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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
8 June 2017 (08.06.2017)

WIPO | PCT

(10) International Publication Number
WO 2017/096038 A1

(51) International Patent Classification:
G01R 31/28 (2006.01) *H01L 21/301* (2006.01)

(21) International Application Number:
PCT/US2016/064412

(22) International Filing Date:
1 December 2016 (01.12.2016)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
6457/CHE/2015 1 December 2015 (01.12.2015) IN
15/130,429 15 April 2016 (15.04.2016) US

(63) Related by continuation (CON) or continuation-in-part (CIP) to earlier application:
US 15/130,429 (CON)
Filed on 15 April 2016 (15.04.2016)

(71) Applicant: **TEXAS INSTRUMENTS INCORPORATED** [US/US]; P.O. Box 655474, Mail Station 3999, Dallas, TX 75265-5474 (US).

(71) Applicant *(for JP only)*: **TEXAS INSTRUMENTS JAPAN LIMITED** [JP/JP]; 24-1, Nishi-Shinjuku 6-chome, Shinjuku-ku Tokyo, 160-8366 (JP).

(72) Inventors: **PAREKHJI, Rubin Ajit**; Flat 131, Sobha Sunflower, No. 173 Vibhuthipura, Basavanagar Main Road, Bangalore 560037 (IN). **MEHENDALE, Mahesh M.**; Ebony-310, Raheja Residency, 8th C Mn, 7th Crs, Koramangala 3rd Blck, Bangalore 560034 (IN). **MENEZES, Vinod**; 26 Lloyd Road, Cooke Town, Bangalore 560005 (IN). **SINGHAL, Vipul K.**; No. 66/3, Bagmane Tech Park, C.V. Raman Nagar Post, Bangalore 560093 (IN).

(74) Agent: **DAVIS, Jr., Michael A.**; Texas Instruments Incorporated, P.O. Box 655474, Mail Station 3999, Dallas, TX 75265-5474 (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,

[Continued on next page]

(54) Title: SYSTEMS AND METHODS OF TESTING MULTIPLE DIES

(57) Abstract: In described examples of a method (800) of testing a semiconductor wafer including a scribe line and multiple dies, the method (800) includes implementing a first landing pad on the scribe line (802), and implementing a first interconnect on the scribe line and between the first landing pad and a first cluster of the plurality of dies (804), thereby coupling the first landing pad to the first cluster of dies. The method (800) further includes performing the testing of the first cluster of dies using automated test equipment (ATE) coupled to a probe tip by contacting the first landing pad with the probe tip and applying an ATE resource to the first cluster of dies (806).

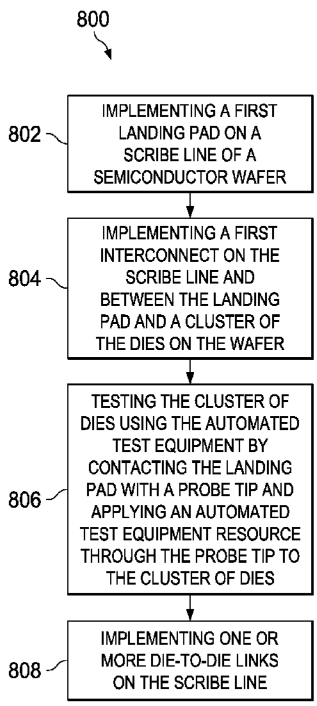

FIG. 8

WO 2017/096038 A1

HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, 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.

(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))*

SYSTEMS AND METHODS OF TESTING MULTIPLE DIES

BACKGROUND

[0001] Automated test equipment (ATE) includes multiple resources (*e.g.*, analog resources and digital resources for test and measurement) that are applied to a device under test (DUT), such as a die or dies on a semiconductor wafer. The resources are applied through an interface including one or more probe heads, where each probe head includes multiple probe tips to provide an electrical contact to landing pads on the DUT.

[0002] Conventional multi-site testing throughput is limited because a total set of one type of ATE resources may be limited to N, where M resources of that type are required to test a die, resulting in the maximum number of dies that may be tested in parallel during each touch-down of the wafer probe being N/M. Further, N/M is an ideal maximum multi-site capability. In practice, the probe card that controls routing to the various probe heads and tips may further constrain the routing density and reduce the amount of possible connections between ATE resources and multiple dies, thereby reducing the attainable multi-site factor.

[0003] In addition to the limitations imposed on the multi-site factor by physical constraints, such as available ATE resources and the design of the probe card, heads, and tips, conventional ATE testing is performed by mapping resources from the ATE onto individual dies on a wafer (or individual packaged parts in final test, where the dies/packaged parts are similarly referred to as DUTs), where all DUTs are tested identically.

[0004] Accordingly, a given test executes on all N/M DUTs where, as above, N/M is the multi-site factor (assuming no additional probe card routing constraints). However, the ATE does not necessarily include an equal number of each type of resource. As a result, the multi-site factor is determined by the resource that is least available from the ATE (*i.e.*, the maximum count of the most constrained resource). Examples of such resource limitations include number of analog channels, number of data logging channels, number of high speed interface channels, and number of clock channels. Although higher parallelism is available for ATE resources greater in number, overall test throughput is impeded by the ATE resources that are lower in number, which results in a longer testing time.

SUMMARY

[0005] In described examples of a method of testing a semiconductor wafer including a scribe line and multiple dies, the method includes implementing a first landing pad on the scribe line, and implementing a first interconnect on the scribe line and between the first landing pad and a first cluster of the plurality of dies, thereby coupling the first landing pad to the first cluster of dies. The method further includes performing the testing of the first cluster of dies using automated test equipment (ATE) coupled to a probe tip by contacting the first landing pad with the probe tip and applying an ATE resource to the first cluster of dies.

BRIEF DESCRIPTION OF THE DRAWINGS

[0006] FIG. 1 shows a schematic of die clusters on a wafer in accordance with example embodiments.

[0007] FIGS. 2a-2b shows a schematic of landing pads implemented in a scribe line on a wafer in accordance with example embodiments.

[0008] FIG. 3 shows an example probe head configuration in accordance with example embodiments.

[0009] FIGS. 4a-4c show example switch configurations for connecting to or isolating dies on a wafer in accordance with example embodiments.

[0010] FIG. 5 shows example test flows in accordance with example embodiments.

[0011] FIG. 6 shows conventional test applications and example multi-content test applications in accordance with example embodiments.

[0012] FIG. 7 shows example probe head configurations for carrying out multi-content test applications in accordance with example embodiments.

[0013] FIG. 8 shows a flow chart of a method in accordance with example embodiments.

[0014] FIG. 9 shows a flow chart of another method in accordance with example embodiments.

DETAILED DESCRIPTION OF EXAMPLE EMBODIMENTS

[0015] In this description, the term “couple” or “couples” means either an indirect or direct wired or wireless connection. For example, if a first device couples to a second device, that connection may be through a direct connection or through an indirect connection via other devices and connections.

[0016] To address the above problems, example embodiments are directed to systems and methods for testing multiple dies on a semiconductor wafer. For example, a scribe line refers to the space between dies on a wafer where a saw can safely cut the wafer without damaging the dies or circuits implemented on the wafer. Conventionally, the scribe line is a non-functional spacing that merely serves to ensure that a saw (e.g., a mechanical saw, a laser-based saw, or other known device for separating dies on a wafer) is able to effectively cut between the dies or circuits.

[0017] However, in accordance with example embodiments, a landing pad and an interconnect coupled to the landing pad are implemented in a scribe line of a wafer. The interconnect couples the landing pad to a cluster of dies on the wafer. A tip of a probe head contacts the landing pad during testing to provide an electrical connection between the probe head and the cluster of dies. Subsequently, the cluster of dies is tested using automated test equipment (ATE) that includes multiple resources as described above, which are applied to the cluster of dies (referred to collectively as a device under test (DUT)) via the landing pads. In this way, the scribe line is used to create landing pads and interconnects that allow a single probe tip to fan out, or electrically contact, a cluster of dies rather than a single die. This results in an increase in the attainable multi-site factor, depending on how many dies an interconnect couples a single landing pad to.

[0018] For example, a multi-site factor is conventionally given by N/M as described above (where N is the number of a particular available ATE resource and M is the number of resources of that type required to test a die). However, example embodiments increase the multi-site factor by a factor of L, where L is the number of dies in the cluster contacted by the scribe line-implemented interconnect and the landing pad coupled thereto. So, in a case where a landing pad is implemented in the scribe line and coupled to a die cluster of size 4 through an interconnect also implemented in the scribe line, the multi-site factor is boosted to 4*N/M, which results in an increase in test throughput and corresponding decrease in time required to test all the dies on a wafer of a given size.

[0019] Example embodiments should not necessarily be limited to implementing only a single landing pad and interconnect on a scribe line of a wafer. Instead, multiple landing pads and interconnects may be implemented on scribe lines between dies, such as by employing tunneling to provide scribe-to-die connectivity across different layers. In these examples, multiple probe tips may each contact a different landing pad in the scribe line during a single touch-down, improving the fan out of ATE resources across the wafer. In fact, in some examples, all of the dies contained on a single wafer may be coupled to landing pads accessible by the probe head in a single touch-down, which may permit a test to be concurrently performed on all the dies of a wafer. A number of dies of a wafer that are testable during a single touch-down may vary in practice, such as based on restrictions of available ATE resources, location of probe heads mounted on the probe card, and the size of die clusters.

[0020] In addition to interconnects between a landing pad and a cluster or plurality of dies, the scribe line may also include one or more die-to-die connections. These die-to-die connections allow one die to test another die. In these examples, dies may be classified as a master (i.e., the die applying the test) or a slave (i.e., the die being tested) and as functional (i.e., enabling testing using functions implemented on the die) or sacrificial (i.e., enabling testing using test-only functions). Sacrificial dies can additionally provide landing pads, improved routing between dies, and other embedded design for testability (DFT) elements to aid in test application and response measurement (e.g., a voltage regulator for providing a reference voltage, current measurement using current mirrors and resistors, and built-in self-test (BIST) controllers). Further, in some examples, DFT structures such as measurement units (e.g., resistor dividers, low-cost analog-digital converters (ADCs), and flash BIST controllers) may be implemented in the scribe line, enabling measurements on individual dies to be performed locally instead of being made using ATE resources.

[0021] In some example embodiments, one or more switches may be implemented on signal and/or power connections, either in a scribe line or within a die itself. These switches permit selective connectivity between components (e.g., landing pads, DFT structures, and the dies themselves), which may assist in the isolation of a die or dies identified as being faulty during an earlier testing procedure.

[0022] Accordingly, example embodiments allow for wafer probing with an improved multi-site factor, within the restrictions of available ATE resources, by enabling probe heads mounted on the probe card to cover or “fan out” to an increased number of dies or die clusters. In some examples, in fact, all dies on a wafer may be tested with a single touch-down event, due to implementing interconnects and landing pads within the scribe line.

[0023] Referring again to FIG. 1, schematic examples of contacting dies on a wafer are shown. In the first example 100, a prober 102 (generally referring to the combination of one or more probe tips on a probe head) is driven by an ATE 104. The prober 102 is an electromechanical assembly that transmits electrical signals from the ATE 104 to an electrical contact on the wafer. In the example 100, the mechanical prober 102, driven by the ATE 104, is configured to test a grouping of nine dies 106 in a single touch-down event. Achieving a high degree of parallelism in testing dies 106 is limited by the fact that the ATE 104 has a finite number of electrical resources (or pins).

[0024] To illustrate this limit, in the second example 150, four separate probers 152 are each driven by a dedicated ATE 104, and thus a grouping of 36 dies may be tested in a single touch-down event. However, this requires four times the resources of the first example 100—or four ATEs 154—to implement, which is not ideal. Alternatively, one ATE 154 may drive four probe heads 152, although ATE 154 resources may be constricted and thus not all 36 dies may be tested at once. Further, it may be desirable to reduce the number of probe heads 152 as well. The probers 102, 152 may include multiple probe heads, each probe head in turn including multiple probe tips. However, the examples 100, 150 are for illustrative purposes to demonstrate the limitations when attempting to expand the parallelism of testing in conventional scenarios.

[0025] Conventionally, landing pads are located on each of the various dies of a wafer (i.e., the landing pads are the pads of the particular die). To address this limitation, FIG. 2a illustrates a semiconductor wafer 200 in accordance with example embodiments, in which landing pads 202 and interconnects 204 are positioned in one or more scribe lines 206 between dies 208 of the wafer 200. The view 200a depicts landing pads 202 coupled to test routes or interconnects 204 in portions of the scribe line area 206. The view 200b depicts the remaining free scribe area 210, along with the landing pads 202 for reference. The view 200c depicts the merging of the landing pads 202, the interconnects 204, and the free scribe area 210.

[0026] FIG. 2b illustrates the landing pads 202 and the interconnects 204 in the scribe lines 206 of FIG. 2a in greater detail. For example, multiple landing pads 202a-d are shown in the scribe line 206. In this example discussion, it may be assumed that each pad 202a-d corresponds to a different ATE resource needed to test a die 208. Each landing pad 202a-d is coupled to a corresponding interconnect 204a-d, respectively, which provides routing to the dies 208. As shown, the dies 208 also include pads 220, which would conventionally be used to make electrical contact with the dies 208.

[0027] However, as described above, contacting the pads 220 limits the amount of parallelism in testing dies 208 that may be achieved. By contrast, example embodiments improve the parallelism in testing the dies 208 by utilizing (in the shown example of FIG. 2b) one pad 202a-d to contact four different dies 208. Thus, in one example where a prober 102 only contains four pins, each of a different, required ATE resource type, the example shown in FIG. 2b allows four dies 208 to be tested in a single touch-down event. Conventional wafers and testing systems and methods would require four separate touch-down events, one to test each of the dies 208 shown.

[0028] FIG. 2b is an example, and the scope of example embodiments is not limited to a 2x2 tile arrangement (nor a 7x7 tile arrangement as shown in the expanded view 250). Instead, FIG. 2b illustrates the improvements to parallelism in testing the dies 208 enabled by example embodiments in which landing pads 202 and interconnects 204 are positioned in the scribe line 206 to “fan out” to a number of dies 208 in the area. The number of landing pads 202 and interconnects 204 able to be positioned in the scribe line 206 may vary based on the width of the scribe line 206 and the width of the interconnects 204 (e.g., signal lines may be narrower, while interconnects 204 carrying a higher current may be thicker). Also, tunneling may be employed to enable scribe-to-die connectivity across different layers of the scribe line 206, as shown by the four interconnects 204a-d in FIG. 2b.

[0029] FIG. 3 illustrates a distributed probe head configuration relative to a single probe head configuration. For example, FIG. 3 shows two example probe heads 300, 350 in accordance with example embodiments. The probe head 300 contains a single prober 302 and pogo connectors 303 (or other spring-loaded devices to allow for compliance when contacting a wafer). The prober 302 includes a plurality of pins 305 that provide electrical connectivity between ATE resources 304 and a pad on the wafer when in contact. The probe head 350 includes similar pogo connectors 353 and several probers 352, which can thus land on a wafer in multiple locations. In some examples, depending on the prober 300, 352, the probe head 302, 352 configuration, how the wafer is laid out, how much current is required to test each die on the wafer, how many pads are required to test each die, the entire wafer may be tested in a single touch-down event because the probers 302, 352 and associated pins (e.g., pins 305) may be “fanned out” across multiple dies as described above.

[0030] FIG. 4a illustrates example switch configuration or topology for connecting to and/or isolating individual (or clusters of) dies. FIG. 4 shows a scribe line 206 between two dies 208, similar to those described above with respect to FIG. 2a. Further, the scribe line contains landing pads 202, depicted as rectangular elements. Also, the dies 208 themselves contain landing pads 402, which may be used before or after the scribe line 206 has been cut. As shown, the landing pads 202 in the scribe line 206 may couple to one or more dies 208 as described above. The scribe line 206 also includes internal connections 404 that enable connectivity between the scribe 206 and the die 208, but are not landing pads 202. The internal connection 404 may comprise a logic element, such as a resistor divider or design for testability (DFT), which may be used to drive one or more of the surrounding dies 208. Alternatively, the internal connection 404 may be more “simple” and merely present a metal-to-metal connection between different layers of the scribe line 206 (e.g., between different interconnects, described above with respect to FIG. 2a).

[0031] FIG. 4b demonstrates that pad landings may occur on either the landing pad 202 on the scribe line 206 or a landing pad 402 on the die itself. For example, the landing pad 202 may be used during a wafer test, whereas (for example) the landing pad 402 may be used after the pad of the die 208 is bonded to a wire to be placed into a packaged part. Further, a switch 410 may allow for isolation of this particular die 208 from the landing pad 202 in the event that the particular die 208 fails a test, but the remaining dies 208 in its cluster are still to be tested.

[0032] FIG. 4c demonstrates a landing pad 402 only on a die, which is coupled to an internal connection 404 by a switch 412. Inclusion of the landing pad 402 on the die is beneficial so that the die may be tested even after the scribe line is sawed through. As described above, the internal connection 404 may comprise a logic element or may simply provide a connection to another layer of the scribe line 206, such as to couple to another landing pad or die. Thus, even in the example of FIG. 4c, although a landing pad 202 itself is not implemented in the scribe line 206, testing of the die 208 and surrounding dies may be improved through their access to a common logic element 404 in the scribe line 206, which would not otherwise exist. For example, either of the switches 410, 412 may be operated, so that after a die 208 is identified as bad or faulty, that die may be isolated from power and input signals or, in the case of 412, isolated from access to the logic resource or connection 404.

[0033] FIG. 5 illustrates various example wafer probe test flows in accordance with example embodiments. From the outset, the “phases” corresponding to columns of the table do not necessarily relate to the time taken to perform the phase. Instead, the phases are intended to show the possible steps that may be taken during a particular test flow.

[0034] With respect to all test flows, a phase 0 scribe characterization test is performed, which is an electrical test that occurs before actual testing of the individual dies on the wafer. Accordingly, upon manufacture of the wafer, a so-called wafer health flow is performed to ensure that the wafer as a whole meets basic requirements. Subsequent to this phase 0 step, the various test flows may differ.

[0035] For example, in the conventional test flow, the first phase includes probing of the dies on the wafer in a conventional manner, such as contacting landing pads on the dies with the probe head and applying ATE resources to those contacted dies. This process may be repeated a number of times until all the dies on the wafer have been suitably tested. Then, the second phase comprises a final test, which may include a test of a die integrated as a packaged part.

[0036] However, in accordance with example embodiments, test flows 1 and 2 begin with a test of the scribe itself. For example, because example embodiments utilize the scribe for connectivity—a landing pad to fan out to multiple dies by various interconnects and/or other internal connections such as logic elements—the scribe itself is tested before beginning testing of any of the associated cluster of dies. Both of test flows 1 and 2 then proceed to test both the scribe and dies, such as using the fanned out scribe-implemented landing pads described above. The fanning out of the scribe-implemented landing pads permits a higher multi-site than would otherwise be achievable within the context of a single touch-down event.

[0037] Test flow 1 continues in phase 3 to test the dies themselves (again, using the fanned out scribe-implemented landing pads in some examples), while test flow 2 includes an optional step of testing a die-to-die connection; test flow 2 also then proceeds to test the dies themselves. After the dies have been tested, both test flows 1 and 2 may proceed to testing dies in the context of being a packaged part.

[0038] The above-described examples improve upon the attainable multi-site factor by fanning out a single landing pad to a plurality of dies, increasing the number of dies reachable in a single touch-down event. In some cases, the method used for testing dies includes mapping ATE resources onto individual dies or DUTs on the wafer or DUTs as packaged parts in a final test. However, it may be that all DUTs are tested identically; that is, a same given text executes on all the DUTs being tested at a given time based on the multi-site factor of the particular testing system, which may waste some resources. For example, resources required exclusively for "Test B" are unutilized while "Test A" is performed. As a result, the overall system multi-site factor is determined by the ATE resource that is least available (i.e., the maximum count of the most constrained ATE resource). Examples of ATE resources include analog channels, digital channels, data logging channels, high-speed interface channels, and clock channels. Commonly, an equivalent number of each resource is not available, which leads to the multi-site factor constraint described above.

[0039] Thus, although higher parallelism is available for ATE resources that are greater in number, the overall test throughput (which requires multiple resources to be applied) may still be somewhat impeded by the constrained number of ATE resources. Accordingly, for constrained ATE resources, the test time increases as tests across multiple DUTs are performed serially. However, in developing the example embodiments, it was determined that a typical maximum ATE resource utilization is approximately 70%. Accordingly, across different applications of tests in a test schedule that utilize a varying set of ATE resources for different tests in the schedule, approximately only 70% of the ATE resources are used on average across the time duration required to run all the tests in the test schedule.

[0040] To address these issues, some example embodiments may test multiple DUTs (e.g., dies on a wafer or packaged parts) using available ATE resources (or resources from similar test processing equipment) with different tests being applied to different DUTs at the same time. For example, different DUTs are tested using different tests (or different ATE resources) at the same time, which results in a summing of ATE resources that are able to be applied at one time. Due to physical constraints (e.g., the arrangement of probe heads, probe tips, and landing pads), the number of different ATE resources applied to DUTs at the same time may not be the entire sum of the numbers of those ATE resources. For example, if 70 resources of type A and 30 resources of type B are available, physical constraints involved in contacting the wafer or DUTs may result in all 70 resources of type A being applied to certain DUTs, while only 20 resources of type B are applied to other DUTs. Regardless, more resources are applied during a single touch-down event than conventional testing using a single set of resources to apply one test per touch-down event.

[0041] For example, multiple DUTs are tested concurrently, although different tests may be applied to different ones of the DUTs at the same time. In the context of testing dies on a wafer, ATE resources are mapped onto different pads on the wafer using various probe head configurations mounted on a probe card. Similarly, in the context of testing packaged parts, ATE resources are mapped onto different packaged device pins on an ATE load board using various routing and relay configurations. The ATE hardware may be configured to allow dynamic allocation of resources connected to various ATE channels to different DUT pads or pins. The particular allocation of ATE resources to the ATE channels is controlled by an internal ATE test program. Thus, example embodiments improve ATE resource utilization by scheduling tests concurrently that leverage otherwise-unused ATE resources. As a result of concurrent testing, overall test throughput is increased while idle ATE resources for a given touch-down event are reduced. The disclosed examples may apply to both ATE-based testing (in which probe cards are used to test dies on a wafer) and board-based testing, in which a load board is used to test packaged dies.

[0042] In some example embodiments, a plurality of dies or DUTs are grouped together to form a “cluster,” where each DUT is part of only one cluster. DUTs within a cluster are tested concurrently, and may be tested with different test content (or different applied ATE resources). However, all clusters on the wafer may be identical and tested concurrently. Thus, a cluster may be viewed as a DUT itself, composed of sub-modules each of which is also a DUT (e.g., an individual die). As a result, the ATE interfaces with the clusters, which make up the wafer, all of the clusters being identical and tested with identical content.

[0043] The described examples overcome conventional bottlenecks in test time, test throughput, attainable multi-site factor (and thus test concurrency), and resource utilization in various scenarios. For example, where constraints exist due to DUT pins because each DUT offers only a limited number of pins for testing, example embodiments may be leveraged in at least one of two ways. First, the number of DUT pins available for test may be increased. Accordingly, the test mode pin-muxing is relaxed, with different pins contacted for different tests. However, this may complicate the probe head/ load board relay designs in order to allocate different ATE resources to different pins of the DUT. Second, the same set of DUT pins may be utilized for application of different tests, where each test may require application of a dedicated pre-amble to internally assign the pins to the relevant module inside the DUT being tested (i.e., apply a pin-muxing preamble). In both the cases, the ATE throughput increases because different ATE resources are used to test different DUTs with different test content.

[0044] Referring again to FIG. 6, for simplicity, each DUT may require two tests, referred to as test A and test B. For test A to be performed on a DUT, two ATE resources of type A are required. Similarly, for test B to be performed on a DUT, two ATE resources of type B are required. The example ATE utilized in FIG. 6 is capable of supplying four resources of type A and two resources of type B. Again, for simplicity, each ATE resource connects to one DUT pin, so: 4 pins of resource type A are available from the ATE, and 2 pins of resource type B are available from the ATE.

[0045] The example 602 illustrates a conventional test flow, in which only one DUT is capable of being tested with resource B at a time, because only two resources of type B are available, and test B requires two resources of type B. Thus, DUT1 testing is complete in 2 cycles, DUT2 testing is complete in 2 cycles, and DUT3 testing is complete in 2 cycles.

[0046] The example 604 illustrates another conventional test flow in which test A is applied to two dies simultaneously (i.e., DUT1 and DUT2 in cycle 1 and DUT3 and DUT4 in cycle 4). This is enabled by the fact that four ATE resources of type A are available, but test A only requires two resources of type A. However, because the amount of resources of type B is constrained, then the application of test B to the DUTs must occur serially.

[0047] In accordance with example embodiments, examples 606 and 608 improve upon the conventional test flows 602 and 604 by applying dissimilar test content to different DUTs concurrently. In the example implementation 606 during cycle 1, test A is applied to DUT1 and test B is applied to DUT2. Continuing on, during cycle 2, test A is applied to DUT 3 and test B is applied to DUT1, while in cycle 3, test A is applied to DUT2 and test B is applied to DUT3. Thus, in the first three cycles, tests A and B have been applied to three DUTs. Cycles 4-6 are similarly used to apply tests A and B to DUTs 4-6. In the example 606, the same test is never concurrently applied to multiple DUTs (i.e., test A because that is the only test capable of concurrent application to more than one DUT). Importantly, however, different types of tests (i.e., test A and test B) are concurrently applied to different DUTs in the same cycle, resulting in the improvement of testing efficiency over conventional examples 602 and 604.

[0048] Example 608 presents similar benefits relative to example 606. However, the test flow illustrated by the example implementation 608 demonstrates additional benefits that may be achieved by example embodiments of multi-content testing. For example, because four resources are of type A, and because applying test A only requires two resources of type A, during cycle 1, test A is applied to both DUT1 and DUT2 and test B is applied to DUT2. Then, during cycle 2, test A is applied to DUT2 (completing DUT2's test process) and test B is applied to DUT1 (completing DUT1's test process). Finally, during cycle 3, test B is applied to DUT3. Cycles 4-6 are similarly utilized for DUTs 4-6. Thus, as was the case with example 606 described above, in the first three cycles, tests A and B have been applied to three DUTs.

[0049] However, as illustrated in example 608, resource A is idle for effectively three cycles' worth of time (i.e., over the three cycles required to test three DUTs). Certain real-world examples may take advantage of this, such as where a DUT requires an additional test A' that also utilizes resource of type A, or where test A requires additional time to complete relative to test B (e.g., double the time of test B in example 608).

[0050] FIG. 7 shows various configurations of probe heads, having varying numbers of probe tips, that may be used to perform the multi-content test in accordance with example embodiments. The example configuration 706a shows one way that the test of example 606 above may be performed using a probe head having 12 probe tips. For example, each DUT has four probe tips dedicated to that die: two to deliver resources of type A and two to deliver resources of type B.

[0051] The configuration 706b represents some examples in which the size of the probe head can be reduced, in this case to a probe head having six probe tips. Each pair of probe tips is dedicated to a DUT, and thus resources of type A and type B are mapped to the same pins of that particular DUT. In this example 706b, a pin-muxing preamble is applied (notated by a double boundary between cycles) to inform the DUT that the ATE is about to switch from application of resources of type A to application of resources of type B (or vice versa in the case of DUT2). This improves the efficiency because in 706a, eight pins were unused every cycle, whereas in 706b, only two pins are unused every cycle.

[0052] The configuration 708 represents some examples in which the probe head may be somewhat reduced in size, but rather than utilizing a pin-muxing preamble as in the example 706b, the probe head itself may be moved between cycles to contact different pins of the DUTs. This reduction in the number of probe pins is achieved by assigning idle probe pins during the testing of one die to the testing of a neighboring die. This is feasible when the topology of the pins within the die has a similar pattern to the topology of the pins across two neighboring dies.

[0053] Referring again to FIG. 8, a method 800 is shown in accordance with example embodiments. The method 800 begins in block 802 with implementing a first landing pad on a scribe line of a semiconductor wafer including a plurality of dies. The method 800 continues in block 804 with implementing a first interconnect on the scribe line and between the landing pad and a cluster of the dies on the wafer. As described above, by implementing such an interconnect between a scribe-based landing pad and a cluster of dies, a single probe tip may be used to apply an ATE resource to multiple dies due to the fact that the interconnect fans out from the landing pad to the cluster of dies. This improves the attainable multi-site factor for a given touchdown event. Thus, the method 800 also continues in block 806 with testing the cluster of dies using the ATE by contacting the landing pad with the probe tip and applying an ATE resource through the probe tip to the cluster of dies.

[0054] In some cases, the method 800 may continue in block 808 with implementing one or more die-to-die links on the scribe line, which enables one die to test another die by the die-to-die link. For example, a receiver circuit of one die can be used to test a transmitter circuit of a neighboring die, and vice versa. In addition to enabling a wider variety of testing flows, this may also help reduce the dependency on the ATE for transmitter-receiver testing.

[0055] In examples, additional landing pads and interconnects may be implemented on the scribe line to enable testing of additional cluster(s) with another probe tips coupled to the ATE in a single touchdown event. In some other examples, the landing pads, interconnects, and probe tips may be configured in a way that enables all of the dies on the wafer to be tested concurrently in a single touchdown event.

[0056] During testing of the one or more clusters, the ATE may monitor a response from the dies as a result of the application of the ATE resource. Further, the testing may indicate the presence of a bad die in one or more of the clusters, which may be isolated (e.g., for further testing or discarding). In some cases, the bad die may be isolated by operating a programmable switch to decouple that die from the rest of the cluster. However, in other cases, such as where such switches are unavailable or where efficiency or timing constraints require, the cluster containing the bad die may be discarded and testing on other clusters of dies on the wafer is performed.

[0057] FIG. 9 shows another method 900 in accordance with example embodiments. The method 900 begins in block 902 with contacting first and second devices under test with probe tips coupled to ATE. Examples of devices under test may include dies on a wafer, clusters of dies on a wafer, and packaged parts. Further, where the device under test comprises a cluster of dies on a wafer, the contact may be by a landing pad implemented on a scribe line and a coupling interconnect, as described above. In some cases, the devices under test may each comprise a cluster of packaged parts that is coupled to pins on the probe head using routing on an ATE hardware board. The method 900 continues in block 904 with applying a first ATE resource to the first device under test while applying a second ATE resource to the second device under test. As described above, in some cases the ATE lacks a sufficient amount of one of the resources to apply that resource to the first and second devices under test concurrently. Thus, but controlling the ATE such that different resources are applied to different devices under test concurrently, an increased test throughput is enabled by utilizing resources that conventionally are wasted.

[0058] In some cases, the method 900 further continues in block 906 with generating a pin-muxing preamble for one of the devices under test to configure that device under test for the ATE resource that is to be applied to that device under test. In this way, multiple types of resources may be mapped to the same pin of a device under test, while allowing for configuration of that pin before the application of a different resource type.

[0059] In the foregoing discussion, various examples utilize particular probe head designs, probe tip numbers and pin constrictions on DUTs.

[0060] Modifications are possible in the described embodiments, and other embodiments are possible, within the scope of the claims.

CLAIMS

What is claimed is:

1. A method of testing a semiconductor wafer comprising a scribe line and a plurality of dies, the method comprising:

implementing a first landing pad on the scribe line;

implementing a first interconnect on the scribe line and between the first landing pad and a first cluster of the plurality of dies, thereby coupling the first landing pad to the first cluster of dies; and

performing the testing of the first cluster of dies using automated test equipment (ATE) coupled to a probe tip by: contacting the first landing pad with the probe tip; and applying an ATE resource to the first cluster of dies.

2. The method of claim 1 further comprising implementing one or more die-to-die links on the scribe line.

3. The method of claim 2 further comprising using a first one of the plurality of dies to perform a test on a second one of the plurality of dies using the die-to-die link implemented on the scribe line.

4. The method of claim 1 further comprising:

implementing a second landing pad on the scribe line;

implementing a second interconnect on the scribe line and between the second landing pad and a second cluster of the plurality of dies, thereby coupling the second landing pad to the second cluster of dies; and

performing concurrent testing of the first and second clusters by: contacting each of the first and second landing pads with probe tips coupled to the ATE; and applying the ATE resource to the first and second clusters of dies.

5. The method of claim 4 further comprising:

implementing additional landing pads and interconnects on the scribe line such that each of the plurality of dies of the wafer is coupled to one of the landing pads; and

testing all of the plurality of dies of the wafer concurrently.

6. The method of claim 1 wherein performing the testing further comprises observing a response from the first cluster of dies as a result of the application of the ATE resource.

7. The method of claim 1 wherein a result of the testing indicates a presence of a bad die in

the first cluster, the method further comprising isolating the bad die from the other dies in the cluster.

8. The method of claim 7 wherein the bad die is isolated from the other dies in the cluster by operating a programmable switch.

9. The method of claim 7 further comprising discarding the first cluster as a result of the presence of the bad die in the first cluster.

10. A system for testing a plurality of semiconductor dies, the system comprising:

a semiconductor wafer comprising: a scribe line between at least some of the plurality of semiconductor dies; a first landing pad on the scribe line; and a first interconnect on the scribe line and between the first landing pad and a first cluster of the plurality of dies, the first interconnect configured to couple the first landing pad to a first cluster of the dies; and

a probe tip coupled to automated test equipment (ATE) and configured to: contact the first landing pad; and apply an ATE resource to the first cluster of dies to test the first cluster of dies.

11. The system of claim 10 wherein the scribe line comprises one or more die-to-die links.

12. The system of claim 11 wherein a first one of the plurality of dies is configured to perform a test on a second one of the plurality of dies using the die-to-die link implemented on the scribe line.

13. The system of claim 10 further comprising:

a second landing pad on the scribe line;

a second interconnect on the scribe line and between the second landing pad and a second cluster of the plurality of dies, the second interconnect configured to couple the second landing pad to the second cluster of dies; and

a second probe tip coupled to the ATE configured to contact the second landing pad and apply the ATE resource to the second clusters of dies to concurrently test the first and second clusters of dies.

14. The system of claim 13 further comprising:

additional landing pads and interconnects on the scribe line such that each of the plurality of dies of the wafer is coupled to one of the landing pads; and

a plurality of probe tips coupled to the ATE, each of the probe tips configured to contact one of the landing pads and apply the ATE resource to each of the plurality of dies to concurrently test all of the plurality of dies of the wafer.

15. The system of claim 10 wherein the probe tip is further configured to receive a response from the first cluster of dies as a result of the application of the ATE resource.

16. The system of claim 10 wherein a result of the test indicates a presence of a bad die in the first cluster, the probe tip configured to isolate the bad die from the other dies in the cluster.

17. The system of claim 16 wherein the semiconductor wafer further comprises one or more switches that, when caused to open, isolate the bad die from the other dies in the first cluster.

18. A method of testing first and second devices under test using automated device equipment (ATE), the method comprising:

performing concurrent testing of the first and second devices under test by: contacting each of the first and second devices under test with probe tips coupled to the ATE; and applying a first ATE resource to the first device under test while applying a second ATE resource to the second device under test;

wherein the ATE lacks a sufficient amount of one of the resources to apply that resource to the first and second devices under test concurrently.

19. The method of claim 18 further comprising generating a pin-muxing preamble for one of the devices under test to configure that device under test for the ATE resource that is to be applied to that device under test.

20. The method of claim 18 wherein the devices under test each comprise one or more dies on a wafer.

21. The method of claim 18 wherein the devices under test each comprise one or more packaged parts.

22. The method of claim 18 wherein the devices under test each comprise a cluster of dies on a wafer, the cluster of dies coupled to a landing pad implemented in a scribe line of the wafer.

23. A system for testing first and second devices under test, the system comprising:

a plurality of probe tips coupled to automated test equipment (ATE) and configured to: contact each of the first and second devices under test; and apply a first ATE resource to the first device under test while applying a second ATE resource to the second device under test;

wherein the ATE lacks a sufficient amount of one of the resources to apply that resource to the first and second devices under test concurrently.

24. The system of claim 23 wherein the probe tip is configured to apply a pin-muxing preamble to one of the devices under test to configure that device under test for the ATE resource that is to

be applied to that device under test.

25. The system of claim 23 wherein the devices under test each comprise one or more dies on a wafer.

26. The system of claim 23 wherein the devices under test each comprise a packaged part.

27. The system of claim 23 wherein the devices under test each comprise a cluster of dies on a wafer, the cluster of dies coupled to a landing pad implemented in a scribe line of the wafer.

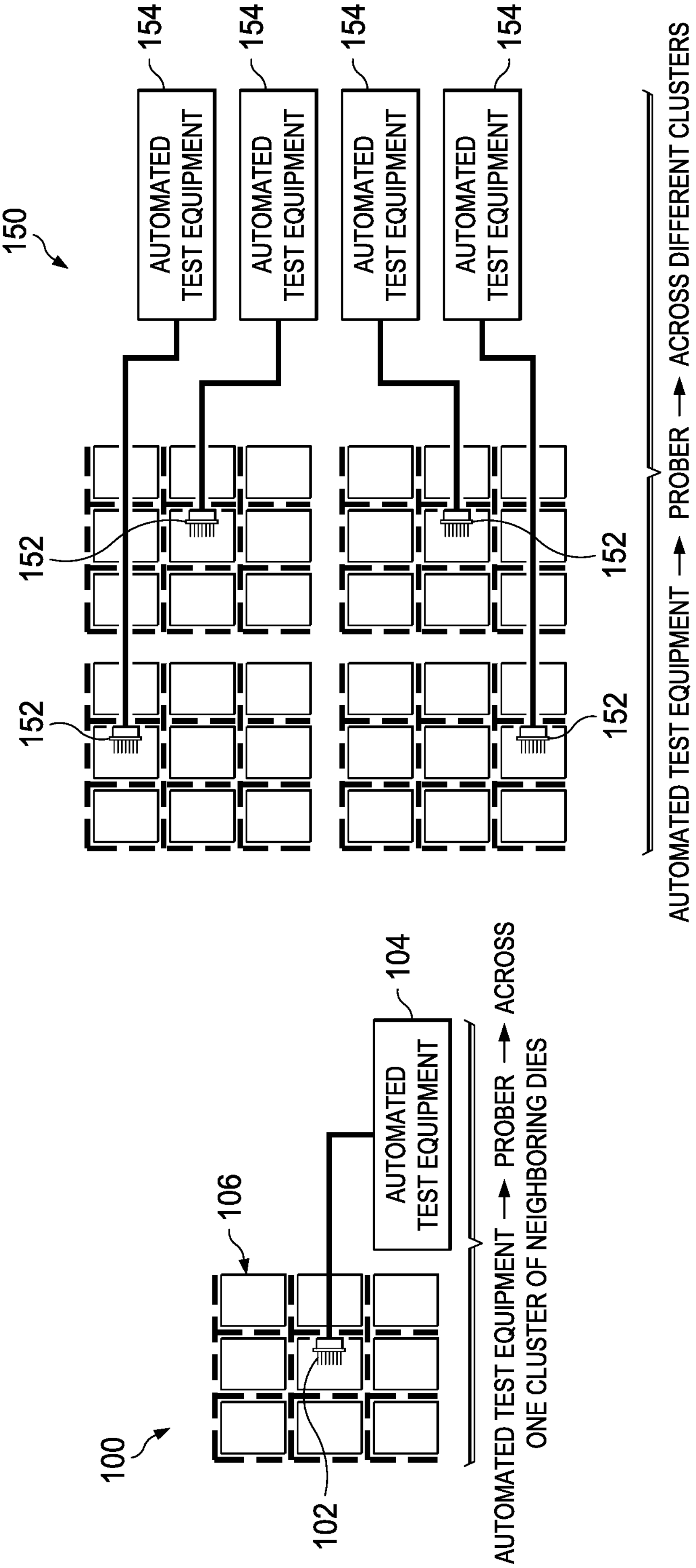
150
154
154
154
154
AUTOMATED TEST EQUIPMENT
AUTOMATED TEST EQUIPMENT
AUTOMATED TEST EQUIPMENT
AUTOMATED TEST EQUIPMENT
152
152
152
152
AUTOMATED TEST EQUIPMENT → PROBER → ACROSS DIFFERENT CLUSTERS
100
106
104
AUTOMATED TEST EQUIPMENT
102
AUTOMATED TEST EQUIPMENT → PROBER → ACROSS ONE CLUSTER OF NEIGHBORING DIES

FIG. 1

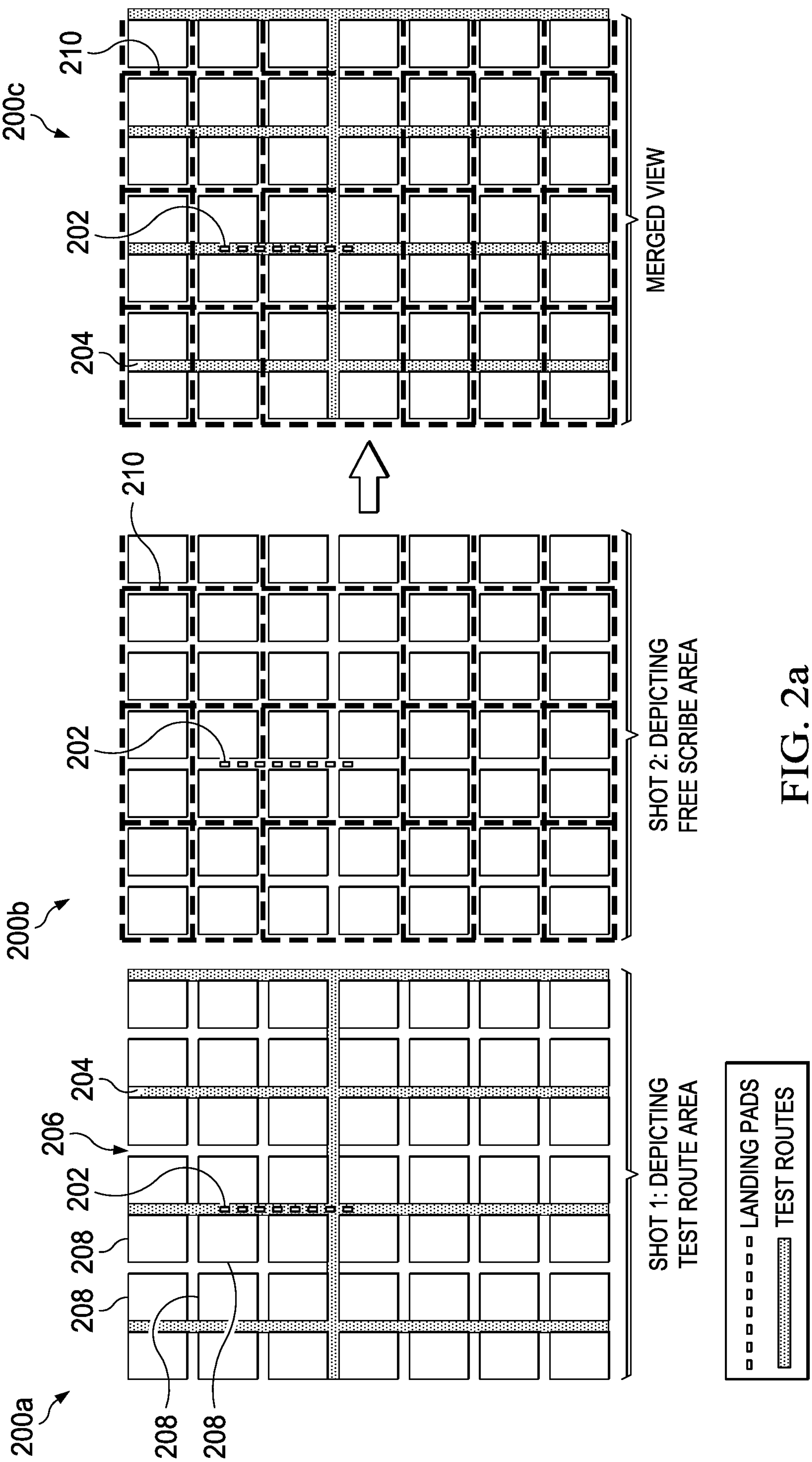
200a
200b
200c
202
204
206
208
210
SHOT 1: DEPICTING TEST ROUTE AREA
SHOT 2: DEPICTING FREE SCRIBE AREA
MERGED VIEW
LANDING PADS
TEST ROUTES

FIG. 2a

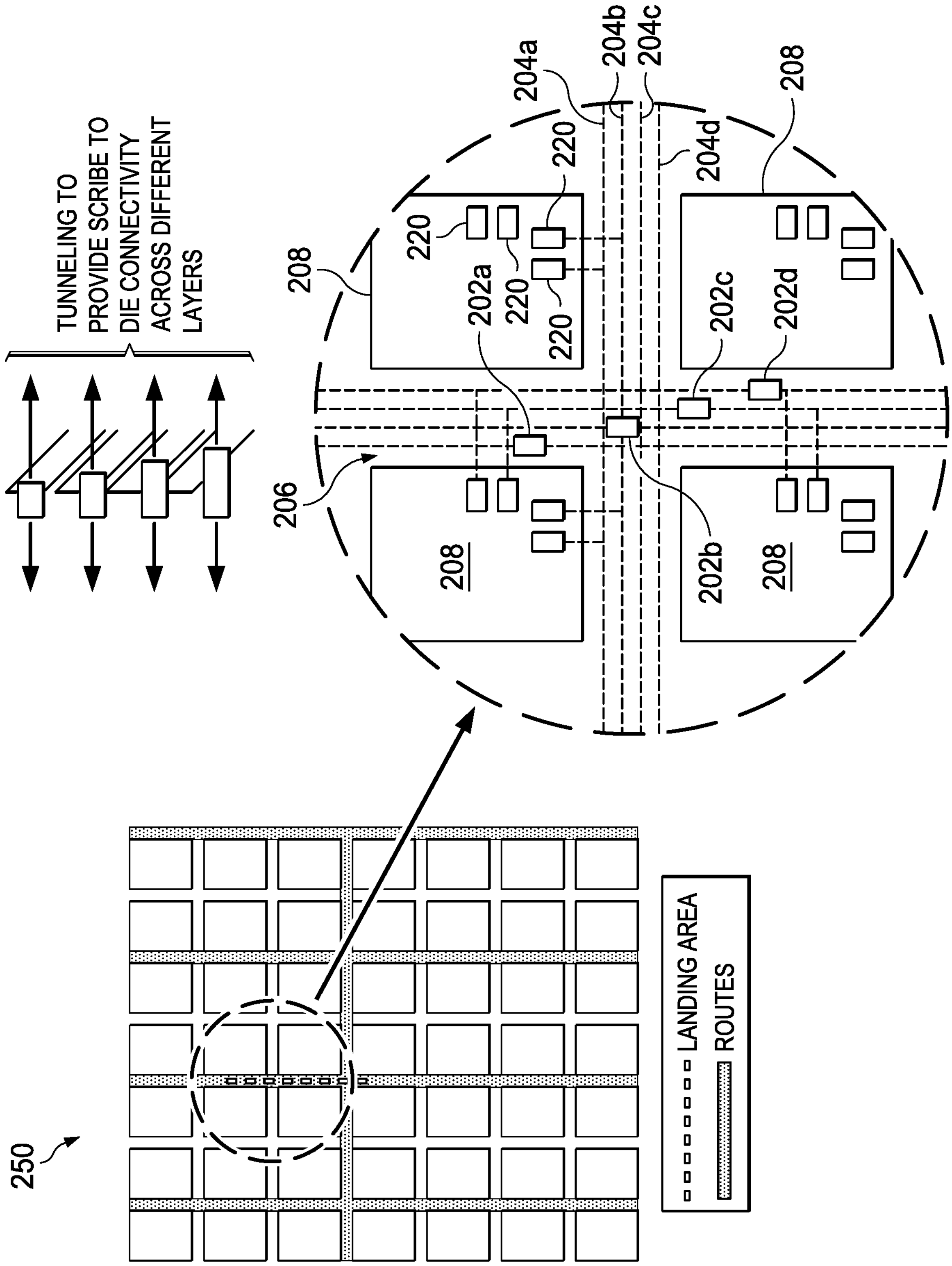
TUNNELING TO PROVIDE SCRIBE TO DIE CONNECTIVITY ACROSS DIFFERENT LAYERS
204a
204b
204c
204d
208
220
202a
202b
202c
202d
206
250
LANDING AREA
ROUTES

FIG. 2b

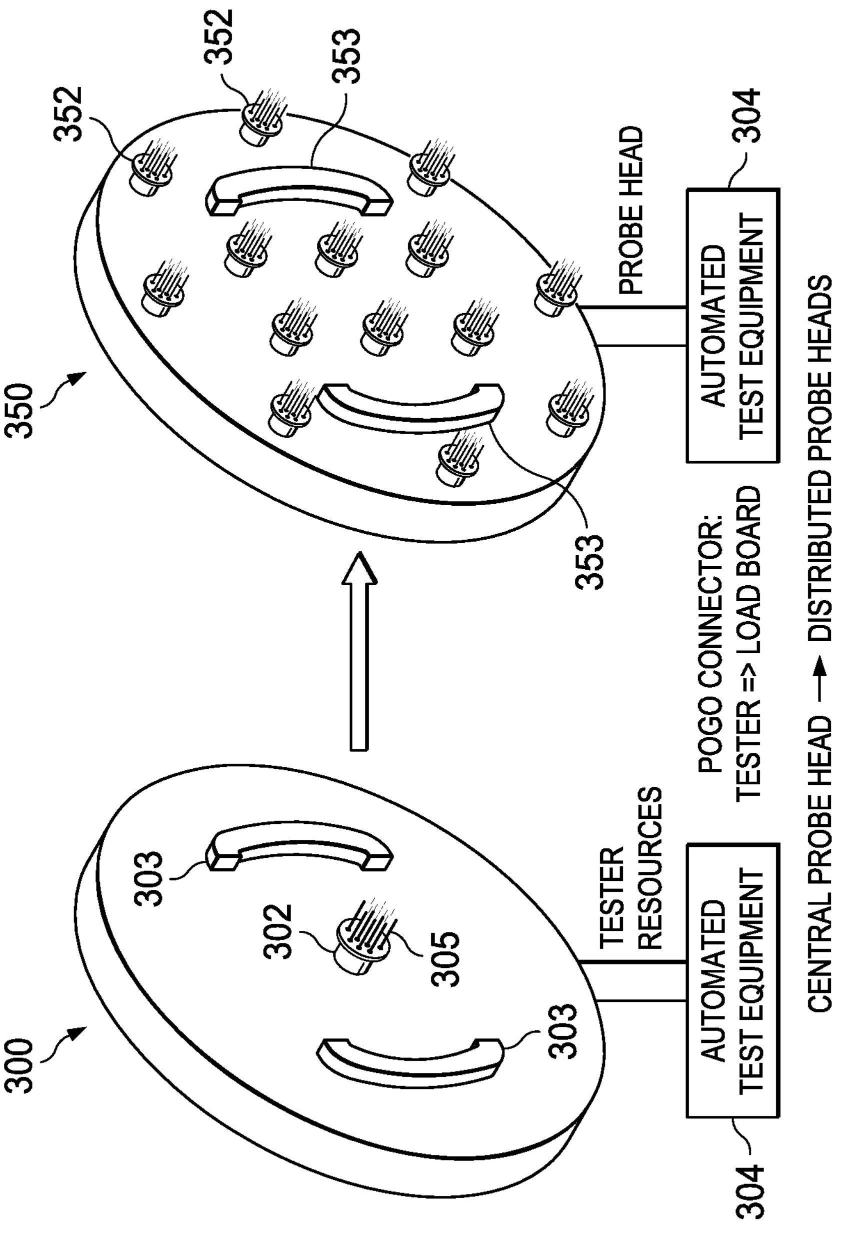
300
302
303
303
305
TESTER RESOURCES
AUTOMATED TEST EQUIPMENT
304
350
352
352
353
353
PROBE HEAD
AUTOMATED TEST EQUIPMENT
304
POGO CONNECTOR: TESTER => LOAD BOARD
CENTRAL PROBE HEAD → DISTRIBUTED PROBE HEADS

FIG. 3

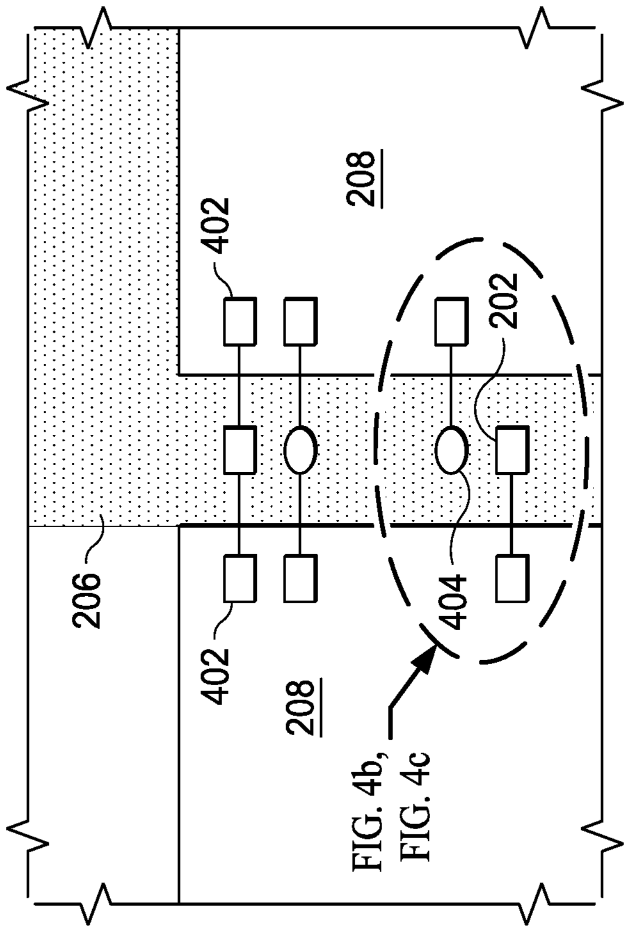

FIG. 4a

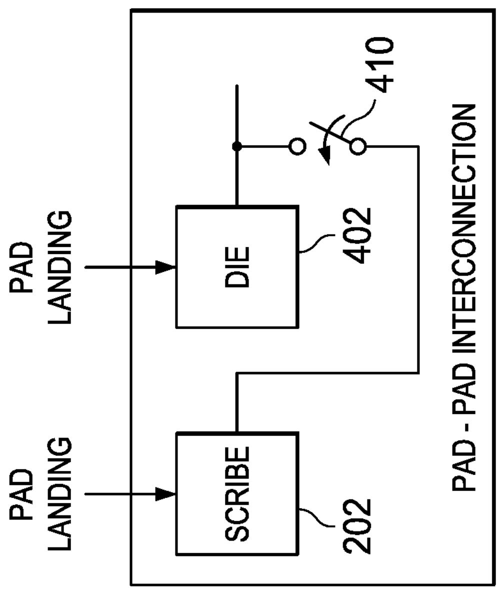

FIG. 4b

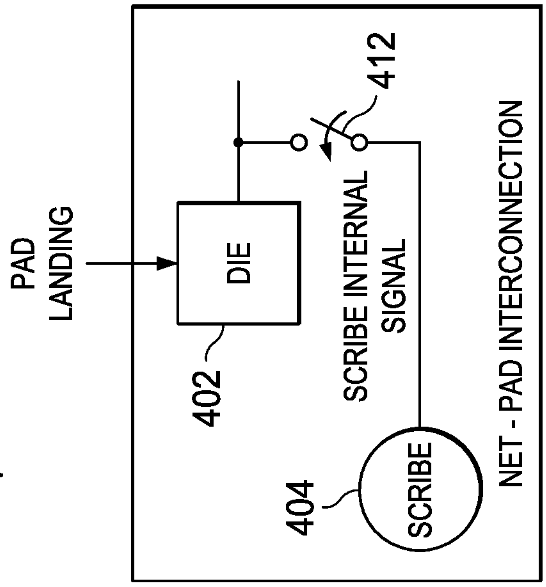

FIG. 4c

FLOW	PHASE 0	PHASE 1	PHASE 2	PHASE 3	PHASE 4	PHASE 5
CONVENTIONAL	SCRIBE CHARACTERIZATION	WAFER PROBE	FINAL TEST	X	X	X
1		SCRIBE	SCRIBE + DIE	DIE	PACKAGE	X
2		SCRIBE	SCRIBE + DIE	DIE + DIE	DIE	PACKAGE

FIG. 5

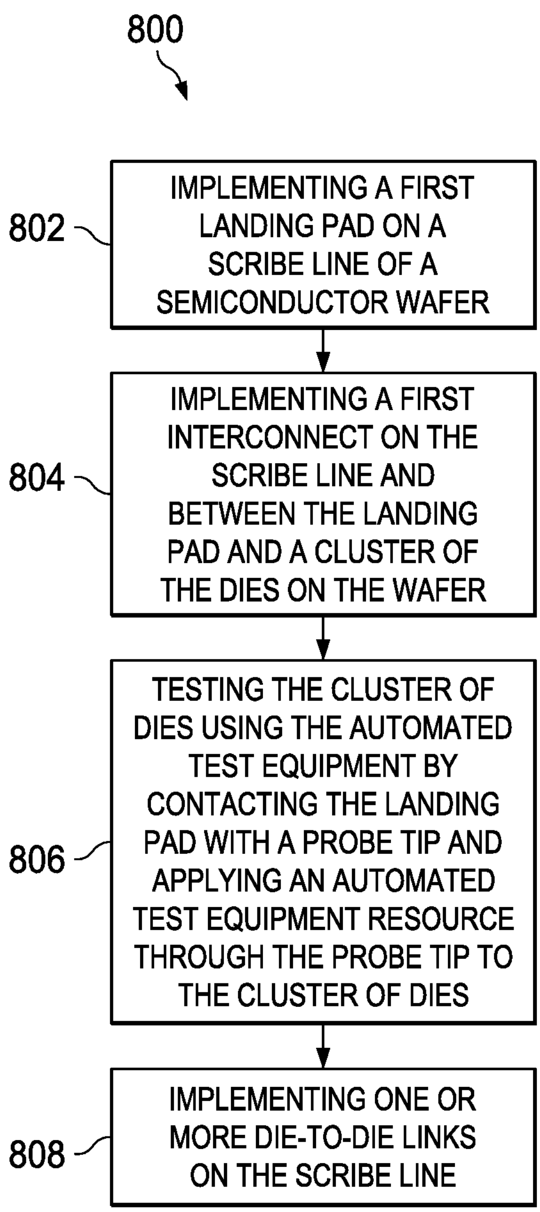

FIG. 8

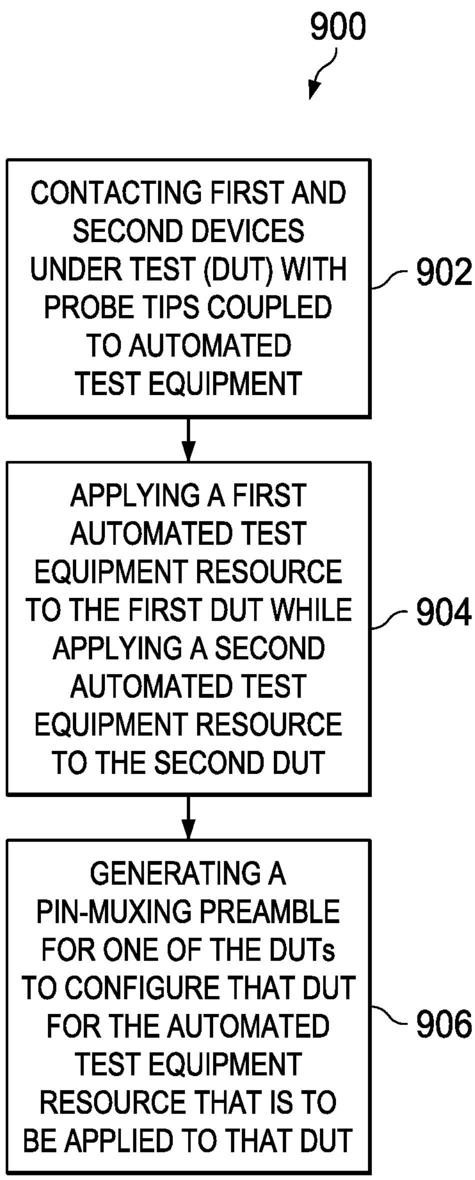

FIG. 9

602

	AUTOMATED TEST EQUIPMENT		
	RESOURCE A 2	RESOURCE A 2	RESOURCE B 2
CYCLE 1		DUT 1	
CYCLE 2			DUT 1
CYCLE 3		DUT 2	
CYCLE 4			DUT 2
CYCLE 5		DUT 3	
CYCLE 6			DUT 3

604

	AUTOMATED TEST EQUIPMENT		
	RESOURCE A 2	RESOURCE A 2	RESOURCE B 2
CYCLE 1	DUT 1	DUT 2	
CYCLE 2			DUT 1
CYCLE 3			DUT 2
CYCLE 4	DUT 3	DUT 4	
CYCLE 5			DUT 3
CYCLE 6			DUT 4

606

	AUTOMATED TEST EQUIPMENT		
	RESOURCE A 2	RESOURCE A 2	RESOURCE B 2
CYCLE 1		DUT 1	DUT 2
CYCLE 2		DUT 3	DUT 1
CYCLE 3		DUT 2	DUT 3
CYCLE 4		DUT 4	DUT 5
CYCLE 5		DUT 6	DUT 4
CYCLE 6		DUT 5	DUT 6

608

	AUTOMATED TEST EQUIPMENT		
	RESOURCE A 2	RESOURCE A 2	RESOURCE B 2
CYCLE 1	DUT 1	DUT 3	DUT 2
CYCLE 2		DUT 2	DUT 1
CYCLE 3			DUT 3
CYCLE 4	DUT 4	DUT 6	DUT 5
CYCLE 5		DUT 5	DUT 4
CYCLE 6			DUT 6

FIG. 6

706a

	12-PROBE TIP HEAD					
	2 RESOURCE A	2 RESOURCE B	2 RESOURCE A	2 RESOURCE B	2 RESOURCE A	2 RESOURCE B
CYCLE 1	DUT 1			DUT 2		
CYCLE 2		DUT 1			DUT 3	
CYCLE 3			DUT 2			DUT 3
CYCLE 4	DUT 4			DUT 5		
CYCLE 5		DUT 4			DUT 6	
CYCLE 6			DUT 5			DUT 6

	6-PROBE TIP HEAD		
	2 RESOURCE A,B	2 RESOURCE A,B	2 RESOURCE A,B
CYCLE 1	DUT 1 (A)	DUT 2 (B)	
CYCLE 2	DUT 1 (B)		DUT 3 (A)
CYCLE 3		DUT 2 (A)	DUT 3 (B)
CYCLE 4	DUT 4 (A)	DUT 5 (B)	
CYCLE 5	DUT 4 (B)		DUT 6 (A)
CYCLE 6		DUT 5 (A)	DUT 6 (B)

706b

	8-PROBE TIP HEAD			
	2 RESOURCE A	2 RESOURCE B	2 RESOURCE A	2 RESOURCE B
CYCLE 1	DUT 1	DUT 2	DUT 3	
CYCLE 2		DUT 1	DUT 2	
CYCLE 3				DUT 3
CYCLE 4	DUT 4	DUT 5	DUT 6	
CYCLE 5		DUT 4	DUT 5	
CYCLE 6				DUT 6

708

FIG. 7

INTERNATIONAL SEARCH REPORT	International application No. PCT/US 2016/064412

A. CLASSIFICATION OF SUBJECT MATTER

G01R31/28 *(2006.01)*
H01L21/301*(2006.01)*

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)

H01L21/00, 21/04, 21/18, 21/30, 21/301, 21/66, 21/71, G01R31/00, 31/26, 31/28, H01L23/00, 23/52, 23/58

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

EAPATIS, ESP@cenet, Google, KIPRIS, PAJ, PatSearch, RUABRU, RUABU1, RUPAT, RUPAT-OLD, RUPTO, USPTO, WIPO

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y A	US 5059899 A (MICRON TECHNOLOGY INC) 22.10.1991, Figs. 1-2, col.2, lin. 19-30, 30-36, 45-55, col.3, lin. 10-16, col.4, lin. 10-25	1-2, 6-8, 10-11, 15-17 3-5, 9, 12-14, 18-27
Y	US 2009239316 A1 (TEXAS INSTRUMENTS INCORPORATED) 24.09.2009, Fig.3, paragraphs [0021], [0032], [0045]	1-2, 6-8, 10-11, 15-17
A	WO 2015103076 A2 (CELERINT LLC) 09.07.2015	1-27
A	US 2003062521 A1 (JANG BYUNG-TAK et al.) 03.04.2003	1-27
A	US 2013 075725 A1 (TAIWAN SEMICONDUCTOR MANUFACTURING COMPANY et al.) 28.03.2013	1-27

☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

*	Special categories of cited documents:	"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
"A"	document defining the general state of the art which is not considered to be of particular relevance	"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
"E"	earlier document 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)	"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
"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	"&"	document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
16 March 2017 (16.03.2017)	30 March 2017 (30.03.2017)
Name and mailing address of the ISA/RU: Federal Institute of Industrial Property, Berezhkovskaya nab., 30-1, Moscow, G-59, GSP-3, Russia, 125993 Facsimile No: (8-495) 531-63-18, (8-499) 243-33-37	Authorized officer M.Adireeva Telephone No. (8-495) 531-65-15