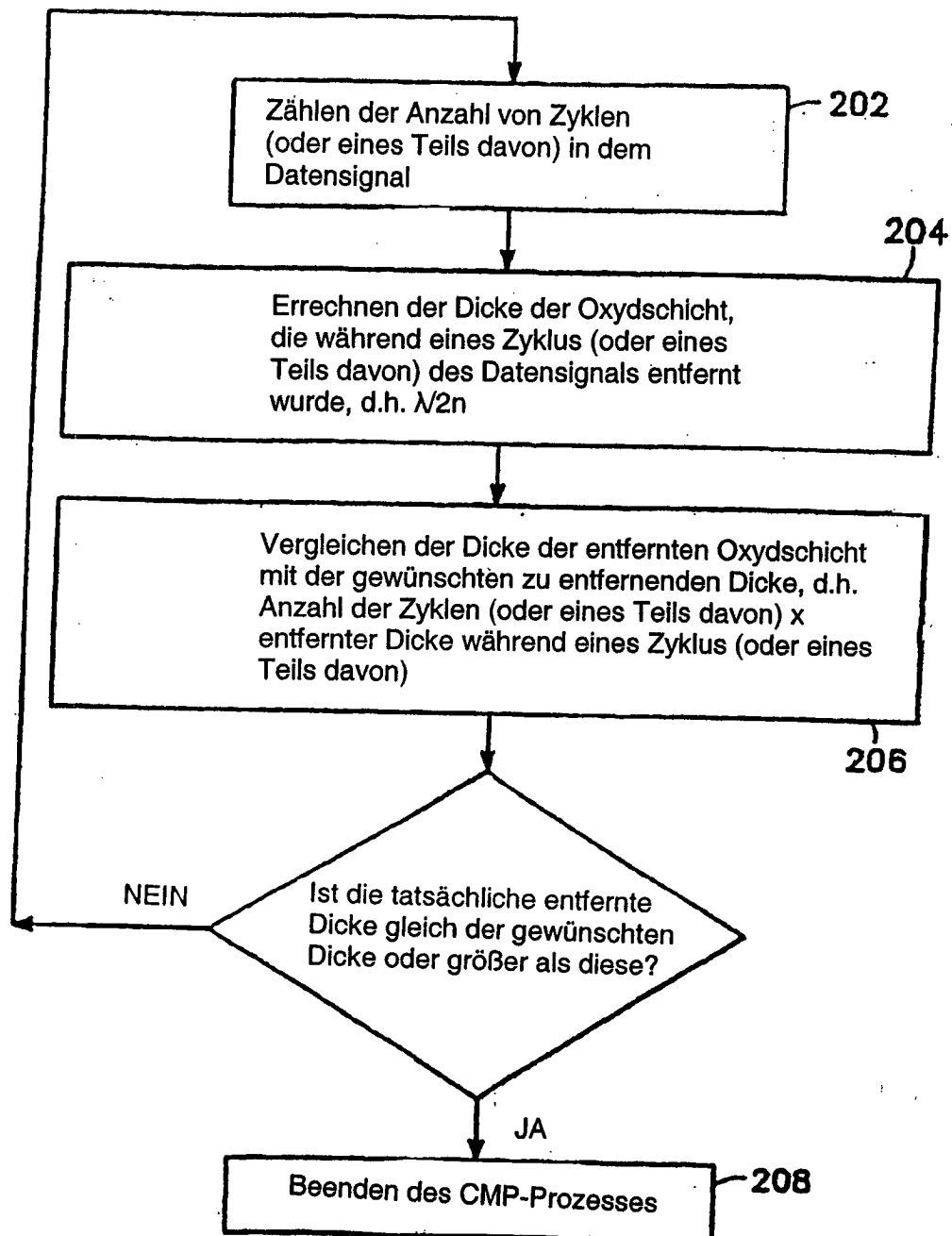
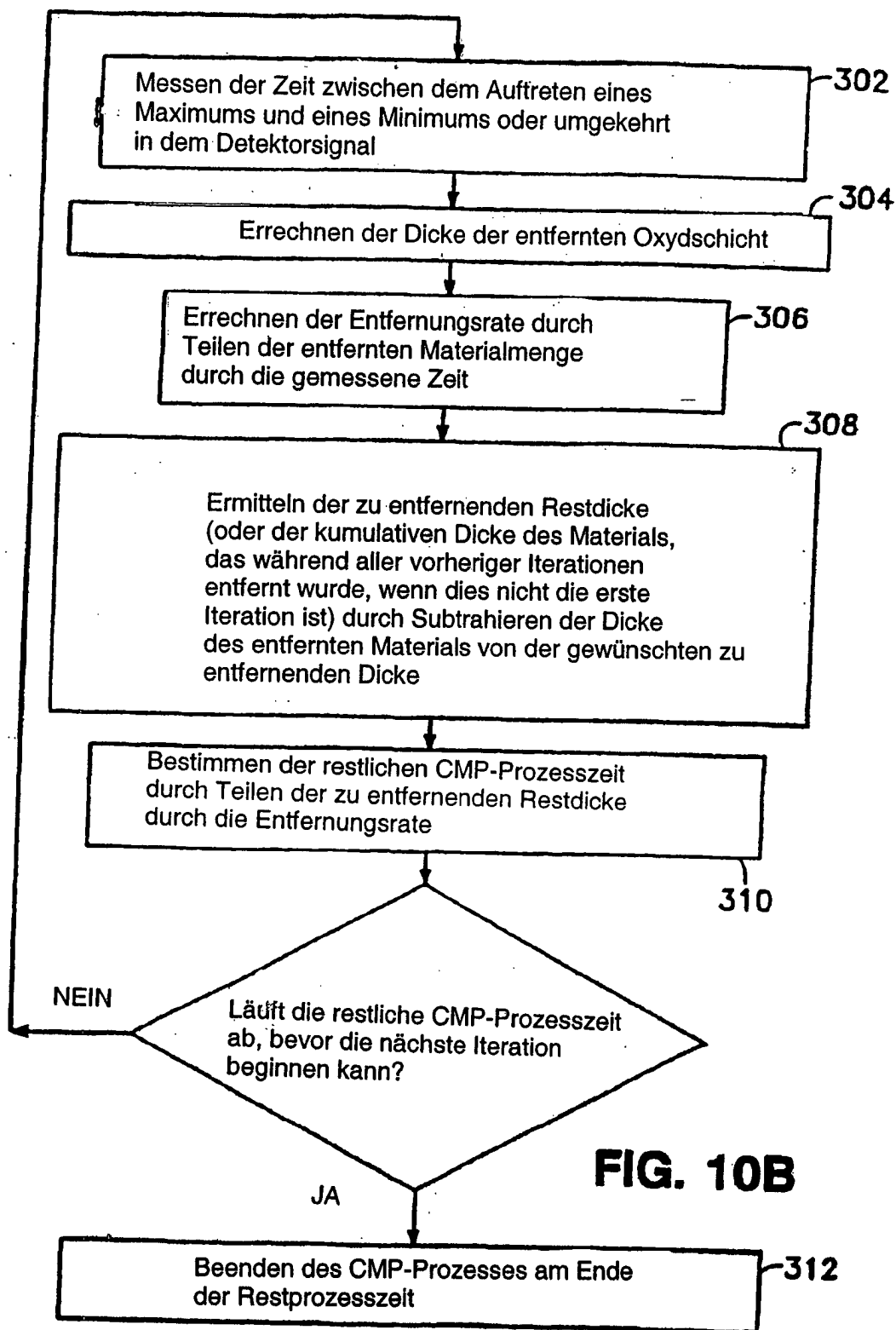


FIG. 9B

**FIG. 10A**



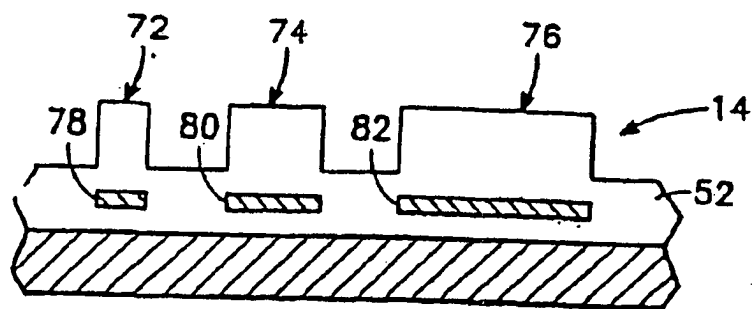


FIG. 11A

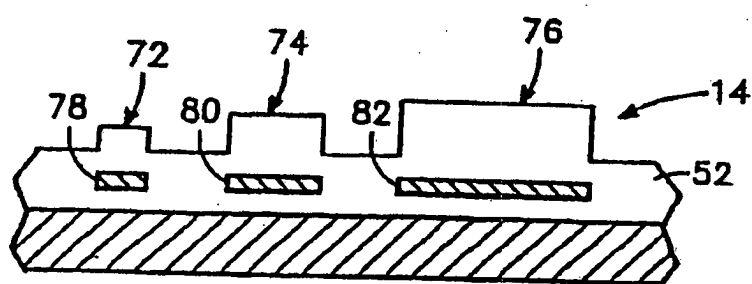


FIG. 11B

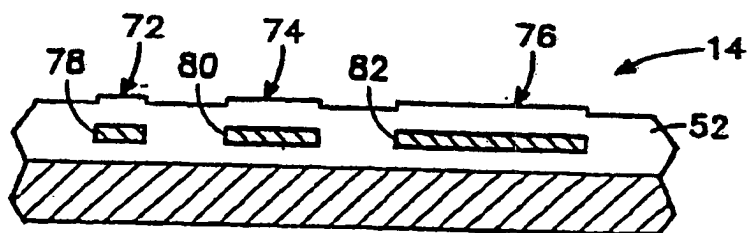


FIG. 11C

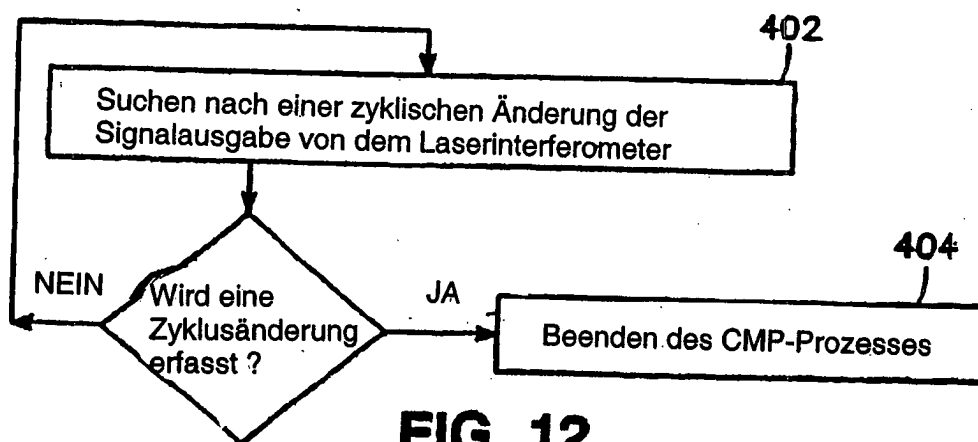


FIG. 12

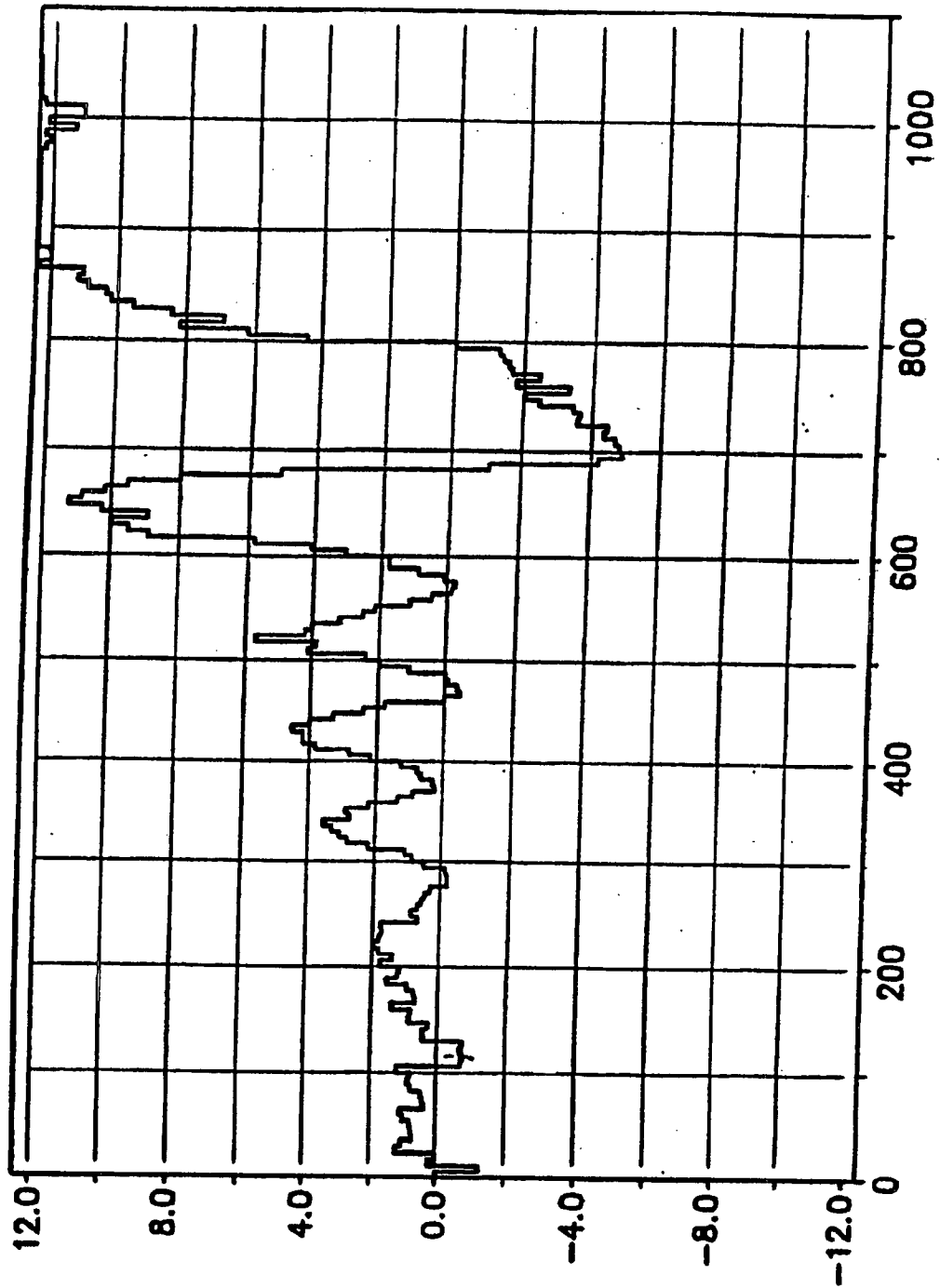


FIG. 13

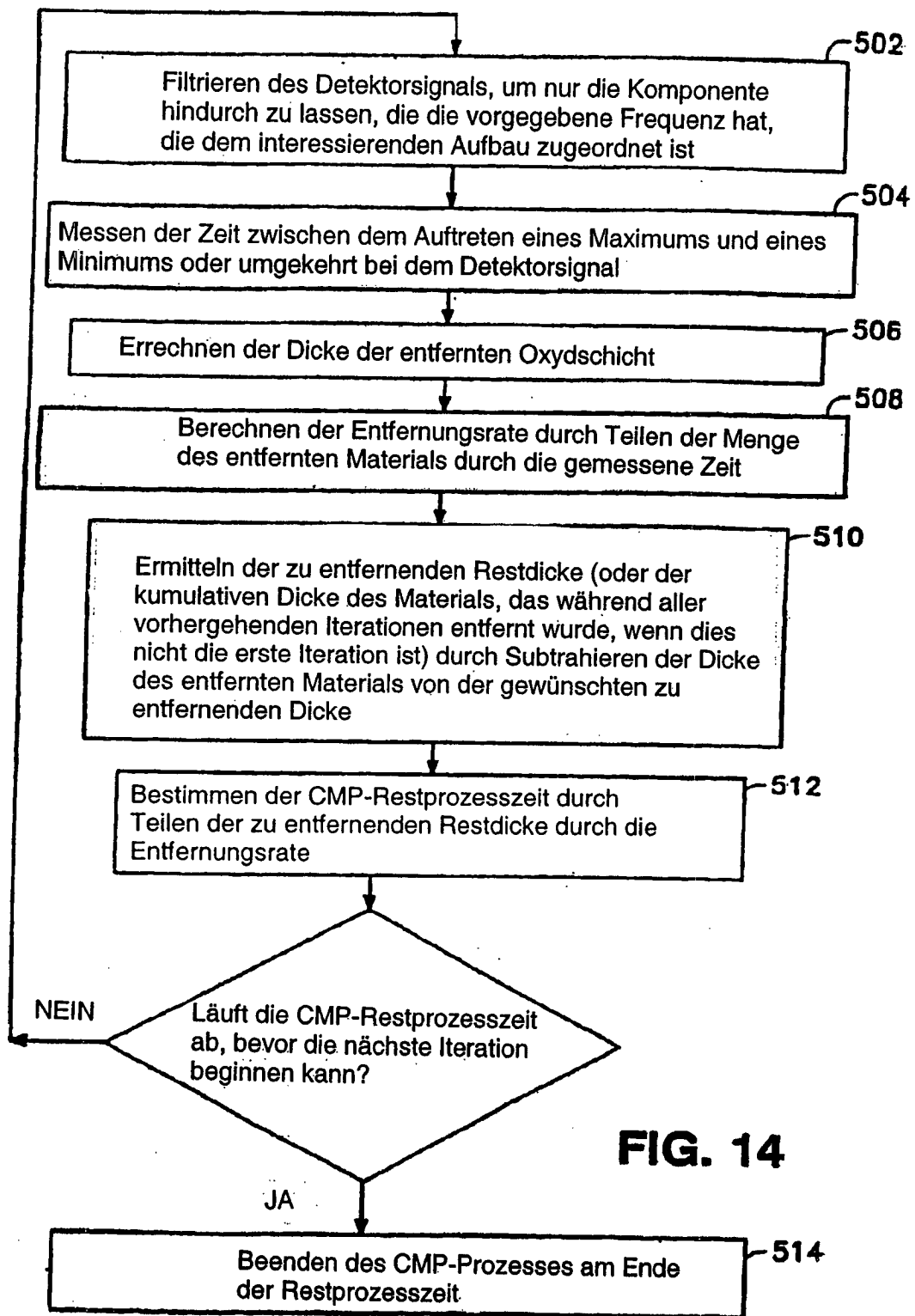
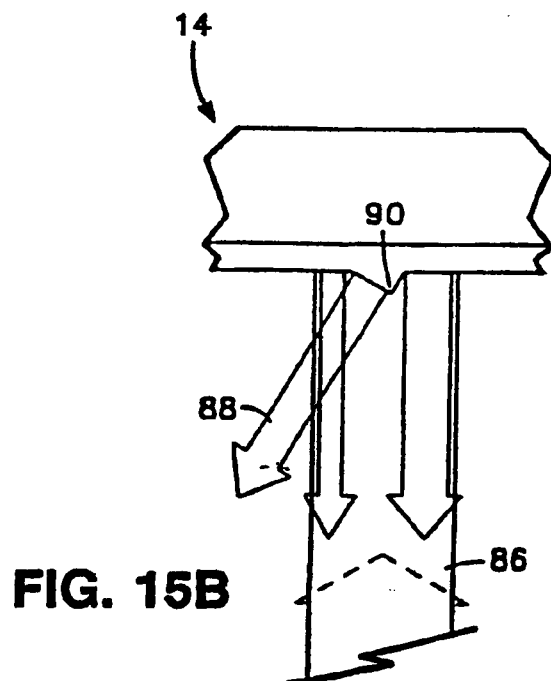
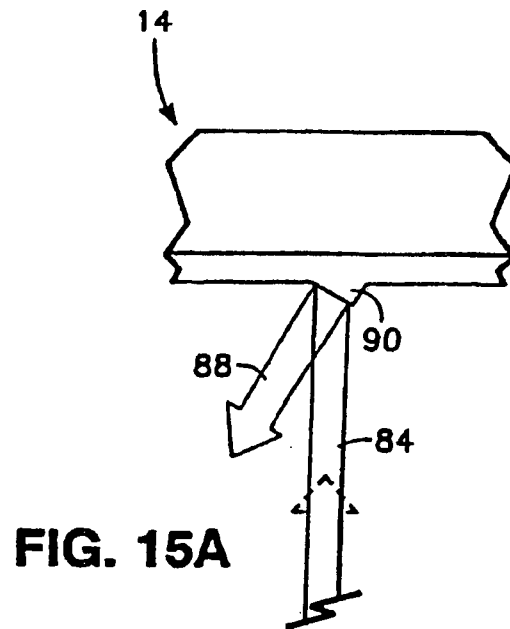
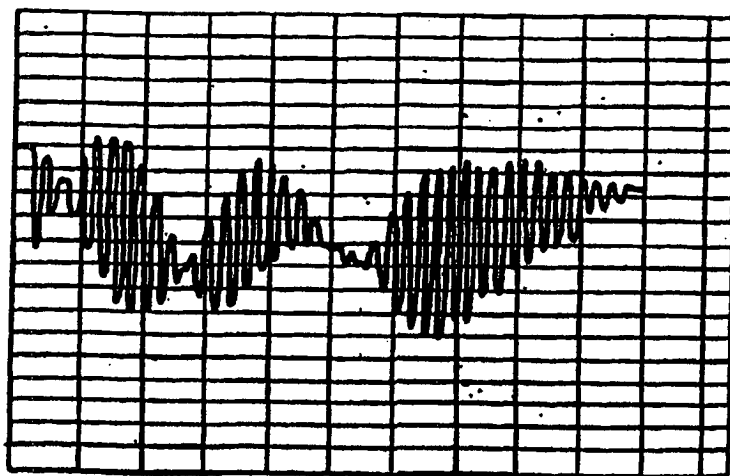
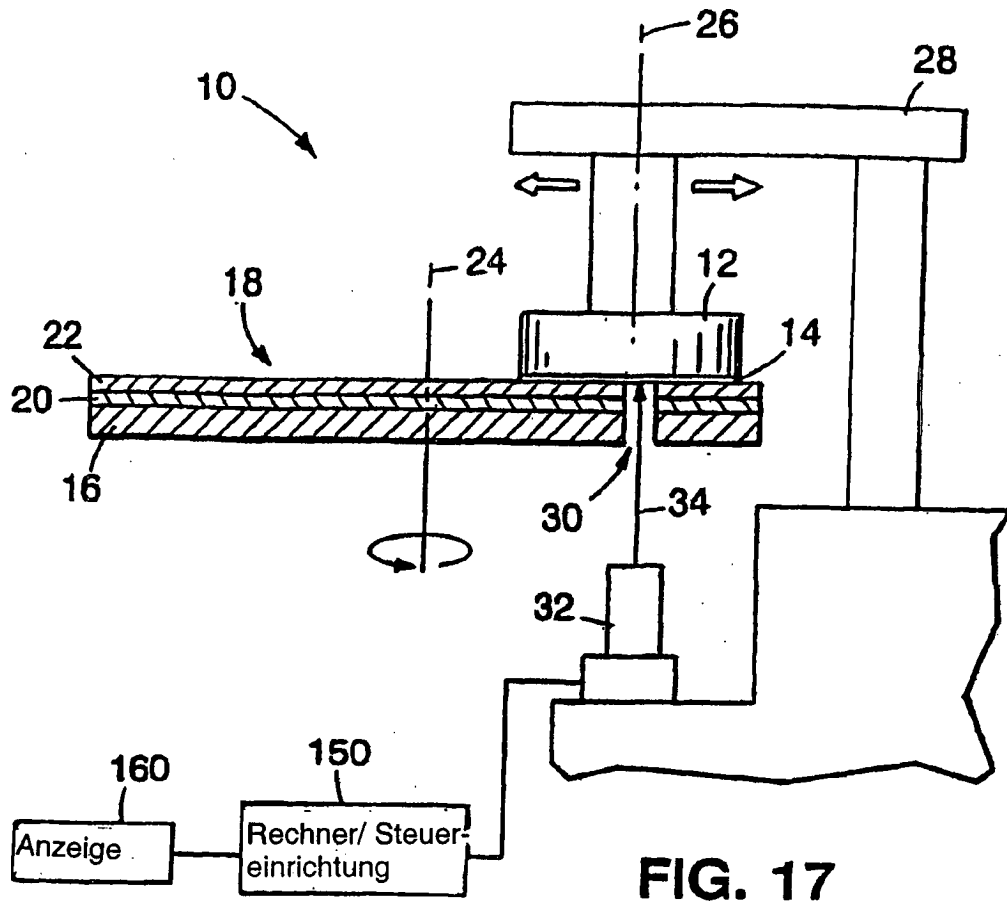
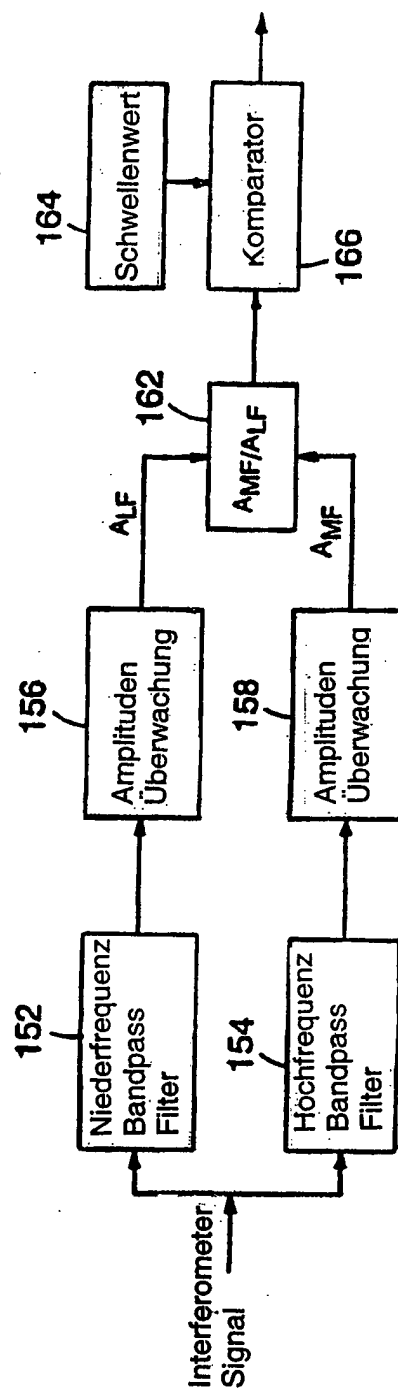
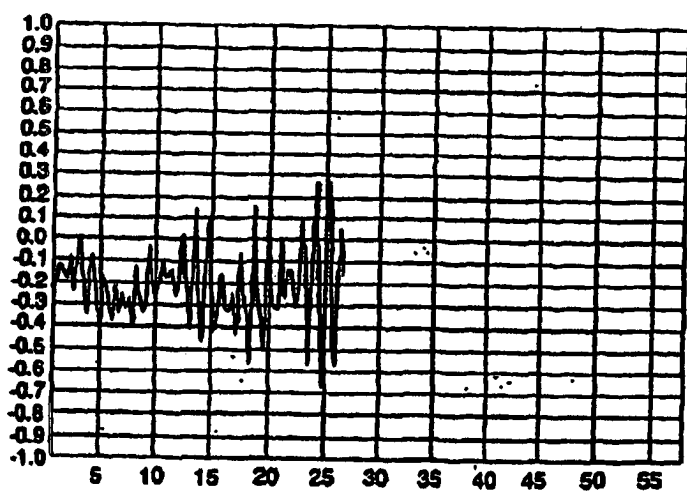
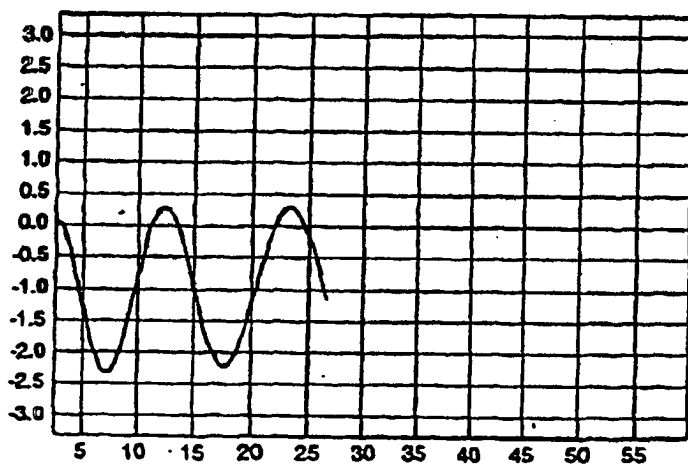
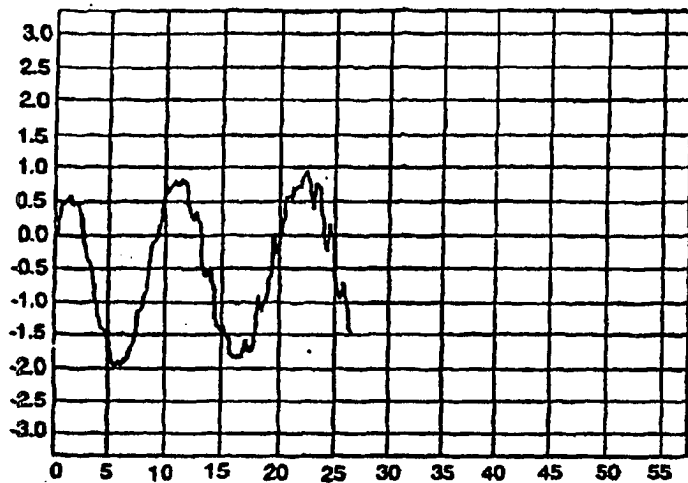


FIG. 14





**FIG. 18**



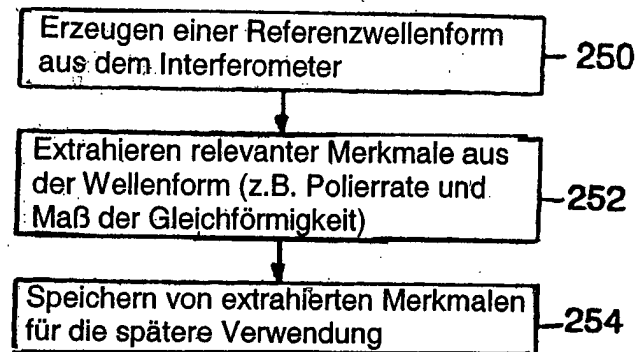


FIG. 20A

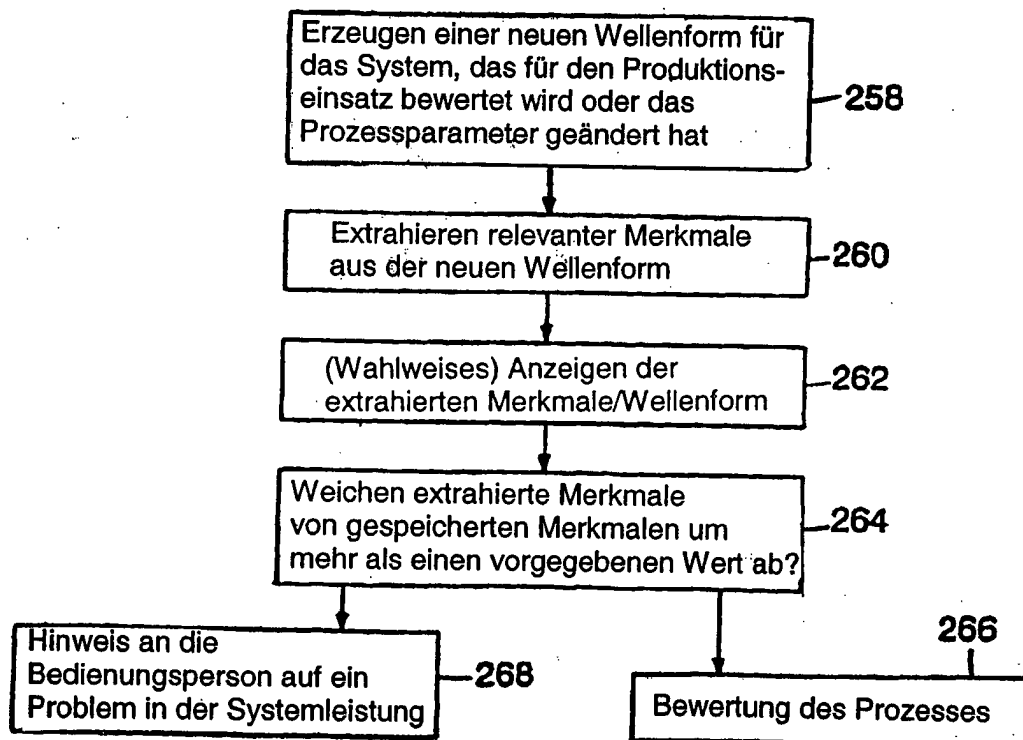


FIG. 20B

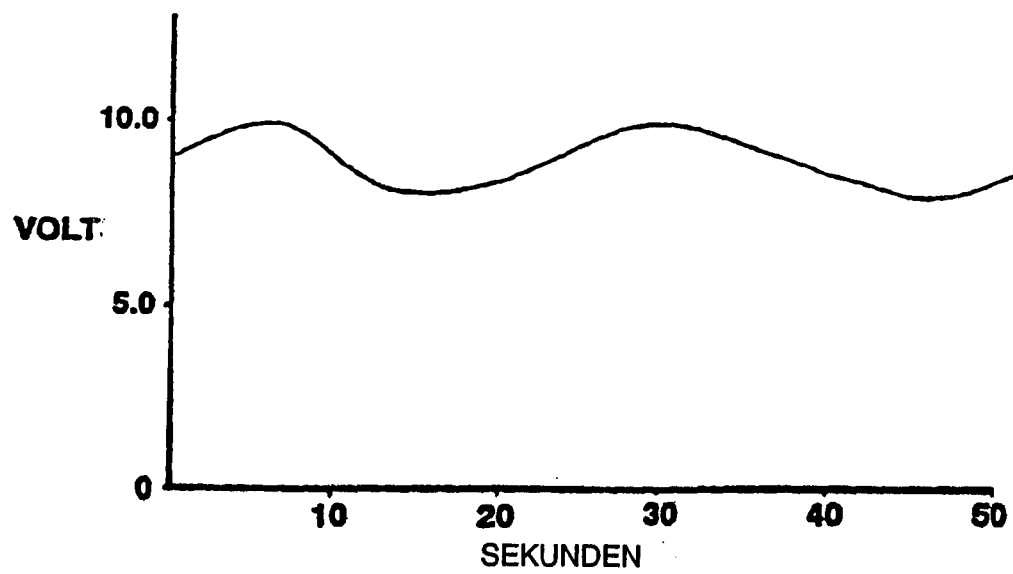
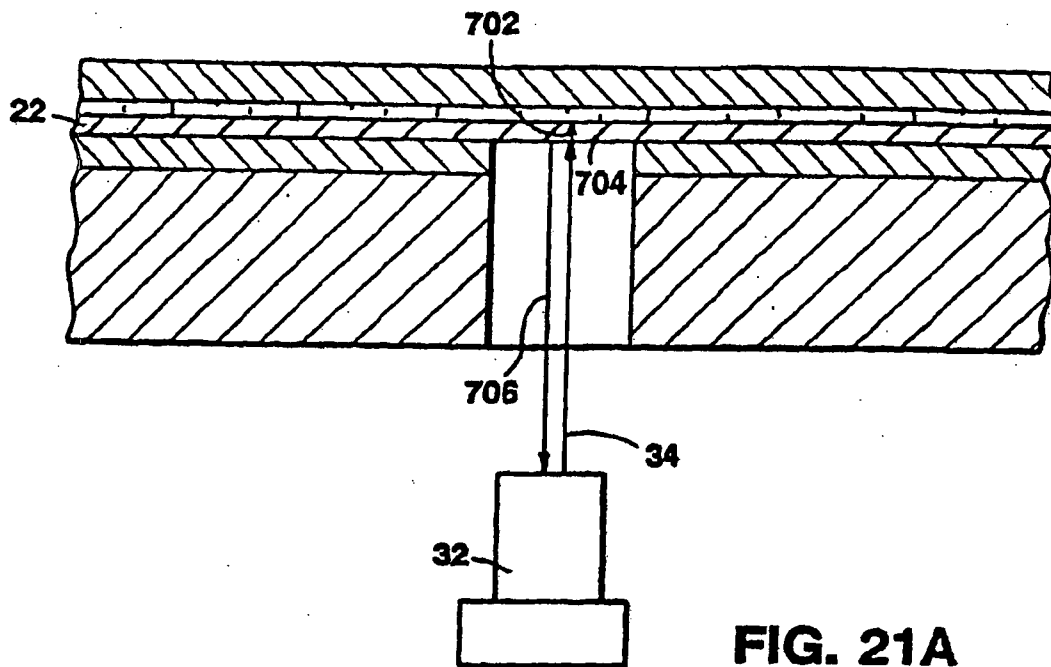


FIG. 21B

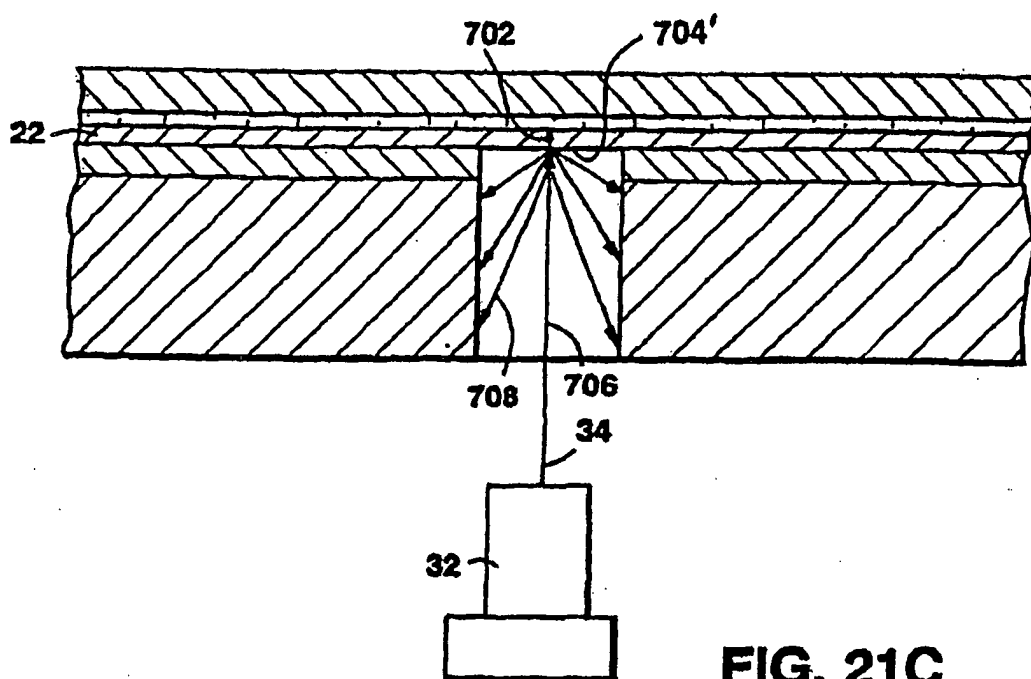


FIG. 21C

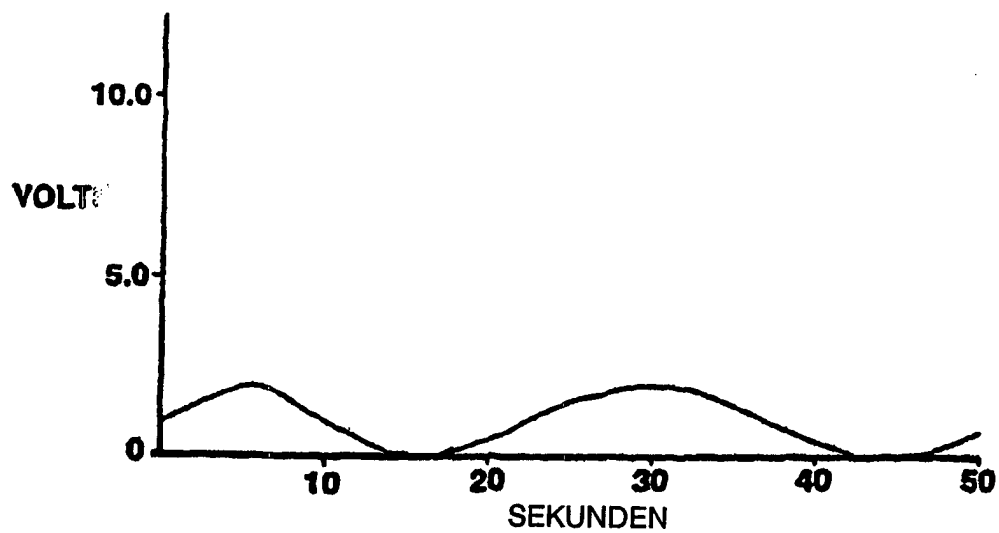


FIG. 21D