

Wafer-level 3D integration technology

An overview of wafer-level three-dimensional (3D) integration technology is provided. The basic reasoning for pursuing 3D integration is presented, followed by a description of the possible process variations and integration schemes, as well as the process technology elements needed to implement 3D integrated circuits. Detailed descriptions of two wafer-level integration schemes implemented at IBM are given, and the challenges of bringing 3D integration into a production environment are discussed.

S. J. Koester
A. M. Young
R. R. Yu
S. Purushothaman
K.-N. Chen
D. C. La Tulipe, Jr.
N. Rana
L. Shi
M. R. Wordeman
E. J. Spriggs

Introduction

The last few decades have seen an astonishing increase in the functionality of computational systems. This capability has been driven by the scaling of semiconductor devices, which has enabled the number of transistors on a single chip to grow at a geometric rate, following Moore's Law [1]. In particular, the scaling of silicon metal-oxide semiconductor field-effect transistors (MOSFETs) [2] drives the effort to continue this trend into the future.

However, several serious roadblocks exist. The first is the difficulty and expense of continued lithographic scaling, which could make it economically impractical to scale devices beyond a certain pitch. The second is that even if lithographic scaling can continue, the power dissipated by the transistors will bring clock frequency scaling to a halt. In fact, it could be argued that clock frequency scaling has already stopped, and microprocessor designs have increasingly relied on new architectures to improve performance. These factors suggest that in the near future, it will no longer be possible to improve system performance through scaling alone, and that additional methods to achieve the desired enhancement will be needed. Three-dimensional (3D) integration technology offers the promise of being a new way of increasing system performance, even in the absence of scaling. This promise is due to a number of characteristic features of 3D integration, including decreased total wiring length (and thus reduced interconnect delay times), a dramatically increased number of interconnects between chips, and the ability to allow dissimilar materials, process technologies, and functions to be integrated.

Motivation for 3D integration

Overall, *3D technology* can be broadly defined as any technology that stacks semiconductor elements on top of each other and uses vertical, as opposed to peripheral, interconnects between the wafers. Under this definition, 3D technology could include simple chip stacks, silicon chip carriers and interposers, chip-to-wafer stacks, and full wafer-level integration. Each of these technologies has benefits for specific applications, and the technology appropriate for a particular application is driven in large part by the required interconnect density. For instance, for wireless communication, only a few through-silicon vias (TSVs) are needed per chip in order to make low-inductance contacts to the backside ground plane. On the other hand, high-performance processors [3] will require extremely high densities ($>10^5$ pins/cm²) of vertical interconnects. Applications such as 3D chips for supply voltage stabilization and regulation reside somewhere in the middle, and a myriad of applications exist that require the full range of interconnect densities possible.

Since this paper deals mainly with technologies for realizing extremely dense interconnects, it is reasonable to explore in more detail the advantages that a large number of vertical interconnects can have for 3D computing systems. Perhaps the most compelling advantage of this capability is that it allows a massive communication bandwidth to be achieved between the processor cores and memory. In typical 2D chips, the fast cache memory is composed of static RAM (SRAM) or embedded DRAM (eDRAM) [4] that is integrated directly onto the chip. Such schemes work well for single-threaded or few-threaded architectures. However, there has been a strong industry trend to move to more highly multithreaded,

©Copyright 2008 by International Business Machines Corporation. Copying in printed form for private use is permitted without payment of royalty provided that (1) each reproduction is done without alteration and (2) the *Journal* reference and IBM copyright notice are included on the first page. The title and abstract, but no other portions, of this paper may be copied by any means or distributed royalty free without further permission by computer-based and other information-service systems. Permission to *republish* any other portion of this paper must be obtained from the Editor.

0018-8646/08/\$5.00 © 2008 IBM

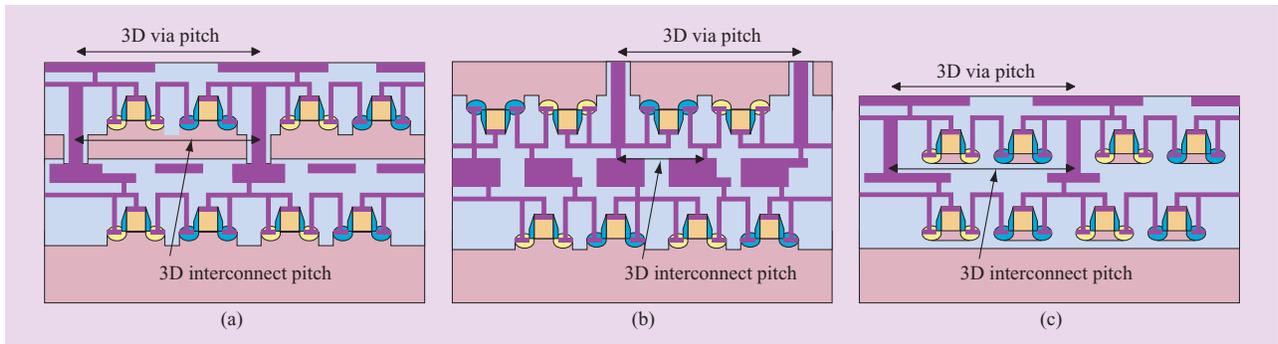


Figure 1

Cross-sections of the three main 3D integration schemes: (a) face to back; (b) face to face; (c) SOI based.

massively parallel computational systems, and these systems require considerably more memory bandwidth than can be accommodated in a 2D architecture [5].

Basic 3D integration approaches

A wide range of 3D integration approaches is possible and these have been reviewed extensively elsewhere [6–9]. These schemes have various advantages and tradeoffs, so a variety of optimized process flows may ultimately be used to meet the needs of the various targeted applications. As a review, the main differences between 3D integration approaches are the assembly orientation (face to back or face to face), the TSV process sequence (vias first or vias last), and whether the substrate is completely removed using a natural etch stop [bulk or silicon on insulator (SOI)].

The face-to-back method is based on bonding the front side of the bottom wafer to the thinned backside of the top wafer. Similar approaches were originally developed at IBM for multichip modules (MCMs) used in IBM S/390* G5 systems [10], and later this same approach was demonstrated at the wafer level for both CMOS (complementary metal-oxide semiconductor) and microelectromechanical systems (MEMS) applications. **Figure 1(a)** depicts a two-layer stack assembled in the face-to-back configuration. The height of the structure and, therefore, the height of the interconnecting via depend on the thickness of the thinned top wafer. If the aspect ratio of the via is limited, the substrate thickness can then limit the number of interconnects between the two wafers. In this scheme, the total number of interconnects between wafers cannot be larger than the number of TSVs. In order to construct such a stack, a handle wafer must be used. This can induce distortions in the top wafer, making it difficult to achieve tight alignment tolerances [11]. In previous work, we have

found that distortions as large as 50 ppm can be induced by glass handle wafers, though strategies such as temperature-compensated bonding can reduce these distortions. For face-to-back schemes, typically, the minimum thickness for the top wafer is on the order of 25–50 μm using conventional grind and polish thinning. For Cu vias, aspect ratios on the order of 5:1 are typically needed for reliable fill, thus limiting the interconnect pitch to values on the order of 10–20 μm . For tungsten vias, much higher aspect ratios are possible, and for these vias, the interconnect pitch is likely to be limited by electrical requirements of the interconnects and bonding alignment tolerances. Of course, advances in wafer thinning, via fill, and wafer alignment technologies could reduce the achievable via pitch in the future.

Figure 1(b) shows the face-to-face approach of joining the front sides of two wafers. IBM originally used this method [12] to create MCMs with an interconnect pitch of less than 20 μm with reduced process complexity compared to the face-to-back scheme. A key potential advantage of face-to-face assembly is the ability to decouple the number of TSVs from the total number of interconnections between the layers. Therefore, it could be possible to achieve much higher interconnect densities than allowed by face-to-back assembly. In this case, the interconnect pitch is limited only by the alignment tolerance of the bonding process (plus the normal overlay error induced by the standard CMOS lithography steps). For typical tolerances of about 1 μm for state-of-the-art aligned bonding systems, it is therefore conceivable that interconnect pitches of $\leq 5 \mu\text{m}$ could be achieved with face-to-face assembly. However, the improved interlevel connectivity can be exploited only for two-layer stacks; for multilayer stacks, the TSVs will still limit the total combined 3D interconnect density.

For the schemes depicted in Figures 1(a) and 1(b), both SOI and bulk wafers can be used, where the process is essentially blind to the presence of the buried oxide. However, it is also possible to use the buried oxide as a way of enhancing the 3D integration process, and this scheme is shown in **Figure 1(c)**. This SOI scheme has been described extensively in previous publications [13–15] and is only briefly summarized here. In the SOI scheme, the SOI layer is used as an etch stop to allow the substrate to be completely removed before the two wafers are combined. As a consequence, the minimum height of the interconnecting via could be as short as 1–2 μm , allowing Cu via dimensions as small as 0.2–0.4 μm [14]. Via chains with an interconnect pitch of 6.7 μm have been demonstrated using this technique [14], and if extremely tight ($<0.5 \mu\text{m}$) wafer-level alignment could also be achieved, then via pitches less than 2 μm could be conceivable, allowing for numerous new system-level opportunities not achievable with looser interconnect pitches. However, there are some potential risks to removing the substrate down to the buried-oxide layer. The bulk silicon etch could lead to a chemical attack of the device layer, now less than 200 nm away. It is also unclear whether the removal of the entire substrate could also affect the strain in the devices, a particularly important factor in advanced CMOS technologies [16]. Guarini et al. [15] demonstrated functionality of CMOS devices after full substrate removal and bonding to a second wafer with little degradation of the device characteristics. However, these devices did not have stress liners, and a detailed analysis of stress effects in thin transferred layers is still needed.

Process components for 3D integration

As described above, there are a wide range of possible 3D integration schemes. However, all 3D integrated circuits (ICs) must have three main process components: a vertical interconnect, aligned wafer bonding, and wafer thinning and backside processing. In this section, these basic process components are described and, in particular, the process aspects that differentiate wafer-scale 3D integration from silicon-carrier and chip stacks.

Vertical interconnect

Perhaps the most important technology element for 3D integration is the vertical interconnect, sometimes referred to as the TSV or the through-silicon interconnect, although in some cases, such as in our SOI 3D scheme to be described later, the via does not need to go through silicon because the substrate is entirely removed. A vertical interconnect is necessary for 3D integration to truly take advantage of 3D for system-level performance, because without it, interconnects would be limited to the periphery of the chip, and in this case, the

interconnect density would be no greater than in conventional planar technology. This interconnection method is essentially the same as a contact hole process, with the difference that a much deeper hole has to be created vertically through the silicon material using a special etching process. In addition to IBM, other companies and institutes, including Elpida Memory, IMEC, Intel Corporation, Samsung, and Tezzaron Semiconductor, are developing TSV methodologies optimized for their applications [17, 18].

A variety of vertical TSV technologies have been developed by IBM for silicon-carrier applications [19–21]. A number of vias-last approaches have been demonstrated, and this process has the advantage that the constituent CMOS technologies in the 3D stack do not need to be modified, because the TSV is formed only after assembly by using backside deep reactive-ion etching. Insulation is subsequently applied to the interior of the via and selectively removed from the base to allow electrical contact. Metallization of large vias prepared in this manner has been demonstrated by using an initial partial fill with plated Cu followed by evaporation of Cr/Cu ball-limiting metallurgy and Pb/Sn solder. However, with the exception of the SOI scheme described later in which the silicon substrate is entirely removed, vias-last schemes in general may not be well suited for high-density 3D integration because of the aspect-ratio requirements of the dielectric and metal fill.

Another approach to TSVs that has been demonstrated at IBM and elsewhere is a vias-first process flow. This approach has the advantage of allowing higher-temperature dielectric and metal fill processes and also allows high-aspect-ratio vias to be formed. One particular vias-first approach that has been described is an annular-via geometry for large-area contacts [21]. This structure uses an etched ring-shaped via geometry such that the ring width is narrow enough to allow complete fill of the structure using a variety of materials, including doped polysilicon, electroplated Cu, or chemical vapor deposition tungsten. However, for higher-density 3D integration applications, annuli with small central cores (the region defined by the inner diameter of the annulus) can be used. The annular region is filled with isolation dielectrics, and the central core is subsequently etched and metallized.

The vias required for 3D integration schemes that entirely remove the substrate can be formed by a vastly simplified process, since isolation does not need to be deposited. Therefore, a vias-last process or a vias-first process can be completed easily. In addition, because the distance between the layers is much smaller than in TSV schemes, the via pitch and size can be dramatically reduced. Thinning the bonded wafer to a membrane of a

few microns could also have a dramatic impact on power distribution because of the reduced via resistance.

Aligned wafer bonding

Aligned wafer bonding is the second key technology needed for 3D integration. Aligned wafer bonding for 3D integration is fundamentally different than blanket wafer-bonding processes that are used, for instance, in SOI substrate manufacturing. There are several differences. First, alignment is required because patterns on each wafer need to be in registration in order to allow the interconnect densities required to take advantage of true 3D system capabilities. Second, wafers for 3D integration typically have significant topography on them, and these surface irregularities can make high-quality bonding significantly more difficult than for blanket wafers, particularly for oxide-fusion bonding. Finally, due to the fact that CMOS circuits, usually with back-end-of-line (BEOL) metallization, already exist on the wafers, the thermal budget restriction on the bonding process can be quite severe, and typically the bonding process needs to be performed at temperatures less than 400°C. In the following subsection, we briefly describe three particular technologies for aligned 3D wafer bonding that have been investigated at IBM: Cu–Cu compression bonding, transfer-join (T–J) bonding (hybrid Cu and adhesive bonding), and oxide-fusion bonding. We also discuss issues associated with the alignment of full wafers and how this process differs from typical lithographic alignments or chip-placement approaches performed on the die level.

Cu bonding

Attachment of two wafers is possible using a thermocompression bond created by applying pressure to two wafers with Cu-metallized surfaces at elevated temperatures. For 3D integration, the Cu–Cu join can serve the additional function of providing electrical connection between the two layers. Optimization of the quality of this bonding process is a key issue being addressed and includes provision of various surface-preparation techniques, postbonding straightening, thermal annealing cycles, as well as use of optimized pattern geometry [22–24]. All of these factors can affect the crystal quality of the Cu, including such properties as the texture and grain size, which can have a direct impact on the bond strength and electrical properties.

The mechanism of thermocompression Cu bonding occurs when, under elevated temperatures and pressures, the microscopic contacts between two Cu regions start to deform, further increase the contact area, and finally diffuse into each other to complete the bonding process. The key parameters of Cu bonding are the temperature and pressure, the bonding duration, and Cu surface

cleanliness. Optimization of all of these parameters is needed to achieve a high-quality bond. The degree of surface cleanliness is related not only to the prebonding surface cleaning, but also to the vacuum condition during bonding [22]. In addition, despite that the bonding temperature is the most significant parameter in determining the bonding quality, the temperature has to be compatible with BEOL process temperatures.

The quality of patterned Cu bonding on the wafer-level scale has been investigated for real device applications [23, 24]. To achieve the best bond quality, the structural design of Cu bond pads determines not only the circuit placement, but also bond quality since it is related to the available area that can be bonded in a local region or across the entire wafer. Cu bond pad size (interconnect size), pad pattern density (total bond area), and seal design have also been studied. According to existing results from several pattern densities ranging from <1% to 35%, higher bond density results in better bond quality. The bonded area rarely fails in dicing tests when pattern density is larger than 13% [24]. Note that the bonding area density requirement does not have a detrimental impact on the real estate of the design, because the bonds associated with the TSVs could be included in this 13% requirement. In other words, if the pad density associated with the TSVs is very high (>13%), then no additional bonding pads are needed. However, if the TSV density is low, then structural bonding pads must be added to ensure adequate bonding yield. In addition, a seal design that would consist of extra Cu bonds surrounding the electrical interconnects, chip edge, and wafer edge could prevent corrosion and provide extra mechanical support [23], though additional studies are needed. Additional discussion of the reliability impact of the various join methodologies is presented in the section “Challenges and limitations for 3D integration” below.

T–J assembly

A variation of the Cu–Cu compression bonding process can be accomplished by using a lock-and-key structure along with an intermediate adhesive layer to improve bond strength. This technology was originally developed for thin-film MCMs and underwent extensive reliability testing during build and qualification [25, 26]. However, this scheme is equally suitable for wafer-level 3D integration and could have significant advantages over direct Cu–Cu-based schemes.

In the T–J assembly scheme, the mating surfaces of the two device wafers that are to be joined are provided with a matching set of protrusions (keys) on one side and receptacles (locks) on the other, as shown in **Figure 2(a)**. The protrusion, also referred to as the *stud*, can be the extension of a TSV or a specially fabricated BEOL Cu

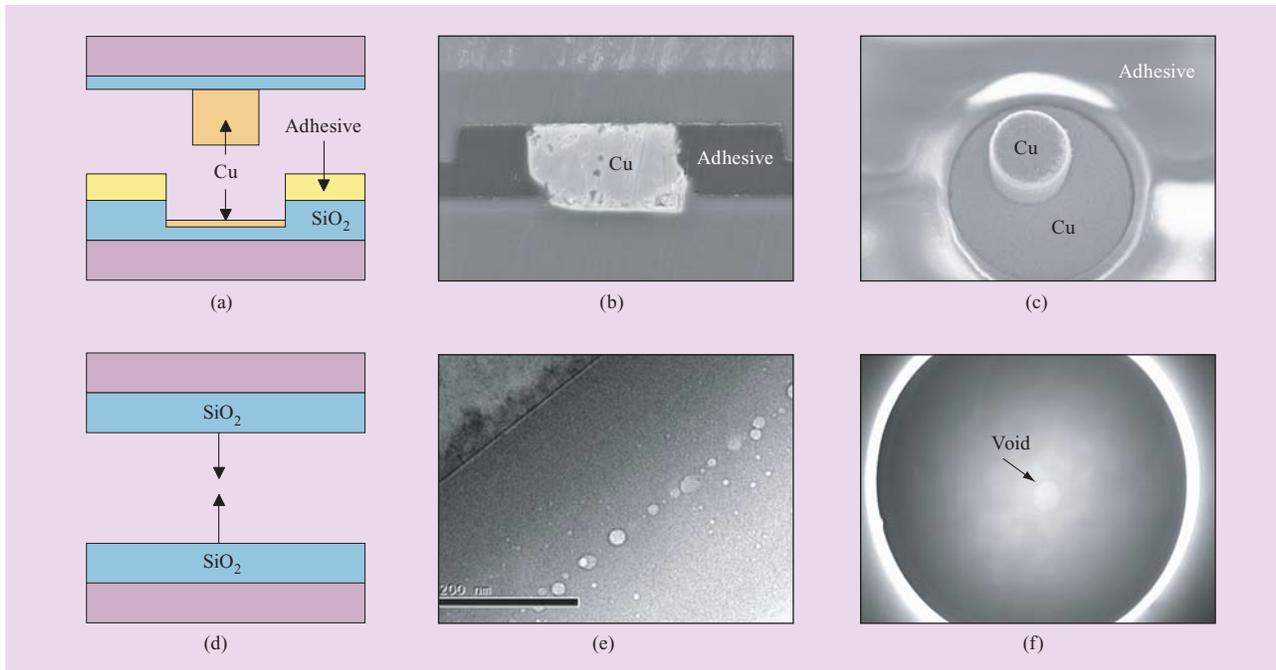


Figure 2

Bonding schemes: (a) cross-section of the T–J bonding scheme; (b) polished cross-section of a completed T–J bond; (c) a top-down scanning electron micrograph view of a T–J bond after delayering, showing lock-and-key structure and surrounding adhesive layer; (d) oxide-fusion bonding scheme; (e) cross-sectional transmission electron micrograph of an oxide-fusion bond; (f) whole-wafer infrared image of two wafers joined using oxide-fusion bonding.

stud. The receptacle is provided at the bottom with a Cu pad that will be bonded to the Cu stud later. At least one of the mating surfaces, for example, the lower one in Figure 2(a), is provided with an adhesive atop the last passivation dielectric layer. Both substrates can be silicon substrates, or optionally, one of them could be a thinned wafer attached to a handle substrate. The studs and pads are optionally connected to circuits within each wafer by means of 2D wiring and TSVs as appropriate.

The substrates can be aligned in the same way for the direct Cu–Cu technique and then bonded together by the application of a uniform and moderate pressure at a temperature in the 350°C to 400°C range. The layer thickness of the stud and the combined thickness of the adhesive and the insulator are adjusted such that the Cu-stud-to-Cu-pad contact is first established during the bonding so that a metal–metal bond is formed. Under the continued bonding pressure, the stud height is compressed, and the adhesive is brought into contact with the opposing insulator surface and bonded. The adhesive material is chosen to have the appropriate rheology required to enable flow, fill, and bonding of the two wafers, filling any gaps between features. Additionally, the adhesive is tailored to be thermally stable at the

bonding temperature and during any subsequent process excursions required, such as additional layer attachment or final BEOL wiring on the 3D stack. Depending on the wafers bonded together, either handle-wafer removal or backside wafer thinning is performed next. The process can be repeated as needed if additional wafer layer attachment is desired. A cross-sectional scanning electron microscope (SEM) view of a Cu–Cu T–J bond with a polymer adhesive interlayer is shown in Figure 2(b). Figure 2(c) shows a top-down SEM view of the alignment of a stud to a pad in a bonded structure after the upper substrate has been delayered for the purpose of constructional analysis.

The lock-and-key T–J approach can be combined with any of the 3D integration schemes described earlier. The fact that the adhesive increases the mechanical integrity of the joined features in the 3D stack means that it avoids the fill factor requirements of direct Cu–Cu bonding. However, it is unclear how much of a real advantage this represents, since due to the low thermal conductivity of typical polymer adhesives, fill shapes may still be desirable to improve the thermal conductivity between layers. A less obvious but significant feature in T–J assembly is that the structure can maintain wafer-to-

wafer registration without actual interface contact between the two bonding wafers. One of the key issues in wafer bonding is the removal of trapped air or water between the wafers. Heat and vacuum are needed to remove such entrapment, but both could lead to wafer drift if used improperly; the T–J technique allows the use of heat and vacuum without the risk of wafer shift.

Oxide-fusion bonding

Oxide-fusion bonding is another technique that can be used to attach two fully processed wafers together. At IBM, we have published extensively on the use of this basic process capability to join SOI wafers in a face-to-back orientation [14], and different schemes for using oxide-fusion bonding in 3D integration have been implemented by others [27]. General requirements include low-temperature bonding-oxide deposition and anneal for compatibility with ICs, extreme planarization of the two surfaces to be joined, and surface activation of these surfaces to provide the proper chemistry to allow robust bonding to take place. An integrated oxide-fusion bonding flow being explored by IBM is described in detail in the section “Oxide-bonding 3D process vehicle” below.

The oxide-bonding process is shown in **Figure 2(d)**, along with a cross-sectional transmission electron micrograph (TEM) of the bonding interface in **Figure 2(e)** and a whole-wafer infrared (IR) image of a typical bonded pair in **Figure 2(f)**. The TEM in **Figure 2(e)** shows a distributed nanoscale void pattern, while the plan-view IR image in **Figure 2(f)** shows that after post-bonding anneals of 150°C and 280°C, excellent bond quality is maintained, although occasional macroscopic voids are observed, as shown in the figure. In order to improve the interface quality, we have investigated UV (ultraviolet)-assisted annealing of the oxide films before bonding [28]. We have found that increasing the temperature of the UV treatment (from 250°C to 350°C) not only concentrates the distribution of the microvoids toward the bonding interface, but also leads to a reduction in the bonding strength. UV treatment also tends to reduce the number of large macroscopic voids observed in bonded layers after annealing but does not consistently eliminate them entirely.

The use of multiple levels of back-end wiring typically leads to significant surface topography. This creates challenges for oxide-fusion bonding, which requires extremely planar surfaces. While it is possible to reduce nonplanarity by aggressively controlling metal-pattern densities in mask design, we have found that process-based planarization methods are also required. As described in Reference [28], typical wafers with back-end metallization have significant pattern-induced topography. We have found that simply depositing a thick SiO₂ layer and planarizing using chemical-

mechanical polishing (CMP) can reduce the roughness by a factor of approximately 4; however, this reduction is still insufficient to prevent pattern-induced voiding at the bonding interface. Therefore, we have developed an optimized CMP protocol that dramatically reduces pattern-dependent variations by a factor of nearly 10 compared to the incoming wafer [28]. The development of this type of advanced planarization technology will be critical to the commercialization of oxide-bonding schemes, because pattern-dependent topographic variations will be encountered on a routine basis.

Wafer-level alignment

Wafer-level alignment is fundamentally different from standard lithographic alignments used in CMOS fabrication, because the alignment must be performed over the entire wafer as opposed to being performed on a die-by-die basis. This requirement makes overlay control much more difficult than in die-level schemes. Distortion due to wafer bow and thermal expansion as well as pattern irregularity due to lithographic skew or run-out can all lead to overlay tolerance errors. In addition, the transparency or opacity of the substrate can also affect the wafer alignment. Several manufacturers are developing alignment tools that have an alignment accuracy of less than 1 μm.

Because of the temperature excursions and potential distortions associated with the bonding process itself, it is standard procedure in the industry to use aligner tools (which have high throughput) and then move wafers to specialized bonding tools with double-sided heating chucks for good control of temperature and pressure of the wafer and the gradient across the stack. The key to good process control is the ability to separate the alignment and prebonding steps from the actual bonding process. Such a methodology allows for better understanding of the final alignment error contributions. That said, the actual bonding process and the technique used can affect overall alignment overlay, and it is critical to understand this issue to choose the proper bonding process.

The alignment issue arises for Cu–Cu bonding when the surrounding dielectric materials from both wafers are recessed. In this scenario, if there is a large misalignment prior to bonding, the wafers can still be clamped for bonding. However, this structure cannot inhibit the thermal misalignment created during thermocompression and is not resistant to additional alignment slip due to shear forces induced during the thermocompression process. One way to prevent those is to use a lock-and-key structure to limit the misalignment within the desired region. Of course, the built-in lock-and-key structure of the T–J assembly procedure can aid the wafer-to-wafer registration by keeping the aligned wafers locked together

during the steps following initial align and placement. This could be a significant factor in maintaining the ability to extend the 3D process to tighter interconnect pitches, as this pitch is likely to be ultimately limited by alignment and bonding tolerances.

Wafer-thinning processes that use handle wafers and lamination can also add distortion to the thinned silicon layer. This distortion can be caused by both a mismatch in coefficients of thermal expansion between materials and by the use of a polymer lamination material with low elastic modulus. For example, if left uncontrolled, the use of glass handle wafers can introduce alignment errors in the range of 5 μm at the edge of a 200-mm wafer, a value that is significantly larger than the errors in more direct silicon-to-silicon alignments. So, in any process using handle wafers, control and correction of these errors is an important consideration. In practice, these distortions can often be well modeled as global magnification errors. This allows the potential to correct most of the wafer-level distortion using methods based on control of temperature, handle-wafer materials, and lamination polymers.

Alignment considerations are somewhat unique in SOI-based oxide-fusion bonding. This process is often practiced when the SOI wafer is laminated to a glass handle wafer and the underlying silicon substrate is removed, leaving a thin SOI layer attached to the glass prior to alignment. Unlike in other cases in which either separate optical paths are used to image the surfaces to be aligned or IR illumination is required to image through the wafer stack, one can see through this SOI on glass at visible wavelengths. This allows very accurate direct optical alignment to an underlying Si wafer in a manner similar to wafer-scale contact aligners. Wafer contact and a preliminary oxide-fusion bond must be initiated in the alignment tool itself, but once this is achieved, there is minimal alignment distortion introduced by downstream processing [29, 30].

Wafer thinning and backside processing

Wafer thinning is a necessary component of 3D integration, as it allows the interlayer distance to be reduced, therefore allowing a high density of vertical interconnects. The greatest challenge in wafer thinning is that the wafer must be thinned to about 5–10% of its original thickness with a required uniformity of $<1\text{--}2\ \mu\text{m}$. In bulk Si, this thinning is especially challenging because there is no natural etch stop. The final thickness depends on the thinning process control capabilities and is limited by the thickness uniformity specifications of the Si removal process (that being mechanical grinding and polishing plus wet or dry etching). Successful thinning to a uniform Si thickness of a few microns has been

demonstrated, but typically thicknesses of 20–40 μm are necessary for a robust process.

Standard process steps for Si thinning are as follows. First, a coarse grinding step is performed in order to thin the wafer from its original thickness ($\sim 700\text{--}800\ \mu\text{m}$) to a thickness of 125–150 μm . This type of process is usually performed using a 400-mesh grinding surface. This step is followed by a fine grinding step to thin to about 100 μm using an 1,800–2,000 mesh surface. Next, a mechanical polishing step can be performed down to the desired thickness of 30–60 μm . For most processes, it is desirable for these grinding steps to be performed only on uniform regions of the silicon, because stresses associated with mechanical grind and polish steps can damage fine features in the silicon. If it is necessary to expose the TSV from the backside, it is typically desirable to complete the thinning process using a plasma-based etching process (such as reactive-ion etching) to expose the TSVs. Limitations on the uniformity of the backside thinning set the practical limit on the TSV depth and, therefore, the via pitch that can be achieved using this type of blind thinning process.

Unlike the bulk-like process, in our SOI-based 3D integration scheme, the buried oxide can act as an etch stop for the final wafer-thinning process, and a purely wet chemical etching process can be used for this process. For instance, tetramethylammonium hydroxide (TMAH) can remove 0.5 $\mu\text{m}/\text{min}$ of silicon with excellent selectivity to SiO_2 . In our process, we typically remove about 600 μm of the silicon wafer by the mechanical techniques described above and then employ a 25% TMAH at 80°C to etch (at a rate of 40 $\mu\text{m}/\text{hr}$) the last 100 μm of silicon down to the buried oxide layer. The buried oxide has a better than 300:1 etch selectivity relative to silicon and, therefore, acts as a very efficient etch stop layer. The overwhelming advantage of such an approach is that all of the Si can be uniformly removed, leaving a very smooth ($<10\ \text{nm}$) surface for the application of bonding-oxide films, and the layer of transferred circuits is automatically a minimal thickness across the wafer, facilitating the fabrication of high-density interlevel wiring later in the process flow.

For face-to-back assembly, the wafer to be thinned must first be transferred to a handle wafer. The temporary bond to the handle wafer must be sufficiently strong to survive the thinning process while still allowing release of the handle wafer after bonding. If glass is used as the handle wafer, it has the advantage of being immune to wet etchants that might be needed for thinning the silicon. However, distortions (bow and expansion errors) can occur, particularly after the wafer is thinned, and these can affect the alignment and subsequent processing after bonding. The distortion of the thin membrane of Si on a temporary handle wafer arises as one of the most

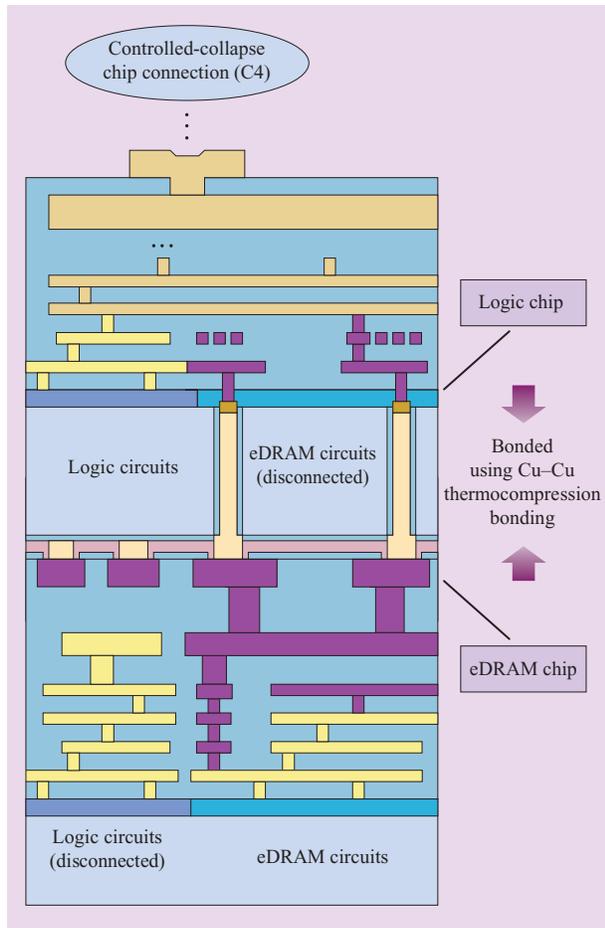


Figure 3

Cu-bonding 3D process test vehicle, with a cross-section of the assembled stack shown at the left.

significant challenges of the face-to-back approach. Wafer thinning using the face-to-face assembly scheme is in some ways simpler in that the distortion error associated with thinning on a handle wafer is eliminated. However, additional difficulties arise due to the fact that the base wafer must remain immune to the grinding and etching solutions used to thin the top wafer.

IBM 3D integration test vehicles

Ultimately, the integration scheme and process steps that are optimal for a particular application will be determined by a number of factors, including the number of interlayer interconnects required, the external I/O capacity needed, the number of layers to be stacked, the power and cooling requirements of the stack, the expected performance, cost, and the choice of technology (e.g., bulk or SOI). Specifically, the choice of a bonding medium, the use of a carrier wafer, the order of

processing steps, the incorporation of additional thermal cycles (such as curing, bond strengthening, and out-gassing), the choice of metallization techniques, the implementation of seals, or any other yield or reliability enhancers (such as via redundancy) are all dictated by application needs.

In the next two sections, we present two 3D test vehicles that have been built at IBM to develop the process and integration technology and we discuss the rationale behind the process options used to build these structures. These constructions were specifically chosen because they represent two distinct cases of different 3D options for high-performance CMOS technology. The first involves technology needed for microprocessor and memory stacking, while the second is a general scheme for ultrahigh-density 3D integration based on SOI technology. Each of these options was investigated to better understand the integration issues associated with each basic scheme as well as to assess 3D integration in general as a potential replacement of traditional planar circuit layout to enable future high-performance system applications.

Cu-bonding 3D process vehicle

In order to develop a general-purpose 3D processing scheme that would be compatible with either bulk or SOI CMOS, a new vehicle was created to develop the 3D process. This process-development vehicle was based on an ASIC (application-specific integrated circuit) microprocessor chip with approximately 95 million transistors and built in our conventional 0.13- μm bulk CMOS process, which includes deep-trench eDRAM. We have partitioned this design so that the eDRAM is on one level and the remaining CMOS logic is on a different level (**Figure 3**). This partitioning does not lead to an increase in performance, but it does represent a straightforward way to test the integration of 3D processing elements with complex circuits. As part of this process, a conventional 2D ASIC chip was fabricated alongside the 3D version of the chip (using a two-die reticle) to test the impact of the 3D process on a known-working chip. The 2D and 3D versions of the chip were intended to be functionally equivalent. In this paper, we report on the process-integration scheme used for this build and the structural results.

The process-integration flow developed for this work is intended to be compatible with standard foundry or ASIC processing, with minimal deviations from the base process. Processing in 3D begins with the fabrication of dielectrics in the top wafer that will isolate the TSVs from the surrounding silicon substrate. These via-isolation dielectrics are embedded into the wafers before front-end-of-line CMOS processing is initiated. The via isolation is formed by etching an annular shape into the top side of

the wafer, which is subsequently filled by thermal oxidation and polysilicon deposition. The wafer is then planarized, and the oxide-polysilicon-oxide via-isolation films appear as shown in the top-down optical micrograph of **Figure 4(a)**. The single-crystal-silicon region that remains in the central core of the annulus will ultimately be replaced by via metallization; however, at this point there is no metal in the structure, and such wafers can be run as normal substrates through conventional CMOS processing. We have shown that adding these via structures into the wafer does not significantly affect the electrical characteristics of downstream CMOS, as can be seen in the performance-screen ring oscillator (PSRO) data shown in **Figure 4(b)**.

Once CMOS processing is completed, additional 3D-specific processes are performed. First, the top wafer is laminated to a glass handle wafer and then thinned to expose the via patterns from the backside. The silicon in the central core of the annular pattern is then removed and replaced with copper using plating and CMP processes. At this point, the top wafer can be bonded to the bottom wafer, and the glass handle wafer removed. The scanning electron micrograph (SEM) in **Figure 4(c)** shows a detail of an interwafer connection after wafer bonding. The location of the copper TSV, the surrounding via-isolation layers (oxide-polysilicon-oxide), and the connection to the bottom wafer are labeled in the figure. We have demonstrated the ability to process full 200-mm wafers in this manner and the ability to build standard controlled-collapse chip connection (C4) metallurgy on such bonded assemblies. It is important to note that the interlayer-connection density using this methodology is on the order of 10 times tighter than can be achieved using conventional bump-based stacking approaches.

Oxide-bonding 3D process vehicle

On the basis of the success of our earlier oxide-bonding 3D integration work, we initiated a project to assess the viability of using this process to stack commercial foundry-based SOI CMOS starting wafers. This project uses a face-to-back integration scheme to assemble the 3D stack; a wafer that is completed through middle-of-the-line processing is transferred on top of a fully fabricated bottom wafer (with four levels of metal) so that both circuits are facing in the same direction. After wafer bonding, the vertical interlayer via is formed as well as two additional levels of metal. A cross-section of the desired final structure is shown in **Figure 5**. For this work, a specific test vehicle was designed by collaborators at the Mayo Foundation (with support from IBM Engineering and Technology Services) using a conventional CMOS design kit supplemented with 3D design rules provided by IBM. Below are additional details on the process flow for

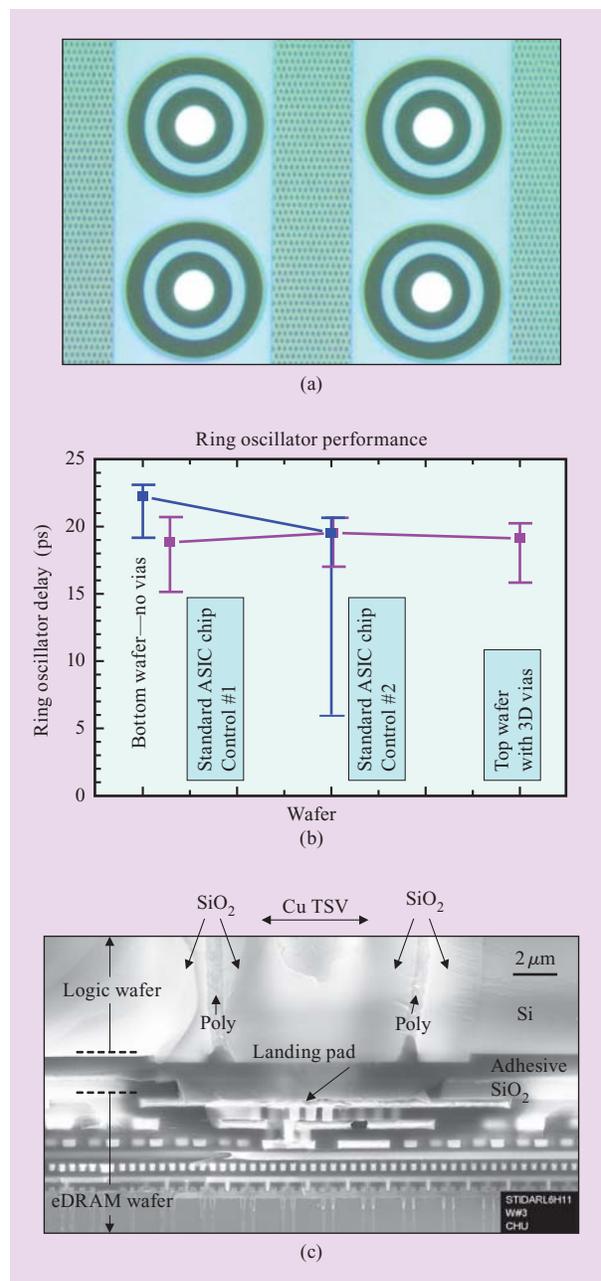


Figure 4

Results of electrical and physical analysis from Cu-bonding 3D process test vehicle. (a) Top-down optical micrograph of through-silicon via (TSV) isolation dielectrics for Cu-bonding 3D process test vehicle. Dark rings: thermal oxide; light rings: polysilicon; white central cores: single-crystal silicon. (b) Performance-screen ring oscillator (PSRO) delay for wafers integrated with TSVs shows no significant differences when compared with two standard wafers used as experimental controls. (c) Cross-sectional scanning electron micrograph of completed Cu-bonding 3D process vehicle showing detail of the top wafer TSV and the electrical connection to the bottom wafer. (ASIC: application-specific integrated circuit.)

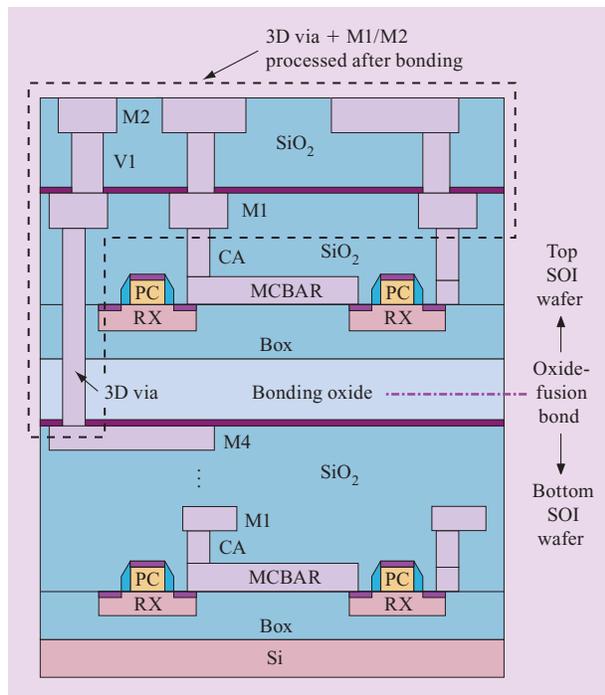


Figure 5

Cross-section of oxide-bonding 3D process test vehicle. The black dashed line indicates structures fabricated after bonding. (PC: gate electrode; RX: isolation; V1: via 1; MCBAR: local interconnect metallization.)

this integration scheme and a description of the structural results of our initial completed hardware.

Starting wafers used were IBM 9S2 (130-nm node) SOI wafers fabricated at IBM Burlington on 200-mm substrates. The bottom wafer was processed up to the fourth metal level (M4), while the top wafer was processed up to the tungsten-contact (CA) level. The top wafer was laminated to a glass handle wafer to allow backside grinding and polishing of the silicon substrate, as described earlier, to a thickness of approximately 100 μm . This part was then etched in TMAH, selectively removing silicon and stopping on the buried oxide of the SOI wafer. At this point, deposition of low-temperature bonding oxide on top of the buried oxide was performed, followed by annealing and CMP. Polishing the backside of the thinned wafer was necessary because of topography associated with the adhesive used to bond the wafer to the glass handle wafer. Because the wafers are simply a thinned membrane at this point, the adhesive layer topography translates through the thin membrane of Si and is seen as roughness on the backside of the SOI. A similar procedure of low-temperature bonding-oxide deposition followed by annealing and CMP was also

applied to the bottom wafer. After chemical surface activation of the planarized bonding surfaces, the wafers were ready for aligned bonding.

Aligned bonding was performed on commercially available tooling; a variety of such options exist. A low-strength, aligned bond was formed on the tool, and the bonded part was then taken through an additional annealing cycle to strengthen the bond at the oxide interface. At this point, the glass handle wafer was removed using laser ablation, and the wafer surface cleaned using oxygen plasma ashing. **Figure 6(a)** shows a cross-sectional SEM of a wafer at this stage, showing the four levels of metal from the bottom wafer and the transistors and contact metallization of the top wafer, all held together by an oxide-fusion bond.

At this point in the process, 3D vias were formed with critical dimensions (CDs) of approximately 0.25 μm and aspect ratios of about 6:1. We have shown the ability to obtain good via-chain yield in these types of structures using fairly standard liner-seed, Cu-plating, and Cu-CMP processes. After the formation of 3D vias, the wafer was then processed using relatively standard CMOS BEOL processing. For our work, a two-level metal dual-damascene Cu BEOL process was employed. **Figure 6(b)** is a cross-section of a completed wafer showing the 3D vias and two additional layers of copper metallization that were fabricated after wafer bonding. Although the specific demonstration discussed here was for a build with four metal wiring levels bonded to two wiring metal levels, in principle, this general methodology could be used to enable any number of metal layers, either before or after the wafer-bonding step.

Challenges and limitations for 3D integration

Clearly, despite the great effort and work scope already covered in the area of 3D integration, several development activities must still be carried out in order to bring 3D integration into a manufacturing environment. In addition, just like 2D systems, 3D ICs will ultimately face roadblocks to their scalability and extendibility, and it is useful to explore the nature of these challenges. In this section, we briefly discuss the manufacturing issues of yield and reliability for wafer-scale 3D systems and then the fundamental limitations of power delivery and cooling.

Numerous reliability concerns exist for wafer-level 3D integration, yet, to date, no comprehensive reliability results have been reported in the literature. For integration schemes that are based on Cu-Cu compression bonding, a major concern arises due to the possibility of exposed Cu at the edges of the bonding interfaces when the surrounding dielectric levels of both layers are recessed. This is different than in standard Cu BEOL, where the Cu has liner all around except for the surface,

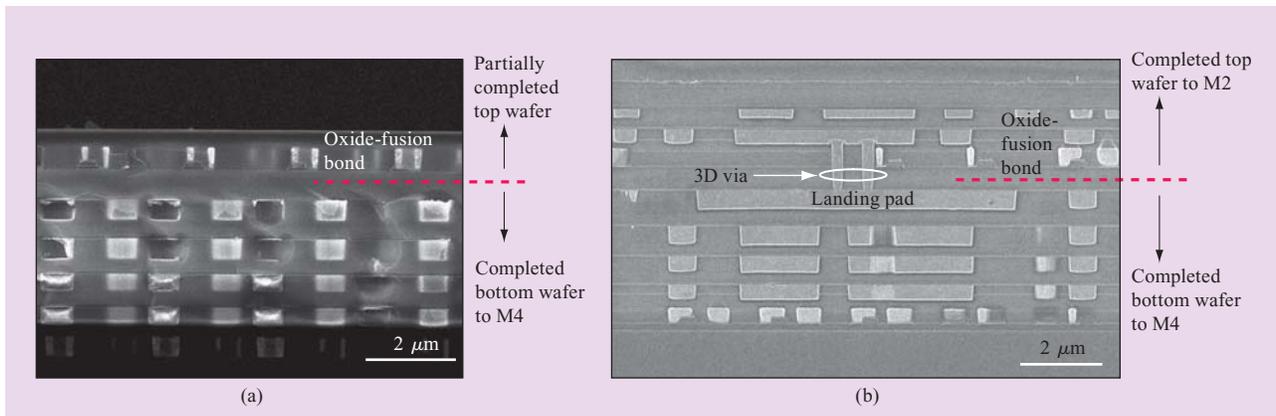


Figure 6

Cross-sectional scanning electron micrograph of oxide-bonding 3D process test vehicle (a) after aligned wafer bonding showing completed bottom wafer attached to a partially completed top wafer by an oxide-fusion bond and (b) after post-bonding fabrication of vertical interconnects and two additional levels of metal.

which is subsequently passivated. Edge seals have been proposed [23]; however, corrosion of the edge seal itself could occur and the use of the edge seal is not a guarantee that further corrosion may not happen.

Electromigration is another concern. Cu is known to have improved electromigration properties compared to other materials, such as solder [31]. However, electromigration studies have yet to be reported on Cu–Cu bonded interfaces, and this is an area where active research is needed. The bonding quality across the wafer is also a significant factor for reliability, as regions with non-ideal bonding quality could facilitate premature failure of the interfacial contact. For all 3D integration schemes, standard reliability metrics for deep thermal cycling, temperature-humidity stressing, and chip–package interactions must also be met.

Wafer-scale 3D integration suffers from a significant drawback compared to chip-scale assembly in the area of yield. Since known-good-die strategies cannot be used, very high die yields for all components in the 3D stack will be critical to achieving a combined yield of the final structure that is comparable to typical 2D yields. It is conceivable that with improved partitioning of technology elements between layers in the stack, such improved yield could occur because the process technology for each layer would be much simpler than in a 2D-only configuration, in which all functions need to be combined in an integrated process. In addition, the use of more robust computer architectures that use multiple cores with redundancy as well as memory architectures with self-repair functionality may be necessary to ensure acceptable yields, particularly in 3D stacks that use advanced CMOS technology.

Looking into the future, a key limitation to the extendibility of 3D integration technology will be the ability to deliver power through the chip stack. The difficulty lies in the fact that power vias need to be wide enough to supply power through the entire chip stack while maintaining low enough resistance to avoid excessive voltage drop on V_{dd} and ground. This problem is exacerbated by the fact that the chip dissipating the most power is likely to be located near the heat sink, farthest from the package, necessitating that power vias run throughout the entire stack. Ultimately, it is conceivable that power delivery considerations could limit the number of chips that it is beneficial to integrate, because a greater percentage of the area allocated to power vias will be needed as the total thickness of the stack increases. Preliminary calculations indicate that for high-performance microprocessor chips, it will be difficult to increase the stack thickness beyond a few hundred microns before the area allocated for power vias becomes too large. This analysis provides a strong motivation for thinning the individual layers in the 3D stack so that more layers can ultimately be integrated. The development of energy-optimized processor designs will also be critical to maintain or decrease the total chip power as additional layers are added.

Along with the issue of power delivery for 3D integration comes the issue of cooling. The main cooling issue arises from the fact that only one chip (usually the microprocessor) can be located directly adjacent to a conventional heat sink. The remaining chips in the stack need to be located between the processor chip and the package and, therefore, necessarily have to operate at a temperature that is higher than the chip located next to

the heat sink. This issue suggests that it will be very difficult to stack two chips that simultaneously dissipate a significant fraction of the total power delivered. Furthermore, it highlights the need to maintain the best thermal conductivity possible between the chips in the stack. The ability to achieve good thermal contact between the layers depends greatly on the integration scheme. In this regard, we note that the metallic connections between the wafers provide the best means for heat transfer between the wafers. Therefore, one would expect that Cu–Cu bonding would have excellent thermal conductivity between the layers, particularly if a large number of dummy metallization layers were present. If an intermediate adhesive layer were used in lieu of dummy bonding pads, then the thermal conductivity between the layers would be decreased. The SOI-based scheme may also suffer from poor thermal conductivity between layers unless very-high-density 3D vias are used.

The use of low- k films in the back-end layers of advanced CMOS chips presents another barrier for thermal transport between the wafers. In this case, the use of thermal vias to ensure adequate conductivity between the layers may need to be investigated. Since these cooling issues could place a significant design restriction on the types of systems that can be realized in a stacked configuration, conventional 2D cooling solutions will ultimately need to be replaced by 3D cooling solutions [32–34]. Such techniques that have been investigated in the literature include double-sided heat sinks, microchannels for fluidic cooling, and interleaved 3D heat-sink geometries. The anticipated cooling issues highlight the fact that additional evaluation of thermal dissipation in bonded structures is critical [35].

Summary

We have presented an overview of wafer-level 3D integration technology. Unlike chip-to-chip and chip-to-wafer integration schemes, wafer-level 3D integration offers the potential for a high-throughput 3D integration scheme that allows high-density interconnects and the flexibility of processing wafers in a semiconductor fabrication line after joining. A motivation for using high-interconnect-density 3D systems has been provided, which mainly relates to the massive bandwidths that can be achieved between microprocessors and memory elements that are stacked in a 3D configuration. The process technology for making such systems has been described and consists of the main elements of vertical interconnects, aligned wafer bonding, and wafer thinning and backside processing. Each of these processes brings about new technology challenges that will need to be addressed in a manufacturing environment. In order to address and clarify the process and integration issues, we

have completed structural builds on two 3D test vehicles, one using an oxide-bonding scheme and one using a Cu–Cu bonding approach. Both of these projects have helped to highlight the process elements and approaches that are most desirable for manufacturing. However, several critical issues need to be studied further; these include power delivery, cooling, and reliability. A thorough assessment of these issues will be needed as 3D technology takes the next steps toward manufacturability.

Acknowledgments

We thank the following people for their contributions to this work: Steven Steen, Cornelia Tsang, Paul Andry, David Frank, Jyotica Patel, James Vichiconti, Deborah Neumayer, Robert Trzcinski, Latha Ramakrishnan, James Tornello, Michael Lofaro, Gill Singco, John Ott, David DiMilia, William Price, Jesus Acevedo, James Lu, Sang Hwui Lee, Ravi Kumaroh, Eric Perfecto, Art Merryman, Mike Berger, Merita Popovic, and John Butler. We also acknowledge the support of the IBM Microelectronics Research Laboratory and Central Scientific Services as well as the staff at EV Group and Suss MicroTec. This project was funded in large part by DARPA under SPAWAR contract numbers N66001-00-C-8003 and N66001-04-C-8032.

*Trademark, service mark, or registered trademark of International Business Machines Corporation in the United States, other countries, or both.

References

1. G. E. Moore, "Cramming More Components onto Integrated Circuits," *Electronics* **38**, No. 8, 114–117 (1965).
2. R. H. Dennard, F. H. Gaensslen, H. N. Yu, V. L. Rideout, E. Bassous, and A. R. LeBlanc, "Design of Ion-Implanted MOSFETs with Very Small Physical Dimensions," *IEEE J. Solid-State Circuits* **9**, No. 5, 256–268 (1974).
3. K. Puttaswamy and G. H. Loh, "Implementing Caches in a 3D Technology for High Performance Processors," *Proceedings of the IEEE International Conference on Computer Design*, San Jose, CA, 2005, pp. 525–532.
4. S. S. Iyer, J. E. Barth, Jr., P. C. Parries, J. P. Norum, J. P. Rice, L. R. Logan, and D. Hoyniak, "Embedded DRAM: Technology Platform for the Blue Gene/L Chip," *IBM J. Res. & Dev.* **49**, No. 2/3, 333–350 (2005).
5. P. Emma, "How is Bandwidth Used in Computers? Why Bandwidth is the Next Major Hurdle in Computer Systems Evolution and What Technologies Will Emerge to Address the Bandwidth Problem," *High-Performance Energy-Efficient Microprocessor Design*, V. G. Oklobdzija and R. K. Krishnamurthy, Eds., Springer Publishing Company, New York, 2006, pp. 235–287.
6. A. W. Topol, D. C. La Tulipe, Jr., L. Shi, D. J. Frank, K. Bernstein, S. E. Steen, A. Kumar, et al., "Three-Dimensional Integrated Circuits," *IBM J. Res. & Dev.* **50**, No. 4/5, 491–506 (2006).
7. J. A. Burns, B. F. Aull, C. K. Chen, C.-L. Chen, C. L. Keast, J. M. Knecht, V. Suntharalingam, K. Warner, P. W. Wyatt, and D.-R. W. Yost, "A Wafer-Scale 3-D Circuit Integration Technology," *IEEE Trans. Elect. Dev.* **53**, No. 10, 2507–2516 (2006).

8. P. R. Morrow, C.-M. Park, S. Ramanathan, M. J. Kobrin, and M. Harmes, "Three-Dimensional Wafer Stacking via Cu-Cu Bonding Integrated with 65-nm Strained-Si/low-k CMOS Technology," *IEEE Elect. Dev. Lett.* **27**, No. 5, 335–337 (2006).
9. M. Koyanagi, T. Nakamura, Y. Yamada, H. Kikuchi, T. Fukushima, T. Tanaka, and H. Kurino, "Three-Dimensional Integration Technology Based on Wafer Bonding with Vertical Buried Interconnections," *IEEE Trans. Elect. Dev.* **53**, No. 11, 2799–2808 (2006).
10. C. Narayan, S. Purushothaman, F. Doany, and A. Deutsch, "Thin Film Transfer Process for Low Cost MCM-D Fabrication," *IEEE Trans. Comp. Pkg. Mfg. Tech. Part B: Adv. Pkg.* **18**, No. 1, 42–46 (1995).
11. D. C. La Tulipe, L. Shi, A. Topol, S. Steen, D. Pfeiffer, D. Posillico, D. Neumayer, et al., "Critical Aspects of Layer Transfer and Alignment Tolerances for 3D Integration Processes," *Proceedings of the International Conference and Exhibition on Device Packaging*, Scottsdale, AZ, 2006.
12. E. D. Perfecto, R. R. Shields, A. K. Malhotra, M. P. Jeanneret, D. C. McHerron, and G. A. Katopis, "MCM-D/C Packaging Solution for IBM Latest S/390 Servers," *IEEE Trans. Adv. Pkg.* **23**, No. 3, 515–520 (2000).
13. K. W. Guarini, A. T. Topol, D. V. Singh, D. C. La Tulipe, L. Shi, A. M. Young, A. Alam, et al., "Process Technologies for Three Dimensional Integration," *Proceedings of the Sixth International Conference on Microelectronics and Interfaces*, Santa Clara, CA, 2005, pp. 212–214.
14. A. W. Topol, D. C. La Tulipe, L. Shi, S. M. Alam, D. J. Frank, S. E. Steen, J. Vichiconti, et al., "Enabling SOI-Based Assembly Technology for Three-Dimensional (3D) Integrated Circuits (ICs)," *Proceedings of the IEEE International Electron Devices Meeting*, Washington, DC, 2005, pp. 352–355.
15. K. W. Guarini, A. W. Topol, M. Jeong, R. Yu, L. Shi, M. R. Newport, D. J. Frank, et al., "Electrical Integrity of State-of-the-Art 0.13 μm SOI CMOS Devices and Circuits Transferred for Three-Dimensional (3D) Integrated Circuit (IC) Fabrication," *Proceedings of the International Electron Devices Meeting*, San Francisco, CA, 2003, pp. 943–945.
16. W. Haensch, E. J. Nowak, R. H. Dennard, P. M. Solomon, A. Bryant, O. H. Dokumaci, A. Kumar, X. Wang, J. B. Johnson, and M. V. Fischetti, "Silicon CMOS Devices Beyond Scaling," *IBM J. Res. & Dev.* **50**, No. 4/5, 339–361 (2006).
17. S. C. Johnson, "3-D TSV Chips Take Off," *Semiconductor International*, July 1, 2007 (available online); see <http://www.semiconductor.net/article/CA6455523.html>.
18. J. Vardaman, "3-D Through-Silicon Vias Become a Reality," *Semiconductor International*, June 1, 2007 (available online); see <http://www.semiconductor.net/article/CA6445435.html>.
19. P. S. Andry, C. Tsang, E. Sprogis, C. Patel, S. L. Wright, and B. C. Webb, "A CMOS-Compatible Process for Fabricating Electrical Through-Vias in Silicon," *Proceedings of the 56th Electronic Components and Technology Conference*, San Diego, CA, 2006, pp. 831–837.
20. C. S. Patel, C. K. Tsang, C. Schuster, F. E. Doany, H. Nyikal, C. W. Baks, R. Budd, et al., "Silicon Carrier with Deep Through-Vias, Fine Pitch Wiring and Through Cavity for Parallel Optical Transceiver," *Proceedings of the 55th Electronic Components and Technology Conference*, Lake Buena Vista, FL, 2005, pp. 1318–1324.
21. C. K. Tsang, P. S. Andry, E. J. Sprogis, C. S. Patel, B. C. Webb, D. G. Manzer, and J. U. Knickerbocker, "CMOS-Compatible Through Silicon Vias for 3D Process Integration," *Mater. Res. Soc. Symp. Proc.* **970**, 145–153 (2007).
22. K. N. Chen, C. S. Tan, A. Fan, and R. Reif, "Morphology and Bond Strength of Copper Wafer Bonding," *Electrochem. Solid-State Lett.* **7**, No. 1, G14–G16 (2004).
23. K.-N. Chen, C. K. Tsang, A. W. Topol, S. H. Lee, B. K. Furman, D. L. Rath, J.-Q. Lu, A. M. Young, S. Purushothaman, and W. Haensch, "Improved Manufacturability of Cu Bond Pads and Implementation of Seal Design in 3D Integrated Circuits and Packages," *Proceedings of the 23rd International VLSI Multilevel Interconnection Conference*, Fremont, CA, 2006, pp. 195–202.
24. K.-N. Chen, S. H. Lee, P. S. Andry, C. K. Tsang, A. W. Topol, Y.-M. Lin, J.-Q. Lu, A. M. Young, M. Jeong, and W. Haensch, "Structure, Design and Process Control for Cu Bonded Interconnects in 3D Integrated Circuits," *Proceedings of the International Electron Devices Meeting*, San Francisco, CA, 2006, pp. 20–22.
25. H. B. Pogge, C. Prasad, and R. Yu, "Bridging the Chip/Package Process Divide," *Proceedings of the Advanced Metallurgy Conference*, Montreal, Canada, 2001, pp. 129–136.
26. R. Yu, "Wafer Level 3-D Integration," *Proceedings of the 24th International VLSI/ULSI Multilevel Interconnection Conference*, Fremont, CA, 2007, pp. 223–230.
27. P. Leduc, F. de Crécy, M. Fayolle, B. Charlet, T. Enot, M. Zussy, B. Jones, et al., "Challenges for 3D IC Integration: Bonding Quality and Thermal Management," *Proceedings of the IEEE International Interconnect Technology Conference*, Burlingame, CA, 2007, pp. 210–212.
28. A. W. Topol, S. J. Koester, D. C. La Tulipe, and A. M. Young, "3-D Fabrication Options for High Performance CMOS Technology," *Wafer Level 3-D ICs Process Technology*, C. S. Tan, R. J. Gutmann, and L. R. Rafael, Eds., Springer, New York, 2008; ISBN 978-0-387-76532-7.
29. S. E. Steen, D. C. La Tulipe, A. W. Topol, D. J. Frank, K. Belote, and D. Posillico, "Wafer Scale 3-D Integration: Overlay as the Key to Drive Potential," *Microelectr. Eng.* **84**, No. 5/6, 1412–1415 (2007).
30. A. W. Topol, D. C. La Tulipe, L. Shi, S. M. Alam, A. M. Young, D. J. Frank, S. E. Steen, et al., "Assembly Technology for Three Dimensional Integrated Circuits," *Proceedings of the 22nd International VLSI Multilevel Interconnection Conference*, Fremont, CA, 2005, pp. 83–88.
31. J. Tao, N. W. Cheung, C. Hu, H.-K. Kang, and S. S. Wong, "Electromigration Performance of Electroless Plated Copper/Pd-Silicide Metallization," *IEEE Elect. Dev. Lett.* **13**, No. 8, 433–435 (1992).
32. C. B. Sobhan and S. V. Garimella, "A Comparative Analysis of Studies on Heat Transfer and Fluid Flow in Microchannels," *Micro. Thermophys. Eng.* **5**, No. 4, 293–311 (2001).
33. B. Dang, M. S. Bakir, and J. D. Meindl, "Integrated Thermal-Fluidic I/O Interconnects for an On-chip Microchannel Heat Sink," *IEEE Elect. Dev. Lett.* **27**, No. 2, 117–119 (2006).
34. D. Gerty, D. W. Gerlach, Y. K. Joshi, and A. Glezer, "Development of a Prototype Thermal Management Solution for 3-D Stacked Chip Electronics by Interleaved Solid Spreaders and Synthetic Jets," *Proceedings of the 13th International Workshop on Thermal Investigation of ICs and Systems*, Budapest, Hungary, 2007; see <http://hal.archives-ouvertes.fr/docs/00/20/25/53/PDF/therm07156.pdf>.
35. R. V. Joshi, T. Smy, K. Banerjee, and A. Topol, "Thermal Dissipation in Bonded Structures," *Proceedings of the SEMATECH Workshop*, Albany, NY, 2007.

Received January 23, 2008; accepted for publication March 26, 2008; Internet publication October 24, 2008

Steven J. Koester *IBM Research Division, Thomas J. Watson Research Center, P.O. Box 218, Yorktown Heights, New York 10598 (koester@us.ibm.com)*. Dr. Koester is a Research Staff Member and Manager of the Exploratory CMOS Technology group at IBM Research. He received B.S.E.E and M.S.E.E. degrees from the University of Notre Dame, and a Ph.D. degree in electrical and computer engineering from the University of California at Santa Barbara, where he performed research involving the fabrication of quantum devices in the InAs/AlSb heterostructure system. He joined IBM in 1995 as a postdoctoral researcher working on the fabrication and characterization of nanostructured devices in Si/SiGe strained-layer materials. Since 1997, he has performed research on a variety of group-IV and III-V heterostructure materials and devices, with an emphasis on strained-layer field-effect transistors and photodetectors. Dr. Koester has authored or coauthored more than 120 technical publications and conference presentations. He holds 18 U.S. patents.

Albert M. Young *IBM Research Division, Thomas J. Watson Research Center, P.O. Box 218, Yorktown Heights, New York 10598*. Dr. Young is a Research Staff Member involved in research and development activities related to 3D integrated circuits. His prior work experience includes 45-nm-node front-end process integration and program management in emerging products at the Semiconductor Research and Development Center at IBM. Prior to joining IBM, he worked for several years in the Solid State Division at Massachusetts Institute of Technology (MIT) Lincoln Laboratory. He received a Ph.D. degree in electrical engineering and computer science from MIT.

Roy R. Yu *IBM Research Division, Thomas J. Watson Research Center, P.O. Box 218, Yorktown Heights, New York 10598 (yur@us.ibm.com)*. Dr. Yu is a Senior Engineer. He received a B.S. degree in nuclear physics from the University of Fudan, China, and a Ph.D. degree in physics from the University of Pennsylvania. His research interests include e-beam/deep UV hybrid lithography for CMOS back-end-of-line scaling and on-wafer level fine-pitch integration. He has more than 40 patents issued or pending and more than 30 technical publications.

Sampath Purushothaman *IBM Research Division, Thomas J. Watson Research Center, P.O. Box 218, Yorktown Heights, New York 10598*. Dr. Purushothaman is a Research Staff Member and Manager of Advanced Interconnect Technology at the IBM T. J. Watson Research Center. He received a B.Tech. degree in metallurgy from the Indian Institute of Technology, India, and M.S. and Eng.Sc.D. degrees in metallurgy and materials science from Columbia University. He has been with IBM since 1979 and has worked in various research areas including advanced packaging interconnects for high-performance bipolar and CMOS server systems; materials and processes for flat panel displays; fabrication and optimization of thin-film transistor devices based on organic semiconductors; and processing and integration of copper wiring with low and ultralow dielectrics for silicon back-end-of-line interconnects. He has authored more than 60 technical publications and holds more than 100 U.S. patents. Dr. Purushothaman is the recipient of the Institute of Electrical and Electronics Engineers (IEEE) EDS Paul Rapport Award for best paper in 2001 and several IBM technical awards.

Kuan-Neng Chen *IBM Research Division, Thomas J. Watson Research Center, P.O. Box 218, Yorktown Heights, New York 10598*. Dr. Chen is a Research Staff Member. He received his M.S.

degree in materials science and engineering and his Ph.D. degree in electrical engineering and computer science, both from Massachusetts Institute of Technology. His studies have included wafer-level 3D device integration, reconfigurable logic application, and phase-change materials. He has authored or coauthored more than 60 technical papers and has 20 U.S. patents or patent applications.

Douglas C. La Tulipe, Jr. *IBM Research Division, Thomas J. Watson Research Center, P.O. Box 218, Yorktown Heights, New York 10598*. Mr. La Tulipe is an Advisory Engineer. Since joining IBM in 1984, he has worked on various projects including exploratory III-V device fabrication, advanced lithographic thin-film imaging, and back-end-of-line low-*k* reactive-ion etch processing. He joined the Silicon Technology Group in 2004 and worked on 3D integration technology until 2007. He currently works in the Process Integration Group at the College of Nanoscale Science and Engineering at the University at Albany, State University of New York, developing advanced process modules for 32-nm and 22-nm CMOS technology. Mr. La Tulipe is the author of several patents and technical papers.

Narender Rana *IBM Research Division, Thomas J. Watson Research Center, P.O. Box 218, Yorktown Heights, New York 10598*. Dr. Rana is a Metrology Engineer. He joined IBM in 2003. He worked on integrated circuit (IC) physical failure analysis, focus ion beam development, and 3D IC integration. In 2007, he joined the IBM Semiconductor Research and Development Center, working on CD reference metrology and metrology development for the 32-nm CMOS node. He received his M.S. degree in physics and his Ph.D. degree in nanoscience and nanoengineering, both from the University at Albany, State University of New York. He holds two U.S. patents.

Leathan Shi *IBM Research Division, Thomas J. Watson Research Center, P.O. Box 218, Yorktown Heights, New York 10598*. Dr. Shi received his M.S. degree in aeronautics and astronautics and his Sc.D. degree in the field of material engineering, both from Massachusetts Institute of Technology. Since joining the IBM T. J. Watson Research Center in 1985, he has worked in a variety of technical areas, including electronic packaging, interconnects, liquid crystal display projection monitors, back-end-of-line, CMOS device fabrication and integration, and wafer bonding. His current major interests are wafer-level 3D device integration and wafer (or substrate) engineering for high-performance electronic devices. Dr. Shi is a member of the IEEE.

Matthew R. Wordeman *IBM Research Division, Thomas J. Watson Research Center, P.O. Box 218, Yorktown Heights, New York 10598*. Dr. Wordeman is a Research Staff Member. After receiving the Eng.Sc.D. degree from Columbia University, he managed a group at IBM responsible for research in MOSFET device design. He received the 1997 IEEE Cleo Brunetti Award for his work on semiconductor miniaturization. He became manager of the IBM Research DRAM Design Group in 1990. In 1992, he joined the IBM Microelectronics Division (now the IBM Systems and Technology Group) in a 256-Mb DRAM development project as a part of the product design department. In 2004, he returned to the IBM Research Division and currently is interested in exploratory interconnects and the design of advanced embedded DRAM circuits and architectures.

Edmund J. Sprogis *IBM Systems and Technology Group, 1000 River Street, Essex Junction, Vermont 05452 (u6474@us.ibm.com)*. Mr. Sprogis holds the B.S.E.E. degree from Worcester Polytechnic Institute and the M.S.E.E. degree from the University of Vermont. He joined IBM in 1978 performing electrical characterization and defect diagnostics in various logic technologies. He then moved into DRAM technology development and worked on advanced cell structures, architectures, and designing novel test vehicles for the characterization of 4-Mb to 1,284-Mb generation DRAM cells. In 1994, he served as lead engineer for several packaging development projects ranging from stacked-die packages to silicon cubes. He is currently a Senior Engineer and, since 2002, has been developing and qualifying TSV technology for high-performance 3D applications.