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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
28 March 2019 (28.03.2019)

WIPO | PCT

(10) International Publication Number
WO 2019/060577 A1

(51) International Patent Classification:
H01L 21/304 (2006.01) *H01L 21/66* (2006.01)
H01L 21/3105 (2006.01) *H01L 21/67* (2006.01)
H01L 21/321 (2006.01)

(21) International Application Number:
PCT/US2018/052001

(22) International Filing Date:
20 September 2018 (20.09.2018)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
62/562,986 25 September 2017 (25.09.2017) US

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

(72) Inventors: **DHANDAPANI, Sivakumar**; 494 White Chapel Avenue, San Jose, California 95136 (US). **QIAN, Jun**; 575 Dublin Way, Sunnyvale, California 94087 (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, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, 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.

(54) Title: SEMICONDUCTOR FABRICATION USING MACHINE LEARNING APPROACH TO GENERATING PROCESS CONTROL PARAMETERS

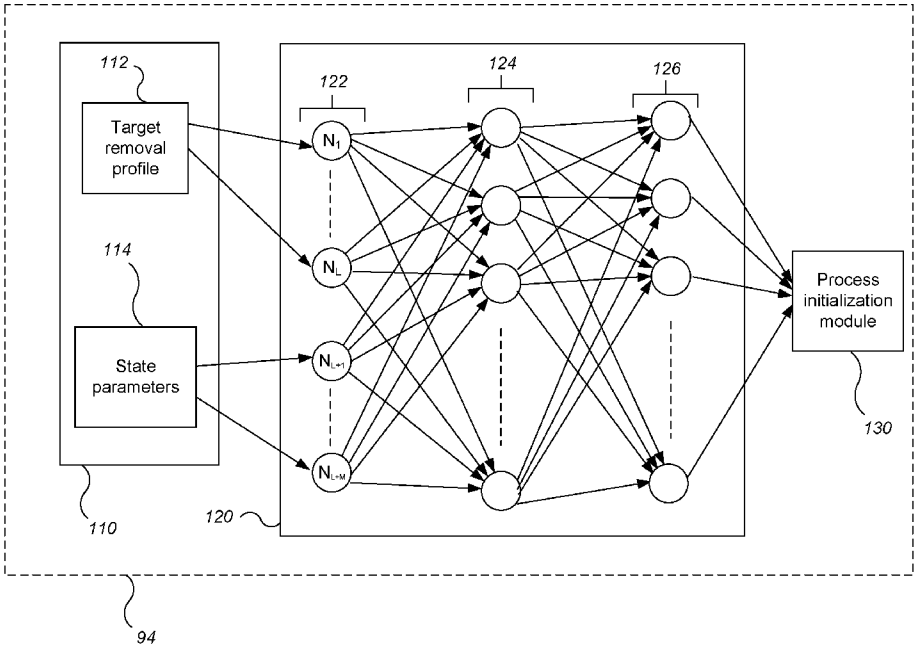


FIG. 3

(57) Abstract: A method for processing substrates includes subjecting each respective first substrate of a first plurality of substrates to a process that modifies a thickness of an outer layer of the respective first substrate, generating a plurality of groups of process parameter values; generating a plurality of removal profiles, training an artificial neural network by backpropagation using the plurality of groups of process parameter and plurality of removal profiles as training data where the artificial neural network has a plurality of input nodes to receive respective removal values from the removal profile and a plurality of output nodes to output control parameter values, for each respective second substrate of a second plurality of substrates determining a target removal profile, determining respective control parameter values to apply by applying the target removal profile to the input nodes, and subjecting each respective second substrate to the process using the respective control parameter values.

[Continued on next page]

WO 2019/060577 A1

(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, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, 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, KM, ML, MR, NE, SN, TD, TG).

Declarations under Rule 4.17:

— *as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(ii))*

— *as to the applicant's entitlement to claim the priority of the earlier application (Rule 4.17(iii))*

Published:

— *with international search report (Art. 21(3))*

— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))*

SEMICONDUCTOR FABRICATION USING MACHINE LEARNING APPROACH TO GENERATING PROCESS CONTROL PARAMETERS

TECHNICAL FIELD

The present disclosure relates to controlling an integrated circuit fabrication process, e.g., a chemical mechanical polishing process.

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 some applications, the filler layer is planarized until the top surface of a patterned layer is exposed. For example, a conductive filler layer 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, the filler layer is planarized until a predetermined thickness is left over an underlying layer. For example, a dielectric layer deposited can be planarized 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 with a durable roughened surface. 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. The processed wafer exhibits a material removal profile, a two dimensional map of the change in a polished layer's thickness after the polishing process.

SUMMARY

In one aspect, a method of processing substrates includes: subjecting each respective first substrate of a first plurality of substrates to a process that modifies a thickness of an

outer layer of the respective first substrate; for each respective first substrate, recording a group of process parameter values used for the process, thus generating a plurality of groups of process parameter values; for each respective first substrate, measuring a removal profile of the outer layer during or after the process with a monitoring system, thus generating a plurality of removal profiles; training an artificial neural network by backpropagation using the plurality of groups of process parameter and plurality of removal profiles as training data, the artificial neural network having a plurality of input nodes to receive respective removal values from the removal profile, a plurality of output nodes to output control parameter values, and a plurality of hidden nodes connecting the input nodes to the output nodes; for each respective second substrate of a second plurality of substrates, determining a target removal profile; for each respective second substrate, determining respective control parameter values to apply to the respective second substrate from the output nodes of the artificial neural network by applying the target removal profile to the input nodes of the artificial neural network; and subjecting each respective second substrate to the process using the respective control parameter values.

In another aspect, a computer program product for controlling processing of a substrate, the computer program product tangibly embodied in a non-transitory computer readable medium, includes instructions for causing a processor to: train an artificial neural network by backpropagation using a plurality of groups of process parameter and a plurality of removal profiles as training data, the artificial neural network having a plurality of input nodes to receive respective removal values from a removal profile, a plurality of output nodes to output control parameter values, and a plurality of hidden nodes connecting the input nodes to the output nodes; obtain a target removal profile for each respective substrate of a plurality of substrates; for each respective substrate, determine respective control parameter values to apply to the respective substrate from the output nodes of the artificial neural network by applying the target removal profile to the input nodes of the artificial neural network; and cause a semiconductor processing system to subject each respective substrate to the process using the respective control parameter values.

In yet another aspect, a polishing system includes: a support to hold a polishing pad; a carrier head to hold a substrate in contact with the polishing pad, the carrier head having a plurality of chambers; a motor to generate relative motion between the support and the carrier

head; and a controller configured to: train an artificial neural network by backpropagation using a plurality of groups of process parameter and a plurality of removal profiles as training data, the artificial neural network having a plurality of input nodes to receive respective removal values from a removal profile, a plurality of output nodes to output control parameter values, and a plurality of hidden nodes connecting the input nodes to the output nodes, wherein the control parameter values include pressures for the chambers in the carrier head; obtain a target removal profile for each respective substrate of a plurality of substrates; for each respective substrate, determine respective control parameter values to apply to the respective substrate from the output nodes of the artificial neural network by applying the target removal profile to the input nodes of the artificial neural network, wherein the respective control parameter values include respective pressures for the chambers in the carrier head; and for each respective substrate, cause the carrier head to apply the respective pressures to the chambers in the carrier head during polishing.

Implementations may include one or more of the following features.

The process may include chemical mechanical polishing. The control parameter values may include pressures of chambers in a carrier head to hold a substrate against a polishing pad.

At least some of the input nodes of the artificial neural network may be configured to receive state parameter values. The state parameter values may include one or more of a retaining ring life, or a polishing pad life. The control parameter values may include one or more of a platen rotation rate, carrier head rotation rate, or polishing time.

The monitoring system may include an in-line metrology system. The in-line metrology system may include an optical monitoring system.

The determining of the target removal profile may include storing a desired thickness profile, receiving a measured thickness profile of the respective second substrate, and determining a difference between the measured thickness profile and the desired thickness profile. The receiving of the measured thickness profile may include measuring a thickness profile of the respective second substrate with the monitoring system.

Certain implementations may have one or more of the following advantages. Control parameter convergence may be achieved more rapidly. Within-wafer thickness non-uniformity and wafer-to-wafer thickness non-uniformity (WIWNU and WTWNU) may be

reduced. Product wafers may be used for model refinement, leading to the actual profile of processed substrates being closer to a desired profile. Complex behavior of the polishing process in response to polishing parameters may be accounted for without explicit knowledge of the functional relationships.

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

DESCRIPTION OF DRAWINGS

FIG. 1 illustrates a schematic cross-sectional view of an example of a polishing apparatus.

FIG. 2 illustrates a schematic top view of a substrate having multiple zones.

FIG. 3 illustrates a neural network used as part of the integrated advanced process control module of the polishing apparatus.

FIG. 4 illustrates an example process for using a neural network model in a CMP process.

FIG. 5 illustrates an integrated advanced process control module that includes multiple neural networks.

Like reference symbols in the various drawings indicate like elements.

DETAILED DESCRIPTION

One challenge in CMP is developing a process model that is able to predict a material removal profile as a function of multiple input parameters. Input parameters may include variations in the initial thickness of a substrate layer, a target material removal profile, the polishing pad condition, the retaining ring condition, the relative speed between the polishing pad and a substrate, and the applied pressure on a substrate. Furthermore, the CMP process is a dynamical system in which surface topology is constantly changing as a function of time during the process, which in turn affects the process behavior. The complex relationship of input parameters with the material removal profile and the dynamic nature of a CMP process contribute to the challenge of developing a process model of the polishing process.

FIG. 1 illustrates an example of a polishing apparatus 20. The polishing apparatus 20 can include a rotatable disk-shaped platen 22 on which a polishing pad 30 is situated. The

platen is operable to rotate about an axis 23. For example, a motor 24 can turn a drive shaft 26 to rotate the platen 22. The polishing pad 30 can be detachably secured to the platen 22, for example, by a layer of adhesive. The polishing pad 30 can be a two-layer polishing pad with an outer polishing layer 32 and a softer backing layer 34.

The polishing apparatus 20 can include a polishing liquid supply port 40 to dispense a polishing liquid 42, such as an abrasive slurry, onto the polishing pad 30. The polishing apparatus 20 can also include a polishing pad conditioning disc to abrade the polishing pad 30 to maintain the polishing pad 30 in a consistent abrasive state.

A carrier head 50 is operable to hold a substrate 10 against the polishing pad 30. Each carrier head 50 also includes a plurality of independently controllable pressurizable chambers, e.g., three chambers 52a-52c, which can apply independently controllable pressurizes to associated zones 12a-12c on the substrate 10 (see FIG. 2). Referring to FIG. 2, the center zone 12a can be substantially circular, and the remaining zones 12b-12c can be concentric annular zones around the center zone 12a.

Returning to FIG. 1, the chambers 52a-52c can be defined by a flexible membrane 54 having a bottom surface to which the substrate 10 is mounted. The carrier head 50 can also include a retaining ring 56 to retain the substrate 10 below the flexible membrane 54. Although only three chambers are illustrated in FIGS. 1 and 2 for ease of illustration, there could be two chambers, or four or more chambers, e.g., five chambers. In addition, other mechanisms to adjust the pressure applied to the substrate, e.g., piezoelectric actuators, could be used in the carrier head 50.

Each carrier head 50 is suspended from a support structure 60, e.g., a carousel or track, and is connected by a drive shaft 62 to a carrier head rotation motor 64 so that the carrier head can rotate about an axis 51. Optionally each carrier head 50 can oscillate laterally, e.g., on sliders on the carousel, by motion along the track; or by rotational oscillation of the carousel itself. In operation, the platen 22 is rotated about its central axis 23, and the carrier head 50 is rotated about its central axis 51 and translated laterally across the top surface of the polishing pad 30.

The polishing apparatus 20 can include an in-line monitoring system for measuring a thickness of a polished layer upon completion of the polishing process. For example, the in-line monitoring system can generate a map of layer thicknesses over the zones 12a-12c. The

measurements from the in-line monitoring system can be communicated to a controller 90. Examples of the in-line monitoring system include an optical monitoring system, e.g., a spectrographic monitoring system. A spectrographic monitoring system can measure a layer thickness value and an associated goodness of fit (GOF) value for each points of the layer thickness map. For example, a broadband light source can be used to illuminate a location on the layer, and the reflection containing an optical interference spectrum created by the layer, the substrate, and any other layers in between the two can be measured. The measured optical interference spectrum can be analyzed, for example, by fitting it with an equation that describes an optical interference spectrum generated by an expected stacking of films. The fitting produces a determination of the thickness of the layer and a GOF value that is indicative of how closely the measured spectrum agrees with the expected film stack. As such, the GOF value can be used as an indicator of reliability of the determined thickness value.

The controller 90 contains a tool control module 92 and an integrated advanced process control module (i-APC) 94. The tool control module 92 and the i-APC module 94 in combination may provide advanced process control functionalities, such as wafer-to-wafer uniformity control. The controller 90 can be a computing device that includes a microprocessor, memory and input/output circuitry, e.g., a programmable computer. Although illustrated with a single block, the controller 90 can be a networked system with functions distributed across multiple computers, and modules 92 and 94 can be located in the same computer or different computers.

The described polishing apparatus has many associated process parameters that control the operation of the polishing apparatus or describe the state of the apparatus or the polishing environment. Process parameters that control the operation of the polishing apparatus (and that can be set, at least initially, by the tool control module 92) ('control parameters') include the following: rotation rate of platen 22; rotation rate of carrier head 50; pressure of the chambers 52a-52c; and polishing time.

Process parameters that reflect the state of the apparatus or polishing environment ('state parameters') include the following: polishing retaining ring life; polishing pad life; polishing pad conditioning disc life; type of polishing pad; and type of polishing liquid ('slurry'). The retaining ring, pad, and conditioning disc are examples of consumable

components ('consumables') within a polishing apparatus. The condition or 'life' of these consumables can be described, for example, as a count of wafers processed; actual wafer polishing time; or total time elapsed since installation.

A typical wafer to be polished by a CMP process has a layer of material with surface topologies to be planarized. A goal of the CMP process is to achieve a desired material removal profile, which is a one or two dimensional map of the change in a polished layer's thickness after the polishing process. The surface topologies vary from wafer-to-wafer due to different die designs having different underlying transistor and interconnect patterns, e.g. due to different pattern densities. These factors interact with the CMP process in a complex manner, which results in different polishing behavior between wafers with different die designs. Furthermore, even wafers with the same die design may have different polishing behaviors due to upstream process variations in deposition or etching. Therefore, different CMP control parameters are typically needed, at least for wafers with different die designs, and possibly for individual wafers with the same die design, to achieve a desired material removal profile. Material removal profiles sometimes have a radial dependence partly due to the axial symmetry and rotation of the polishing head. Accordingly, one control parameter that is often tuned to achieve a desired radial profile is the pressures of the radial chambers 52a-52c.

The polishing apparatus 20 can implement a wafer-to-wafer control. The wafer-to-wafer control can provide improved likelihood of achieving target material removal profiles over a wide range of designs and wafer non-uniformities. A wafer-to-wafer feedback control method uses information about previously processed substrates to improve processing of a subsequent substrate. The wafer-to-wafer feedback control method can be implemented by the i-APC module 94.

In an example implementation of the controller 90, the i-APC module 94 generates an initial set of control parameter values based on an initial process model, and provides the initial set of control parameters values to the tool control module 92. The tool control module 92 can then control the polishing system using the received control parameter values.

After processing of one or more substrates over a user defined period of time, data about the substrates can be used by the i-APC module 94 to improve the initial process model by generating a new process model or updating the initial process model. For example, the

user defined period of time for generation of a new process model or updating of the initial process model can be measured in number of substrates polished, such as 5, 10, 25, or 100 wafers, or a "lot" of wafers (wafers are typically transported and processed in "lots"; a typical lot size is 25 wafers). As another example, the user defined period of time may be defined as a portion of a lifetime of a consumable, such as a polishing pad lifetime. Such generation or updating of the model can be performed off-line, or independent of the operation of the polishing apparatus.

The improved process model can then be used to generate an improved set of control parameter values for a subsequent substrate. Improvement of the process model by the i-APC module 94 can lead to improved polishing uniformity control by minimizing error between the target material removal profile and the actual profile realized on the substrate.

The i-APC module 94 performs tasks including collecting and processing of data from processed wafers to improve processing of future wafers. Collected data can include upstream and downstream metrology data of the product wafers from various wafer metrology tools ("monitoring systems"). Upstream metrology data may include thickness map and associated GOF values of a deposited layer. Downstream metrology data may include a thickness map and associated GOF values of the polished layer, or surface roughness values. The GOF values can be used by the i-APC module 94 to determine whether the thickness value are reliable enough to be used in the process model development. The i-APC module 94 pairs these data with the control and state parameters used during the processing of a particular wafer, and stores it in a data log. The data log may be organized in a variety of ways, including grouping by wafer ID, design ID, lot ID, tool ID, etc. This data log is typically used to monitor trends and drifts in the behavior of a CMP process and to take corrective actions. Due to their size, data logs may be stored in one or more servers that are a part of the controller 90.

In some implementations, the i-APC module 94 stores a desired thickness profile for each wafer to be processed. Using the desired thickness profile, the i-APC module 94 can generate the target material removal profile by subtracting a desired thickness profile from the thickness map of the deposited layer.

A process model can be generated in various ways. For example, a process model for describing the effects of chamber pressures can be generated by processing multiple blanket

wafers, which are un-patterned wafers with a uniform layer of a film to be polished, as a proxy for product wafers that have patterns. By polishing the blanket wafers for a fixed amount of time and measuring the resulting material removal profile, a relationship between a set of chamber pressures and a material removal rate can be determined using the Preston equation.

The Preston equation states:

Material removal rate (MRR) = K_p*V*P Equation 1

where V is the velocity of the polishing pad surface with respect to the substrate surface being polished, P is the pressure applied to a radial zone of a wafer to be polished against the polishing pad, and K_p is a proportionality constant known as the Preston coefficient. The velocity, pressure, and polish time are known and controlled variables, so K_p of each zone can be determined with algebraic manipulations. The resulting set of Preston equations containing the Preston coefficients from polishing of blanket wafers is an example of a baseline process model. The i-APC module 94 can then use this baseline process model to generate a set of initial control parameters for processing of subsequent product wafers. While a process model for modeling effects of chamber pressures are discussed, process models for other control parameters can be generated in this manner.

Due to the differences in surface topologies between a blanket wafer and a product wafer, however, the baseline process model might not achieve the target material removal profile on the polished product wafer. In such case, the i-APC 94 attempts to find, through multiple iterations over polishing of multiple product wafers, a set of correction factors ("offsets") for improving the achieved material removal profile on the product wafers. The i-APC module 94 eventually determines, or converges on, offsets that can be used with the baseline process model for polishing the product wafer. However, such the determination of the offsets can take polishing of multiple product wafers, during which the polished wafers are less likely to meet the target removal profile thus more likely to be rejected, adversely impacting yield of the polishing process.

Consumption of product wafers for offset convergence should be minimized as they are valuable, expensive, and limited in quantity. One way of reducing the number of rejected product wafers until offset convergence is reached is to provide an improved process model that more completely and accurately captures the effects of various process parameters. In

some implementations, the i-APC module 94 is configured to provide the improved process model. One way of improving the quality of the process model is to implement a process model that takes in as inputs (1) target material removal profile and (2) state parameters that reflect the conditions of the polishing apparatus, and outputs a set of estimated control parameters (e.g. chamber pressures) that would achieve the target material removal profile.

The material removal rate is influenced by state parameters 114 of the polishing apparatus such as the life, or age, of various consumables due to their wear. For example, the polishing pad may experience a reduction in elasticity or a reduction in surface roughness due to aging. The retaining ring may get polished away during the CMP process, becoming thinner. The polishing pad conditioning disc may become dull and have reduced capacity for pad conditioning. These changes to characteristics of the consumables can affect the polishing behavior. Therefore, wafers with an identical design may require different sets of control parameters for different state parameters to achieve the target material removal profile. Therefore, incorporating state parameters into the process model can lead to better control parameter estimation.

A neural network can be used to model behavior (e.g., removal profile) of a semiconductor processing system (e.g., a CMP system) in response to the process parameters. The neural network is based on a large collection of simple neural units (artificial neurons). Each individual neural unit computes a respective output, or activation, using a summation function. There may be a threshold function or limiting function on each connection and on the unit itself, such that the signal must surpass the limit before propagating to other neurons. Such threshold function may include a hyperbolic tangent or a sigmoid function.

A neural network is trained, rather than explicitly programmed. In other words, a neural network does not require a priori knowledge of the functional dependence of the process being modeled. Instead, the neural network can be trained by providing it with a sufficiently large and representative set of training data; this technique is called supervised machine learning. Because a neural network can excel at modeling complex relationships between inputs and the outputs that are difficult to express in functional relationships, it can provide significant advantages for control of a semiconductor processing system, such as a CMP system.

FIG. 3 illustrates the functional blocks implemented by the i-APC module 94. The i-APC module 94 includes neural network input variable set 110, which includes target removal profile 112 and state parameters 114, a neutral network 120, and a process initialization module 130 that initializes the tool control module 92 of a semiconductor processing system, e.g., a CMP tool, with the output of the neural network. These functional blocks can be distributed across multiple computers.

The neutral network 120 includes a plurality of input nodes 122 for each input variable from neural network input variable set 110, a plurality of hidden nodes 124 (also called "intermediate nodes" below), and a plurality output nodes 126 that will generate the process parameters to be used for initialization. In a neural network having a single layer of hidden nodes, each hidden node 124 can be coupled to each input node 122, and each output node 126 can be coupled to each hidden node 124.

In general, a hidden node 124 outputs a value that a non-linear function of a weighted sum of the values from the input nodes 122 to which the hidden node is connected.

For example, the output of a hidden node 124, designated node k, can be expressed as:

$$\tanh\left(0.5*a_{k1}(\mathbf{I}_1)+a_{k2}(\mathbf{I}_2)+ \ldots + a_{kM}(\mathbf{I}_M)+b_k\right) \qquad \text{Equation 2}$$

where tanh is the hyperbolic tangent, a_{kx} is a weight for the connection between the k^{th} intermediate node and the x^{th} input node (out of M input nodes), and I_M is the value at the M^{th} input node. However, other non-linear functions can be used instead of tanh, such as a rectified linear unit (ReLU) function and its variants, or a sigmoid function.

FIG. 4 illustrates an example process 400 for training and using a neural network model in a CMP process. The neutral network 120 need to be configured before being used to generate baseline control parameters. The neural network 120 may be configured through training by using various methods of supervised learning. One example of supervised learning method is backpropagation, the use of forward stimulation to reset weights on the 'front' neural units, or input nodes 122.

Before the neural network can be trained, the i-APC control module 94 collects sufficient amount of training data (410). The training data comprises a set of input and output pairs that have been measured, and therefore known to be true. For example, data log maintained by the i-APC control module 94 can be used to retrieve for each substrate, the

state parameters 114 during the process; the target removal profile 112; the control parameters used during the process; and the measured removal profile. The retrieved information is then analyzed to determine its suitability as a training example. In the current example, the measured removal profile is compared with the target profile. If the difference between the two is sufficiently small to be within the specification of the process, the data is determined to be a suitable training example. However, it should be noted that the CMP polish behavior may vary substantially between substrates with different designs. Therefore, it may be beneficial to further categorize the training data into different die designs to be used for creating design-specific neural networks. Once a sufficient amount of such training examples have been collected, the training begins.

The i-APC control module 94 trains the neural network 120 (420), e.g., by backpropagation. Backpropagation is a two phase cycle of propagation and weight update. During the propagation cycle, the input vector (e.g. target removal profile and state parameters from a training example) presented to the network is propagated forward through the network, layer by layer, until it reaches the output layer. The output of the network is then compared to the recorded output (e.g. control parameter from a training example) using a cost function, and an error value is calculated for each of the neurons in the output layer. A cost function or loss function is a function that maps values of one or more variables onto a real number intuitively representing some 'cost' associated with the event. The error values are then propagated backwards, starting from the output, until each neuron has an associated error value which roughly represents its contribution to the original output. Backpropagation uses these error values to calculate the gradient of the loss function with respect to the weights in the network. Then in the weight update phase, back propagation method uses these error values to update the weights, in an attempt to minimize the loss function.

The i-APC control module 94 determines whether the trained neural network is accurate (430). After the training phase 420 is completed, the neural network's accuracy should be verified. Verification of the accuracy, for example, may include operating the neural network in its inference mode, feeding it with previously used training input, and determining whether the output is in agreement, or is sufficiently close to the correct output in the training data. If the neural network output is not sufficiently accurate, then the process

returns to step 410 to collect additional training data. If the output is accurate, then the neural network is ready to be used.

The i-APC control module 94 uses the trained neural network (440) to generate process parameters. In inference mode, the neural network outputs process parameters to be used for initialization in response to target material removal profile 112 and state parameters 114.

In some implementations, individual neural network is generated and trained for each wafer design. In such implementations, the i-APC control module 94 maintains a library of neural networks, and the controller selects a neural network trained for the current wafer design to initialize the process.

The architecture of the neural network 120 can vary in depth and width. For example, although the neural network 120 is shown with a single column of intermediate nodes 124, it can include multiple columns. The number of intermediate nodes 124 can be equal to or greater than the number of input nodes 122.

In some implementations, referring to FIG. 5, the i-APC control module 94 can be configured to have multiple neural networks 120. The neural networks 120 can operate in parallel to generate the chamber pressure values for each zones. The number of neural networks can match the number of zones. For example, there can be a first neural network 120a for the first zone 12a, a second neural network 120b for the second zone 12b and a third neural network for the third zone 12c. The output of each neural network 120 can be fed to the process initialization module 130.

The training of the neural network can be computationally intensive. Accordingly, in some implementations, training of the neural network may be done off-line, e.g., while the CMP tool is undergoing maintenance. In other implementations, the training can be done on a separate server that is part of the controller 90. In some other implementations, the training can be done in a server that is not part of the controller 90, which provides the pre-trained neural network to the controller.

The behavior of a CMP process may slowly change over time for reasons that are not factored in (e.g. wear of parts that are not considered consumables) by the input variables of the neural network. Therefore, in some implementations, a periodic retraining of the neural network may be performed to overcome such changes.

Embodiments 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 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. The polishing system can be a linear polishing system, 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 relative

orientation or positioning of the components; it should be understood that the polishing surface and substrate can be held in a vertical orientation or some other orientation with respect to gravity.

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

WHAT IS CLAIMED IS:

1. A method of processing substrates, comprising:

subjecting each respective first substrate of a first plurality of substrates to a process that modifies a thickness of an outer layer of the respective first substrate;

for each respective first substrate, recording a group of process parameter values used for the process, thus generating a plurality of groups of process parameter values;

for each respective first substrate, measuring a removal profile of the outer layer during or after the process with a monitoring system, thus generating a plurality of removal profiles;

training an artificial neural network by backpropagation using the plurality of groups of process parameter and plurality of removal profiles as training data, the artificial neural network having a plurality of input nodes to receive respective removal values from the removal profile, a plurality of output nodes to output control parameter values, and a plurality of hidden nodes connecting the input nodes to the output nodes;

for each respective second substrate of a second plurality of substrates, determining a target removal profile;

for each respective second substrate, determining respective control parameter values to apply to the respective second substrate from the output nodes of the artificial neural network by applying the target removal profile to the input nodes of the artificial neural network; and

subjecting each respective second substrate to the process using the respective control parameter values.

2. The method of claim 1, wherein the process comprises chemical mechanical polishing.

3. The method of claim 1, wherein at least some of the input nodes of the artificial neural network are configured to receive state parameter values.

4. The method of claim 3, wherein the process comprises chemical mechanical polishing.

5. The method of claim 4, wherein the control parameter values comprise pressures of chambers in a carrier head to hold a substrate against a polishing pad.

6. The method of claim 5, wherein the control parameter values comprise one or more of a platen rotation rate, carrier head rotation rate, or polishing time.

7. The method of claim 4, wherein the state parameter values comprise one or more of a retaining ring life, or a polishing pad life.

8. The method of claim 1, wherein determining the target removal profile comprises storing a desired thickness profile, receiving a measured thickness profile of the respective second substrate, and determining a difference between the measured thickness profile and the desired thickness profile.

9. The method of claim 1, wherein receiving the measured thickness profile comprises measuring a thickness profile of the respective second substrate with the monitoring system.

10. A computer program product for controlling processing of a substrate, the computer program product tangibly embodied in a non-transitory computer readable medium and comprising instructions for causing a processor to:

train an artificial neural network by backpropagation using a plurality of groups of process parameter and a plurality of removal profiles as training data, the artificial neural network having a plurality of input nodes to receive respective removal values from a removal profile, a plurality of output nodes to output control parameter values, and a plurality of hidden nodes connecting the input nodes to the output nodes;

obtain a target removal profile for each respective substrate of a plurality of substrates;

for each respective substrate, determine respective control parameter values to apply

to the respective substrate from the output nodes of the artificial neural network by applying the target removal profile to the input nodes of the artificial neural network; and

cause a semiconductor processing system to subject each respective substrate to the process using the respective control parameter values.

11. The computer program product of claim 10, wherein at least some of the input nodes of the artificial neural network are configured to receive state parameter values.

12. The computer program product of claim 11, wherein the control parameter values represent pressures of chambers in a carrier head to hold a substrate against a polishing pad, and the state parameter values represent one or more of a retaining ring life, or a polishing pad life.

13. A polishing system, comprising:

a support to hold a polishing pad;

a carrier head to hold a substrate in contact with the polishing pad, the carrier head having a plurality of chambers;

a motor to generate relative motion between the support and the carrier head; and

a controller configured to

train an artificial neural network by backpropagation using a plurality of groups of process parameter and a plurality of removal profiles as training data, the artificial neural network having a plurality of input nodes to receive respective removal values from a removal profile, a plurality of output nodes to output control parameter values, and a plurality of hidden nodes connecting the input nodes to the output nodes, wherein the control parameter values comprise pressures for the chambers in the carrier head;

obtain a target removal profile for each respective substrate of a plurality of substrates;

for each respective substrate, determine respective control parameter values to apply to the respective substrate from the output nodes of the artificial neural network by applying the target removal profile to the input nodes of the artificial neural network, wherein the respective control parameter values comprise respective pressures for the

chambers in the carrier head; and

for each respective substrate, cause the carrier head to apply the respective pressures to the chambers in the carrier head during polishing.

14. The polishing system of claim 13, wherein at least some of the input nodes of the artificial neural network are configured to receive state parameter values.

15. The polishing system of claim 14, wherein the control parameter values represent pressures of the plurality of chambers in the carrier head, and the state parameter values represent one or more of a retaining ring life, or a polishing pad life.

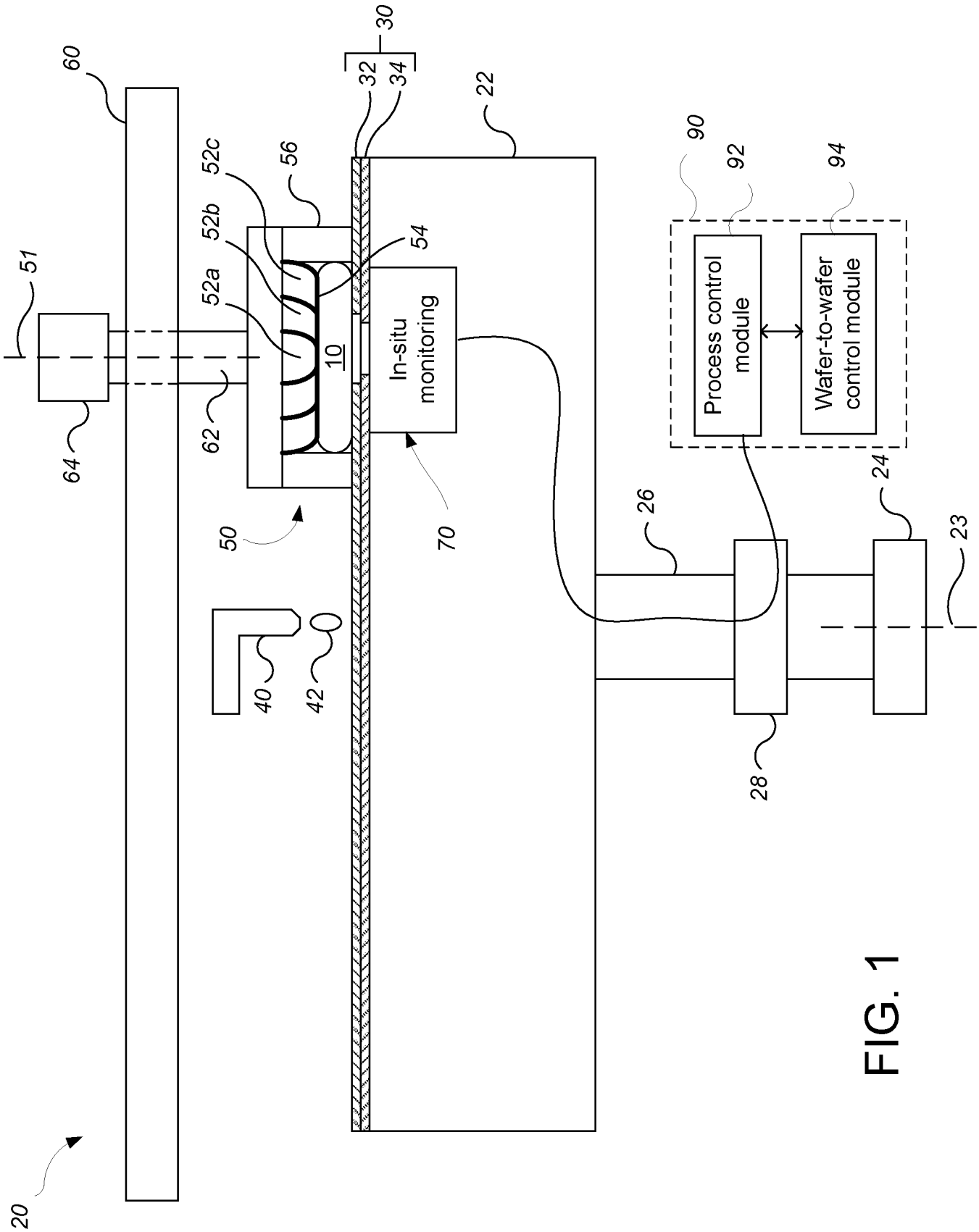

60
30
32
34
22
90
92
94
52c
56
52b
54
51
52a
10
In-situ monitoring
Process control module
Wafer-to-wafer control module
64
62
50
70
26
24
23
40
42
28
20

FIG. 1

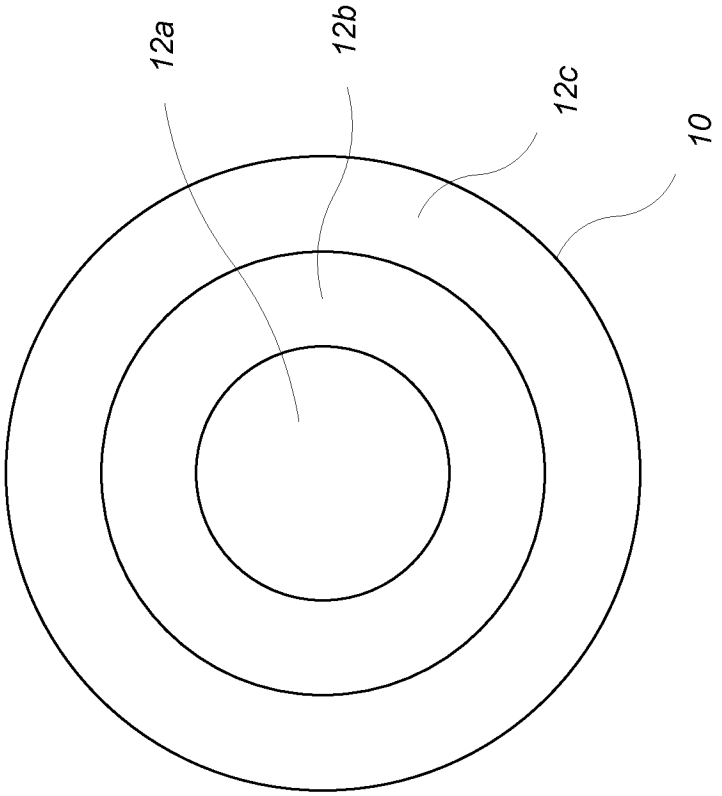


FIG.2

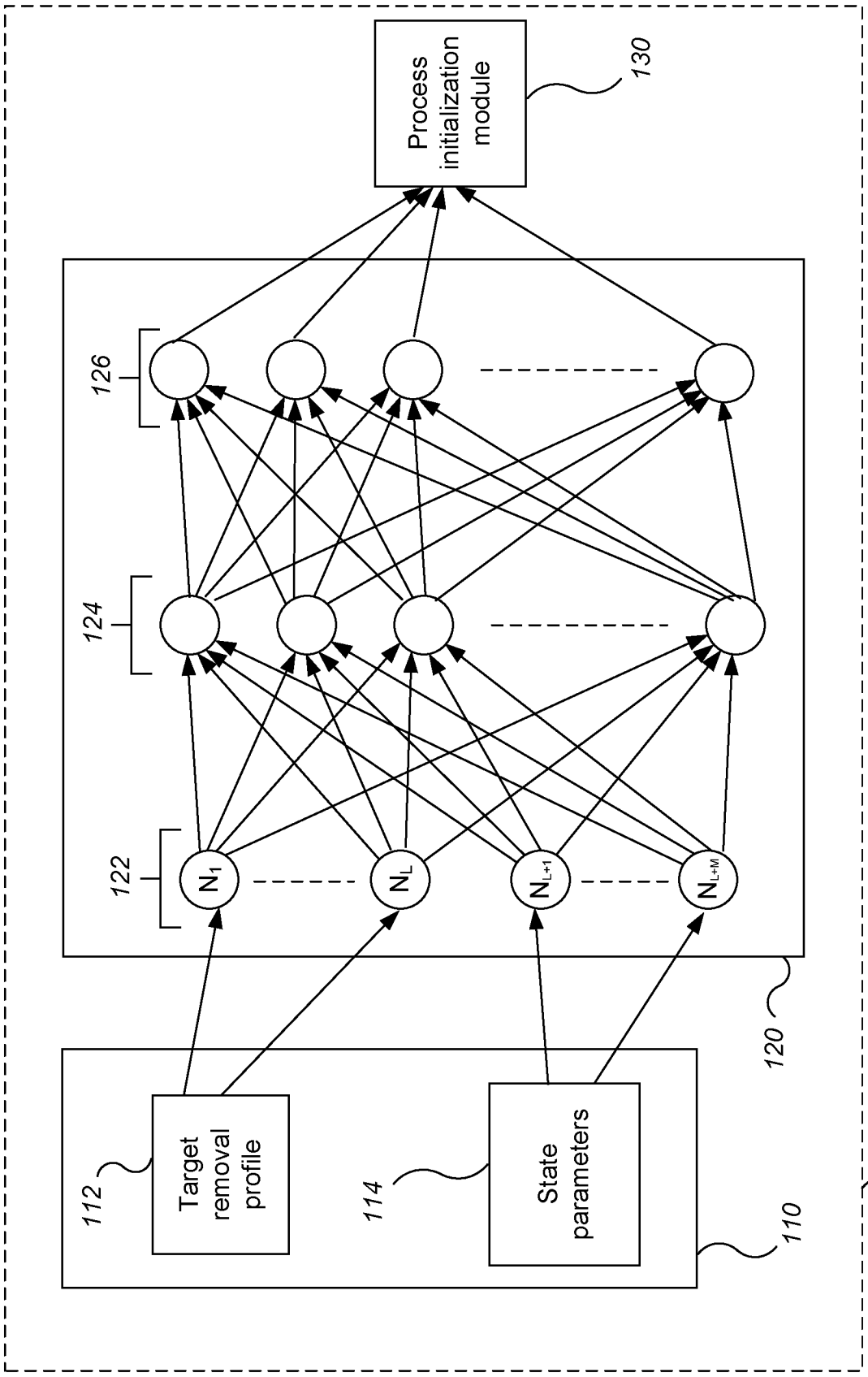

Process initialization module
130
126
124
122
N_1
N_L
N_{L+1}
N_{L+M}
120
Target removal profile
112
State parameters
114
110
94

FIG. 3

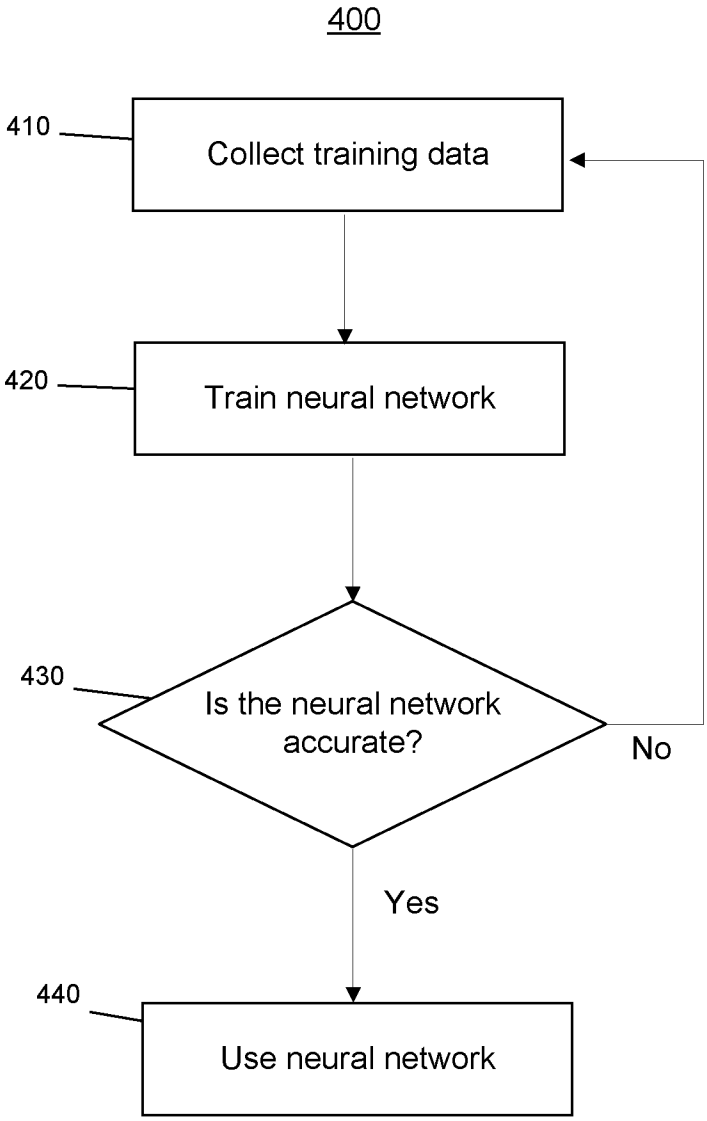

400
410
Collect training data
420
Train neural network
430
Is the neural network accurate?
No
Yes
440
Use neural network

FIG. 4

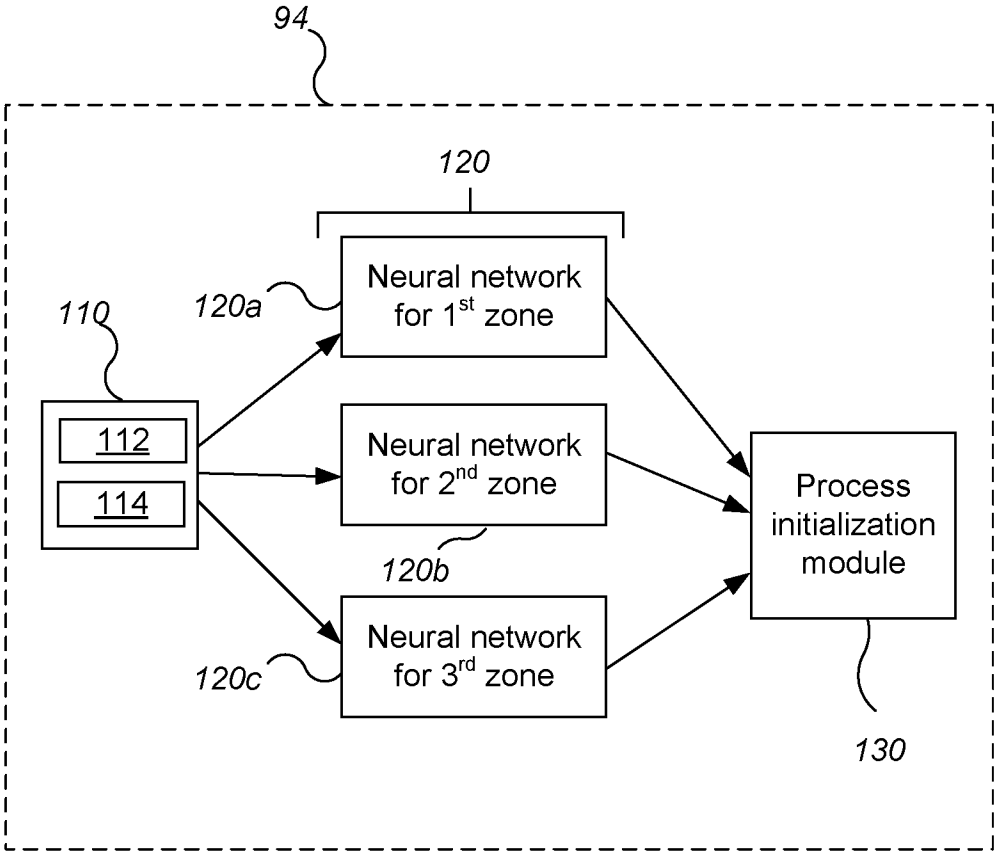

94
120
Neural network for 1st zone
120a
110
112
114
Neural network for 2nd zone
120b
Neural network for 3rd zone
120c
Process initialization module
130

FIG. 5

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US2018/052001

A. CLASSIFICATION OF SUBJECT MATTER

H01L 21/304(2006.01)i, H01L 21/3105(2006.01)i, H01L 21/321(2006.01)i, H01L 21/66(2006.01)i, H01L 21/67(2006.01)i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
H01L 21/304; B24B 49/00; B24B 49/12; B24B 51/00; G06F 17/00; G06N 99/00; H01L 21/66; H01L 21/3105; H01L 21/321; H01L 21/67

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched
Korean utility models and applications for utility models
Japanese utility models and applications for utility models

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
eKOMPASS(KIPO internal) & keywords: training data, control parameter values, CMP, target removal profile

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 7001243 B1 (JINGANG YI et al.) 21 February 2006 See column 1, line 56 - column 2, line 11, column 10, line 60 - column 20, line 50, claims 1-3 and figures 9-16.	1-15
A	US 2007-0224915 A1 (JEFFREY DRUE DAVID et al.) 27 September 2007 See paragraphs [0007], [0033]-[0060] and figure 1.	1-15
A	US 7037172 B1 (CHARLES J. MOLNAR) 02 May 2006 See column 32, lines 5-24 and figures 3-4.	1-15
A	US 2010-0129939 A1 (JEFFREY DRUE DAVID et al.) 27 May 2010 See paragraphs [0073]-[0115] and figures 2-5.	1-15
A	KR 10-2012-0123493 A (TOKYO ELECTRON LIMITED) 08 November 2012 See claim 1 and figures 5-23.	1-15

☐ Further documents are listed in the continuation of Box C.

☒ See patent family annex.

* Special categories of cited documents:
"A" document defining the general state of the art which is not considered to be of particular relevance
"E" earlier application or patent but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed
"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents,such combination being obvious to a person skilled in the art
"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
17 January 2019 (17.01.2019)	**17 January 2019 (17.01.2019)**

Name and mailing address of the ISA/KR
International Application Division
Korean Intellectual Property Office
189 Cheongsa-ro, Seo-gu, Daejeon, 35208, Republic of Korea
Facsimile No. +82-42-481-8578

Authorized officer
CHOI, Sang Won
Telephone No. +82-42-481-8291

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No.

PCT/US2018/052001

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 7001243 B1	21/02/2006	None	
US 2007-0224915 A1	27/09/2007	CN 101242931 A	13/08/2008
		CN 101242931 B	27/03/2013
		CN 101391397 A	25/03/2009
		CN 101391397 B	03/10/2012
		CN 102626895 A	08/08/2012
		CN 102626895 B	25/05/2016
		CN 104526536 A	22/04/2015
		CN 104526536 B	22/09/2017
		CN 105773398 A	20/07/2016
		CN 106272030 A	04/01/2017
		CN 106272030 B	13/11/2018
		CN 107738177 A	27/02/2018
		JP 2009-505847 A	12/02/2009
		JP 2012-256911 A	27/12/2012
		JP 2012-256912 A	27/12/2012
		JP 2012-256913 A	27/12/2012
		JP 2012-507146 A	22/03/2012
		JP 2013-048259 A	07/03/2013
		JP 2015-077684 A	23/04/2015
		JP 2015-079984 A	23/04/2015
		JP 5534672 B2	02/07/2014
		JP 5622806 B2	12/11/2014
		JP 5622807 B2	12/11/2014
		JP 5624099 B2	12/11/2014
		JP 5774482 B2	09/09/2015
		JP 6047327 B2	21/12/2016
		JP 6052906 B2	27/12/2016
		KR 10-1324644 B1	01/11/2013
		KR 10-1398567 B1	22/05/2014
		KR 10-1398570 B1	22/05/2014
		KR 10-1423579 B1	25/07/2014
		KR 10-1521414 B1	19/05/2015
		KR 10-1593927 B1	15/02/2016
		KR 10-1616024 B1	28/04/2016
		KR 10-2008-0042895 A	15/05/2008
		KR 10-2011-0090965 A	10/08/2011
		KR 10-2013-0041396 A	24/04/2013
		KR 10-2013-0041397 A	24/04/2013
		KR 10-2013-0042059 A	25/04/2013
		KR 10-2014-0019484 A	14/02/2014
		KR 10-2014-0140598 A	09/12/2014
		TW 200717637 A	01/05/2007
		TW 201022870 A	16/06/2010
		TW 201103085 A	16/01/2011
		TW I361454 B	01/04/2012
		TW I366872 B	21/06/2012
		TW I492005 B	11/07/2015

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No.

PCT/US2018/052001

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
		US 2007-0039925 A1	22/02/2007
		US 2007-0042675 A1	22/02/2007
		US 2007-0042680 A1	22/02/2007
		US 2007-0042681 A1	22/02/2007
		US 2007-0218812 A1	20/09/2007
		US 2008-0009227 A1	10/01/2008
		US 2008-0102734 A1	01/05/2008
		US 2009-0017726 A1	15/01/2009
		US 2009-0036026 A1	05/02/2009
		US 2010-0035519 A1	11/02/2010
		US 2010-0075582 A1	25/03/2010
		US 2010-0103422 A1	29/04/2010
		US 2010-0105288 A1	29/04/2010
		US 2010-0217430 A1	26/08/2010
		US 2010-0284007 A1	11/11/2010
		US 2011-0209412 A1	01/09/2011
		US 2012-0100642 A1	26/04/2012
		US 2012-0323355 A1	20/12/2012
		US 2013-0344625 A1	26/12/2013
		US 2014-0039660 A1	06/02/2014
		US 2014-0316550 A1	23/10/2014
		US 2015-0364390 A1	17/12/2015
		US 2017-0125313 A1	04/05/2017
		US 7226339 B2	05/06/2007
		US 7306507 B2	11/12/2007
		US 7406394 B2	29/07/2008
		US 7409260 B2	05/08/2008
		US 7614933 B2	10/11/2009
		US 7614936 B2	10/11/2009
		US 7651385 B2	26/01/2010
		US 7657342 B2	02/02/2010
		US 7764377 B2	27/07/2010
		US 7774086 B2	10/08/2010
		US 7931522 B2	26/04/2011
		US 7938714 B2	10/05/2011
		US 8088298 B2	03/01/2012
		US 8260446 B2	04/09/2012
		US 8392012 B2	05/03/2013
		US 8518827 B2	27/08/2013
		US 8554351 B2	08/10/2013
		US 8815109 B2	26/08/2014
		US 8874250 B2	28/10/2014
		US 9117751 B2	25/08/2015
		US 9583405 B2	28/02/2017
		WO 2007-024807 A2	01/03/2007
		WO 2007-024807 A3	12/07/2007
		WO 2007-024807 A9	03/09/2009
		WO 2010-062497 A2	03/06/2010
		WO 2010-062497 A3	05/08/2010

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No.

PCT/US2018/052001

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 7037172 B1	02/05/2006	US 2001-0027018 A1	04/10/2001
		US 2002-0002026 A1	03/01/2002
		US 2004-0115841 A1	17/06/2004
		US 2007-0095726 A1	03/05/2007
		US 2008-0057830 A1	06/03/2008
		US 2008-0306624 A1	11/12/2008
		US 2011-0294399 A1	01/12/2011
		US 2013-0189801 A1	25/07/2013
		US 6204181 B1	20/03/2001
		US 6267644 B1	31/07/2001
		US 6283829 B1	04/09/2001
		US 6291349 B1	18/09/2001
		US 6293851 B1	25/09/2001
		US 6346202 B1	12/02/2002
		US 6390890 B1	21/05/2002
		US 6413153 B1	02/07/2002
		US 6428388 B2	06/08/2002
		US 6435948 B1	20/08/2002
		US 6541381 B2	01/04/2003
		US 6551933 B1	22/04/2003
		US 6568989 B1	27/05/2003
		US 6634927 B1	21/10/2003
		US 6641463 B1	04/11/2003
		US 6656023 B1	02/12/2003
		US 6719615 B1	13/04/2004
		US 6739947 B1	25/05/2004
		US 6796883 B1	28/09/2004
		US 6986698 B1	17/01/2006
		US 7008300 B1	07/03/2006
		US 7131890 B1	07/11/2006
		US 7156717 B2	02/01/2007
		US 7220164 B1	22/05/2007
		US 7377836 B1	27/05/2008
		US 7572169 B1	11/08/2009
		US 7575501 B1	18/08/2009
		US 7878882 B2	01/02/2011
		US 7991499 B2	02/08/2011
		US 8353738 B2	15/01/2013
US 2010-0129939 A1	27/05/2010	JP 2012-510169 A	26/04/2012
		JP 5583137 B2	03/09/2014
		KR 10-1715726 B1	13/03/2017
		KR 10-2011-0102376 A	16/09/2011
		US 2010-0130100 A1	27/05/2010
		US 2010-0297916 A1	25/11/2010
		US 2012-0028377 A1	02/02/2012
		US 8039397 B2	18/10/2011
		US 8292693 B2	23/10/2012
		US 8579675 B2	12/11/2013
		US 8679979 B2	25/03/2014

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No.
PCT/US2018/052001

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
		WO 2010-062910 A2	03/06/2010
		WO 2010-062910 A3	12/08/2010
KR 10-2012-0123493 A	08/11/2012	CN 102016827 A	13/04/2011
		CN 102016827 B	19/06/2013
		CN 102449623 A	09/05/2012
		CN 102449623 B	01/07/2015
		CN 103295065 A	11/09/2013
		CN 103295065 B	28/09/2016
		JP 2011-517807 A	16/06/2011
		JP 2012-522307 A	20/09/2012
		JP 2013-518449 A	20/05/2013
		JP 2014-013581 A	23/01/2014
		JP 5542697 B2	09/07/2014
		JP 5544413 B2	09/07/2014
		JP 5732495 B2	10/06/2015
		JP 5841071 B2	06/01/2016
		KR 10-1609017 B1	04/04/2016
		KR 10-1611628 B1	26/04/2016
		KR 10-1755746 B1	07/07/2017
		KR 10-2011-0052534 A	18/05/2011
		KR 10-2011-0133497 A	12/12/2011
		SG 182776 A1	27/09/2012
		TW 200945220 A	01/11/2009
		TW 201104452 A	01/02/2011
		TW 201202876 A	16/01/2012
		TW I446280 B	21/07/2014
		TW I524189 B	01/03/2016
		TW I525407 B	11/03/2016
		US 2009-0240366 A1	24/09/2009
		US 2010-0138026 A1	03/06/2010
		US 2011-0131162 A1	02/06/2011
		US 2012-0209798 A1	16/08/2012
		US 2013-0151447 A1	13/06/2013
		US 2014-0229409 A1	14/08/2014
		US 8190543 B2	29/05/2012
		US 8396582 B2	12/03/2013
		US 8725667 B2	13/05/2014
		US 8744607 B2	03/06/2014
		US 9275335 B2	01/03/2016
		US 9424528 B2	23/08/2016
		WO 2009-114387 A1	17/09/2009
		WO 2010-114641 A1	07/10/2010
		WO 2011-094592 A1	04/08/2011