

Method for bonding substrates

The invention relates to a method for bonding a first substrate with a second substrate.

Substrates have been aligned with one another and joined with one another for many years in the semiconductor industry. The joining, the so-called bonding, serves to build up a multi-substrate stack. In such a multi-substrate stack, functional units, in particular memory chips, microprocessors, MEMS etc. are connected to one another and thus combined with one another. Diverse possible applications arise from these combination options.

The density of the functional units is increasing year by year. As a result of advancing technological development, the size of the functional units is becoming smaller and smaller. The increasing density is therefore accompanied by a greater number of functional units per substrate. This increasing number of units is primarily responsible for the reduction in parts costs.

The drawback with increasingly small functional units primarily consists in the increasing difficulty of producing an error-free, in particular also complete superposition of all the functional units along the bond interface of the two substrates.

The greatest problem in present-day alignment technology, therefore, does not always consist in aligning two substrates with one another, in particular two wafers, with the aid of alignment marks, but rather in producing an error-free, in particular complete correlation of points of the first substrate with points of a second substrate, said correlation thus extending over the entire area of the substrate. Experience shows that the structures on the surfaces of the substrates after the bonding process are generally not congruent with one another. A general, in particular overall, alignment and a following bonding step of two substrates are thus not always sufficient to obtain a complete and error-free congruence of the desired points at every point of the substrate surfaces.

There are two fundamental problems in the prior art that stand in the way of a straightforward overall alignment and a subsequent bonding step.

In the first place, the positions of the structures of the first and/or second substrates are generally subject to a deviation from the theoretical positions. There may be a number of reasons for this deviation.

It would be conceivable, for example, that the actual produced structures deviate from their ideal positions because the production processes were defective or at least involved a tolerance. An example of this would be the repeated use of lithography by means of a step-and-repeat process, which during each translational displacement of the stamp involves a small, but significant error in the position.

A further, less trivial reason would be the deformation of the substrate due to mechanical, but in particular thermal loading. A substrate has a defined temperature for example at the time of the production of the structures. This temperature is generally not maintained for the entire process sequence of the substrate, but on the contrary changes. The temperature change is accompanied by a thermal expansion and therefore in the most ideal case a change in diameter, in the most unfavourable case a complex thermal deformation.

In the second place, even two substrates which have error-free, in particular full-area congruence, i.e. overlapping of all structures, shortly before the contacting and the actual bonding process can lose this congruence during the bonding process. The bonding process itself thus has a decisive influence in the production of an error-free substrate stack, i.e. having perfect congruence of the structures.

In the third place, layers and structures that are applied on the substrates may generate stresses in a substrate. The layers may for example be insulation layers, the structures may be through-silicon-vias (TSVs). One of the greatest technical problems with the permanent bonding of two substrates is the alignment accuracy of the functional units between the individual substrates. Although the substrates can be very precisely aligned with respect to one another by means of alignment equipment, distortions of the substrates can arise during the bonding process itself. As a result of the distortions thus arising, the functional units will not necessarily be correctly aligned with one another at all positions. The alignment inaccuracy at a specific point on the substrate may be a result of a distortion, a scaling error, a lens error (magnification or reduction error) etc. In the semiconductor industry, all subject areas dealing with such problems are combined under the term "overlay". A suitable introduction to this subject can be found for example in: *Mack, Chris*.

Fundamental Principles of Optical Lithography - The Science of Microfabrication. WILEY, 2007, Reprint 2012.

Each functional unit is designed in the computer before the actual production process. For example, strip conductors, microchips, MEMS, or any other structure producible with the aid of microsystem technology, are designed in a CAD (computer aided design) program. During the production of the functional units, it can however be seen that there is always a deviation between the ideal functional units designed on the computer and the real functional units produced in the clean room. The differences are primarily due to limitations of the hardware, i.e. engineering-related problems, but very often physical limitations. Thus, the resolution accuracy of a structure that is produced by a photolithographic process is limited by the size of the apertures of the photomask and the wavelength of the light used. Mask distortions are directly transferred to the photoresist. Linear motors of machines can only approach reproducible positions within a given tolerance, etc. It is not therefore surprising that the functional units of a substrate cannot be exactly identical to the structures designed on the computer. Even before the bonding process, all substrates thus have a non-negligible divergence from the ideal state.

If the positions and/or shapes of two opposite-lying functional units of two substrates are compared on the assumption that neither of the two substrates is distorted by a bonding process, it is found that in general there is already an imperfect congruence of the functional units, since the latter diverge from the ideal computer model due to the errors described above. The most frequent errors are represented in figure 8 (*Copied from: http://commons.wikimedia.org/wiki/File:Overlay_typical-model_terms_DE.svg, 24.05.2013 and Mack, Chris. Fundamental Principles of Optical Lithography - The Science of Microfabrication. Chichester: WILEY, p. 312, 2007, Reprint 2012).*

According to the illustrations, a rough distinction can be made between overall and local as well as symmetrical and asymmetrical overlay errors. An overall overlay error is homogeneous, therefore independent of location. It produces the same divergence between two opposite-lying functional units irrespective of the position. The conventional overall overlay errors are errors I. and II., which arise due to a translation or rotation of the two substrates with respect to one another. The translation or rotation of the two substrates produces a corresponding translational or rotational error for all the functional units lying respectively opposite one another on the substrates. A local overlay error arises in a location-

dependent manner, mainly due to elasticity and/or plasticity problems and/or preliminary processes, in the present case primarily caused by the continuously propagating bonding wave. Of the represented overlay errors, errors III. and IV. are in particular referred to as "run-out" errors. This error arises primarily due to a distortion of at least one substrate during a bonding process. As a result of the distortion of at least one substrate, the functional units of the first substrate are also distorted in respect of the functional units of the second substrate. Errors I. and II. can however also arise due to a bonding process, but they are usually superimposed by errors III. and IV. to such a marked extent that it is difficult to detect or measure them. This applies to bonders, in particular fusion bonders of the latest design, which have an extremely accurate capability for x- and/or y- and/or rotation correction.

There is already a plant in the prior art, with the aid of which local distortions can be reduced at least partially. It concerns here a local distortion due to the use of active control elements (WO2012/083978A1).

Initial approaches to a solution for correcting "run-out" errors exist in the prior art. US20120077329A1 describes a method for obtaining a desired alignment accuracy between the functional units of two substrates during and after the bonding, whereby the lower substrate is not fixed. The lower substrate is thus not subjected to any boundary conditions and can bond freely to the upper substrate during the bonding process. An important feature in the prior art is, in particular, the flat fixing of a substrate, usually by means of a vacuum device.

In most cases, the arising "run-out" errors become more intensified in a radially symmetrical manner around the contact point, for which reason they increase from the contact point to the periphery. In most cases, it involves a linearly increasing intensification of the "run-out" errors. Under special conditions, the "run-out" errors can also increase non-linearly.

Under particularly optimum conditions, the "run-out" errors can be ascertained not only by suitable measuring devices (EP2463892), but can also be described by mathematical functions. Since the "run-out" errors represent translations and/or rotations and/or scaling between well-defined points, they are preferably described by vector functions. Generally, this vector function is a function $f: \mathbb{R}^2 \rightarrow \mathbb{R}^2$, i.e. a mapping rule, which maps the two-dimensional definition range of the position coordinates onto the two-dimensional value range of "run-out" vectors. Although an exact mathematical analysis of the corresponding vector

fields has not yet been able to be carried out, assumptions are made concerning the function properties. The vector functions are, with a high degree of probability, at least C^n $n \geq 1$ functions, i.e. at least continuously differentiable. Since the "run-out" errors increase from the contact point to the edge, the divergence of the vector function will probably be different from zero. With a high degree of probability, therefore, the vector field is a source field.

The "run-out" error is best ascertained in relation to structures. A structure is understood to mean any element of a first or second substrate which is to be correlated with a structure on the second or respectively first substrate. In the case of a structure, it thus relates for example to

- alignment marks
- corners and/or edges, in particular corners and edges of functional units
- contact pads, in particular Through Silicon Vias (TSVs) or Through Polymer Vias (TPVs)
- strip conductors
- recesses, in particular holes or depressions

The "run-out" error is generally position-dependent and, in the mathematical sense, a displacement vector, between a real and an ideal point. Since the "run-out" error is generally position-dependent, it is ideally specified by vector fields. In the subsequent text, the "run-out" error will, unless stated otherwise, be regarded solely as punctiform in order to facilitate the description.

"Run-out" error R comprises two partial components.

First partial component R_1 describes the intrinsic part of the "run-out" error, i.e. the part that can be traced back to the defective production of the structures or a distortion of the substrate. It is therefore inherent in the substrate. It should be noted that a substrate can also have an intrinsic "run-out" error when the structures have been correctly produced at a first temperature, but the substrate is subject to a temperature change to a second temperature prior to the bonding process and thermal expansions thus occur, which distort the entire substrate and therefore also the structures located thereon.

Temperature differences of a few Kelvin, sometimes even a tenth of a Kelvin, are already sufficient to generate such distortions.

Second partial component R2 describes the extrinsic part of the "run-out" error, i.e. the part that is not caused until the bonding process. It is not present before the bonding process. This includes primarily local and/or overall distortions of the first and/or second substrate due to forces acting between the substrates, which can lead to a deformation in the nanometre range.

Embodiments of the present invention provide a method for bonding two substrates, with which the bonding accuracy is increased as far as possible at every position of the substrates. Embodiments of the invention provide a method with which an error-free, in particular full-area congruence of the structures of two substrates can be produced.

According to a first aspect of the present invention, there is provided a device for bonding a first substrate with a second substrate, including: a measuring means configured to: detect a percentage amount of a bonded surface of the first and second substrates with respect to a non-bonded surface of the first and second substrates during contact of the first and second substrates; and monitor an advance of a bonding wave between the first and second substrates via the detected percentage amount to enable control of the bonding wave.

According to a second aspect of the present invention, there is provided a device for bonding a first substrate with a second substrate at respective contact faces of the substrates, the device including: a first sample holder configured to hold the first substrate to a surface of the first sample holder with a holding force FH1 applied by a first vacuum of the first sample holder; a second sample holder configured to hold the second substrate to a surface of the second sample holder with a holding force FH2 applied by a second vacuum of the second sample holder; and a measuring means configured to: detect a percentage amount of a bonded surface of the first and second substrates with respect to a non-bonded surface of the first and second substrates during contact of the first and second substrates; and monitor an advance of a bonding wave between the first and second substrates via the detected percentage amount to enable control of the bonding wave.

According to a third aspect of the present invention, there is provided a method for bonding a first substrate with a second substrate, including: initiating contact between the first substrate and the second substrate; detecting a percentage amount of a bonded surface of the contacted first and second substrates with respect to a non-bonded surface of the contacted first and second substrates; monitoring an advance of a bonding

wave between the contacted first and second substrates via the detected percentage amount to enable control of the bonding wave; and reducing first and second vacuums respectively included in first and second sample holders to respectively reduce a hold of the contacted first and second substrates by the first and second sample holder and control the bonding wave therebetween.

All combinations of at least two features given in the description, in the claims and/or the drawings also fall within the scope of the invention. In the stated value ranges, values lying inside the stated limits are also deemed to be disclosed as limiting values and can be claimed in any combination. If an individual or a plurality of method steps can be performed on different devices or modules, the latter are in each case disclosed separately as an independent method.

Embodiments of the present invention reduce a heating temperature T_H already during the bonding and to switch off heating during the bonding. Heating temperature T_H serves in particular to produce a temperature sufficient for the bonding at a bonding face of the substrates. An embodiment of the invention consists in removing the fixing of a substrate, in particular during the bonding, in order to permit free deformability of the substrate stack to the bonded. A further embodiment of the invention consists in the possibility of the substrate stack, in particular its interface, being ventilated or subjected to pressure during the bonding.

As a first and/or second substrate wafers are particularly qualified.

A characteristic process according to embodiments of the invention during the bonding, in particular permanent bonding, preferably fusion bonding, is the most concentric, punctiform possible contacting of the two substrates. In particular, the contacting of the two substrates can also take place in a non-concentric manner. The bonding wave being propagated from a non-concentric contact point would reach different sides of the substrate edge at different times. The complete mathematical-physical description of the