(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

PCT

(43) International Publication Date
26 April 2012 (26.04.2012)

(10) International Publication Number
WO 2012/054263 A2

(51) International Patent Classification:
H01L 21/304 (2006.01) *H01L 21/66* (2006.01)

(21) International Application Number:
PCT/US2011/055814

(22) International Filing Date:
11 October 2011 (11.10.2011)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
61/405,110 20 October 2010 (20.10.2010) US

(71) Applicant *(for all designated States except US)*: **APPLIED MATERIALS, INC.** [US/US]; 3050 Bowers Avenue, Santa Clara, California 95054 (US).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: **ZHANG, Jimin** [US/US]; 398 Huckleberry Dr., San Jose, California 95123 (US). **WANG, Zhihong** [CN/US]; 100 Buckingham Drive Apt 234, Santa Clara, California 95051 (US). **LEE, Harry Q.** [US/US]; 1501 Ben Roe Drive, Los Altos, California 94024 (US). **TU, Wen-Chiang** [US/US]; 777 West Middlefield Road, Mountain View, California 94043 (US).

(74) Agent: **GOREN, David J.**; Fish & Richardson P.C., P.O. Box 1022, Minneapolis, Minnesota 55440-1022 (US).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

— *without international search report and to be republished upon receipt of that report (Rule 48.2(g))*

(54) Title: MULTIPLE MATCHING REFERENCE SPECTRA FOR IN-SITU OPTICAL MONITORING

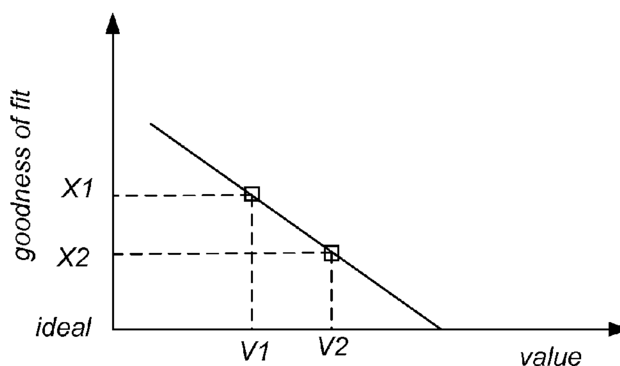

FIG. 13

(57) Abstract: A method of controlling polishing includes storing a plurality libraries, each library including a plurality of reference spectra, polishing a substrate, measuring a sequence of spectra of light from the substrate during polishing, and for each measured spectrum of the sequence of spectra, finding a best matching first reference spectrum from a first library from the plurality of libraries and finding a best matching second reference spectrum from a different second library from the plurality of libraries, determining a first value associated with the best matching first reference spectrum and determining a second value from the best matching second reference spectrum, and calculating a third value from the first value and the second value to generate a sequence of calculated third values. At least one of a polishing endpoint or an adjustment for a polishing rate can be determined based on the sequence of calculated third values.

MULTIPLE MATCHING REFERENCE SPECTRA FOR IN-SITU OPTICAL MONITORING

TECHNICAL FIELD

The present disclosure relates to optical monitoring, e.g., during chemical mechanical polishing of substrates.

BACKGROUND

An integrated circuit is typically formed on a substrate by the sequential deposition of conductive, semiconductive, or insulative layers on a silicon wafer. One fabrication step involves depositing a filler layer over a non-planar surface and planarizing the filler layer. For certain applications, the filler layer is planarized until the top surface of a patterned layer is exposed. A conductive filler layer, for example, can be deposited on a patterned insulative layer to fill the trenches or holes in the insulative layer. After planarization, the portions of the conductive layer remaining between the raised pattern of the insulative layer form vias, plugs, and lines that provide conductive paths between thin film circuits on the substrate. For other applications, such as oxide polishing, the filler layer is planarized until a predetermined thickness is left over the non planar surface. In addition, planarization of the substrate surface is usually required for photolithography.

Chemical mechanical polishing (CMP) is one accepted method of planarization. This planarization method typically requires that the substrate be mounted on a carrier head. The exposed surface of the substrate is typically placed against a rotating polishing pad. The carrier head provides a controllable load on the substrate to push it against the polishing pad. A polishing liquid, such as a slurry with abrasive particles, is typically supplied to the surface of the polishing pad.

One problem in CMP is determining whether the polishing process is complete, i.e., whether a substrate layer has been planarized to a desired flatness or thickness, or when a desired amount of material has been removed. Variations in the initial thickness of the substrate layer, the slurry composition, the polishing pad condition, the relative speed between the polishing pad and the substrate, and the load on the substrate can cause variations in the material removal rate. These variations cause variations in the time needed

to reach the polishing endpoint. Therefore, it may not be possible to determine the polishing endpoint merely as a function of polishing time.

In some systems, a substrate is optically monitored in-situ during polishing, e.g., through a window in the polishing pad. However, existing optical monitoring techniques may not satisfy increasing demands of semiconductor device manufacturers.

SUMMARY

In some optical monitoring processes, a spectrum measured in-situ, e.g., during the polishing processes, is compared to multiple libraries of reference spectra to find the best matching reference spectrum. Unfortunately, computer processing power places a practical limit on the number of reference libraries to which a measured spectrum can be compared while still operating with sufficient speed for in-situ monitoring, and with a limited number of reference libraries, there may still be room for improvement in the accuracy and reliability of the endpoint detection and process control. One approach is to find the best matching reference spectrum from each of two (or more) reference libraries, determine values associated with the best matching spectra, and interpolate between or extrapolate from the values, e.g., using a weighted average based on the goodness of fit of the measured spectrum to the best matching reference spectrum, to generate a calculated value that can be used in endpoint detection and process control.

In one aspect, a method of controlling polishing includes storing a plurality libraries, each library including a plurality of reference spectra, polishing a substrate, measuring a sequence of spectra of light from the substrate during polishing, and for each measured spectrum of the sequence of spectra, finding a best matching first reference spectrum from a first library from the plurality of libraries and finding a best matching second reference spectrum from a different second library from the plurality of libraries, determining a first value associated with the best matching first reference spectrum and determining a second value from the best matching second reference spectrum, and calculating a third value from the first value and the second value to generate a sequence of calculated third values. At least one of a polishing endpoint or an adjustment for a polishing rate can be determined based on the sequence of calculated third values.

Implementations can include on or more of the following features.

Calculating the third value from the first value and the second value may include interpolating between the first value and the second value. Calculating the third value from the first value and the second value may include calculating a weighted average of the first value and the second value. Calculating the weighted average of the first value and the second value may include calculating a first goodness of fit between the best matching first reference spectrum and the measured spectrum, calculating a second goodness of fit between the best matching second reference spectrum and the measured spectrum, and calculating weights for the weighted average based on the first goodness of fit and the second goodness of fit. The goodness of fit may be a sum of squared differences, a sum of absolute differences, or a cross-correlation. Calculating the third value V3 may include calculating $V3=(V1*W1+V2*W2)/(W1+W2)$, where V1 is the first value, V2 is the second value, W1 is a first weight, and W2 is a second weight, and W1 and W2 are calculated based on the first goodness of fit and the second goodness of fit. W1 may equal $1-(X1/(X1+X2))$ and W2 may equal $1-(X2/(X1+X2))$, where X1 is the first goodness of fit and X2 is the second goodness of fit.

Calculating the third value from the first value and the second value may include extrapolating from the first value and the second value. Extrapolating from the first value and the second value may include calculating a first goodness of fit between the best matching first reference spectrum and the measured spectrum, and calculating a second goodness of fit between the best matching second reference spectrum and the measured spectrum, and extrapolating based on the first value, the first goodness of fit, the second value and the second goodness of fit. The goodness of fit may be a sum of squared differences, a sum of absolute differences, or a cross-correlation. Calculating the third value V3 may include calculating $V3=V1-X1*(V1-V2)/(X1-X2)$, where V1 is the first value, V2 is the second value, X1 is the first goodness of fit and X2 is the second goodness of fit.

Whether to interpolate or extrapolate in calculating the third value may be determined. A first goodness of fit between the best matching first reference spectrum and the measured spectrum may be calculated, a second goodness of fit between the best matching second reference spectrum and the measured spectrum may be calculated, and a third goodness of fit between the best matching first reference spectrum and the best

matching second reference spectrum may be calculated. Determining whether to interpolate or extrapolate may include comparing the third goodness of fit to the fist goodness of fit and the second goodness of fit. Calculating the third value may include interpolating between the first value and the second value if the third goodness of fit is worse than the first goodness of fit and the second goodness of fit. Calculating the third value may include extrapolating from the first value and the second value if the third goodness of fit is better than the first goodness of fit or the second goodness of fit.

Finding a best matching first reference spectrum from a first library from the plurality of libraries may include searching only the first library, and finding the best matching second reference spectrum from the different second library from the plurality of libraries may include searching only the second library. The plurality of libraries may include three libraries. Finding a best matching first reference spectrum from a first library from the plurality of libraries may include searching at least two of the three libraries. Finding a best matching first reference spectrum from a first library from the plurality of libraries may include searching the three libraries. The first value and the second value may be index values or thickness values. A linear function may be fit to the sequence of third values. Polishing may be halted when the linear function matches or exceeds a target value. The substrate may include a second layer overlying a first layer, the first layer having a different composition than the second layer, the second layer being polished in the polishing step. The second layer may be a barrier layer and the first layer may be a dielectric layer.The first library may include spectra for the substrate having the first layer of a first thickness and the second library may include spectra for a substrate having the first layer of a different second thickness. Measuring the sequence of spectra of light from the substrate may include making a plurality of sweeps of a sensor across the substrate. Each spectrum from the sequence of spectra may correspond to a single sweep of the sensor from the plurality of sweeps.

The substrate may include a plurality of zones, and a polishing rate of each zone may be independently controllable by an independently variable polishing parameter. A sequence of spectra from each zone may be measured during polishing. For each measured spectrum in the sequence of spectra for each zone, a best matching first reference spectrum from a first library from the plurality of libraries may be found and a best matching second reference spectrum from a different second library from the plurality of libraries may be found, a first

value associated with the best matching first reference spectrum may be determined and a second value from the best matching second reference spectrum may be determined, and a third value may be calculated from the first value and the second value to generate a sequence of calculated third values for each zone. Based on the sequence of calculated third values for each zone, the polishing parameter for at least one zone may be adjusted to adjust the polishing rate of the at least one zone such that the plurality of zones have a smaller difference in thickness at the polishing endpoint than without such adjustment.

In another aspect, a computer program product, tangibly embodied in a machine readable storage device, includes instructions to carry out the method.

Implementations may optionally include one or more of the following advantages. The optical monitoring system can be less sensitive to variations in thickness in layers underlying the layer being polished. Reliability of the endpoint system to detect a desired polishing endpoint can be improved, and within-wafer and wafer-to-wafer thickness non-uniformity (WIWNU and WTWNU) can be reduced.

The details of one or more embodiments are set forth in the accompanying drawings and the description below. Other features, aspects, and advantages will become apparent from the description, the drawings, and the claims.

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. 1A and 1B are schematic cross-sectional views of a substrate before and after polishing.

FIG. 2 illustrates a schematic cross-sectional view of an example of a polishing apparatus.

FIG. 3 illustrates a schematic top view of a substrate having multiple zones.

FIG. 4 illustrates a top view of a polishing pad and shows locations where in-situ measurements are taken on a substrate.

FIG. 5 illustrates a measured spectrum from the in-situ optical monitoring system.

FIG. 6 illustrates a library of reference spectra.

FIG. 7 illustrates a value trace.

FIG. 8 illustrates a value trace having a linear function fit to values collected after clearance of an overlying layer is detected.

FIG. 9 is a flow diagram of an example process for fabricating a substrate and detecting a polishing endpoint.

FIG. 10 illustrates a plurality of value traces.

FIG. 11 illustrates a calculation of a plurality of desired slopes for a plurality of
5 adjustable zones based on a time that a value trace of a reference zone reaches a target value.

FIG. 12 illustrates a calculation of an endpoint for based on a time that a value trace of a reference zone reaches a target value.

FIG. 13 illustrates calculation of a combined value for use in a value trace.

Like reference numbers and designations in the various drawings indicate like
10 elements.

DETAILED DESCRIPTION

One optical monitoring technique is to measure spectrum of light reflected from a substrate during polishing, and identify a best matching reference spectrum from a library of reference spectra. A sequence of best matching reference spectra provide a series of values,
15 e.g., index values, and a function, e.g., a line, is fit to the series of values. The projection of the function to a target value can be used to determine endpoint or to change a polishing rate.

For polishing of some substrates, e.g., where there variations in the thickness of one or more layers underlying the layer being polished, the matching algorithm may be unreliable. Without being limited to any particular theory, if a thickness of an underlying
20 layer , i.e., a layer underlying the layer being polished, varies from the thickness used for generation of the reference spectra, then none of the reference spectra may provide a good match. One technique to counter this effect is to use multiple libraries of reference spectra, with the different libraries containing spectra representing different thicknesses of the underlying layer. Unfortunately, as noted above, computer processing power places a
25 practical limit on the number of reference libraries (and thus the number of different thicknesses of the underlying layer) to which a measured spectrum can be compared while still operating with sufficient speed for in-situ monitoring. With the ever increasing needs of semiconductor device manufacturers, there may still be room for improvement in the accuracy and reliability of the endpoint detection and process control.

30 However, it may be possible to improve accuracy and reliability of the endpoint detection and process control while maintaining practical computer processing loads by

finding the best matching reference spectrum from each of two (or more) reference libraries. That is, for each measured spectrum, a first best matching reference spectrum can be determined from a first library containing reference spectra representing a first thicknesses of the underlying layer, and a second best matching reference spectrum can be determined from a second library containing reference spectra representing a different second thicknesses of the underlying layer.

A “calculated” value can be calculated from the values associated with the best matching reference spectra, e.g., by interpolating between or extrapolating from the values. For example, a weighted average of the values can be calculated with the weights based on the goodness of fit of the measured spectrum to the best matching reference spectrum. For example, a first value associated with the first best matching reference spectrum and a second value associated with the second best matching reference spectrum can be determined, and a weighted average of the first value and the second value can be calculated to provide the calculated value. The sequence of pairs of best matching reference spectra provide a sequence of calculated values, and a function, e.g., a line, is fit to the series of calculated values. The projection of the function to a target value can be used to determine endpoint or to change a polishing rate.

As an example, referring to FIG. 1A, a substrate 10 can include a patterned first layer 12 (this layer can also be referred to as an underlying layer) of a first dielectric material, e.g., a low-k material, e.g., carbon doped silicon dioxide, e.g., Black Diamond™ (from Applied Materials, Inc.) or Coral™ (from Novellus Systems, Inc.). Disposed over the first layer 12 is a second layer 16 (this layer can also be referred to as an overlying layer) of a different second dielectric material, e.g., a barrier layer, e.g., a nitride, e.g., tantalum nitride or titanium nitride. Optionally disposed between the first layer and the second layer are one or more additional layers 14 of another dielectric material, different from both the first and second dielectric materials, e.g., a low-k capping material, e.g., a material formed from tetraethyl orthosilicate (TEOS). Together, the first layer 12 and the one or more additional layers 14 provide a layer stack below the second layer. Disposed over the second layer (and in trenches provided by the pattern of the first layer) is a conductive material 18, e.g., a metal, e.g., copper.

Chemical mechanical polishing can be used to planarize the substrate until the first layer of the first dielectric material is exposed. For example, referring to FIG. 1B, after planarization, the portions of the conductive material 18 remaining between the raised pattern of the first layer 12 form vias and the like. In addition, it is sometimes desired to remove the first dielectric material until a target thickness remains or a target amount of material has been removed.

One method of polishing is to polish the conductive material on a first polishing pad at least until the second layer, e.g., the barrier layer, is exposed. In addition, a portion of the thickness of the second layer can be removed, e.g., during an overpolishing step at the first polishing pad. The substrate is then transferred to a second polishing pad, where the second layer, e.g., the barrier layer is completely removed, and a portion of the thickness of the underlying first layer, e.g., the low-k dielectric, is also removed. In addition, if present, the additional layer or layers, e.g., the capping layer, between the first and second layer can be removed in the same polishing operation at the second polishing pad.

FIG. 2 illustrates an example of a polishing apparatus 100. The polishing apparatus 100 includes a rotatable disk-shaped platen 120 on which a polishing pad 110 is situated. The platen is operable to rotate about an axis 125. For example, a motor 121 can turn a drive shaft 124 to rotate the platen 120. The polishing pad 110 can be a two-layer polishing pad with an outer polishing layer 112 and a softer backing layer 114.

The polishing apparatus 100 can include a port 130 to dispense polishing liquid 132, such as a slurry, onto the polishing pad 110 to the pad. The polishing apparatus can also include a polishing pad conditioner to abrade the polishing pad 110 to maintain the polishing pad 110 in a consistent abrasive state.

The polishing apparatus 100 includes one or more carrier heads 140. Each carrier head 140 is operable to hold a substrate 10 against the polishing pad 110. Each carrier head 140 can have independent control of the polishing parameters, for example pressure, associated with each respective substrate.

In particular, each carrier head 140 can include a retaining ring 142 to retain the substrate 10 below a flexible membrane 144. Each carrier head 140 also includes a plurality of independently controllable pressurizable chambers defined by the membrane, e.g., 3 chambers 146a-146c, which can apply independently controllable pressurizes to associated

zones 148a-148c on the flexible membrane 144 and thus on the substrate 10 (see FIG. 3). Referring to FIG. 2, the center zone 148a can be substantially circular, and the remaining zones 148b-148e can be concentric annular zones around the center zone 148a. Although only three chambers are illustrated in FIGS. 2 and 3 for ease of illustration, there could be one or two chambers, or four or more chambers, e.g., five chambers.

Returning to FIG. 2, each carrier head 140 is suspended from a support structure 150, e.g., a carousel, and is connected by a drive shaft 152 to a carrier head rotation motor 154 so that the carrier head can rotate about an axis 155. Optionally each carrier head 140 can oscillate laterally, e.g., on sliders on the carousel 150; or by rotational oscillation of the carousel itself. In operation, the platen is rotated about its central axis 125, and each carrier head is rotated about its central axis 155 and translated laterally across the top surface of the polishing pad.

While only one carrier head 140 is shown, more carrier heads can be provided to hold additional substrates so that the surface area of polishing pad 110 may be used efficiently. Thus, the number of carrier head assemblies adapted to hold substrates for a simultaneous polishing process can be based, at least in part, on the surface area of the polishing pad 110.

The polishing apparatus also includes an in-situ optical monitoring system 160, , e.g., a spectrographic monitoring system, which can be used to determine whether to adjust a polishing rate or an adjustment for the polishing rate as discussed below. An optical access through the polishing pad is provided by including an aperture (i.e., a hole that runs through the pad) or a solid window 118. The solid window 118 can be secured to the polishing pad 110, e.g., as a plug that fills an aperture in the polishing pad, e.g., is molded to or adhesively secured to the polishing pad, although in some implementations the solid window can be supported on the platen 120 and project into an aperture in the polishing pad.

The optical monitoring system 160 can include a light source 162, a light detector 164, and circuitry 166 for sending and receiving signals between a remote controller 190, e.g., a computer, and the light source 162 and light detector 164. One or more optical fibers can be used to transmit the light from the light source 162 to the optical access in the polishing pad, and to transmit light reflected from the substrate 10 to the detector 164. For example, a bifurcated optical fiber 170 can be used to transmit the light from the light source 162 to the substrate 10 and back to the detector 164. The bifurcated optical fiber an include a

trunk 172 positioned in proximity to the optical access, and two branches 174 and 176 connected to the light source 162 and detector 164, respectively.

In some implementations, the top surface of the platen can include a recess 128 into which is fit an optical head 168 that holds one end of the trunk 172 of the bifurcated fiber. The optical head 168 can include a mechanism to adjust the vertical distance between the top of the trunk 172 and the solid window 118.

The output of the circuitry 166 can be a digital electronic signal that passes through a rotary coupler 129, e.g., a slip ring, in the drive shaft 124 to the controller 190 for the optical monitoring system. Similarly, the light source can be turned on or off in response to control commands in digital electronic signals that pass from the controller 190 through the rotary coupler 129 to the optical monitoring system 160. Alternatively, the circuitry 166 could communicate with the controller 190 by a wireless signal.

The light source 162 can be operable to emit white light. In one implementation, the white light emitted includes light having wavelengths of 200-800 nanometers. A suitable light source is a xenon lamp or a xenon mercury lamp.

The light detector 164 can be a spectrometer. A spectrometer is an optical instrument for measuring intensity of light over a portion of the electromagnetic spectrum. A suitable spectrometer is a grating spectrometer. Typical output for a spectrometer is the intensity of the light as a function of wavelength (or frequency).

As noted above, the light source 162 and light detector 164 can be connected to a computing device, e.g., the controller 190, operable to control their operation and receive their signals. The computing device can include a microprocessor situated near the polishing apparatus, e.g., a programmable computer. With respect to control, the computing device can, for example, synchronize activation of the light source with the rotation of the platen 120.

In some implementations, the light source 162 and detector 164 of the in-situ monitoring system 160 are installed in and rotate with the platen 120. In this case, the motion of the platen will cause the sensor to scan across each substrate. In particular, as the platen 120 rotates, the controller 190 can cause the light source 162 to emit a series of flashes starting just before and ending just after the optical access passes below the substrate 10. Alternatively, the computing device can cause the light source 162 to emit light continuously

starting just before and ending just after each substrate 10 passes over the optical access. In either case, the signal from the detector can be integrated over a sampling period to generate spectra measurements at a sampling frequency.

In operation, the controller 190 can receive, for example, a signal that carries information describing a spectrum of the light received by the light detector for a particular flash of the light source or time frame of the detector. Thus, this spectrum is a spectrum measured in-situ during polishing.

As shown by in FIG. 4, if the detector is installed in the platen, due to the rotation of the platen (shown by arrow 204), as the window 108 travels below a carrier head, the optical monitoring system making spectra measurements at a sampling frequency will cause the spectra measurements to be taken at locations 201 in an arc that traverses the substrate 10. For example, each of points 201a-201k represents a location of a spectrum measurement by the monitoring system (the number of points is illustrative; more or fewer measurements can be taken than illustrated, depending on the sampling frequency). The sampling frequency can be selected so that between five and twenty spectra are collected per sweep of the window 108. For example, the sampling period can be between 3 and 100 milliseconds.

As shown, over one rotation of the platen, spectra are obtained from different radii on the substrate 10. That is, some spectra are obtained from locations closer to the center of the substrate 10 and some are closer to the edge. Thus, for any given scan of the optical monitoring system across a substrate, based on timing, motor encoder information, and optical detection of the edge of the substrate and/or retaining ring, the controller 190 can calculate the radial position (relative to the center of the substrate being scanned) for each measured spectrum from the scan. The polishing system can also include a rotary position sensor, e.g., a flange attached to an edge of the platen that will pass through a stationary optical interrupter, to provide additional data for determination of which substrate and the position on the substrate of the measured spectrum. The controller can thus associate the various measured spectra with the controllable zones 148b-148e (see FIG. 2) on the substrates 10a and 10b. In some implementations, the time of measurement of the spectrum can be used as a substitute for the exact calculation of the radial position.

Over multiple rotations of the platen, for each zone, a sequence of spectra can be obtained over time. Without being limited to any particular theory, the spectrum of light

reflected from the substrate 10 evolves as polishing progresses (e.g., over multiple rotations of the platen, not during a single sweep across the substrate) due to changes in the thickness of the outermost layer, thus yielding a sequence of time-varying spectra. Moreover, particular spectra are exhibited by particular thicknesses of the layer stack.

5 In some implementations, the controller, e.g., the computing device, can be programmed to compare a measured spectrum to multiple reference spectra and to determine which reference spectra provide the best match. In particular, the controller can be programmed to compare each spectrum from a sequence of measured spectra from each zone to multiple reference spectra to generate a sequence of best matching reference spectra for 10 each zone.

As used herein, a reference spectrum is a predefined spectrum generated prior to polishing of the substrate. A reference spectrum can have a pre-defined association, i.e., defined prior to the polishing operation, with a value representing a time in the polishing process at which the spectrum is expected to appear, assuming that the actual polishing rate 15 follows an expected polishing rate. Alternatively or in addition, the reference spectrum can have a pre-defined association with a value of a substrate property, such as a thickness of the outermost layer.

A reference spectrum can be generated empirically, e.g., by measuring the spectra from a test substrate, e.g., a test substrate having a known initial layer thicknesses. For 20 example, to generate a plurality of reference spectra, a set-up substrate is polished using the same polishing parameters that would be used during polishing of device wafers while a sequence of spectra are collected. For each spectrum, a value is recorded representing the time in the polishing process at which the spectrum was collected. For example, the value can be an elapsed time, or a number of platen rotations. The substrate can be overpolished, 25 i.e., polished past a desired thickness, so that the spectrum of the light that reflected from the substrate when the target thickness is achieved can be obtained.

In order to associate each spectrum with a value of a substrate property, e.g., a thickness of the outermost layer, the initial spectra and property of a “set-up” substrate with the same pattern as the product substrate can be measured pre-polish at a metrology station. 30 The final spectrum and property can also be measured post-polish with the same metrology station or a different metrology station. The properties for spectra between the initial spectra

and final spectra can be determined by interpolation, e.g., linear interpolation based on elapsed time at which the spectra of the test substrate was measured.

In addition to being determined empirically, some or all of the reference spectra can be calculated from theory, e.g., using an optical model of the substrate layers. For example, and optical model can be used to calculate a reference spectrum for a given outer layer thickness D. A value representing the time in the polishing process at which the reference spectrum would be collected can be calculated, e.g., by assuming that the outer layer is removed at a uniform polishing rate. For example, the time Ts for a particular reference spectrum can be calculated simply by assuming a starting thickness D0 and uniform polishing rate R ($Ts = (D0-D)/R$). As another example, linear interpolation between measurement times T1, T2 for the pre-polish and post-polish thicknesses D1, D2 (or other thicknesses measured at the metrology station) based on the thickness D used for the optical model can be performed ($Ts = T2-T1*(D1-D)/(D1-D2)$).

Referring to FIGS. 5 and 6, a measured spectrum 300 (see FIG. 5) can be compared to reference spectra 320 from two or more libraries 310 (see FIG. 6). As used herein, a library of reference spectra is a collection of reference spectra which represent substrates that share a property in common. However, the property shared in common in a single library may vary across multiple libraries of reference spectra. For example, different libraries can include reference spectra that represent substrates with different underlying layer thicknesses. For a given library of reference spectra, variations in the upper layer thickness, rather than other factors (such as differences in wafer pattern, underlying layer thickness, or layer composition), can be primarily responsible for the differences in the spectral intensities.

Reference spectra 320 for different libraries 310 can be generated by polishing multiple "set-up" substrates with different substrate properties (e.g., underlying layer thicknesses, or layer composition) and collecting spectra as discussed above; the spectra from one set-up substrate can provide a first library and the spectra from another set-up substrate with a different underlying layer thickness can provide a second library. Alternatively or in addition, reference spectra for different libraries can be calculated from theory, e.g., spectra for a first library can be calculated using the optical model with the underlying layer having a first thickness, and spectra for a second library can be calculated using the optical model with the underlying layer having a different second thickness.

It is desirable for at least two of the libraries to effectively span the expected variations in thickness of the underlying layer in the product substrates being polished. To achieve this, one library can be based on a substrate with an underlying layer having near the maximum expected thickness, and another library can be based on a substrate with an underlying layer having near the minimum expected thickness. If the reference spectra are generated based on an optical model, then the minimum expected thickness of the underlying layer can be used in the optical model to generate one library, and the maximum expected thickness of the underlying layer can be used in the optical model to generate the other library.

If thickness measurements of the underlying layer are available for a large number of set-up substrates, then it is possible to simply pick the two set-up substrates having the largest and smallest underlying layer thickness. Spectra collected from the substrate with the largest underlying layer thickness during the set-up process can become one library, and spectra collected from the substrate with the smallest underlying layer thickness during the set-up process can become the other library.

If thickness measurements are not available, it may still be possible to select two substrates. Assuming a group of set-up substrates has been polished, then two set-up substrates can be selected from the group. In particular, for each set-up substrate, the sequence of spectra from the set-up substrate can be assumed to provide an assumed library. Endpoint times can be calculated for all of the other set-up substrates (i.e., the set-up substrates other than the one providing the assumed library) based on the assumed library, and an average endpoint time calculated. This is performed for each substrate, so that an average endpoint time of the other substrates is calculated for each substrate. The two set-up substrates that result in the largest and smallest average endpoint times can be used as the two set-up substrates to provide the two libraries.

Each reference spectrum is associated with a value. In some implementations, each reference spectrum 320 is assigned an index value 330. In general, each library 310 can include many reference spectra 320, e.g., one or more, e.g., exactly one, reference spectra for each platen rotation over the expected polishing time of the substrate. This index 330 can be the value, e.g., a number, representing the time in the polishing process at which the reference spectrum 320 is expected to be observed. The spectra can be indexed so that each

spectrum in a particular library has a unique index value. The indexing can be implemented so that the index values are sequenced in an order in which the spectra of a test substrate were measured. An index value can be selected to change monotonically, e.g., increase or decrease, as polishing progresses. In particular, the index values of the reference spectra can be selected so that they form a linear function of time or number of platen rotations (assuming that the polishing rate follows that of the model or test substrate used to generate the reference spectra in the library). For example, the index value can be proportional, e.g., equal, to a number of platen rotations at which the reference spectra was measured for the test substrate or would appear in the optical model. Thus, each index value can be a whole number. The index number can represent the expected platen rotation at which the associated spectrum would appear. Alternatively, in some implementations, each reference spectrum 320 is assigned a thickness value 330.

The reference spectra and associated index values can be stored in a reference library. For example, each reference spectrum 320 and its associated index value 330 can be stored in a record 340 of database 350. The database 350 of reference libraries of reference spectra can be implemented in memory of the computing device of the polishing apparatus.

As noted above, the controller 190 can be programmed to, for each zone of the substrate, compare each measured spectrum of the sequence of measured spectra to a plurality of reference spectra from each of a plurality of libraries of reference spectra. The controller finds, for each measured spectrum of the sequence of measured spectra, a best matching first reference spectrum from a first library from the plurality of libraries and finds a best matching second reference spectrum from a different second library from the plurality of libraries. The first library can include spectra representing substrates having a first underlying layer thickness, and the first library can include spectra representing substrates having a different second underlying layer thickness.

In some implementations, each measured spectrum of the sequence of measured spectra is compared to exactly two libraries, i.e., only the first library and the second library. In this case only the first library need be searched to find the best matching first reference spectrum, and only the second library need be searched to find the best matching second reference spectrum. These implementations can be particularly useful if the first underlying layer thickness and the second underlying layer thickness are sufficiently far apart to reliably

span the expected variations in thickness of the underlying layer in the product substrates being polished. In such a case, comparing the measured spectra to just the two libraries can reduce computational load.

In some implementations, each measured spectrum of the sequence of measured spectra is compared to three (or more) libraries. To find at least one of the best matching reference spectra, e.g., the best matching first spectrum, at least two of the three libraries can be searched. In some implementations, three or more libraries are searched, and the best matching reference spectrum of any of the libraries is used as the best matching first reference spectrum. Then the best matching reference spectrum from any of the remaining libraries, i.e., excluding the first library, is used as the second reference spectrum. These implementations can be particularly useful if there is large wafer-to-wafer variation in underlying layer thickness and additional libraries are needed to provide good matching reference spectra.

In short, whether the measured spectrum is compared to two libraries or to three or more libraries, the best matching spectra from two different libraries are determined.

In some implementations, a best matching reference spectrum can be determined by calculating, for each reference spectrum, a sum of squared differences between the measured spectrum and the reference spectrum. The reference spectrum with the lowest sum of squared differences has the best fit. Other techniques for finding a best matching reference spectrum are possible, e.g., lowest sum of absolute differences, lowest sum of derivative differences, or greatest cross-correlation.

A method that can be applied to decrease computer processing is to limit the portion of the library that is searched for matching spectra. The library typically includes a wider range of spectra than will be obtained while polishing a substrate. During substrate polishing, the library searching is limited to a predetermined range of library spectra. In some embodiments, the current rotational index N of a substrate being polished is determined. For example, in an initial platen rotation, N can be determined by searching all of the reference spectra of the library. For the spectra obtained during a subsequent rotation, the library is searched within a range of freedom of N. That is, if during one rotation the index number is found to be N, during a subsequent rotation which is X rotations later, where the freedom is Y, the range that will be searched from $(N + X) - Y$ to $(N + X) + Y$.

For each measured spectrum in the sequence, a goodness of fit is calculated between the measured spectrum and each of the best matching spectra. Thus, a first goodness of fit between the measured spectrum the best matching first spectrum can be calculated, and a second goodness of fit between the measured spectrum and the best matching second reference spectrum can be calculated. The goodness of fit can be calculated using a sum of squared differences between the measured spectrum and the best matching reference spectrum, but other techniques, e.g., lowest sum of absolute differences, lowest sum of derivative differences, or greatest cross-correlation, are possible. In some implementations, the goodness of fit is the same value that was used to determine the best matching reference spectrum. In some implementations, the goodness of fit is calculated using a different algorithm than the one used for the determination of the best matching spectrum.

For measured spectrum in the sequence, the values associated with the best matching reference spectra are determined. For example, to determine the value for a best matching reference spectrum, the stored value from the record in the database associated with the best matching reference spectra can be retrieved. A first value associated with the best matching first reference spectrum can be determined, and a second value from the best matching second reference spectrum can be determined.

The values associated with the best matching reference spectra are combined to generate a combined value (which is also described below as a "third value"). The combined value can be calculated from the values associated with the best matching reference spectra, e.g., by interpolating between the first value and second value.

For example, in order to interpolate between the values, a weighted average of the values can be calculated. In order to perform the calculation of the combined value, a weight can be calculated for each best matching spectra. A first weight can be calculated for the best matching first spectrum, and a second weight can be calculated for the best matching second reference spectrum. The weight can be calculated from the goodness of fit values. However, in some situations, it may not be necessary to calculate weights. For example, if the first value is equal to the second value then the third value can be simply be equal to the first value and the second value.

In some implementations, a weighted average of the values can be calculated to provide the combined value. For example, a third value can be calculated as a weighted

average of the first value and the second value. For example, a third value V3 can be calculated as

$$V3= (W1*V1+W2*V2)(W1+W2)$$

where V1 is the first value, V2 is the second value, W1 is a first weight and W2 is a second
5 weight. However, the first weight and second weight could be used in other calculations of the third value.

The weights, e.g., the first weight and the second weight, used in calculation of the weighted average can be based on the goodnesses of fit of the measured spectrum to the best matching reference spectra. There are variety of ways to calculate the weights, and they can
10 depend on the format of the goodness of fit.

In one implementations, where the goodness of fit ranges from 0 (for an ideal goodness of fit) upward, e.g., for a goodness of fit calculated as a sum of squared deviations, the first weight can simply be the second goodness of fit, and the second weight can simply be the first goodness of fit, i.e.,

15 $$W1 = X2 \qquad W2=X1$$

In another implementation, where the goodness of fit ranges from 0 (for an ideal goodness of fit) upward, e.g., for a goodness of fit calculated as a sum of squared deviations, the first weight and the second weight can be calculated as

$$W1 = 1-(X1/(X1+X2)) \qquad W2=1-(X2/(X1+X2))$$

20 In another implementation, where the goodness of fit ranges from 1 (for an ideal goodness of fit) downward, e.g., for a goodness of fit calculated as a cross-correlation, the first weight and the second weight can be calculated as

$$W1 = X1/(X1+X2) \qquad W2=X2/(X1+X2)$$

As polishing progresses, a calculated third value can be generated for each measured
25 spectrum of the sequence of spectra, thus generating a sequence of calculated third values. A polishing endpoint or an adjustment for a polishing rate can thus be based on the sequence of calculated third values.

In addition, in some situations it may be possible to extrapolate from rather than interpolate between the first value and the second value (in which case the third value would
30 not be intermediate between the first value and the second value).

Without being limited to any particular theory, in general, at least for some types of substrates, the goodness of fit between a measured spectrum and a reference spectrum can be a linear function of the difference in the thickness of the underlying layer. Where the goodness of fit ranges from 0 (for an ideal goodness of fit) upward, e.g., for a goodness of fit calculated as a sum of squared deviations, the higher the goodness of fit, the greater the underlayer thickness difference.

In order to determine whether to perform interpolation or extrapolation to calculate the combined value, a first goodness of fit between the best matching first reference spectrum and the measured spectrum can be calculated, a second goodness of fit between the best matching second reference spectrum and the measured spectrum can be calculated, and a third goodness of fit between the best matching first reference spectrum and the best matching second reference spectrum can be calculated. The third goodness of fit is compared to the fist goodness of fit and the second goodness of fit. If the third goodness of fit is worse than both the first goodness of fit and the second goodness of fit, then interpolation can be performed, e.g., as described above. On the other hand, if the third goodness of fit is better than either the first goodness of fit or the second goodness of fit, then extrapolation can be performed, e.g., as described below. What constitutes "worse" will depend on the format of the goodness of fit. Where the goodness of fit ranges from 0 (for an ideal goodness of fit) upward, e.g., for a goodness of fit calculated as a sum of squared deviations, then interpolation can be performed if the third goodness of fit is larger than the first goodness of fit and the second goodness of fit. Where the goodness of fit ranges from 1 (for an ideal goodness of fit) downward, e.g., for a goodness of fit calculated as a cross-correlation, then interpolation can be performed if the third goodness of fit is smaller than the first goodness of fit and the second goodness of fit.

As noted above, the goodness of fit between a measured spectrum and a reference spectrum can be a linear function of difference in thickness of the underlying layer, and thus can be a linear function of the difference in values. As shown in FIG. 13, one technique to perform extrapolation is to fit a line to the pair of points provided by the first value and first goodness of fit, and the second value and second goodness of fit. The value where the line intersects the ideal goodness of fit, e.g., 0 for a goodness of fit calculated as a sum of squared deviations, provides the third value.

In some implementations, this can be simplified to calculating the third value V3 as follows

$$V3=V1-X1*(V1-V2)/(X1-X2)$$

where V1 is the first value, V2 is the second value, X1 is the first goodness of fit and X2 is
5 the second goodness of fit.

Alternatively, since often an exact match with a reference spectra is not possible, the value where the value where the line intersects a goodness of fit slightly offset from the ideal goodness of fit, e.g., 0.005 to 0.01 for a goodness of fit calculated as a sum of squared deviations, provides the third value.

10 The goodness of fit for the extrapolation can be calculated using a sum of squared differences between the measured spectrum and the best matching reference spectrum, but other techniques, e.g., lowest sum of absolute differences, lowest sum of derivative differences, or greatest cross-correlation, are possible. In some implementations, the goodness of fit is for the extrapolation the same value that was used to determine the best
15 matching reference spectrum. In some implementations, the goodness of fit for the extrapolation is calculated using a different algorithm than the one used for the determination of the best matching spectrum.

Referring to FIG. 7, which illustrates the results for only a single zone of a single substrate, the third value calculated from the each pair of best matching spectra for each
20 measured spectrum in the sequence can be determined to generate a time-varying sequence of values 212. This sequence of values can be termed a value trace 210 (where the first value and second value are index values, the trace can be termed an index trace, and where the first value and second value are thickness values, the trace can be termed an thickness trace) . In general, the value trace 210 can include one, e.g., exactly one, value per sweep of the optical
25 monitoring system below the substrate.

For a given value trace 210, where there are multiple spectra measured for a particular zone in a single sweep of the optical monitoring system (termed “current spectra”), a best match can be determined between each of the current spectra and the reference spectra of the two or more libraries. In some implementations, each selected current spectra is
30 compared against each reference spectra of the selected library or libraries. Given current spectra e, f, and g, and reference spectra E, F, and G, for example, a matching coefficient

could be calculated for each of the following combinations of current and reference spectra: e and E, e and F, e and G, f and E, f and F, f and G, g and E, g and F, and g and G. Whichever matching coefficient indicates the best match, e.g., is the smallest, determines the best-matching reference spectrum for that library. Alternatively, in some implementations, the current spectra can be combined, e.g., averaged, and the resulting combined spectrum is compared against the reference spectra to determine the best match. The same round-robin comparison can be made for each library, and the two reference spectra with the best match from two different libraries are used as the best matching first reference spectrum and the best matching second reference spectrum.

In summary, the value trace includes a sequence 210 of values 212, with each particular value 212 of the sequence being generated by combining the values of the two best matching reference spectra from different libraries. The time value for each index of the index trace 210 can be the same as the time at which the measured spectrum was measured.

As shown in FIG. 8, a function, e.g., a polynomial function of known order, e.g., a first-order function (e.g., a line 214) is fit to the sequence of values, e.g., using robust line fitting. Other functions can be used, e.g., polynomial functions of second-order, but a line provides ease of computation. Polishing can be halted at an endpoint time TE that the line 214 crosses a target index IT.

In some implementations, the line is fit to the values after time TC and values for spectra collected before the time TC are ignored. An in-situ monitoring technique can be used to detect clearing of the second layer and exposure of the underlying layer or layer structure. For example, exposure of the first layer at a time TC can be detected by a sudden change in the motor torque or total intensity of light reflected from the substrate, or from dispersion of the collected spectra as discussed in greater detail below.

FIG. 9 shows a flow chart of a method of polishing a product substrate. The product substrate can have at least the same layer structure (but not layer thicknesses) and the same pattern, as the test substrates used to generate the reference spectra of the library.

A sequence of measured spectra are obtained during polishing (step 902), e.g., using the in-situ monitoring system described above.

The measured spectra are analyzed to generate a sequence of values, and a function is fit to the sequence of values. In particular, for each measured spectrum in the sequence of

measured spectra, the two best matching reference spectra from different libraries are found (step 904). The two values associated with the two best matching reference spectra are determined (step 906) , two goodnesses of fit of the measured spectrum to the two best matching reference spectra are calculated (step 908), and a third value is calculated from the two values and the two goodnesses of fit (step 910), e.g., by a weighted average.

A function, e.g., a linear function, is fit to the sequence of values (step 912). As noted above, in some implementations values collected before the time TC, e.g., a time at which clearance of the second layer is detected, are not used in the calculation of the function.

Polishing can be halted once the value (e.g., a calculated value generated from the linear function fit to the sequence of values) reaches a target value (step 914). The target value IT can be set by the user prior to the polishing operation and stored. Alternatively, a target amount to remove can be set by the user, and a target value IT can be calculated from the target amount to remove. For example, an index difference ID can be calculated from the target amount to remove, e.g., from an empirically determined ratio of amount removed to the index (e.g., the polishing rate), and adding the index difference ID to the index value IC at the time TC that clearance of the overlying layer is detected (see FIG. 8).

It is also possible to use the function fit to the values to adjust the polishing parameters, e.g., to adjust the polishing rate of one or more zones on a substrate to improve polishing uniformity.

Referring to FIG. 10, a plurality of traces is illustrated. As discussed above, a value trace can be generated for each zone. For example, a first sequence 210 of values 212 (shown by hollow circles) can be generated for a first zone, a second sequence 220 of values 222 (shown by hollow squares) can be generated for a second zone, and a third sequence 230 of values 232 (shown by hollow triangles) can be generated for a third zone. Although three zones are shown, there could be two zones or four or more zones. All of the zones can be on the same substrate, or some of the zones can be from different substrates being polished simultaneously on the same platen.

For each substrate index trace, a polynomial function of known order, e.g., a first-order function (e.g., a line) is fit to the sequence of values of spectra, e.g., using robust line fitting. For example, a first line 214 can be fit to values 212 for the first zone, a second line

224 can be fit to the values 222 of the second zone, and a third line 234 can be fit to the values 232 of the third zone. Fitting of a line to the values can include calculation of the slope S of the line and an x-axis intersection time T at which the line crosses a starting value, e.g., 0. The function can be expressed in the form $I(t) = S\cdot(t - T)$, where t is time. The x-axis
5 intersection time T can have a negative value, indicating that the starting thickness of the substrate layer is less than expected. Thus, the first line 214 can have a first slope S1 and a first x-axis intersection time T1, the second line 224 can have a second slope S2 and a second x-axis intersection time T2, and the third line 234 can have a third slope S3 and a third x-axis intersection time T3.

10 At some during the polishing process, e.g., at a time T0, a polishing parameter for at least one zone is adjusted to adjust the polishing rate of the zone of the substrate such that at a polishing endpoint time, the plurality of zones are closer to their target thickness than without such adjustment. In some embodiments, each zone can have approximately the same thickness at the endpoint time.

15 Referring to FIG. 11, in some implementations, one zone is selected as a reference zone, and a projected endpoint time TE at which the reference zone will reach a target value IT is determined. For example, as shown in FIG. 11, the first zone is selected as the reference zone, although a different zone and/or a different substrate could be selected. The target thickness IT is set by the user prior to the polishing operation and stored.
20 Alternatively, a target amount to remove TR can be set by the user, and a target value IT can be calculated from the target amount to remove TR. For example, a value difference ID can be calculated from the target amount to remove, e.g., from an empirically determined ratio of amount removed to thevalue (e.g., the polishing rate), and adding the value difference ID to the value IC at the time TC that clearance of the overlying layer is detected.

25 In order to determine the projected time at which the reference zone will reach the target value, the intersection of the line of the reference zone, e.g., line 214, with the target value, IT, can be calculated. Assuming that the polishing rate does not deviate from the expected polishing rate through the remainder polishing process, then the sequence of values should retain a substantially linear progression. Thus, the expected endpoint time TE can be
30 calculated as a simple linear interpolation of the line to the target value IT, e.g., $IT = S\bullet(TE -$

T). Thus, in the example of FIG. 11 in which the first zone is selected as the reference zone, with associated first line 214, IT = S1•(TE –T1), i.e., TE = IT/S1-T1.

One or more zones, e.g., all zones, other than the reference zone (including zones on other substrates) can be defined as adjustable zones. Where the lines for the adjustable zones meet the expected endpoint time TE define projected endpoint for the adjustable zones. The linear function of each adjustable zone, e.g., lines 224 and 234 in FIG. 11, can thus be used to extrapolate the value, e.g., EI2 and EI3, that will be achieved at the expected endpoint time ET for the associated zone. For example, the second line 224 can be used to extrapolate the expected value, EI2, at the expected endpoint time ET for the second zone, and the third line 234 can be used to extrapolate the expected value, EI3, at the expected endpoint time ET for the third zone.

As shown in FIG. 11, if no adjustments are made to the polishing rate of any of the zones after time T0, then if endpoint is forced at the same time for all zones, then each zone can have a different thickness (which is not desirable because it can lead to defects and loss of throughput).

If the target value will be reached at different times for different zones (or equivalently, the adjustable zones will have different expected indexes at the projected endpoint time of the reference zone), the polishing rate can be adjusted upwardly or downwardly, such that the zones would reach the target value (and thus target thickness) closer to the same time than without such adjustment, e.g., at approximately the same time, or would have closer to the same value (and thus same thickness), at the target time than without such adjustment, e.g., approximately the same value (and thus approximately the same thickness).

Thus, in the example of FIG. 11, commencing at a time T0, at least one polishing parameter for the second zone is modified so that the polishing rate of the zone is increased (and as a result the slope of the index trace 220 is increased). Also, in this example, at least one polishing parameter for the third zone is modified so that the polishing rate of the third zone is decreased (and as a result the slope of the trace 230 is decreased). As a result the zones would reach the target index (and thus the target thickness) at approximately the same time (or if pressure to the zones halts at the same time, the zones will end with approximately the same thickness).

In some implementations, if the projected index at the expected endpoint time ET indicate that a zone of the substrate is within a predefined range of the target thickness, then no adjustment may be required for that zone. The range may be 2%, e.g., within 1%, of the target index.

5 The polishing rates for the adjustable zones can be adjusted so that all of the zones are closer to the target index at the expected endpoint time than without such adjustment. For example, a reference zone of the reference substrate might be chosen and the processing parameters for all of the other zone adjusted such that all of the zones will endpoint at approximately the projected time of the reference substrate. The reference zone can be, for 10 example, a predetermined zone, e.g., the center zone 148a or the zone 148b immediately surrounding the center zone, the zone having the earliest or latest projected endpoint time of any of the zones of any of the substrates, or the zone of a substrate having the desired projected endpoint. The earliest time is equivalent to the thinnest substrate if polishing is halted at the same time. Likewise, the latest time is equivalent to the thickest substrate if 15 polishing is halted at the same time. The reference substrate can be, for example, a predetermined substrate, a substrate having the zone with the earliest or latest projected endpoint time of the substrates. The earliest time is equivalent to the thinnest zone if polishing is halted at the same time. Likewise, the latest time is equivalent to the thickest zone if polishing is halted at the same time.

20 For each of the adjustable zones, a desired slope for the trace can be calculated such that the adjustable zone reaches the target value at the same time as the reference zone. For example, the desired slope SD can be calculated from (IT-I)=SD*(TE-T0), where I is the value (calculated from the linear function fit to the sequence of values) at time T0 when the polishing parameters are to be changed, IT is the target value, and TE is the calculated 25 expected endpoint time. In the example of FIG. 11, for the second zone the desired slope SD2 can be calculated from (IT-I2)=SD2*(TE-T0), and for the third zone the desired slope SD3 can be calculated from (IT-I3)=SD3*(TE-T0).

Alternatively, in some implementations, there is no reference zone, and the expected endpoint time can be a predetermined time, e.g., set by the user prior to the polishing process, 30 or can be calculated from an average or other combination of the expected endpoint times of two or more zones (as calculated by projecting the lines for various zones to the target index)

from one or more substrates. In this implementation, the desired slopes are calculated substantially as discussed above, although the desired slope for the first zone of the first substrate must also be calculated, e.g., the desired slope SD1 can be calculated from (IT-I1)=SD1*(TE'-T0).

5 Alternatively, in some implementations, there are different target values for different zones. This permits the creation of a deliberate but controllable non-uniform thickness profile on the substrate. The target values can be entered by user, e.g., using an input device on the controller. For example, the first zone can have a first target value, the second zone can have a second target value, and the third zone can have a third target value.

10 For any of the above methods described above, the polishing rate is adjusted to bring the slope of the trace closer to the desired slope. The polishing rates can be adjusted by, for example, increasing or decreasing the pressure in a corresponding chamber of a carrier head. The change in polishing rate can be assumed to be directly proportional to the change in pressure, e.g., a simple Prestonian model. For example, for each zone of each substrate, 15 where zone was polished with a pressure Pold prior to the time T0, a new pressure Pnew to apply after time T0 can be calculated as Pnew = Pold*(SD/S), where S is the slope of the line prior to time T0 and SD is the desired slope.

For example, assuming that pressure Pold1 was applied to the first zone of the first substrate, pressure Pold2 was applied to the second zone of the first substrate, pressure Pold3 20 was applied to the first zone of the second substrate, and pressure Pold4 was applied to the second zone of the second substrate, then new pressure Pnew1 for the first zone of the first substrate can be calculated as Pnew1 = Pold1*(SD1/S1), the new pressure Pnew2 for the second zone of the first substrate clan be calculated as Pnew2 = Pold2*(SD2/S2), the new pressure Pnew3 for the first zone of the second substrate clan be calculated as Pnew3 = 25 Pold3*(SD3/S3), and the new pressure Pnew4 for the second zone of the second substrate clan be calculated as Pnew4 = Pold4*(SD4/S4).

The process of determining projected times that the substrates will reach the target thickness, and adjusting the polishing rates, can be performed just once during the polishing process, e.g., at a specified time, e.g., 40 to 60% through the expected polishing time, or 30 performed multiple times during the polishing process, e.g., every thirty to sixty seconds. At a subsequent time during the polishing process, the rates can again be adjusted, if

appropriate. During the polishing process, changes in the polishing rates can be made only a few times, such as four, three, two or only one time. The adjustment can be made near the beginning, at the middle or toward the end of the polishing process.

Polishing continues after the polishing rates have been adjusted, e.g., after time T0, the optical monitoring system continues to collect spectra for at least the reference zone and determine values for the reference zone. In some implementations, the optical monitoring system continues to collect spectra and determine values for each zone. Once the index trace of a reference zone reaches the target index, endpoint is called and the polishing operation stops.

For example, as shown in FIG. 12, after time T0, the optical monitoring system continues to collect spectra for the reference zone and determine values 312 for the reference zone. If the pressure on the reference zone did not change (e.g., as in the implementation of FIG. 11), then the linear function can be calculated using data points from both before T0 (but not before TC) and after T0 to provide an updated linear function 314, and the time at which the linear function 314 reaches the target value IT indicates the polishing endpoint time. On the other hand, if the pressure on the reference zone changed at time T0, then a new linear function 314 with a slope S' can be calculated from the sequence of values 312 after time T0, and the time at which the new linear function 314 reaches the target value IT indicates the polishing endpoint time. The reference zone used for determining endpoint can be the same reference zone used as described above to calculate the expected endpoint time, or a different zone (or if all of the zones were adjusted as described with reference to FIG. 11, then a reference zone can be selected for the purpose of endpoint determination). If the new linear function 314 reaches the target value IT slightly later (as shown in FIG. 12) or earlier than the projected time calculated from the original linear function 214, then one or more of the zones may be slightly overpolished or underpolished, respectively. However, since the difference between the expected endpoint time and the actual polishing time should be less than a couple seconds, this need not severely impact the polishing uniformity.

In some implementations, e.g., for copper polishing, after detection of the endpoint for a substrate, the substrate is immediately subjected to an overpolishing process, e.g., to remove copper residue. The overpolishing process can be at a uniform pressure for all zones

of the substrate, e.g., 1 to 1.5 psi. The overpolishing process can have a preset duration, e.g., 10 to 15 seconds.

In addition, although the discussion above assumes a rotating platen with an optical endpoint monitor installed in the platen, system could be applicable to other types of relative motion between the monitoring system and the substrate. For example, in some implementations, e.g., orbital motion, the light source traverses different positions on the substrate, but does not cross the edge of the substrate. In such cases, the collected spectra can still be grouped, e.g., spectra can be collected at a certain frequency and spectra collected within a time period can be considered part of a group. The time period should be sufficiently long that five to twenty spectra are collected for each group.

As used in the instant specification, the term substrate can include, for example, a product substrate (e.g., which includes multiple memory or processor dies), a test substrate, a bare substrate, and a gating substrate. The substrate can be at various stages of integrated circuit fabrication, e.g., the substrate can be a bare wafer, or it can include one or more deposited and/or patterned layers. The term substrate can include circular disks and rectangular sheets.

Embodiments of the invention and all of the functional operations described in this specification can be implemented in digital electronic circuitry, or in computer software, firmware, or hardware, including the structural means disclosed in this specification and structural equivalents thereof, or in combinations of them. Embodiments of the invention can be implemented as one or more computer program products, i.e., one or more computer programs tangibly embodied in a machine-readable storage media, for execution by, or to control the operation of, data processing apparatus, e.g., a programmable processor, a computer, or multiple processors or computers. A computer program (also known as a program, software, software application, or code) can be written in any form of programming language, including compiled or interpreted languages, and it can be deployed in any form, including as a stand-alone program or as a module, component, subroutine, or other unit suitable for use in a computing environment. A computer program does not necessarily correspond to a file. A program can be stored in a portion of a file that holds other programs or data, in a single file dedicated to the program in question, or in multiple coordinated files (e.g., files that store one or more modules, sub-programs, or portions of code). A computer

program can be deployed to be executed on one computer or on multiple computers at one site or distributed across multiple sites and interconnected by a communication network.

The processes and logic flows described in this specification can be performed by one or more programmable processors executing one or more computer programs to perform functions by operating on input data and generating output. The processes and logic flows can also be performed by, and apparatus can also be implemented as, special purpose logic circuitry, e.g., an FPGA (field programmable gate array) or an ASIC (application-specific integrated circuit).

The above described polishing apparatus and methods can be applied in a variety of polishing systems. Either the polishing pad, or the carrier heads, or both can move to provide relative motion between the polishing surface and the substrate. For example, the platen may orbit rather than rotate. The polishing pad can be a circular (or some other shape) pad secured to the platen. Some aspects of the endpoint detection system may be applicable to linear polishing systems, e.g., where the polishing pad is a continuous or a reel-to-reel belt that moves linearly. The polishing layer can be a standard (for example, polyurethane with or without fillers) polishing material, a soft material, or a fixed-abrasive material. Terms of relative positioning are used; it should be understood that the polishing surface and substrate can be held in a vertical orientation or some other orientation.

Particular embodiments of the invention have been described. Other embodiments are within the scope of the following claims.

What is claimed is:

CLAIMS

1. A method of controlling polishing, comprising:

storing a plurality libraries, each library including a plurality of reference spectra;

polishing a substrate;

measuring a sequence of spectra of light from the substrate during polishing;

for each measured spectrum of the sequence of spectra, finding a best matching first reference spectrum from a first library from the plurality of libraries and finding a best matching second reference spectrum from a different second library from the plurality of libraries;

for each measured spectrum of the sequence of spectra, determining a first value associated with the best matching first reference spectrum and determining a second value from the best matching second reference spectrum;

for each measured spectrum of the sequence of spectra, calculating a third value from the first value and the second value to generate a sequence of calculated third values; and

determining at least one of a polishing endpoint or an adjustment for a polishing rate based on the sequence of calculated third values.

2. The method of claim 1, wherein calculating the third value from the first value and the second value comprises interpolating between the first value and the second value.

3. The method of claim 2, wherein calculating the third value from the first value and the second value comprises calculating a weighted average of the first value and the second value.

4. The method of claim 3, wherein calculating the weighted average of the first value and the second value comprises calculating a first goodness of fit between the best matching first reference spectrum and the measured spectrum, calculating a second goodness of fit between the best matching second reference spectrum and the measured spectrum, and calculating weights for the weighted average based on the first goodness of fit and the second goodness of fit.

5. The method of claim 1, wherein calculating the third value from the first value and the second value comprises extrapolating from the first value and the second value.

6. The method of claim 5, wherein extrapolating from the first value and the second value comprises calculating a first goodness of fit between the best matching first reference spectrum and the measured spectrum, and calculating a second goodness of fit between the best matching second reference spectrum and the measured spectrum, and extrapolating based on the first value, the first goodness of fit, the second value and the second goodness of fit.

7. The method of claim 1, further comprising determining whether to interpolate or extrapolate in calculating the third value.

8. The method of claim 7, further comprising calculating a first goodness of fit between the best matching first reference spectrum and the measured spectrum, calculating a second goodness of fit between the best matching second reference spectrum and the measured spectrum, and calculating a third goodness of fit between the best matching first reference spectrum and the best matching second reference spectrum , and wherein determining whether to interpolate or extrapolate comprises comparing the third goodness of fit to the fist goodness of fit and the second goodness of fit.

9. The method of claim 8, further comprising interpolating between the first value and the second value if the third goodness of fit is worse than the first goodness of fit and the second goodness of fit.

10. The method of claim 8, further comprising extrapolating from the first value and the second value if the third goodness of fit is better than the first goodness of fit or the second goodness of fit.

11. The method of claim 1, wherein finding a first best matching reference spectrum from a first library from the plurality of libraries consists of searching only the first library, and wherein finding the best matching second reference spectrum from a different second library from the plurality of libraries comprises searching only the second library.

12. The method of claim 1, further comprising fitting a linear function to the sequence of third values.

13. The method of claim 12, further comprising halting the polishing when the linear function matches or exceeds a target value.

14. The method of claim 1, wherein the substrate includes a second layer overlying a first layer, the first layer having a different composition than the second layer, the second layer being polished in the polishing step, and wherein the first library comprises spectra for the substrate having the first layer of a first thickness and the second library comprises spectra for a substrate having the first layer of a different second thickness.

15. The method of claim 1, wherein the substrate includes a plurality of zones, and a polishing rate of each zone is independently controllable by an independently variable polishing parameter, and further comprising:

measuring a sequence of spectra from each zone during polishing;

for each measured spectrum in the sequence of spectra for each zone, finding a first best matching reference spectrum from a first library from the plurality of libraries and finding a best matching second reference spectrum from a different second library from the plurality of libraries;

for each measured spectrum of the sequence of spectra for each zone, determining a first value associated with the best matching first reference spectrum and determining a second value from the best matching second reference spectrum;

for each measured spectrum of the sequence of spectra for each zone, calculating a third value from the first value and the second value to generate a sequence of calculated third values for each zone; and

based on the sequence of calculated third values for each zone adjusting the polishing parameter for at least one zone to adjust the polishing rate of the at least one zone such that the plurality of zones have a smaller difference in thickness at the polishing endpoint than without such adjustment.

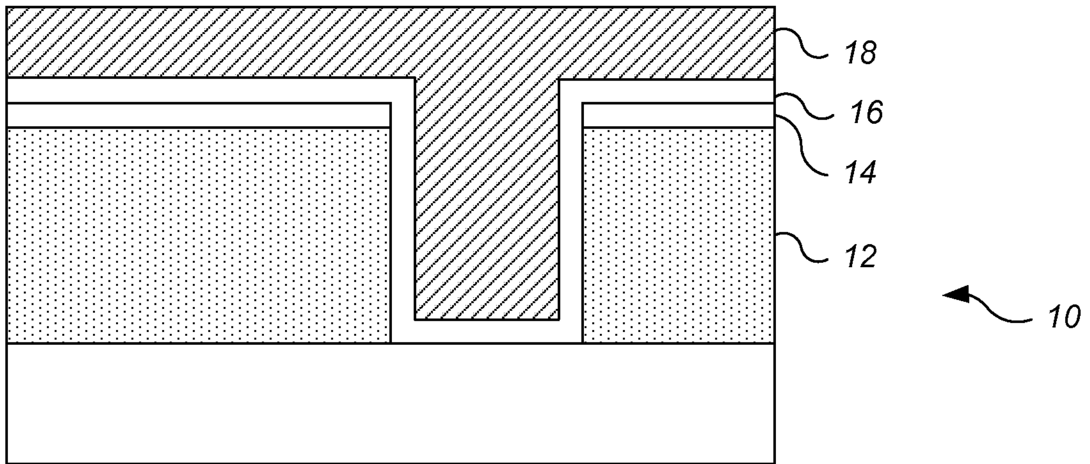

FIG. 1A

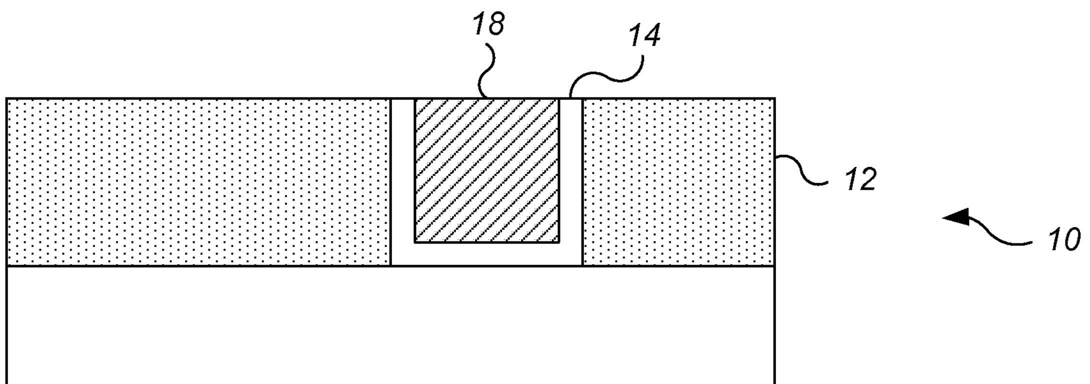

FIG. 1B

Controller

FIG. 2

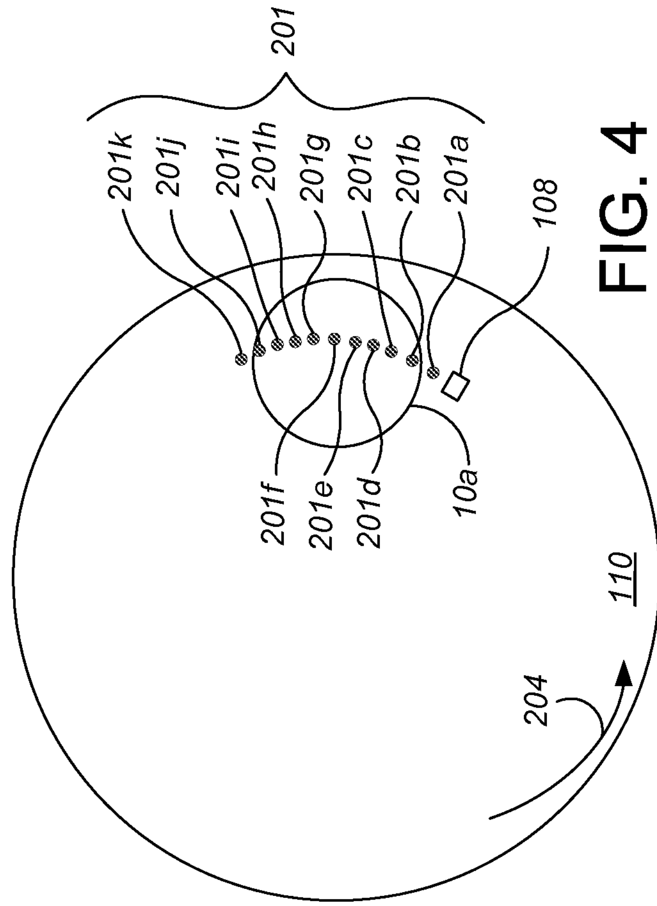

FIG. 4

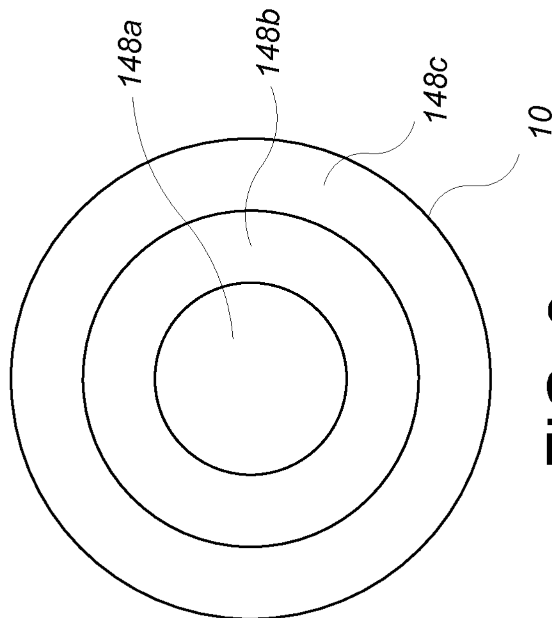

FIG. 3

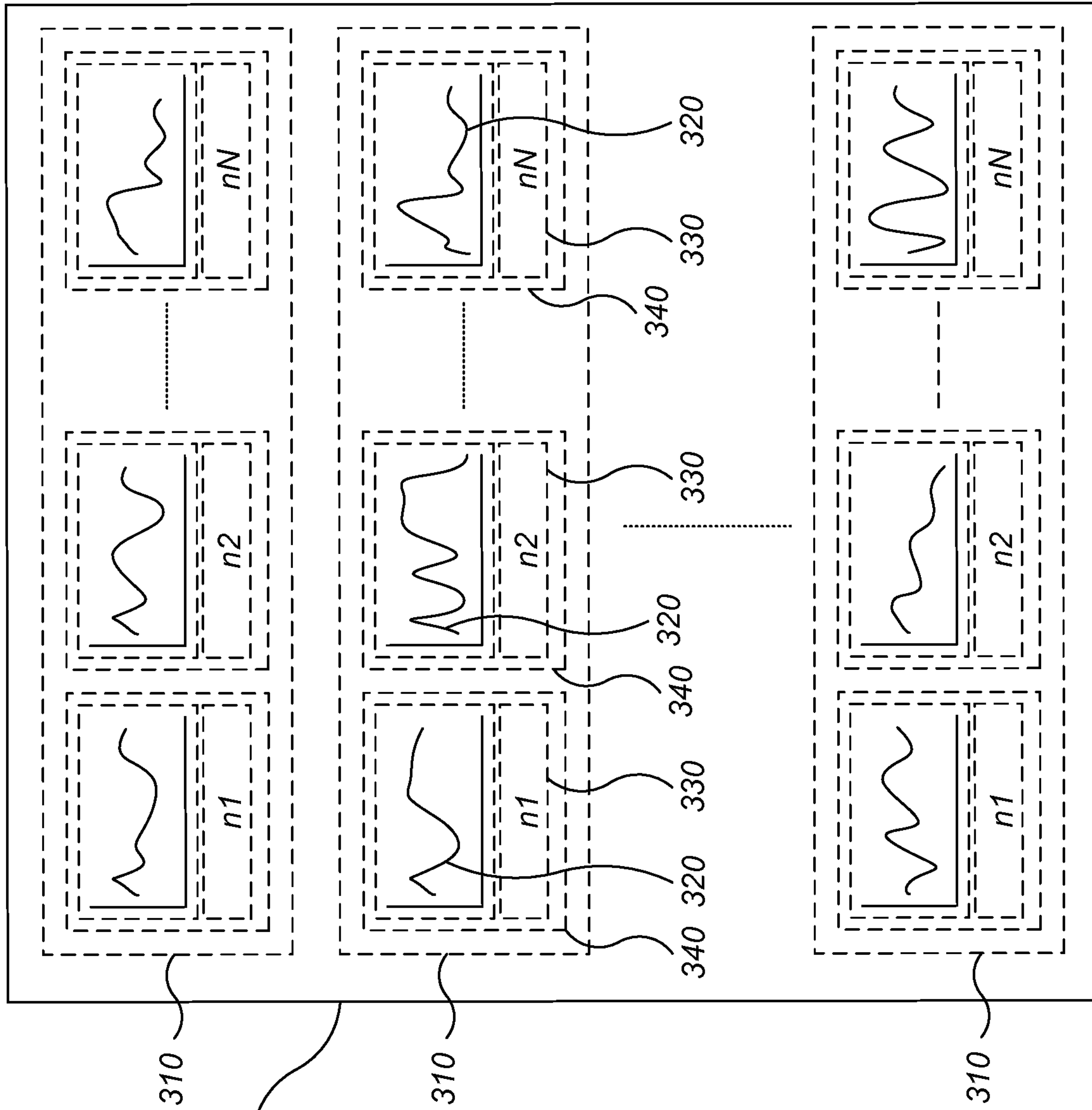

FIG. 6

300
wavelength
intensity

FIG. 5

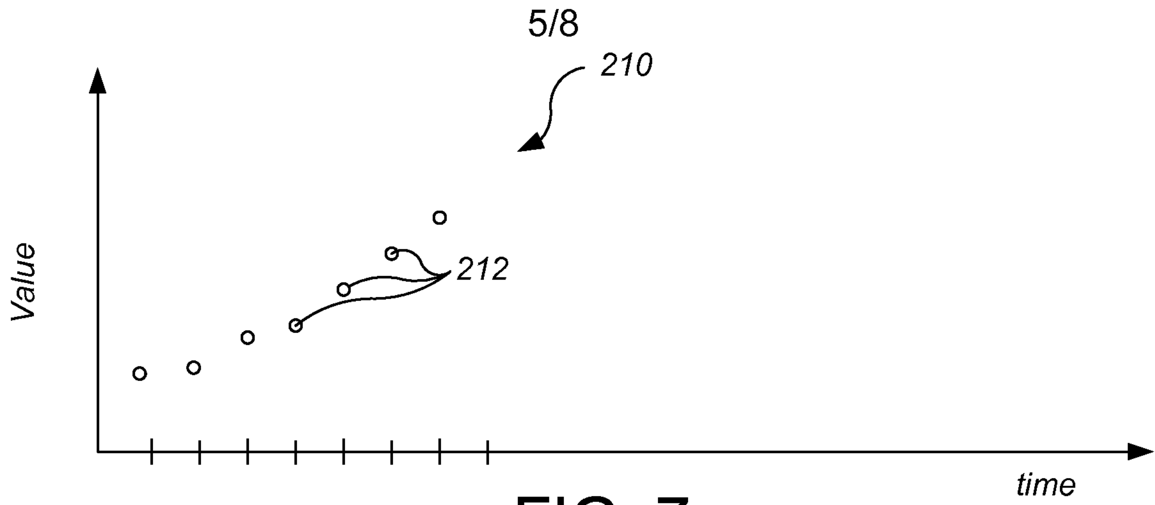

FIG. 7

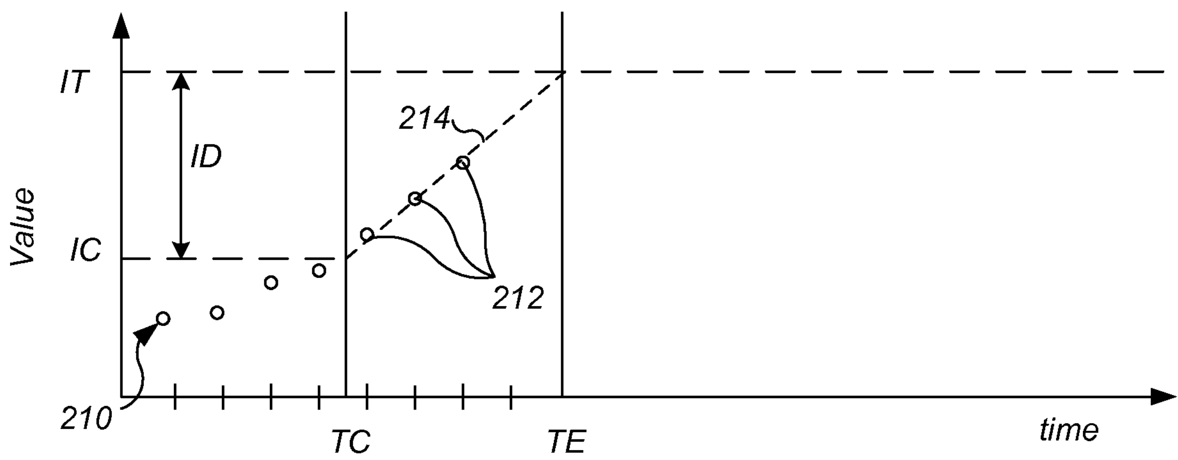

FIG. 8

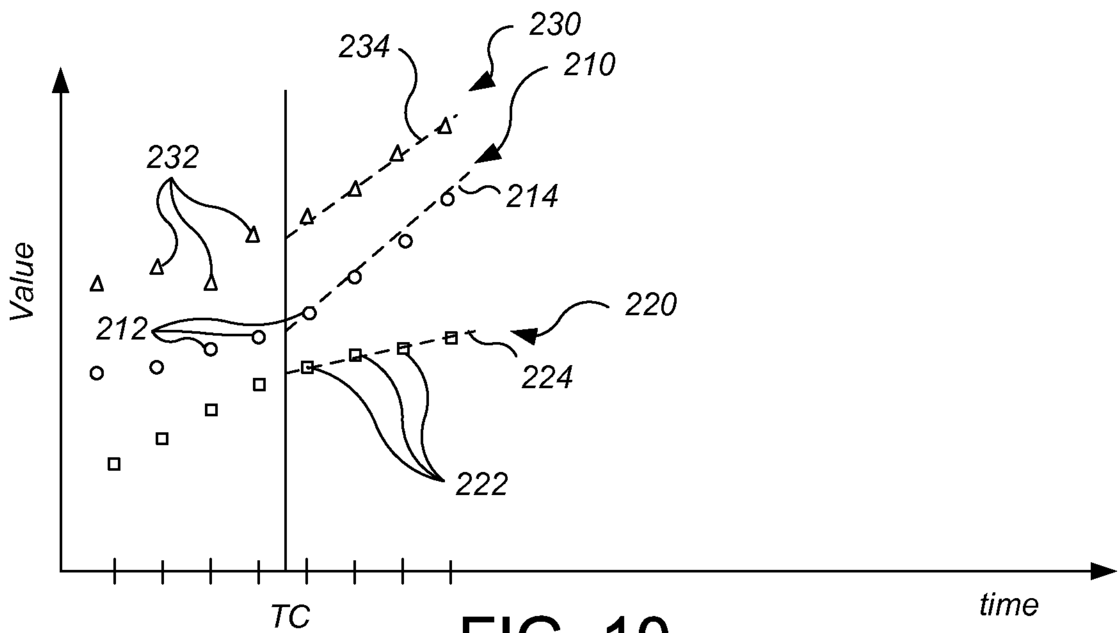

FIG. 10

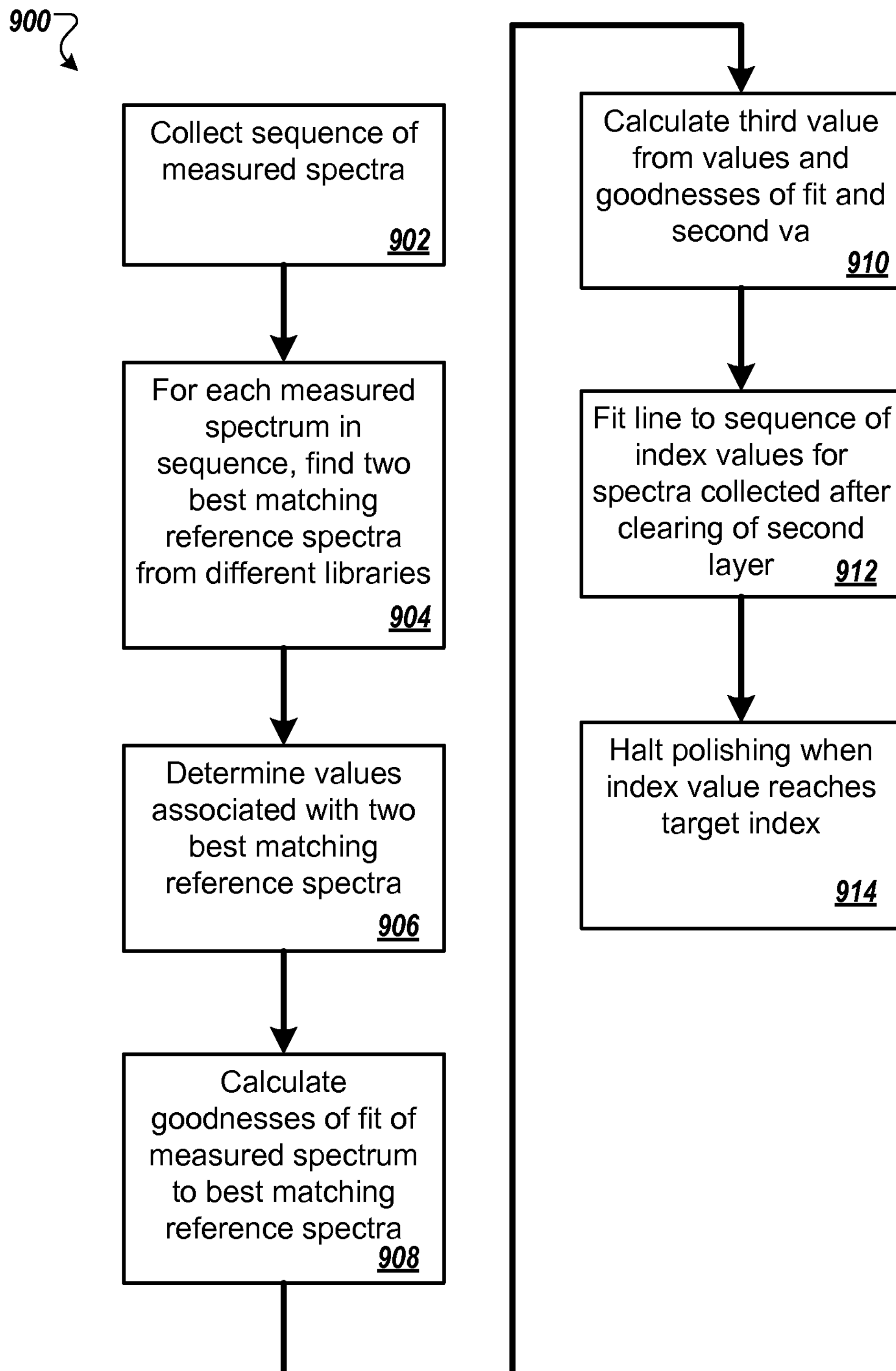

FIG. 9

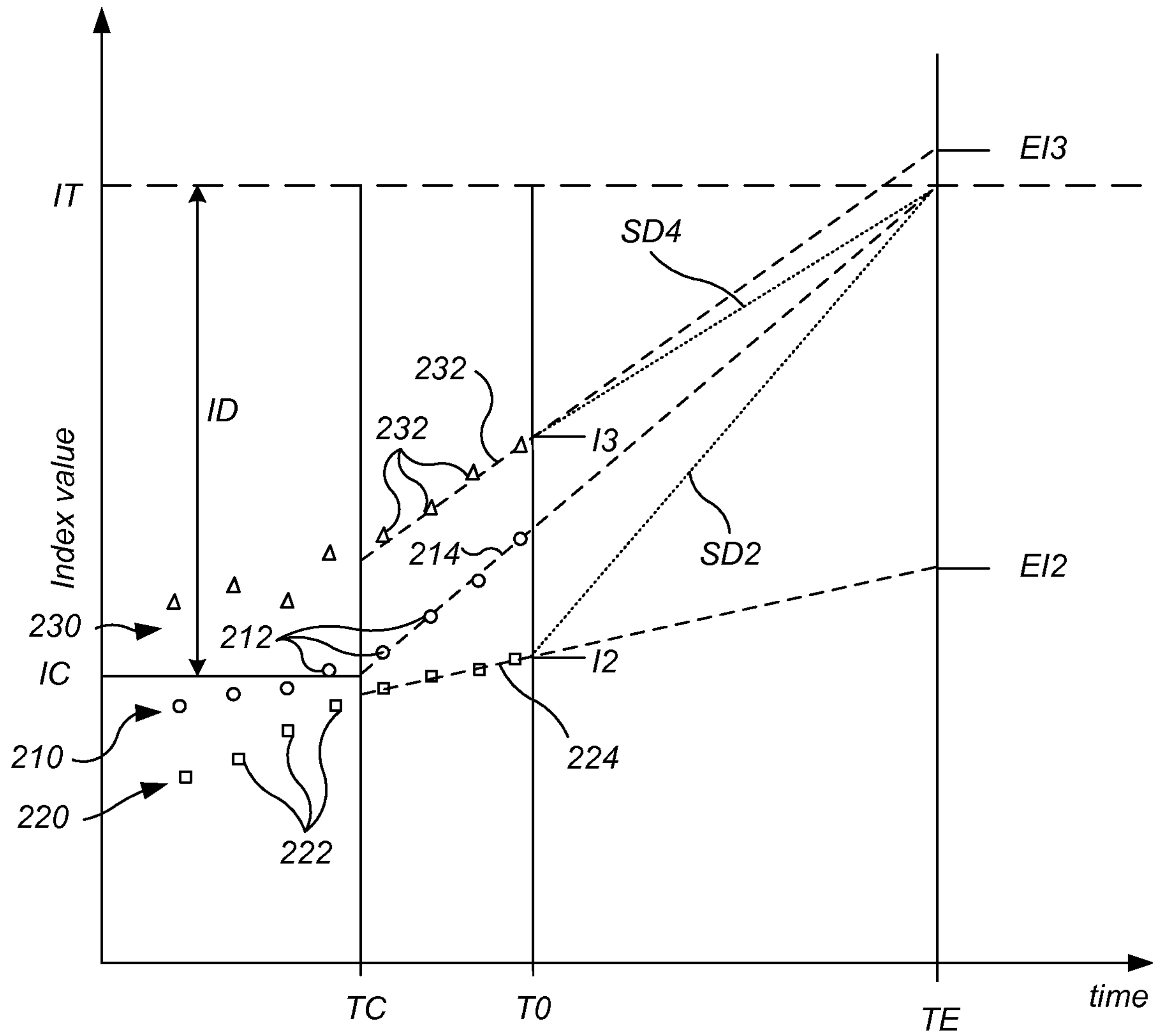

FIG. 11

FIG. 12

FIG. 13