AFRL-RY-WP-TR-2017-0168

HETEROGENEOUS INTEGRATION TECHNOLOGY

Dr. Burhan Bayraktaroglu

Devices for Sensing Branch
Aerospace Components & Subsystems Division

AUGUST 2017
Final Report

AIR FORCE RESEARCH LABORATORY
SENSORS DIRECTORATE
WRIGHT-PATTERSON AIR FORCE BASE, OH 45433-7320
AIR FORCE MATERIEL COMMAND
UNITED STATES AIR FORCE

NOTICE AND SIGNATURE PAGE

Using Government drawings, specifications, or other data included in this document for any purpose other than Government procurement does not in any way obligate the U.S. Government. The fact that the Government formulated or supplied the drawings, specifications, or other data does not license the holder or any other person or corporation; or convey any rights or permission to manufacture, use, or sell any patented invention that may relate to them.

This report was cleared for public release by the USAF 88th Air Base Wing (88 ABW) Public Affairs Office (PAO) and is available to the general public, including foreign nationals. Copies may be obtained from the Defense Technical Information Center (DTIC) (http://www.dtic.mil).

AFRL-RY-WP-TR-2017-0168 HAS BEEN REVIEWED AND IS APPROVED FOR PUBLICATION IN ACCORDANCE WITH ASSIGNED DISTRIBUTION STATEMENT.

*//Signature//

BURHAN BAYRAKTAROGLU, Program Manager
Devices for Sensing Branch
Aerospace Component & Subsystems Division

*//Signature//

ROSS W. DETTMER

Chief
Devices for Sensing Branch
Aerospace Component & Subsystems Division

This report is published in the interest of scientific and technical information exchange, and its publication does not constitute the Government's approval or disapproval of its ideas or findings.

*Disseminated copies will show "//Signature//" stamped or typed above the signature blocks.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*
August 2017	Final	September 1, 2016 – May 1, 2017

4. TITLE AND SUBTITLE

HETEROGENEOUS INTEGRATION TECHNOLOGY

5a. CONTRACT NUMBER
In-house

5b. GRANT NUMBER
N/A

5c. PROGRAM ELEMENT NUMBER
N/A

6. AUTHOR(S)

Dr. Burhan Bayraktaroglu

5d. PROJECT NUMBER
N/A

5e. TASK NUMBER
N/A

5f. WORK UNIT NUMBER
N/A

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Devices for Sensing Branch (RYDD)
Aerospace Components & Subsystems Division
Air Force Research Laboratory, Sensors Directorate
Wright-Patterson Air Force Base, OH 45433-7320
Air Force Materiel Command, United States Air Force

8. PERFORMING ORGANIZATION REPORT NUMBER
AFRL-RY-WP-TR-2017-0168

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Air Force Research Laboratory
Sensors Directorate
Wright-Patterson Air Force Base, OH 45433-7320
Air Force Materiel Command
United States Air Force

10. SPONSORING/MONITORING AGENCY ACRONYM(S)
AFRL/RYDD

11. SPONSORING/MONITORING AGENCY REPORT NUMBER(S)
AFRL-RY-WP-TR-2017-0168

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES
PAO case number 88ABW-2017-2617; Clearance Date 24 May 2017. Report contains color.

14. ABSTRACT

This report provides a review of the heterogeneous integration of different types of devices and materials for the purpose of increasing the functional density and therefore the performance of electro-optic systems for sensor applications. A categorization of heterogeneous integration methods is provided to provide a structure for this review. The history and the current status of integration technologies in each category are examined and product examples are provided. Finally, potential applications are discussed where heterogeneous integration may provide an advantage for electronic system applications.

15. SUBJECT TERMS
heterogeneous integration, electro-optic systems, sensor applications

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT:	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON (Monitor)
a. REPORT	b. ABSTRACT	c. THIS PAGE			Burhan Bayraktaroglu
Unclassified	Unclassified	Unclassified	SAR	108	19b. TELEPHONE NUMBER *(Include Area Code)* N/A

Table of Contents

Section **Page**

Section **Page**

List of Figures

Figure **Page**

Figure **Page**

Figure **Page**

Figure **Page**

List of Tables

Table **Page**

1. EXECUTIVE SUMMARY

This report provides a review of the heterogeneous integration of different types of devices and materials for the purpose of increasing the functional density and the performance of electro-optic systems for sensor applications.

The importance of heterogeneous integration can be best understood if we first consider the electronic systems alone. The impressive integration levels achieved on a single chip for electronic systems is continuing thanks to ever shrinking transistor feature sizes and the improvements in the interconnect technology. Almost all electronic applications (digital, memory, analog, power etc.) benefit from this improvement in integration density and transistor improvements. However, the design, fabrication and testing maturity are not the same for all applications and therefore different technology nodes are employed for each application at a given time. Integrating several functions on the same chip requires the use of the lowest common technology node rather than the best available one for each. This is where heterogeneous integration first comes in. Rather than using the lowest common technology node for all applications, heterogeneous integration allows the use of the best available technology node for each application to maintain maximum performance. Multiple application components can be separately fabricated and then heterogeneously integrated with each other on a common platform to preserve the compactness while increasing the performance and the functionality density.

The functional density of electronic systems can be further enhanced by integrating different types of devices made from advanced materials. Such devices can provide functions that cannot be obtained from the baseline technology (mostly refers to Si complementary metal-oxide-semiconductor (CMOS)). These added functions can be imaging, optical signal processing, mechanical functions (i.e. inertial) or high frequency amplification. Co-integration of such functions increases the functional density even further.

We will examine the concept of heterogeneous integration first and determine how this technology may help in product development. To provide a structure for this review, we will categorize heterogeneous integration techniques. This is a fast moving technology area driven by commercial products and the technology front is rapidly changing. Categories are kept at a high level to avoid obsolescence because of this rapid change. The history and the current status of integration technologies in each category are examined and product examples are provided. Finally, we discuss potential applications where heterogeneous integration may provide an advantage for electronic system applications.

This report provides only a technology review. It is not intended for identifying existing problems and proposing new solutions. Different readers may benefit from this document by concentrating on different sections. Those interested in understanding the purpose of heterogeneous integration, will benefit from sections 2-6. Categorization of heterogeneous integration technologies can be found in sections 7-8. Those interested in the technical details of various integration methods can find them in sections 9-12. If you are only interested in the examples of heterogeneous integration, they are in section 13. Summary and recommendations are in section 14.

2. INTRODUCTION

By word definition, "heterogeneous integration" (HI) refers to the integration of dissimilar components on a common platform. The term is extensively used in very diverse applications to encompass efforts to make previously separate functions to operate together by an intimate fusion of components. It can mean a seamless integration of previously incompatible software, database, drugs or machine parts. The particular definition that is of significance for us is the integration of dissimilar sensor components onto a common substrate to make new compact components that provide enhanced characteristics. The concept of heterogeneous integration that is most closely aligned to our needs is the one used by the Institute of Electrical and Electronics Engineers (IEEE) as formulized in the International Technology Roadmap for Semiconductors (ITRS 2.0) documents. We will focus our assessment by concentrating on this specific version. However, a broader definition will also be used in connection with some types of HI.

IEEE defines Heterogeneous Integration to be "the integration of separately manufactured components into a higher-level assembly that, in the aggregate, provides enhanced functionality and improved operating characteristics". In this definition, components should be taken to mean any unit whether individual die, micro-electromechanical systems (MEMS) device, passive component and assembled package or sub-system that are integrated into a single package. The operating characteristics should also be taken in its broadest meaning including characteristics such as system level performance and cost of ownership [1].

Heterogeneous integration includes a variety of technology components ranging from transistors and their fabrication methods to testing and packaging. Therefore, the heterogeneous integration roadmap activities are sponsored by more than a single IEEE chapter. Although the lead IEEE Chapter is the IEEE Components, Packaging and Manufacturing Technology Society (CPMT), other IEEE Societies such as IEEE Electron Devices Society (EDS) and IEEE Photonics Society are involved as well as the Semiconductor Equipment and Materials International (SEMI) organization. The intention is to expand the roadmap collaboration to other IEEE Technical Societies that share interest in the Heterogeneous Technology Roadmap as well as to organizations outside IEEE that share this common vision for the roadmap [2].

When we think about the evolution of electronic systems, we often form opinions by relying solely on the progress of "digital technologies" whose evolution is often governed by the so-called "Moore's Law". This empirical rule predicts that the number of transistors of an (digital) electronic system will double every 2 years. The transistor geometries are expected to continue shrinking to accommodate ever higher density circuits following the well-known technology roadmaps such as the ITRS. These roadmaps help synchronize technology development efforts by providing guidance to research communities and funding agencies. So far, the use of such roadmaps have been very successful in improving resource efficiencies and in providing predictions in the capabilities of future systems.

Until recently, compliance with the Moore's Law used to require simple geometrical shrinkage (geometric scaling) of transistor feature sizes so that more transistors can be accommodated on a given chip size. Smaller transistor gate lengths also improved the device performance by reducing the transit time for electrons between electrodes. This trend has continued unbroken

until about 2004, when it became apparent that simply cramming more transistors in an area is not the best way to improve (digital) system performance. Delays due to metal interconnects and the formation of thermal hot spots have severely limited the performance gains obtained by density improvements [3]. Currently, the circuit density is still increasing but the operation frequency is saturating, as shown in Figure 1.

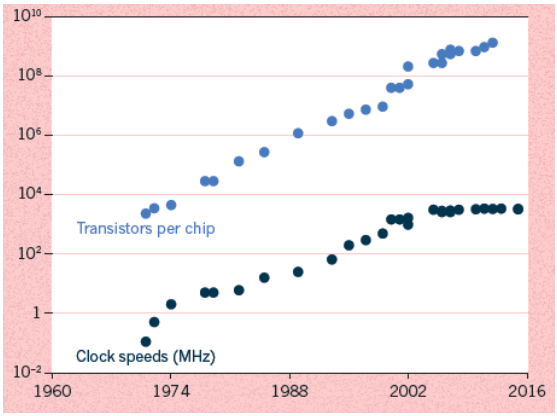


Figure 1: Number of Transistors per Chip has doubled every 2 Years and this Trend is continuing as the Transistor Feature Sizes Shrink
But the speed of chips saturated since 2004 due to high temperature limits [Source: Intel].

Other scaling approaches are being investigated to increase the chip performance without necessarily reducing transistor size, in which case the circuit density increase will saturate. This new reality can be seen by comparing the device feature size reduction expectations of the 2013 ITRS roadmap with the more recent 2015 ITRS roadmap expectations as shown in Figure 2 [4]. While previously it was expected that the physical gate length shrinkage would continue until at least 2028, now it is believed that it will saturate in 2021 when the 9nm node it reached. Any further increase in circuit density will now come from 3D architectures. The circuit density improvements by 3D stacking of similar circuits (functions) is “homogeneous integration”.

The same approach can be applied to the integration of digital circuits with other types of circuits. These circuits may include functions that can be best implemented by a different technology node. They may include functions that are implemented in technologies that do not follow the Moore’s Law (such as analog/radio frequency (RF) circuits). They may indeed include functions that are not even electronic (such as optical, mechanical or acoustic). The integration of these diverse function circuits is “heterogeneous integration”.

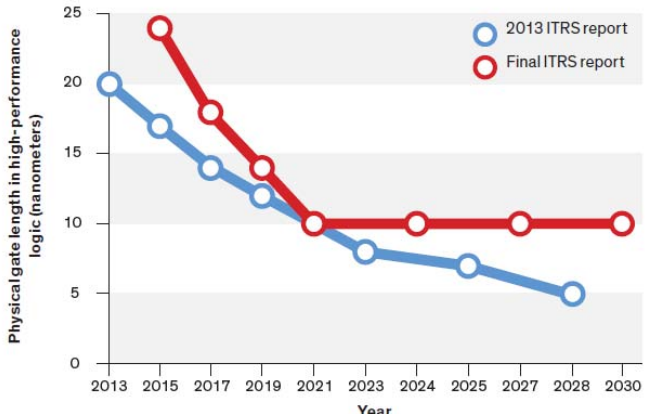


Figure 2: Physical Gate Length Shrinkage of Transistors in 2013 and 2015 [4]

3. HETEROGENEOUS INTEGRATION

Before we define what heterogeneous integration means and what technologies it contains, we must clarify what it does not mean. Heterogeneous integration is not a continuation of the Moore's Law through 3D integration. Moore's Law is strictly for the rate of increase in digital circuit integration density. Heterogeneous integration is for the integration of different functions. Both activities appeared to merge recently with the introduction of 3D integration at the chip level. This is, however, is only an appearance based on the similarities in the integration methods. While, integration density increases are still implemented at the chip level, functional density increases are implemented both at the chip level and in packages.

As mentioned above, the integration density improvements are beginning to saturate when confined to 2D surfaces, and therefore the third dimension will be used to continue improvements. This trend has already started with the stacking of memory chips and will spread to processor chips. This new path to integration density increase is called the **More Moore (MM)** roadmap.

System in Package (SiP) integration has always been about functionality integration and it was used to integrate multi-functional chips in one location. There are many ways to increase the density of functions in a package, as we will review in detail below. Some of these involve stacking chips on each other whereas others involve stacking packages on top of each other. There are also many choices in electrical signal routing between chips involving bond wires, ball grid arrays, through Si vias (TSV), interposer layers etc. The increase in functionality density through heterogeneous integration is called **More than Moore (MtM)**.

Figure 3 is a simplified diagram that shows the use of 3D integration for both integration and functional density improvements. When similar technologies are integrated, such as dynamic random-access memory (DRAM) chips, the final assembly has the same function as the individual chips but the integration density is increased. This is called 3D-Integrated Circuit (3D-IC). The integration density increase follows the MM roadmap. When diverse technology components are integrated, heterogeneous integration results. The evolution of this integration approach is MtM.

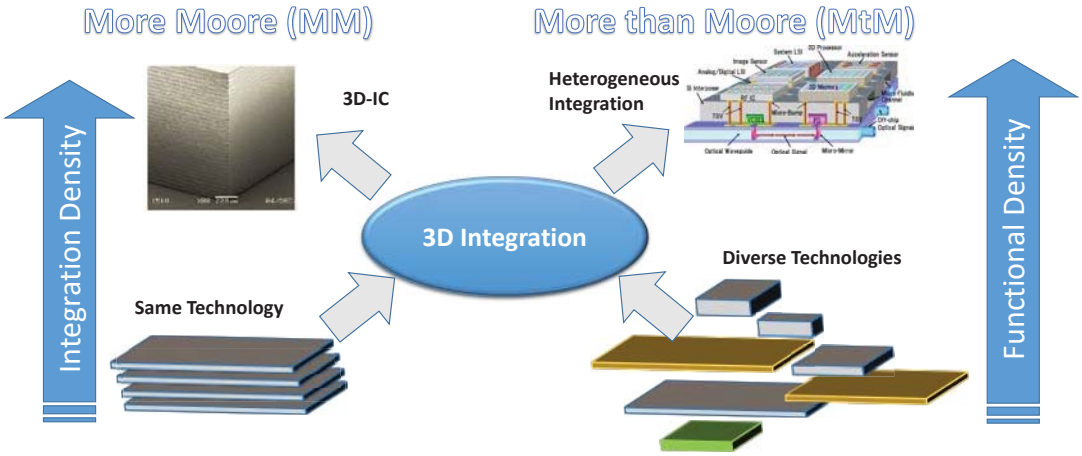


Figure 3: 3D Integration of Similar or Diverse Technology Components follows MM and MtM Roadmaps

4. WHY DO WE NEED HETEROGENEOUS INTEGRATION?

In 2016, ITRS 2.0 was introduced to capture this new reality in technology outlook. This is the first time ITRS is not based on the expected improvements in the transistor technology underscoring the anticipation that the microelectronics product performance improvement will no longer be paced solely by gate length shrinkage. Most commercial (outside of data center applications) and military applications will benefit from the integration of "non-digital" functions with digital circuits. In this new approach, **more functions** are offered per given chip area in contrast to the previous approach where only increased performance of the **same function** was offered from the same chip area as the technology evolved. The road mapping activities that govern such highly integrated heterogeneous systems is known as "More-than-Moore" (MtM). MtM shifts the intellectual epicenter of product development from the transistor level innovations to system-level design innovations. Heterogeneous Integration plays an important role in the implementation of MtM. This is especially true for military applications where it takes longer to field technology innovations in products.

Military microelectronic sensor systems are complex systems that require the use of a range of different technologies, as indicated in Figure 4. For various valid reasons, such as reliability, testing, re-work and availability, components made using these technologies are integrated on printed circuit boards (PCBs). Sometimes, several components can be integrated inside modules, which are custom packages. A sensor system constructed in this fashion has the usual communication bandwidth problems due to time delay caused by component interface networks. To reduce delay, components must be placed closer to each other, and if possible, direct connections must be made (rather than using bond wires).

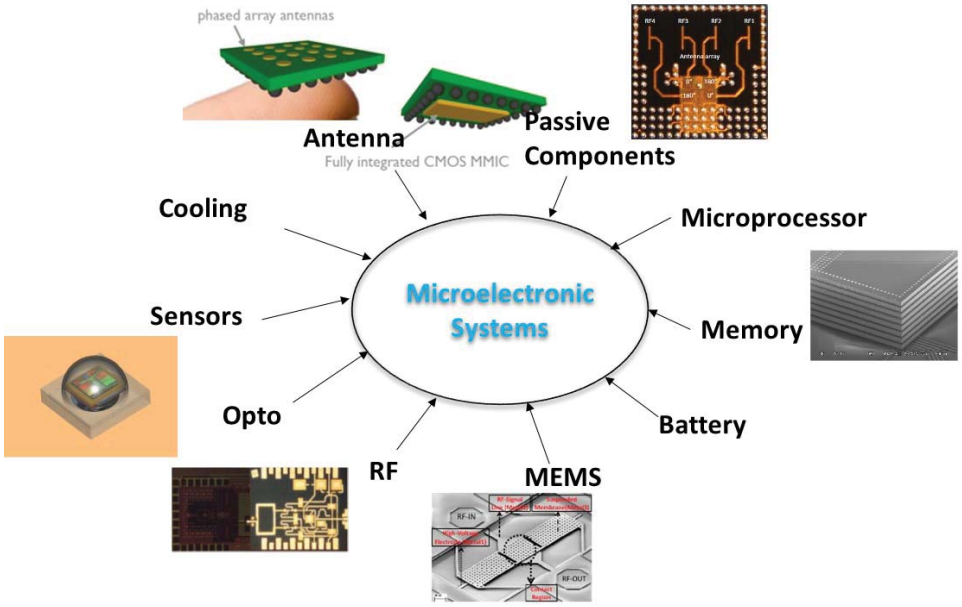


Figure 4: Many Different Technologies are used in the Implementation of Modern MEMS which can benefit from HI

There are a number of drivers for integrating components together into a common package. These include:

- Compact and disposable sensors
- Higher communication bandwidth
- Integrated data processing
- Improved size, weight, power, and cost (SWAP-C)
- Simplified supply chain management

Many of the technology solutions in the form of heterogeneous integration on wafers or in packages developed in response to these drivers for commercial applications are also suitable for some military systems. We will review all technology approaches currently available in open literature and provide examples of improvements achieved by HI so far. The information referenced in this document are from open literature. No proprietary information is included.

First, we examine the potential impact of the HI on a typical product development strategy.

5. IMPACT OF HETEROGENEOUS INTEGRATION ON PRODUCT DEVELOPMENT

The impact of heterogeneous integration on future products of large organizations with their independent research and development (R&D) capabilities can be visualized by examining the schematic representation of the technology food chain for products, as shown in Figure 5. The science and technology (S&T) base on the left side includes a variety of components ranging from materials, device concepts, fabrication methodology to circuit designs. This is an ever evolving area for many organizations with several diverse ideas being pursued in parallel. In the case for commercial applications, Si digital IC's make up a significant portion of this base. The Si IC technology is evolving as fast as it can (Moore's Law and More Moore) for commercial applications thanks to huge investments made by several multi-international companies. Most organizations no longer maintain independent development in this area, but instead rely on technology roadmaps to anticipate the availability of performance levels at a given future time without necessarily investing in its evolution. Organizations may maintain some basic S&T development activities in materials, devices and fabrication techniques to differentiate their products.

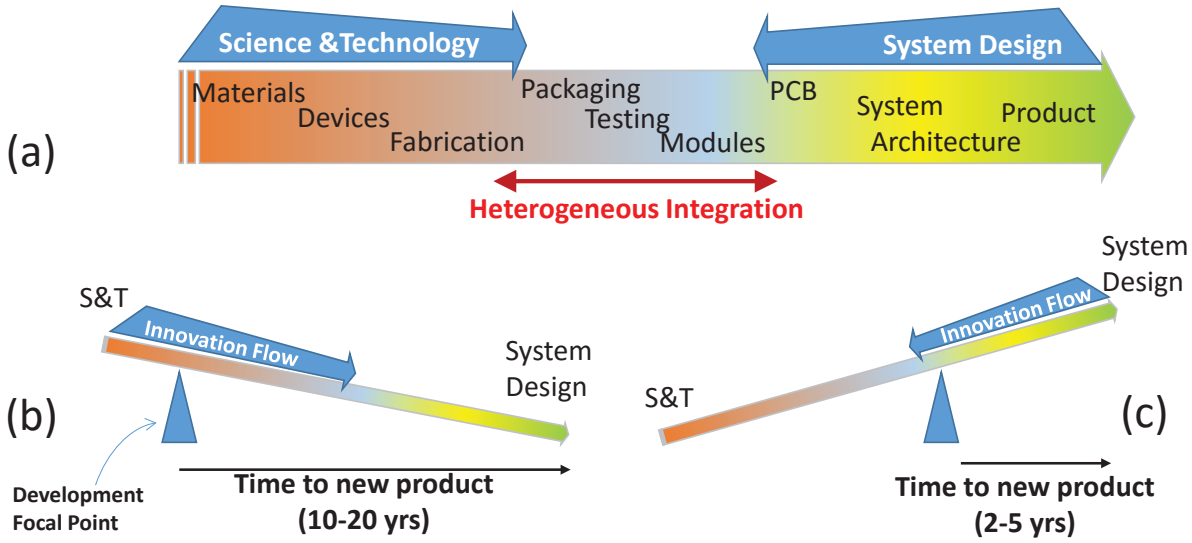


Figure 5: Schematic Representation of the Technology Food Chain for Products
(a) Technology food-chain for typical microelectronics products development. Science and technology push and system design pull on this food chain has limited ranges. (b) If the development focal point is closer to the S&T end of the chain, the innovation flow will be toward system designs. It may take a long time for this innovation to reach systems. (b) If the development focal point is closer to the system design end of the chain, innovation flow will be from the systems to components and the activities in the focal point can be influenced. New product development time can be reduced substantially.

The evolution of any new device technology goes through several other stages including manufacturability, design rules, circuit designs and packaging before they are ready to include in modules. Module designs are often specific to each new technology and new module designs are needed for each innovative device technology. Modules are used in sub-systems, whose design

critically depend on the module characteristics. The final product design integrates sub-systems for a system.

The problem with this approach is that system architecture innovations are not the primary drivers of the product characteristics. The requirements flow-down can rarely penetrate through the module level to reach the S&T crucible where many of the innovations are taking place. Therefore, most product enhancements come from innovations initiated at the lower end of the food chain. The worst part is that the bottom-up technology push of this type takes decades of effort to mature. By the time the new basic technology is accepted by the system designers, the range of innovations that can be introduced at the system level are limited. This situation represents a "technology push" condition.

An example of this problem can be observed in the development of commercial microelectronic products. For decades, the computer performance was paced by the innovations at the chip-level technologies. The computer architecture has not changed substantially since from the beginning. Computer designers were in no position to flow-down requirements for an innovative architecture to the chip manufacturers because of the cost and the uncertainty of implementing new basic transistor designs. ITRS and other technology roadmaps were used by the industry to forecast the improvements in chip performance for system designers to implement their limited innovations based on these forecasts. In most cases, the system design improvements were simply in the form of faster computers that are smaller and cheaper but not different in operation than their predecessors.

Heterogeneous integration or more broadly speaking MtM can change this dynamic by moving the development focal point closer to the system designs, as illustrated in Figure 5b and Figure 5c. If the development focal point is near the bottom of the food chain, innovations will flow along the chain to systems designs. This flow can take 10-20 years. This is the current situation. On the other, if the development focal point is moved up to modules by establishing a core competency in heterogeneous integration, the system insertion time can be cut back to 2-5 years. In this case, innovations in system designs can reach the core activities at the focal point of research. This will be a "systems-pull" approach.

A disregard for innovation flow results in a linear (flat) approach of Figure 5a. Without any particular bias for the "innovation flow", the respective influences of "technology-push" and "systems-pull" extend from the ends of the chain toward the middle. The result is an accumulation of unfinished "product components" that can neither be directly inserted in systems nor can they be sustained technologically. If the "product component" is a result of technology push only, it struggles to fit into a conservative system environment. If the "product component" is a result of systems pull only, with no particular regard to technology trends, its manufacturability may be questionable and immediate obsolesce may result. In some simpler application areas such as automobile electronics, where the technology food chain is shorter, the respective influences of S&T and systems designs can extend past each other to reach ends of the chain. For most modern complex systems, the chain is far too long for this to happen.

However, by moving the innovation center to the middle or closer to the systems end of the chain, and making use of the heterogeneous integration innovations, the development dynamics can be modified. There are 2 reasons for this outcome; a) systems innovations can be accommodated because the requirements flow down is more effective (shorter flow), b) basic technology innovations can be readily incorporated into heterogeneous integration technology (by definition, heterogeneous integration is about incorporating different technologies). Therefore, HI offers a good midpoint of a long technology chain and can serve as a development focal point to both absorb new technologies and then pass them on to the system designs.

If a core competency is centered on HI, the technology food chain can be responsive to innovations both from S&T and systems designs, as illustrated in Figure 6. HI core competency can ensure heterogeneous integration of dissimilar technologies developed by the S&T community in a way that is also responsive to the requirements of innovative system designs. This approach represents a win-win situation by accommodating innovations from both ends of the technology/product spectrum.

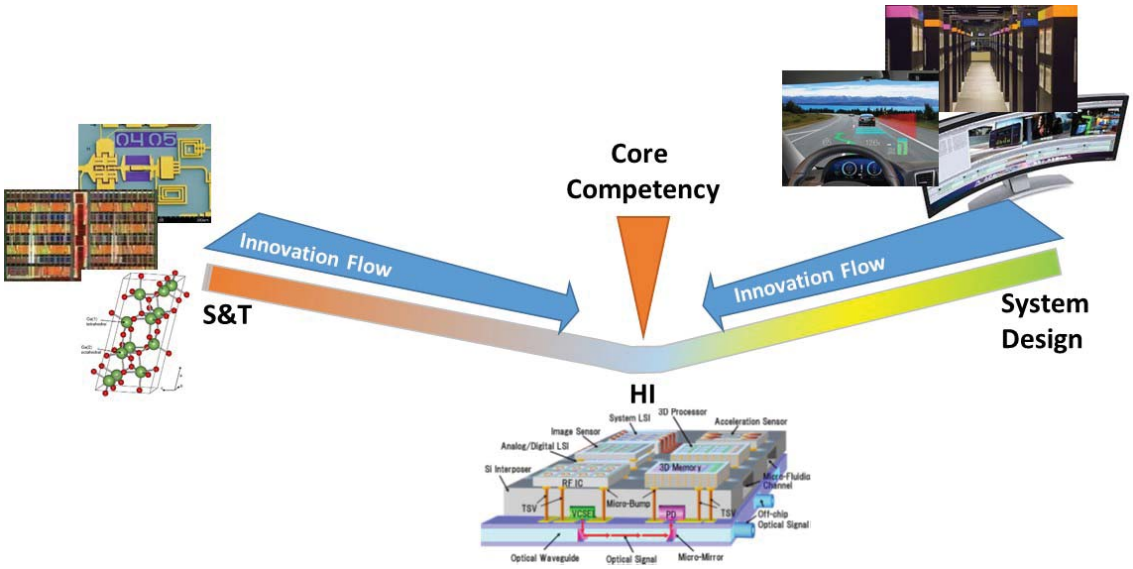


Figure 6: Flexible Technology/Product Food Chain with a Strong Core Competency in HI

In the following sections, we will review the current status and the future trends in HI technologies within the context of MtM roadmaps.

6. HETEROGENEOUS INTEGRATION ROADMAPS

Heterogeneous integration roadmaps include technology developments both at the wafer-level and the packaging-level. Initially, heterogeneous integration simply meant the integration of different device technologies on the same Si substrate, such as bipolar and field-effect transistor integration for Bi-CMOS circuits [5]. Later SiGe heterojunction bipolar transistors were also integrated on Si for higher performance circuits [6-8]. A similar approach was also used for advanced semiconductors based on GaAs to integrate heterojunction bipolar transistors (HBTs) with field-effect transistors [9-18] and also with PIN diodes [17]. These types of integration, which are now a well-established part of the modern semiconductor technology, rely on the fabrication of the active devices directly on the substrate surface. The active devices are fabricated side-by-side on the same plane, and therefore the integration type is 2-dimensional (2D). Other heterogeneous integration approaches make use of the third dimension either as a part of MM or MtM approaches described above. These integration approaches broadly fall under 3D integration. However, 3D integration methods are not easy to categorize because the integration can be accomplished directly on the wafer (system on chip (SoC)) or in the package (SiP), packages can be integrated on the chip or chips can be integrated in the package, chips can be integrated on chips or packages can be integrated in or on other packages etc. There are many existing packaging types and also many types of chip mounting and interconnect technologies that support 3D integration. All these choices of integration produce many technology pathways to future products.

Figure 7 shows the significant commercial products introduced over the last 10 years using both MM and MtM integration approaches. These new products increase both the integration density and the functional density, but there is a lack of unified approach to integration. Each product has a unique form factor that is driven by the application as well as the integration approach employed. While high density memory products employ chip stacking approaches, multi-function products (e.g. graphics processors) employ SiP approaches. It is difficult to find a baseline integration technology evolution used for the fabrication of these state-of-the art products. The seemingly directionless evolution is mostly due to the fact that the product designs now have significant impact on the integration technology used for those products (i.e. systems design pull). This can be contrasted to prior generation products whose integration were similar to each other and the influence of the baseline technologies could be identified (i.e. technology push).

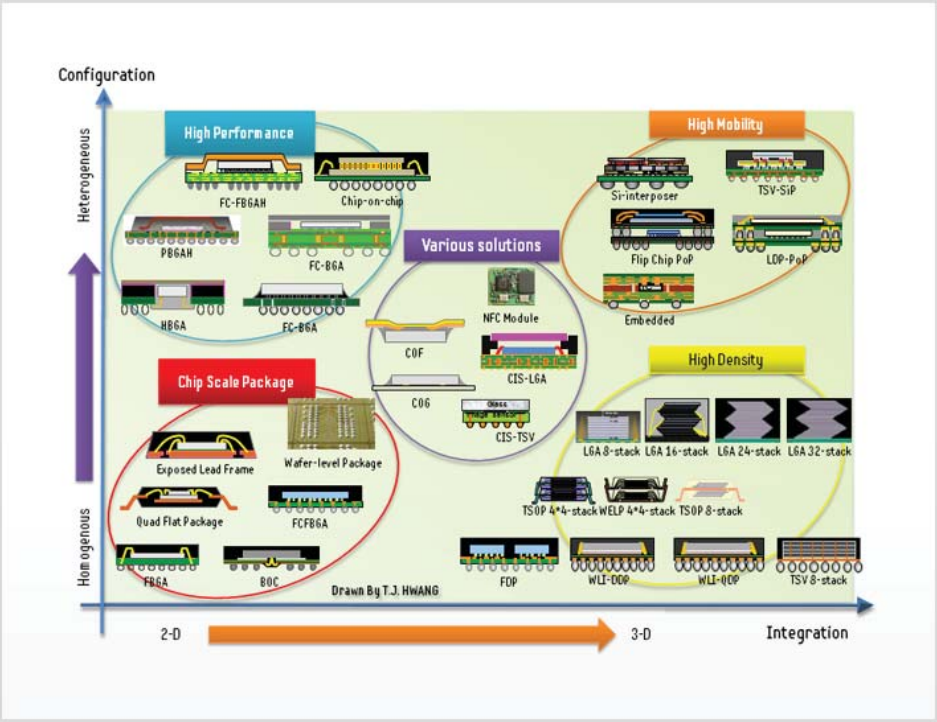


Figure 7: Recent Electronic Products Employing Homogeneous and Heterogeneous Integration Approaches

Integration methods are highly varied and seem to be only driven by the application (http://www.samsung.com/semiconductor/support/package-information/overview/).

The current situation with highly application-driven integration technology makes the categorization difficult. In a technology push environment, the evolution of technology is easier to track since it starts evolving from a few original sources. In an application driven environment, as is the case for heterogeneous integration, there are multiple points of entry for many technologies. Several simplifications are often used to fit various technology paths into artificial roadmaps. An example of such a technology roadmap is shown in Figure 8. In this roadmap, generated by Texas Instruments, various enabling technologies are put to use depending on specific applications. It is clear that the technology evolution tracks closely the application evolution. The current integration technologies such as die stacking, wafer thinning, flip chip (FC), package on package (PoP), flip chip scale packaging (FC CSP), wafer level packaging (WLP) etc. are used for the current applications such as computers, cell phones, automotive electronics. Future integration technologies such as Si embedding, face-to-face (F2F) interconnects, ultra-thin TSVs are reserved for anticipated wearable electronic products. However, this is a technology evolution view of one company. Other electronics giants such as Samsung, Global Foundries, and Fujitsu etc. all have completely different roadmaps describing the future evolution of heterogeneous integration. Some of these roadmaps are shown in the Appendix. We will attempt to categorize heterogeneous integration approaches in the next section to aid in the technology review process.

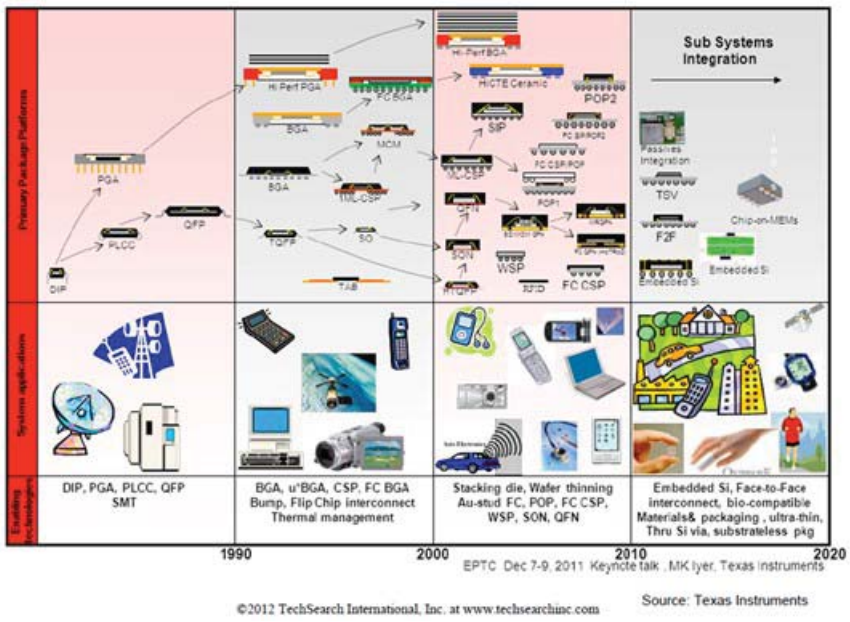


Figure 8: Integration Roadmap for Texas Instruments showing the Evolution of HI
Initial evolution from simple single-chip packages continued until 2000. Since then, a variety of packaging techniques were introduced. Some of these techniques are not evolutionary but revolutionary.

7. HETEROGENEOUS INTEGRATION CATEGORIES

In this section, we will attempt to simplify the types of heterogeneous integration technologies in use today. Homogeneous integration based on a single technology (e.g. memory) and monolithic integration of active and passive components will not be included. These technologies are not regarded as heterogeneous integration and therefore are outside the scope of this analysis.

Figure 9 illustrates the basic categories of heterogeneous integration techniques. The evolution of integration has been from the initial 2-dimensional integration toward 3-dimensional integration. This figure identifies the designations of integration level and the integration location for both chip level (SoC) and package level (SiP) integration approaches. Although the IEEE definition of heterogeneous integration now only refers to integration in packages, we will include the prior and the current integration efforts at the chip level for completeness of technology assessment.

The heterogeneous integration evolution on chip and in packages followed almost independent roadmaps in the past. The integration level designations i.e. 2D, 2.5D and 3D, therefore have different meanings for on-chip and in-package integration. We will first identify these differences.

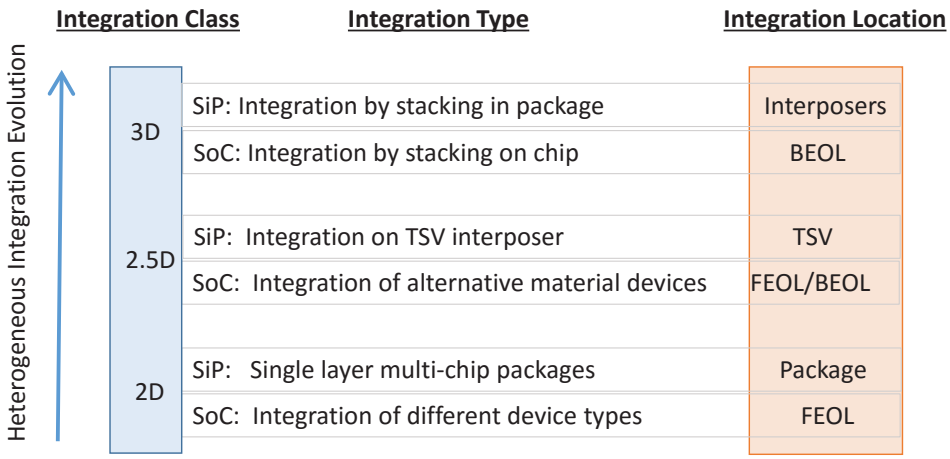


Figure 9: HI Categories

7.1 Integration Categories for SoC

As mentioned above, heterogeneous integration started as an integration of different device types on the Si surface (2D). These active devices could be connected at the lowest interconnect level (M0) and therefore this integration represents the most intimate connection between different devices. When different devices made from alternative materials were integrated on the Si wafer, a similar integration function was achieved. But because of additional processing levels that are required to produce active device layers, which are different than the substrate, connections between these devices moved up to local or intermediate level of wiring. We designate this integration as 2.5D. Later, even more ambitious integration approaches were developed to integrate completely different classes of devices on the same wafer. Examples of this include the integration of MEMS and optical devices on Si or GaAs wafers. This integration type typically takes place at the back end of line (BEOL) part of processing. Different device types can either

be fabricated as a part of a unified wafer fabrication in one facility, or they can be added on the wafer later at a different facility. Either way, the third dimension is used to facilitate this type of integration (3D). 3D integration can also include specialized packaging that may be essential for the add-on devices. For example, some MEMS devices require vacuum inside individual packaging. These custom packages may be best produced around the device. In 3D integration, we also include stacked chips of different technologies. Figure 10 shows some examples of the classes of integration on chip.

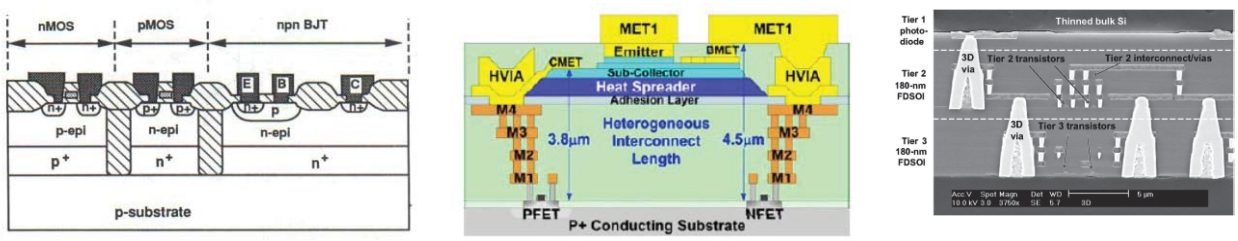


Figure 10: Illustration of On-chip HI Evolution from 2D to 3D

7.2 Integration Categories for SiP

Most electronic system packages are now ball grid array (BGA) type. The old lead frame (LF) type packages are also still used but they are evolving into more flat type packages such as dual flat no-lead (DFN) or quad flat no-lead (QFN) type plastic packages. The current generation of packages are the decendents of these basic package types. However, for the purpose of classifying heterogenerous integration in packages, we concentrate only on their use of dimensions. The simplest integration type is the one where 2 or more different technology chips are packaged side by side on the same surface. This is designated as 2D integration, as shown in Figure 11. The 2.5D integration commonly refers to the use of TSV as the integration platform inside the package. TSV or more generally interposer layers can contain metallized vias through the layer to route interconnections from the chip to the package. Initially TSVs were “fan-in” type, meaning that the pitch density on the top and the bottom surfaces were the same. Other TSVs can be of the “fan-out” type meaning that small pitch connections on the chip side can be spread out to larger pitch sizes on the package side. The first use 2.5D integration in package by Xilinx resulted in a custom field-programmable gate array (FPGA) chip (XC7VH580T) that integrated three active die: an 8 x 28Gb/s transceiver IC and two FPGA ICs known as Super Logic Regions (SLR) on a passive silicon interposer [19]. As we will see later, TSV type interposers come in 2 styles: passive and active. Only passive type TSVs are used in 2.5D integration.

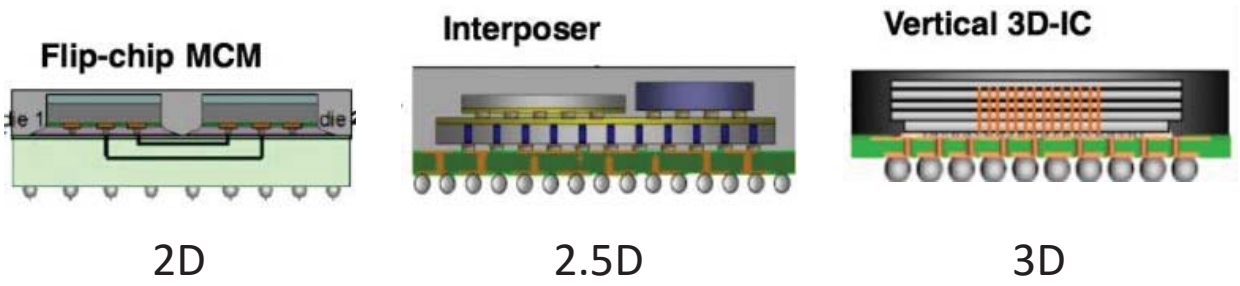


Figure 11: Illustration of In-package HI Evolution from 2D to 3D

The 3D SiP integration includes stacked chips or indeed stacked packages with interposer layers. Although 3D integration has the most promise for future highly integrated systems, which are more commonly referred to as "integrated modules" in both commercial and military applications, the technology maturation is still far behind other integration approaches. Both the passive and active type TSVs are used for these 3D IC packages. Figure 12 illustrates the technology roadmap for both type of interposers [20]. It is expected that most 3D SiP packaging will continue to use passive interposers while active interposers will only be employed in special applications, such as CMOS image sensors (CIS) [21, 22].

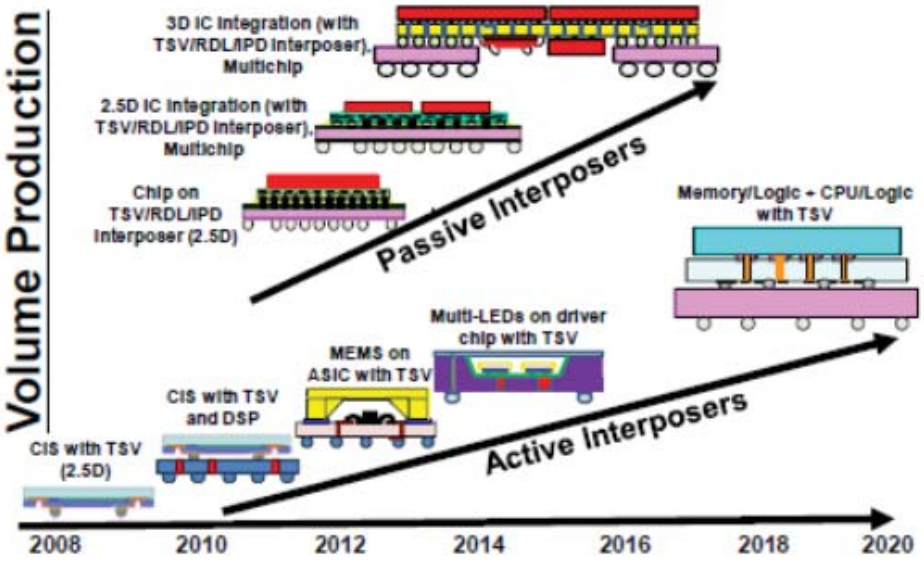


Figure 12: 3D SiP IC Integration Roadmap [20]

8. INTERCONNECT STRATEGIES FOR INTEGRATION

Before we discuss in more detail heterogeneous integration technologies in all categories, it is useful to first examine the *interconnect strategies* that are commonly used. The interconnect types are different for on-wafer and in-package type integation. Here we will only point out the categories of interconnects and the design strategies for using each category.

8.1 Interconnects for On-Chip Integration

One useful method of distinguishing among SoC type integration methods is to specify the location of integration in the process. For this purpose, we use the front and the back end of line process (FEOL and BEOL) definitions. FEOL processes involve the fabrication of active devices and some passive components at or close to the wafer surface, as shown in Figure 13. BEOL includes the fabrication of the interconnect wiring layers and the final passivation layer. These wiring layers are broken into several categories including local wiring, intermediate wiring and global wiring layer. There can be 10-15 wiring levels. While 2D and 2.5D integration takes place at or close to FEOL, 3D integration is at the BEOL. When chips are stacked, they can share global wiring and/or passivation layers.

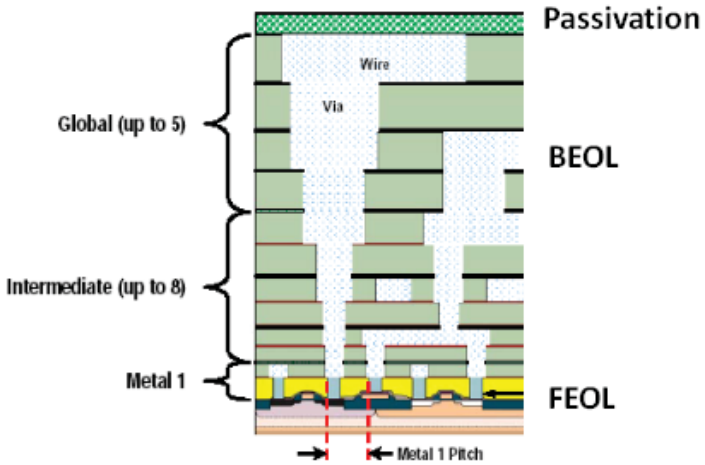


Figure 13: Semiconductor Wafer Process Delineation showing the Location of FEOL and BEOL Layers

In modern chip-level electronic circuits, the interconnect structure is strongly hierarchical. Short and thin interconnects are used at the transistor level (local interconnects). Longer and larger cross-section interconnects are used for connecting blocks of transistors (intermediate interconnects). At a higher level, large circuit blocks also called “IP blocks” or “cores”, are connected by longer and larger diameter wires (global interconnects). Package-level connections are made using bond pads at the top level. This level of interconnect is also used for making direct contact between chips when they are stacked. To complement the wiring size hierarchy, the interconnect density changes exponentially decreasing from the local interconnect to bond pad I/O level interconnects. The exact value of wire dimensions and interconnect density is specific to the technology node (measured typically by the transistor gate size). The interconnect pitch and density scaling for the 14-nm technology node is illustrated in Figure 14.

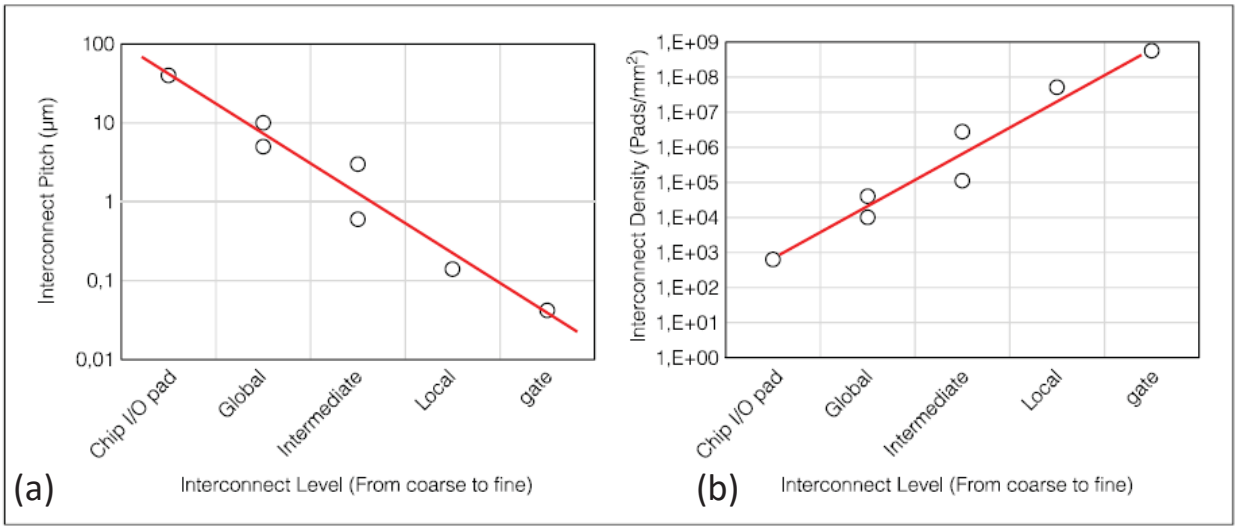


Figure 14: (a) Interconnect Pitch and (b) Interconnect Density for 14-nm CMOS Technology [23]

Various chip-level integration schemes are illustrated in Figure 15 as defined by their integration level. On the one hand, the stacked IC (SIC) technologies are implemented at the global wiring level resulting in lower density integration with contact pitches varying from 40 to 5µm. On the other hand, 3D-IC integration (continuation of the 2D integration in the 3rd dimension) is implemented at the FEOL level with contact pitch of 100nm or less. Other 3D-SoC integration approaches take place at the semi-global, intermediate or local interconnect levels with pitch densities ranging from 5µm to 100nm.

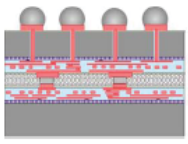
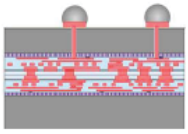
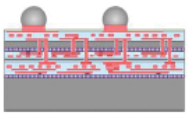
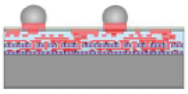
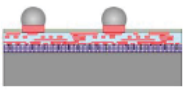
	3D-SIC	3D-SOC			3D-IC
3D-Wiring level	Global	Semi-global	Intermediate	Local	FEOL
Partitioning	Die	**blocks** of standard cells		**Standard cells**	**Transistors**
3D Technology	Die stacking Die-to-Wafer stacks Die-to-Si-interposer	Parallel FEOL wafer processing Wafer-to-Wafer bonding		Sequential FEOL processing Active layer bonding or deposition	
2-tier stack schematic				2nd FEOL after stacking 	Multi-tier FEOL 
Characteristic	Known Good Die 3D stacks or Si-interposer stacking	BEOL between 2 FEOL layers *Overlay 2nd tier defined by W2W alignment/bonding*		*Overlay 2nd tier defined by litho scanner alignment*	FEOL/FEOL stack
Contact Pitch *Relative density:*	40 ⇒ 20 ⇒ 10µm ⇒ *5µm* 1/16 ⇒ ¼ ⇒ 1 ⇒ 4	5 ⇒ 1µm 4 ⇒ 100	2 µm ⇒ 0,5 µm 50 ⇒ 400	200 ⇒ 100 nm 5000 ⇒ 10000	< 100 nm > 10000

Figure 15: On-chip 3D Integration Approaches Defined by Integration Location in the Wiring Hierarchy [23]

8.2 Interconnects for in-Package Integration

The 3D integration in packages is a fast-growing technology area. Although most packages are a variation of BGA packages that have been around for 20 years, 3D integration within these packages have been making fast progress. A variety of new packaging concepts have evolved for new compact system applications. An example of the 3D SiP approaches is shown in Figure 16. Unlike SoC 3D integration, the interconnect strategy for 3D SiP integration is not hierarchical. The existing connection techniques developed for single die packages, i.e. FC and wire bonding (WB), are also used for 3D integration. In addition, a relatively new interconnect technology, TSV is also finding widespread use.

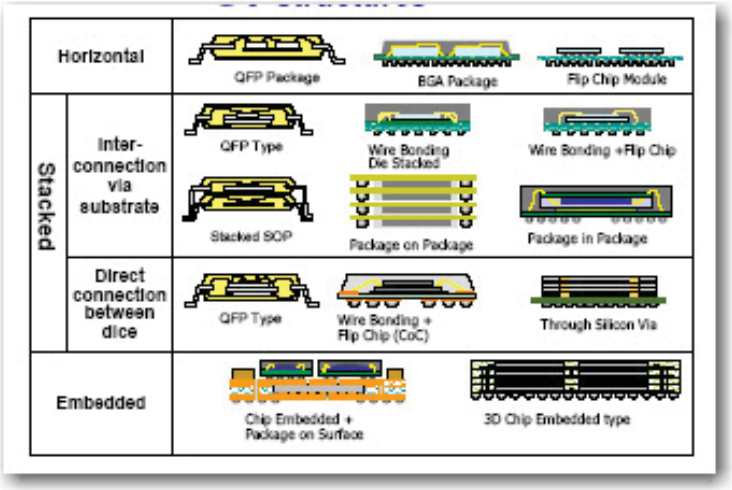


Figure 16: Examples of 3D SiP Integration Concepts

In the selection of an interconnect technology for 3D SiP, several factors are considered including technical or performance advantages. However, the most important factors are application (business) related. The packaging step in an electronic system production is close to the end point and therefore the packaging choices are closely related to the end product application environment. Many business-related decisions, such as time-to-market, development cost, supply chain, production volume, and product cost determine not only the integration level but also the wiring strategies.

The most commonly used wiring technology is the oldest one i.e. wire bonding. This is followed by the flip-chip technology. Figure 17 illustrates the use of FC and WB interconnects for BGA type packages. In the simple single chip package, the chip is FC bonded to the laminate substrate directly. Chips stacked in the package can be WB connected to the substrate. Or a combination of the previous two approaches can be used in PoP approaches. Recently, there has been a surge in the TSV technology to accommodate wiring among several chips. We will examine TSV technology in more detail below. Figure 18 shows and example where the use of TSV interposer in a package to integrate different types of chips [23].

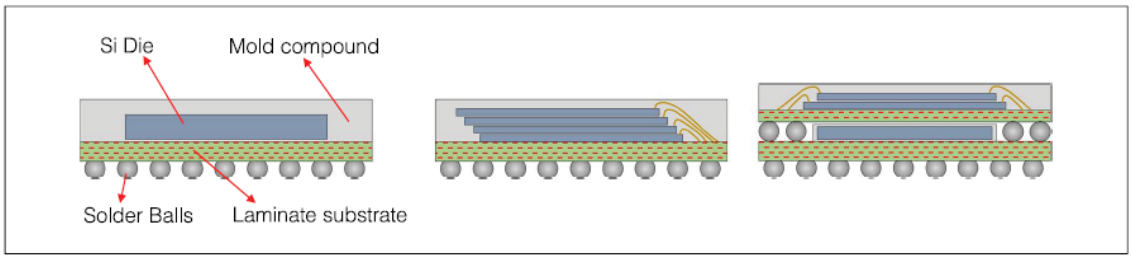


Figure 17: Wiring Choices in a BGA Type Package [23]

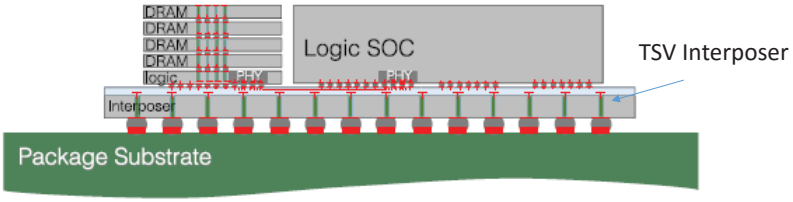


Figure 18: The use of TSV for 3D Integration in SiP Applications [23]

Another popular interconnect technology for in-package integration is the redistribution line (RDL) layer that acts like an integrated interposer between chips. RDL is also similar to the BEOL interconnect stack of Si ICs but employs organic (e.g. polyimide) insulator layers rather than SiO_2. It can be fabricated on Si interposers that may or may not have TSVs. RDLs are also an intrinsic part of the WLP – also known as chip scale packaging (CSP). We will see below in 2.5D SiP and 3D SiP sections some of the unique applications of the RDL technology.

Figure 19 is a useful top-level look at the package level integration approaches and the interconnect technologies that are applicable to each packaging concept.

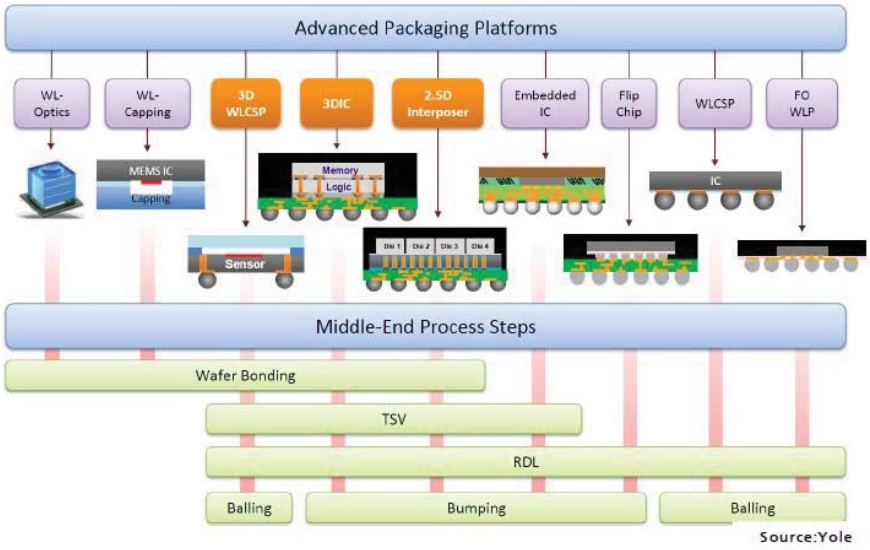


Figure 19: Top-level look at the Package-scale Integration Choices and the use of Various Interconnect Technologies Applicable to each Approach

A new and highly useful interconnect technology developed specifically for heterogeneous integration in packages is the “Cu lateral interconnects”. This approach allows the fabrication of thick Cu lines that can be defined by photolithography and fabricated directly on the assembled boards [24]. The top and side views of this type of interconnect are shown in Figure 20. Cu lateral interconnects resemble wire bond interconnects except that they are conformal and are fabricated in a batch process. This technology was successfully applied to chip heights up to 100 µm. For even thicker chips another conformal wiring technology that is a cross between TSV and Cu lateral interconnects was used [25].

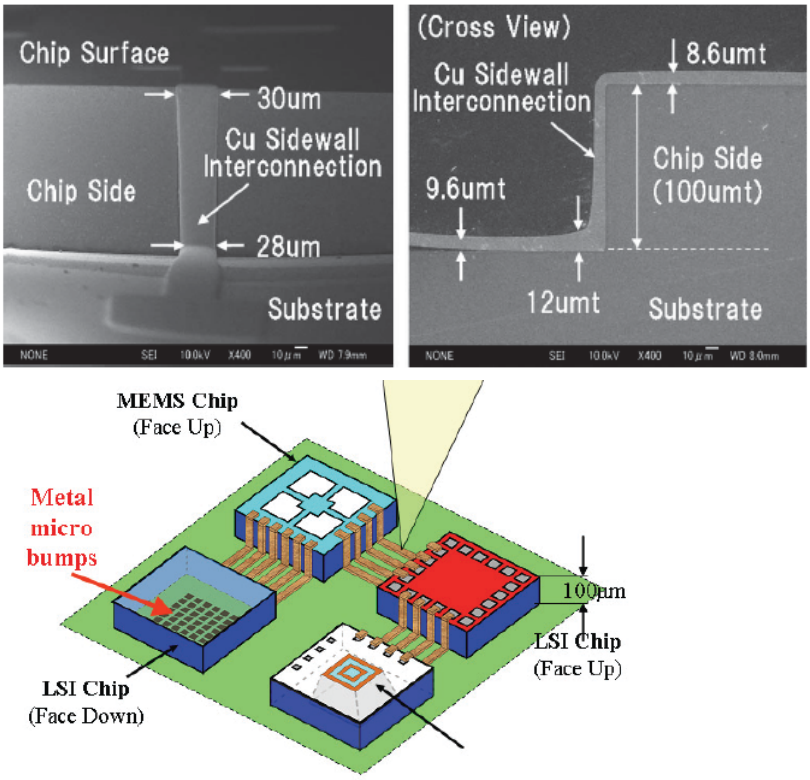


Figure 20: Cu Lateral Interconnects [24]

9. 2D INTEGRATION

9.1 2D SoC Heterogeneous Integration

Definition: 2D SoC heterogeneous integration refers to the integration of different device types on the same chip surface for the purpose of enhancing circuit performance. This type of integration is distinguished from ordinary integration (more same type devices per unit area) by the fact that different device types are integrated in close proximity. In its simplest form, different devices are fabricated on the same wafer surface, side-by-side, and a common FEOL and BEOL wiring is applied to all devices.

Advantages: This integration brings together within the same component different devices with performance advantages over other devices in some parameters. Close proximity of such different devices minimizes communication delays compared to circuits using multiple chips with segregated device types (see Figure 21).

Disadvantages: Additional device fabrication adds to the fabrication complexity. There is little or no overall chip area reduction except perhaps due to overall bond pad number reduction.

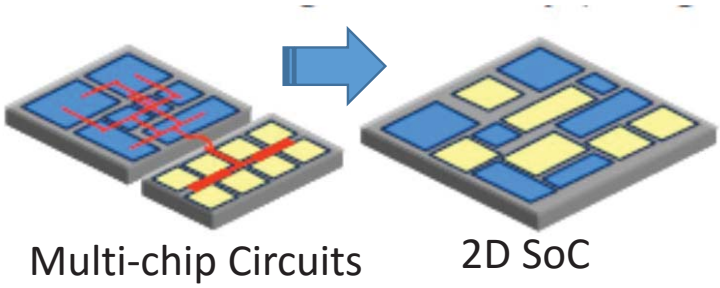


Figure 21: Circuits Fabricated with Different Device Types on Separate Chips are Combined on the Same Chip in 2D SoC [26]

9.1.1 FET-BT-HBT Integration

The semiconductor electronics started with the use of bipolar transistors (BTs) but field effect transistors (FETs) were later introduced for digital applications. As expected, initially each integrated circuit was confined to a single transistor type; BT for analog/RF applications and CMOS for digital and memory applications. The advantages of having both the BT and FET on the same wafer became obvious and BiCMOS technology was introduced for mixed signal and high speed digital circuits [5, 27]. This integration approach can be regarded as the first example of heterogeneous integration. Since all the integration is on the surface of the Si substrate, it is an example of 2D integration.

The Si-based BiCMOS technology has remained as the electronics technology's workhorse for decades and its supremacy in manufacturing maturity and circuit performance went unchallenged. However, in 1990's, the performance of BiCMOS circuits received a boost by the addition of a small amount of germanium (Ge) in to the base layer of the bipolar transistor. The new technology is referred to as SiGe BiCMOS [6, 7]. The use of SiGe in the base layer

technically makes the bipolar transistor a HBT but the base-emitter junction modification is only incidental and not critical to the device operation. The real performance advantage is gained by the graded base layer by the presence of a small percentage of Ge. Nevertheless, the SiGe BiCMOS technology is another example of 2D SoC integration because all active devices remain within the FEOL wiring.

Similar 2D integration approaches were followed in other semiconductor technologies. For example, FET was the first high speed transistor to be developed on GaAs substrates. It provided high RF gain and low noise operation. But soon the advantages of heterojunction bipolar transistor became obvious for power amplifier applications [18]. For complex circuit applications that require both high efficiency power and low noise amplifiers, such as microwave transceivers, the fabrication of both types of devices on the same wafer offered advantages. A range of processing innovations were introduced to make these different devices compatible [9, 12-16, 28, 29]. Some of these innovations include selective area re-growth [13, 29-35], stacked epitaxy layers [16, 36], and planar structures [14, 15]. Figure 22 shows and example of "stacked-layer" growth method of integrating HBT and high electron mobility transistor (HEMT) structures. This technique relies on the growth of all epitaxial layers on top of each other in a single growth run, and the selective removal of unwanted layers from the device areas. Although this approach makes the material preparation more complex and expensive, devices can be placed in close proximity of each other in circuits to make them work together with minimal interconnect delays.

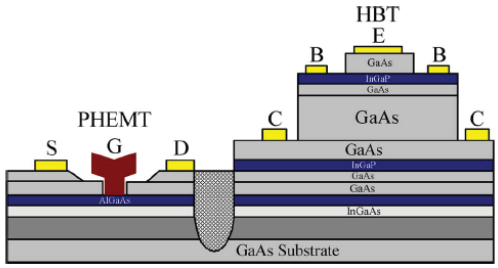


Figure 22: Integrated HBT and HEMT Structures on GaAs Substrates Prepared by Stacked Layer Growth Technique in MBE [37]

The applications of HBT-FET integration on GaAs or InP substrates can be found in digital memory [14], microwave and millimeter-wave front-end circuits [29-34], opto-electronic receivers [36], and handset power amplifiers (PAs) [15]. An ultra-wideband amplifier was fabricated using such an HBT-FET integration covering a frequency range of DC-43GHz, as shown in Figure 23 [37]. The use of 2D SoC heterogeneous integration is important in this case to achieve the ultra-wide bandwidth. Transistors are an intrinsic part of the artificial transmission lines of the distributed amplifier network and their exact placement in the circuit is critical, which makes integration necessary.

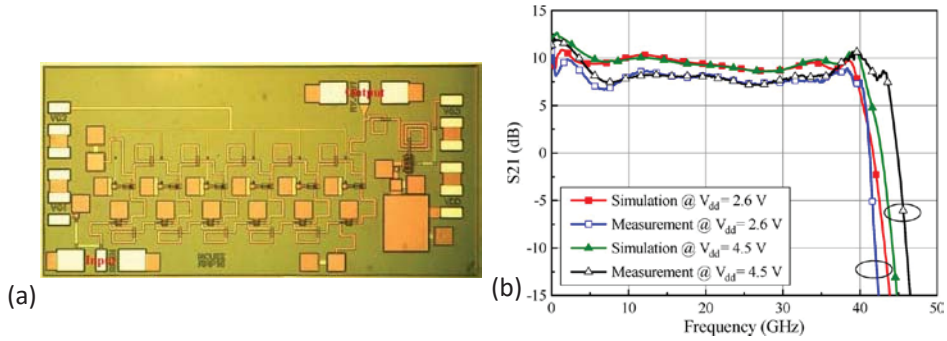


Figure 23: (a) Photograph of the Six-stage HEMT–HBT Cascode Distributed Amplifier Chip and (b) Simulated and Measured Small-signal Gains [37]

Other examples of 2D heterogeneous integration can be found in the InP semiconductor technology, which provides devices for high speed and opto-electronic applications. Heterogeneous integration opportunities therefore present themselves in HBT-PIN diode circuits used for high speed photo-receiver circuits [38-43]. It is important to note that many of the integration examples cited above were made possible by Department of Defense (DoD) investments including those from the Air Force Research Laboratory (AFRL) [14, 34], and the Defense Advanced Research Projects Agency (DARPA) [41].

9.1.2 Early GaAs-Si Integration

Before the advent of the SiGe HBT, there was a concentrated effort in early 1980's to integrate GaAs HBTs with Si CMOS or BiCMOS circuits. The objective of these efforts was twofold: 1) Add a high speed transistor or opto-electronic device (light emitting diode (LED) and laser) to increase functionality, 2) Provide a low cost substrate alternative for GaAs devices (the GaAs wafers at the time were expensive and had non-standard sizes). The second objective was more important than the first because the substrate development for any new material system is very expensive. The government investment alone in this field was not thought to be enough for the development of IC fabrication compatible, semi-insulating substrates. To be sure, highly doped irregular shape GaAs substrates were available for LED and laser applications. But such substrates were not suitable for microwave electronics.

DoD (Air Force Office of Scientific Research (AFOSR) and Army Research Office (ARO)) has invested heavily in the development of direct GaAs growth on Si using molecular beam epitaxy (MBE) [44-47]. Commercial development efforts were mostly undertaken in Japan and using growth methods that were more suitable for higher volume production (metal-organic chemical vapor deposition (MOCVD)) [48, 49]. Both the majority carrier and minority carrier devices fabricated on GaAs layers grown on Si wafers had performance levels comparable to those fabricated on GaAs substrates [50-52]. Based on this initial success, GaAs metal-semiconductor field-effect transistors (MESFETs) [28, 53, 54], HEMTs [55] and LEDs [56] were integrated with Si CMOS circuits.

9.1.3 DARPA DAHI Programs

The GaAs-on-Si approach was abandoned at the height of its technology success when the DARPA Monolithic Microwave Integrated Circuit (MMIC) program started [57]. Under this large DoD investment, high quality semi-insulating GaAs wafers were developed for integrated circuit applications. The success of this and other DoD programs [58] stunted the heterogeneous integration activities for over 2 decades. A new interest developed in heterogeneous integration when DARPA invested in the Compound Semiconductor Materials on Silicon (COSMOS) program [59], which is part of the Diverse Accessible Heterogeneous Integration (DAHI) program [60, 61]. The objective of DAHI is to bring together both high performance and highly integrated circuits for even higher level of performance. As shown in Figure 24 [62], Si metal-oxide semiconductor field-effect transistor (MOSFET) enjoys the highest level of integration while III-V based semiconductor technology have the highest Johnson Figure of Merit (FOM) [63] but low levels of integration. This DARPA program specifically addresses the integration of InP or GaN and Si CMOS technologies using approaches that are both 2D SoC and 2.5D SoC. We will first review 2D SoC approaches in this section.

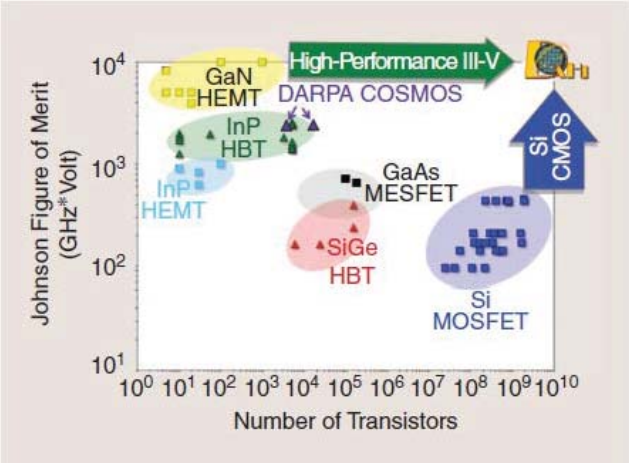


Figure 24: DARPA DAHI Programs Aim to Integrate High FOM Devices with Highly Integrated Si MOSFETs on the same Wafer [62]

Raytheon integrated high performance InP HBTs into a Si CMOS FEOL process at the intermediate wiring level. As shown in Figure 25a, an "engineered Si wafer" was used to accommodate the preparation of layers suitable for both InP and Si [64, 65]. Since high frequency InP circuits require Au-based metallization, and Si FEOL processing is incompatible with Au impurities, the Si FEOL process was completed in a Si facility and the remaining fabrication steps were undertaken in a III-V facility. Figure 25 also show the fabricated InP HBTs and CMOS circuits in close proximity as well as the final circuit demonstration in the form of a differential amplifier [66, 67].

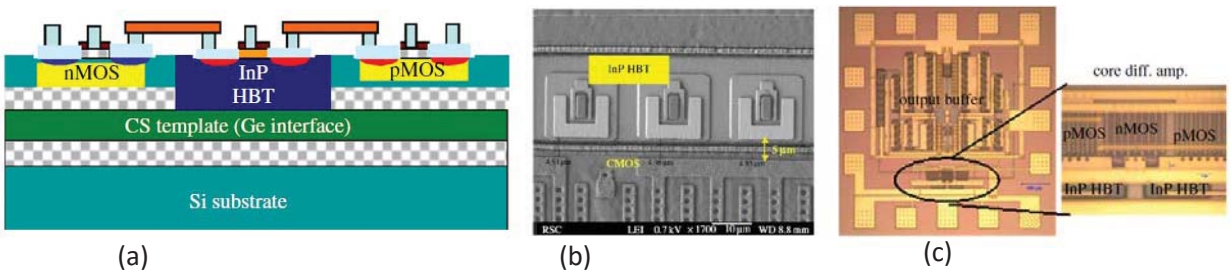


Figure 25: InP Integration with Si CMOS Devices
(Developed under DARPA COSMOS) (a) Cross-sectional drawing, (b) scanning electron microscopy (SEM) picture of integrated devices showing close proximity of different device types, and (c) final demonstration circuit [66].

A similar approach was used to integrate GaN HEMTs with Si CMOS under another DARPA program [68, 69]. Both selective area re-growth and stacked layer approaches were investigated. The selective area re-growth is the same as that developed under the DAHI program, whereas the stacked layer approach is similar to the 2D SoC GaAs FET-HBT integration discussed above. Figure 26 show the cross-sectional drawings of the 2 approaches. As a demonstration vehicle, an integrated GaN amplifier was demonstrated with pMOS control circuits [69]. The remaining DAHI program technical approaches fall under 2.5D SoC integration since the active devices are integrated at different planes (not directly on the wafer surface). We will review these approaches below in the 2.5D SoC section.

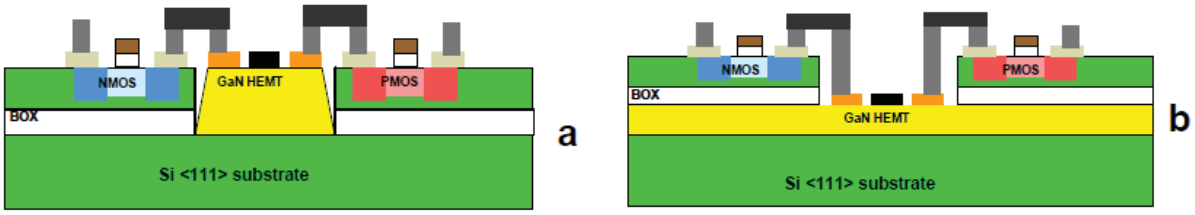


Figure 26: Cross-sectional Drawings of GaN-Si CMOS Integration Approaches [68]

9.1.4 Other Examples of 2D-SoC Heterogeneous Integration

MIT and Nitronex demonstrated a similar GaN-Si pMOS integration before the DARPA COSMOS program [70]. The integration approach taken in this case involved a layer transfer technology to prepare the Si and GaN layer stack on Si substrates. The Si devices were fabricated first. The GaN devices were then fabricated in the windows opened in the Si layer. A picture of the fabricated devices is shown in Figure 27. The minimum proximity of integrated devices was 4μm. No integrated circuit examples were published.

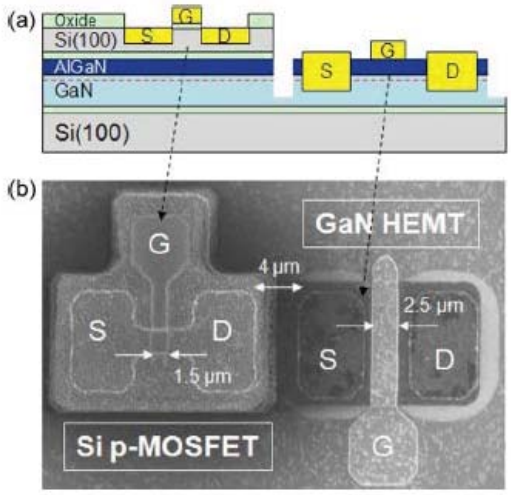


Figure 27: Cross-sectional Drawing and the Top View of Integrated p-MOSFET and GaN HEMT Transistor [70]

In an Army sponsored program, AlGaInP LEDs were integrated on CMOS compatible substrate using SOLES integration platform [71]. The cross-sectional drawing of the integration is shown in Figure 28.

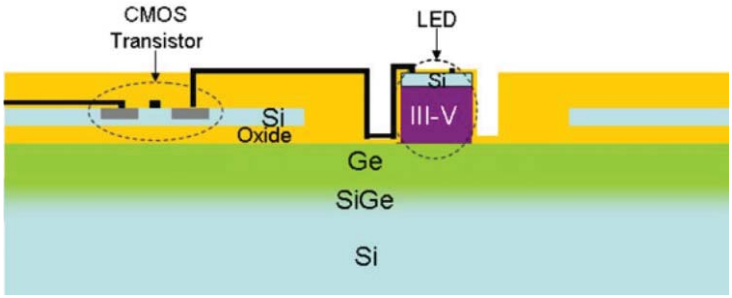


Figure 28: Cross-sectional Drawing of III-V-based (AlGaInP) LEDs with CMOS Transistors on Si Wafers [71]

Recently, indium gallium zinc oxide (IGZO) thin film transistors (TFTs) were integrated with GaN HEMTs on the same wafer at AFRL. The thin film transistor fabrication using metal-oxide semiconductors has a low thermal budget, which makes it highly compatible with the fabrication of other device types. As shown in Figure 29, IGZO TFT was fabricated side-by-side with GaN HEMT [72]. Both transistors were connected in a cascode mode. The use of IGZO in this application enables an enhancement mode operation while preserving the high breakdown voltage advantages of the GaN HEMT.

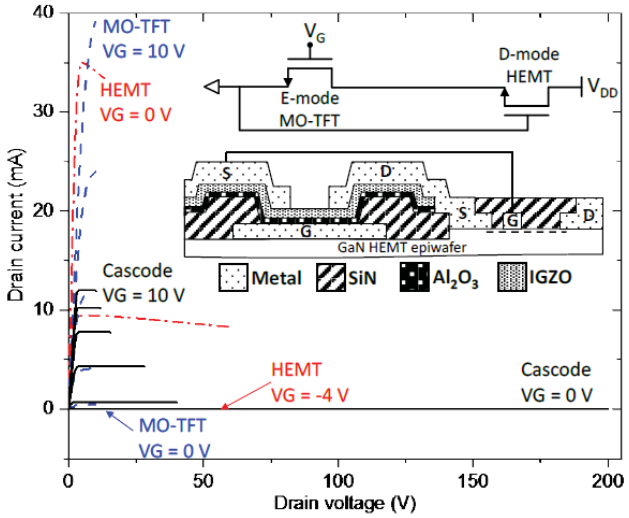


Figure 29: Heterogeneous Integration of IGZO TFT and GaN HEMT in a Cascade Configuration [72]

9.1.5 MEMS Integration with Si CMOS

MEMS are non-semiconductor type active devices used in switching, acoustics, resonators, filters and many other application [73-77]. Since microelectronics fabrication techniques can also be used for their fabrication, MEMS can be integrated with microwave or digital electronic components in systems. Although the fabrication technologies for electronics and MEMS are compatible, the sequencing of fabrication steps require special attention to constraints in thermal budgets, materials selections, and planarity requirements for fine line lithography. Many high volume MEMS applications are satisfied with direct integration on the same chip, while specialized or lower volume applications can be best addressed with in-package integration. Therefore, the MEMS-electronics integration spans over several categories. On-chip MEMS-electronics integration can fall under 2D SoC heterogeneous integration if some of the chip area is dedicated exclusively to the MEMS fabrication. Whereas MEMS circuits integrated on top of the electronics circuits can be regarded as 3D SoC heterogeneous integration technology. Multi-chip integration in packages can also range from 2D SiP to 3D SiP. We will examine each type of integration in the following sections.

The fabrication sequence of MEMS devices with respect to the fabrication of electronic devices is commonly used to differentiate the integration type for SoC approaches. There are 3 options: *"MEMS-first", "MEMS-middle",* and *"MEMS-last"* processes [77, 78]. When MEMS devices are fabricated first, as in accelerometer and pressure sensor applications, the area occupied by the MEMS devices are no longer suitable for electronic device fabrication. This approach, sometimes called monolithic integration, leads to 2D SoC type integration. MEMS-middle approach also leads to 2D SoC integration since MEMS devices are fabricated within the BEOL wiring stack and the Si below the MEMS device is removed. *MEMS-last* approach accommodate MEMS devices on the top of the BEOL stack and therefore provide overlap of heterogeneous device types. Those applications are consistent with the 3D SoC heterogeneous integration and will be reviewed separately.

In some high volume applications, MEMS devices are embedded inside the Si substrate [73, 79, 80]. For example, in the "MEMS first" process used for the lateral accelerometer shown in Figure 30, the free-moving parts of the MEMS are suspended over a cavity in the Si substrate [81]. There are no electronics circuits on or below the MEMS devices. Similarly, "*MEMS-middle"* process approach used for the low noise accelerometers, and millimeter-wave switches, shown in Figure 31, results in the removal Si directly below the moving parts [82, 83]. The MEMS components, in this case, are fabricated using the BEOL layers of the CMOS process. Suspended metal structures like these are important for the fabrication of high-Q inductors in RF circuits also [84].

The main disadvantage of 2D SoC integration as applied to MEMS-CMOS integration is the inefficient use of the Si surface area. The cost of bare Si wafer is not an issue for the fabrication of standalone MEMS devices. But, CMOS wafers are much more expensive and the area allocated to MEMS become an issue especially for high-end (i.e. small gate length transistors with multiple BEOL wiring) CMOS wafers [78].

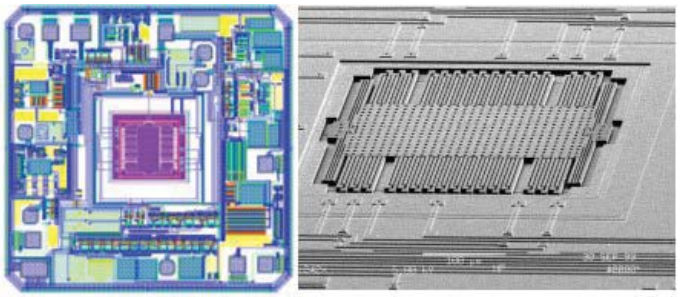


Figure 30: Top View and SEM Picture of Integrated Lateral Accelerometer Fabricated with "MEMS-first" process [81]

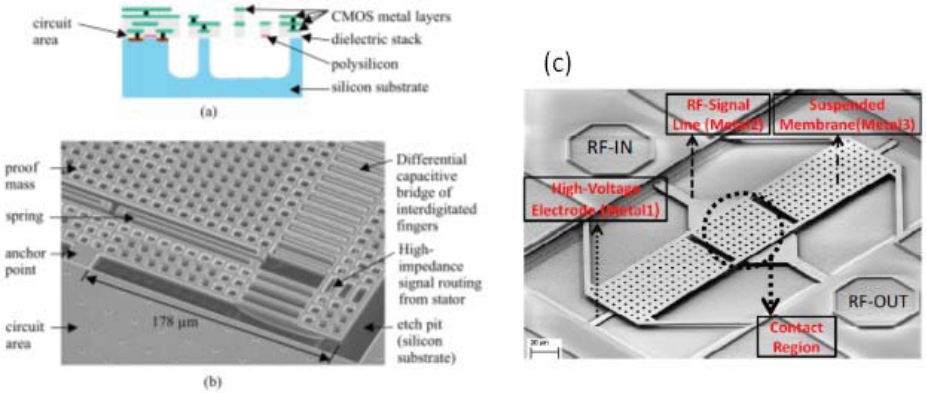


Figure 31: (a) and (b) Integrated Accelerometer [82] and (c) Millimeter-Wave Switch [83] Fabricated using "MEMS-middle" Process

9.1.5 MS Integration with RF Circuits

MEMS switches, resonators and high-Q passive components are finding some critical applications in RF circuits. The monolithic integration of MEMS on GaAs substrates is quite straightforward [76, 85, 86]. The same type of integration can be applied to GaAs MMIC circuits containing other active devices. In an effort sponsored by DARPA (Intelligent RF Front End (IRFFE)), Raytheon developed a 2-18 GHz tunable amplifier, as shown in Figure 32. This single stage amplifier was designed for 0.5 W wideband performance and 30%-55% power added efficiency [87]. The band switching was accomplished using 5 MEMS switches embedded in the input matching network and 5 MEMS switches in the output matching network. On-chip partial packaging was developed to protect MEMS switches. In another effort, Rockwell integrated MEMS switches with pHEMT-based LNAs for 9 GHz applications [88]. Because the MEMS and active GaAs devices are on the same chip but do not share the same surface area, this type of integration falls under 2D SoC integration.

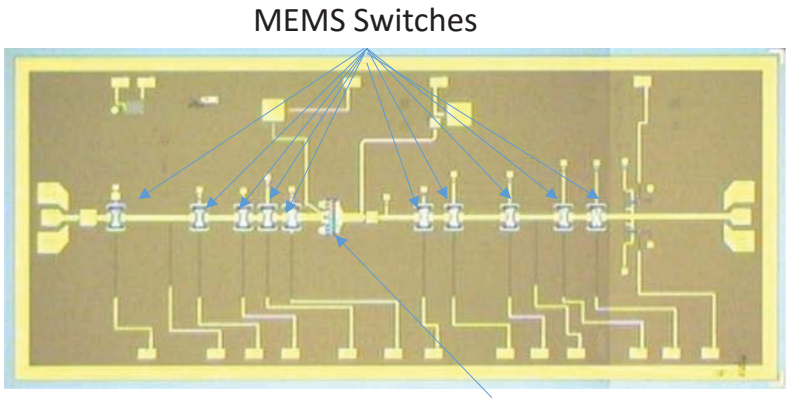


Figure 32: MEMS Integration with pHEMT on GaAs Substrate for 2-18 GHz Wideband Amplifier Applications [87]

Under the AFRL Adaptable RF Criteria-High Performance Cell Array (ARCHIPELA) program, Raytheon integrated RF MEMS capacitive switches with GaN transistors for re-configurable RF amplifiers [66]. In this demonstration, the MEMS switches were placed on top of the passivation layer for the GaN transistor, and therefore at a plane higher than the GaN surface. In that respect, it may be regarded as an example of 2.5D SoC. However, apart from demonstration purposes, the placement of MEMS at that level does not serve any other purpose. This example may also be classified in 2D SoC integration because of low density integration of the integrated chip (see Figure 33). For demonstration purposes, the GaN transistor was fabricated on Si substrates for the implication that in the future both GaN transistor and MEMS switches may be fabricated on CMOS wafers.

In other examples, GaN micromechanical resonators were integrated with GaN HEMTs on Si substrates for timing applications [89] and micromechanical resonators were integrated with GaN transistors [90, 91] for oscillator applications.

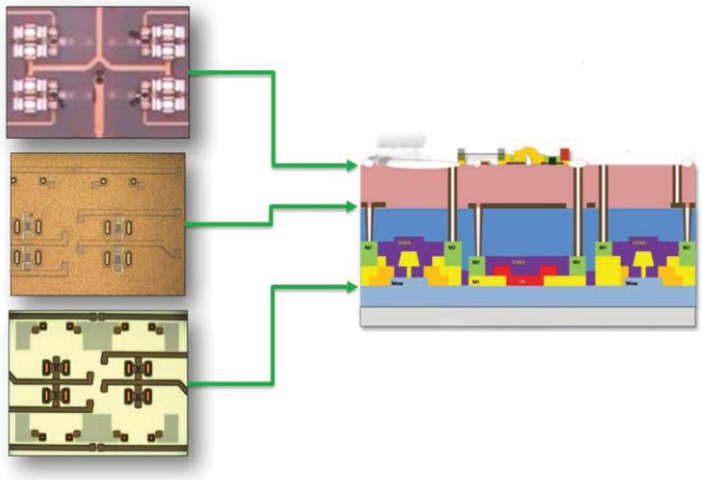


Figure 33: Top Views and the Cross-sectional Drawings of GaN-MEMS Integration on Si Substrate [66]

9.2 2D SiP Heterogeneous Integration

Definition: 2D SiP heterogeneous integration refers to the integration of multiple chips with different device technologies in a simple package. Chips are placed side-by-side in the package on the same surface. Connections between the chips are accommodated in the laminate substrates running under the chips. This type of integration is also called "multi-chip packaging" for digital applications and "microwave modules" in analog applications.

Advantages: Compared to single chip packages, the use of multi-chip packages provide performance advantages by combining specialized chips in a single package. Product speed can be improved by connecting chips to each other with shorter wires. Compared to stacked chip packages, this approach provides better thermal management. Microwave multi-chip packages are heavily used in military and miniature commercial microwave module applications.

Disadvantages: The package size and the number of I/O's increase with the number of chips integrated. This is an old technology for digital applications and has been superseded by more advanced packaging technologies. The microwave version still has many current uses.

2D SiP type integration is one of the oldest and the most common type of in-package integration technique applied to digital, mixed signal and microwave electronic systems. It is not always possible or economical to integrate all system functions on the same chip. Separately fabricated chips can be integrated side-by-side in a package using interconnects running under the chips. It is easier to appreciate the need for such an integration when each chip is manufactured using different device technologies (for performance advantage). The need for multi-chip packaging is less obvious for digital systems that are implemented by the same Si digital technology except for those cases where time-to-market considerations dictate the use of existing chips. All-Si but different technology node components can be integrated for performance advantage.

The key to understanding the need for multi-chip modules for digital electronic systems is to consider the IP maturity level of technology components as they each pass through the process nodes, as shown in Figure 34. The availability of each IP for manufacturing at a given process node is different. For example, state-of-the art logic circuits can be implemented at the 14nm node but embedded flash or DRAM memory is not yet available at this node. Similarly, image sensors are only available at the 90nm node. If single chip solution is attempted for all functions, the integrated chip must be fabricated at the lowest common denominator process node. This would degrade the overall performance of the system due to slower operation of many components. The heterogeneous integration in package for digital systems can be represented as the integration of chips fabricated using different process nodes, as indicated in Figure 35.

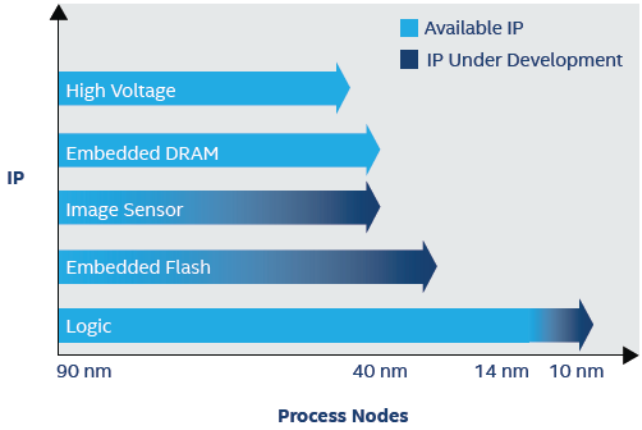


Figure 34: IP Maturity vs. Process Node Example

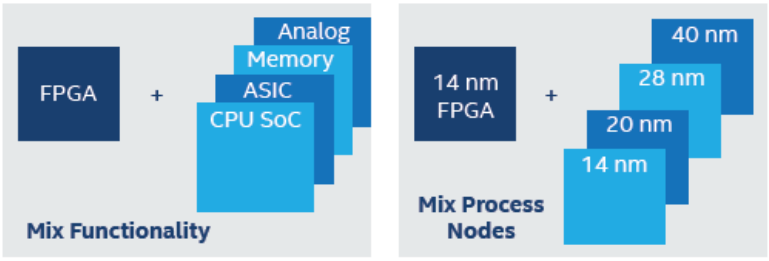


Figure 35: 2D SiP Integration for Digital Electronic Systems can include Mixed Functions where each Function Chip is fabricated at a Different Process Node for Optimum Performance

A similar technology node mismatch often produces complications in the integration of circuits made on advanced semiconductors such as GaAs, InP and GaN. The gate length of FETs or the emitter width of HBTs are the technology nodes for these technologies. Higher speed circuits require shorter gate lengths. Higher power devices can be implemented using wider gate lengths. The epitaxial layer designs are modified according to each technology node. Also, depending on the application (power amplifier, low noise amplifier, digital etc.), different epitaxial layer stacks are preferred for the best performance. When attempting to integrate different circuit functions on the same wafer, a problem similar to technology node mixing in Si is observed.

9.2.1 Managing I/O complexity

The 2D SiP was originally suggested to overcome the I/O bond pad number increase problem in VLSI technology as the chip sizes decreased while the circuit complexity increased [92]. The number of bond pads that can be placed in a single row around the chip did not keep up with the increase in I/O connections. By breaking up the chip into functional blocks and placing them in the same package allows the fabrication of more I/O pads and their efficient connection to the package terminals. For example, the 3-chip module shown in Figure 36 contains a central processing unit (CPU), memory management unit (MMU), and math accelerator unit (MAU). Most of the connection between chips were in the laminate substrate under the chips [93]. Each chip performance was optimized by the selection of the best technology for that function.

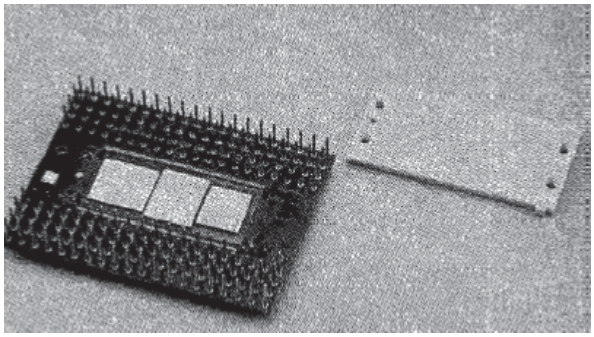


Figure 36: 160 I/O pin grid array package for 3-chip assembly [93]

9.2.2 Microwave Modules

The 2D SiP heterogeneous integration is heavily used in military and commercial microwave systems, where they are called microwave modules [94-98]. Such modules are usually hermetically sealed ceramic packages that contain multiple chips including MMICs, digital control circuits, and power management circuits. The chips are mostly mounted on heatsinks on a 2D surface and connections between chips are made using microwave transmission lines and bond wires (or solder bumps). Because the microwave signal transmission between chips is an essential part of the package design, and such transmission lines can be long (they are also a part of the impedance matching circuits), the packages (modules) tend to be larger than those for digital system applications. For example, the module shown in Figure 37, 3 different GaN chips were packaged in a package measuring 45mm x 15.5mm. Low temperature co-fired ceramic (LTCC) technology is the preferred approach for these modules [99, 100].

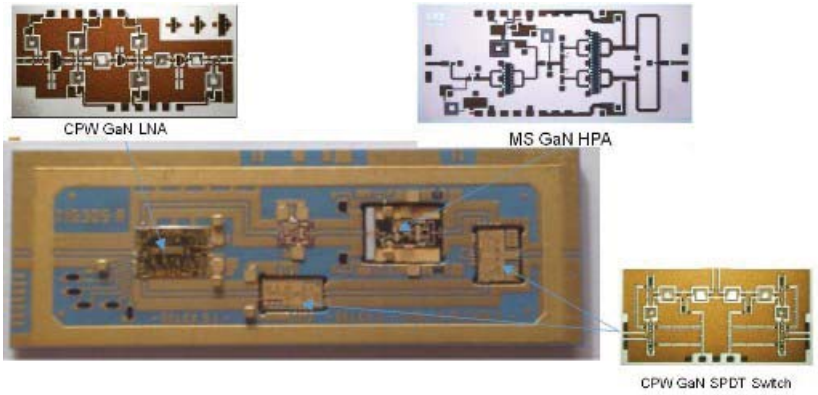


Figure 37: Photograph of the Fully Integrated GaN-based Module Fabricated by the Italian Company SELEX [99]

9.2.3 Highly Miniaturized RF Front End Modules (FEMs)

FEMs also make use 2D SiP technology for commercial applications. The need for in-package integration for this application comes from the fact that different device technologies are needed to address the performance (efficiency being the most important parameter for hand-held systems) improvements. Integration of these diverse device technologies are most efficiently addressed using SiP approaches on multi-layer laminate (organic) substrates in plastic packages [101, 102]. In a typical RF FEM for cell phone applications, power amplifiers based on GaAs HBT technology, low noise amplifiers (LNAs) and switches based on GaAs HEMT technology, filters based on surface acoustic devices (SAWs), control circuits based on Si CMOS and various surface mount high-Q discrete passive components are integrated on the top surface of a laminate substrate. Connections between chips are provided on the top surface as well as within the laminate substrate. Power amplifiers usually require thermal heat sinks, which are implemented in the form of "thermal vias", as shown in Figure 38a. Figure 38b and c show typical RF FEMs [103, 104]. The 2D SiP integration provides design flexibility for the highly competitive commercial cell phone applications where RF FEMs are addressing multiple carrier frequencies and modulation schemes [105, 106]. This approach can be scaled for higher frequencies including mm-wave bands [101].

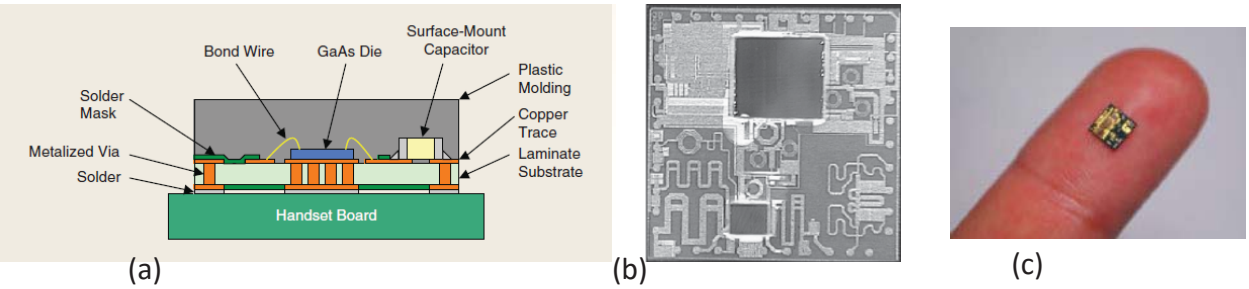


Figure 38: RF FEMs Implemented in Organic Laminate Substrates in Plastic Packages

10. 2.5D INTEGRATION

10.1 2.5D SoC Heterogeneous Integration

Definition: 2.5 D SoC heterogeneous integration is similar to 2D SoC and also refers to the integration of different devices on the same chip for the purpose of enhancing circuit performance. It is distinguished from 2D SoC integration by the fact that the integration plane is higher than the substrate surface. It is distinguished from the 3D SoC by the fact that the integrated device footprints either do not overlap or the integration density is so low that such an overlap does not contribute significantly to chip compactness (functional density). This integration method is useful for bringing closer together diverse functions using electronic, optical and mechanical devices.

Advantages: This type of integration brings together within the same circuit different devices to increase functional integration level. Compactness achieved with this type of integration is an enabler for large multi-sensor arrays.

Disadvantages: Device fabrication is complicated due to material incompatibilities. Transfer of processing technologies between foundries can be challenging due to highly specialized nature of integrated technologies. The circuit integration density is low, especially for microwave applications.

10.1.1 DARPA DAHI Programs

Most of the DARPA DAHI program approaches fall under this heterogeneous integration category. For example, InP chips were integrated on CMOS circuits using flip chip bonding [59, 60, 107, 108], as shown in Figure 39. The integration method used here is a sub-category of the F2F wafer stacking approach. Instead of stacking full wafers (wafer-to-wafer (W2W)), here only fully processed chips are stacked on CMOS wafers (D2W). Because of the stacking of chips rather than wafers, TSVs are not employed for package interface. For the implementation of this integration, a fully processed 0.18μm CMOS wafer with 6 BEOL interconnect levels was further processed in a non-Si facility to produce small diameter Au bumps at the top surface. A separately processed InP wafer containing 0.45μm emitter InP circuits was thinned to 55μm, attached to a glass carrier and broken into separate chips (with glass carrier attached). InP chips were then flip chip bonded on the CMOS circuit and the glass carrier was removed. The integration resulted in 1.33-Gsps digital-to-analog converter (DAC) with 10 dB improvement in the noise performance compared to single technology approaches [108].

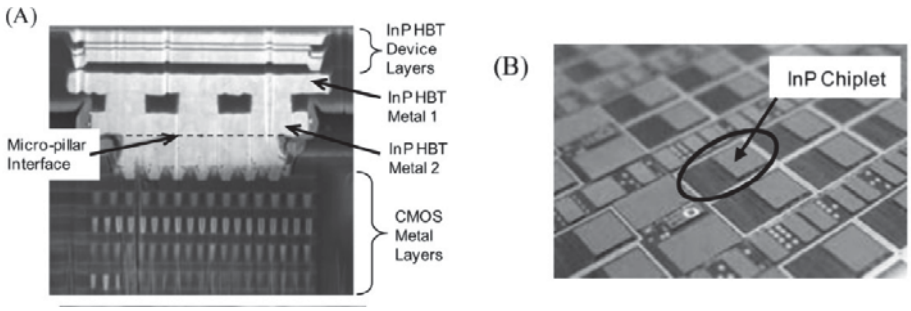


Figure 39: Cross-sectional and Top View of 3D Integrated InP/CMOS Circuits [108]

One of the advantages of stacking individual chips rather than full wafers is the flexibility of integrating several technologies on the same interface surface simultaneously. This was demonstrated by integrating InP HBT and GaN HEMT circuits on a CMOS wafer [109]. The InP chip was stacked using F2F approach, whereas GaN HEMT chip was stacked using B2F approach. The Q-band integrated circuit shown in Figure 40 demonstrated a 2 GHz voltage controlled oscillator (VCO) tuning range and 15dB amplifier gain, with a total power consumption of 1.68W.

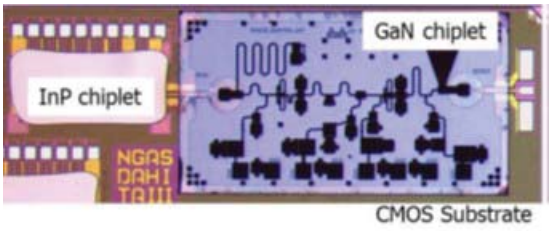


Figure 40: Photo of the DAHI Q-band VCO-amplifier Chain [109]

The next 2 examples of DAHI COSMOS programs used 3D SoC-like integration approaches to integrate InP HBTs with Si CMOS circuits. They are classified under 2.5D SoC for now because the integration density is too low compared to the functional densities expected of 3D SoC integration. Instead of bonding the completely processed InP chips on CMOS circuits, this approach only transferred the InP epitaxial layers to the top of the fully processed CMOS wafer, using F2F wafer bonding, as shown in Figure 41 [110]. The InP transistor is then fabricated and connected to the CMOS circuit at the global wiring level. In both of these approaches, there is no common BEOL wiring fabrication. The heterogeneous components are simply attached to each other at the global wiring level. In comparison, the previously examined InP/CMOS integration was at the FEOL level and contained common BEOL wiring [67]. Impressive W-band LNA performance was demonstrated using this heterogeneous integration platform [111].

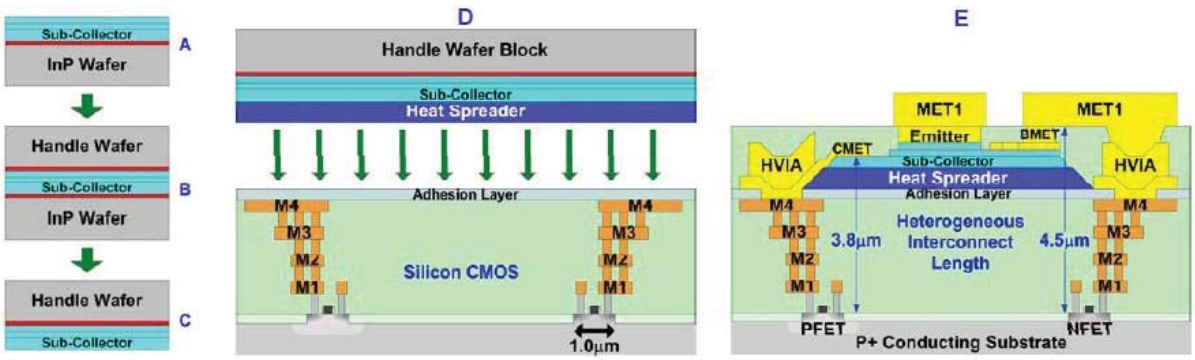


Figure 41: InP/Si CMOS 3D SoC Integration using Epitaxial Layer Transfer and Subsequent Processing to Connect InP Transistors to CMOS Circuits at the Global Wiring Level [110]

Heterogeneous integration of InP HBT and Si CMOS circuits was undertaken at Teledyne [112, 113] using a wafer stacking and substrate removal approach. A partially processed Si CMOS wafer was mounted on a processed InP wafer F2F using low temperature bonding. In this "CMOS-on-the-top" approach, the excess heat is removed from the InP side, while the circuit I/O connections were made from the Si wafer side. A cross-sectional view of the integrated chip is shown in Figure 42. The interconnect pitch at the plane of interface was 5μm, which means only global wiring level integration was achieved.

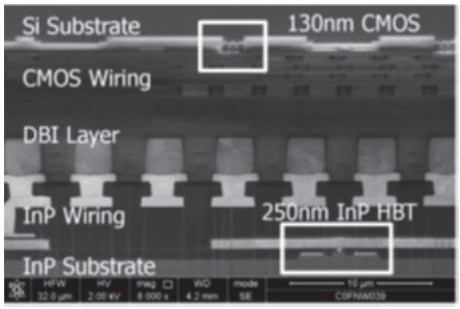


Figure 42: FIB/SEM Cross-section of 3D Integration of 250nm InP HBTs and 130nm CMOS [113]

10.1.2 Commercial 2.5D-SoC Heterogeneous Integration Examples

InP DHBT circuits were integrated with Si BiCMOS circuits also by Ferdinand-Braun-Institut (FBH), Berlin [114-117]. Fully processed wafers (Si BiCMOS with 6 layer BEOL and InP wafer with only M1 level metallization) were bonded F2F at elevated temperatures and in vacuum. After bonding the InP substrate was removed and interconnect metallization was carried out to complete the integration of InP and BiCMOS circuits. This approach is "InP-on-top" type and the heat removal is through the Si wafer. A cross-sectional drawing of the integrated chip is shown in Figure 43.

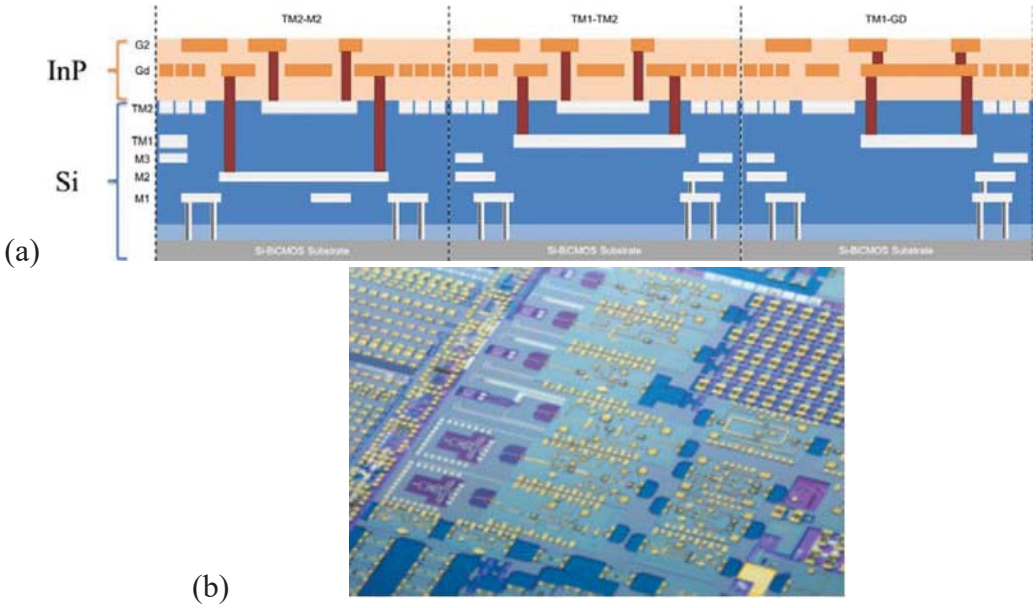


Figure 43: (a) Cross-sectional Drawing of the Integrated InP-BiCMOS Chip and (b) Top View of the mm-wave Power Source Circuits [115, 117]

It was shown that the performance of 0.8μm emitter InP HBTs maintained their high frequency performance levels (f_T and f_{max} values of 400 and 350 GHz, respectively) after the integration [118]. The advantages of this integration was demonstrated by a series of circuit designs for generating millimeter-wave signal sources. In these designs, the BiCMOS devices were used to fabricate VCO circuits to generate the fundamental frequency signal, and the InP devices were used to double [119], triple [120] or quadruple this signal to achieve up to 330GHz [121].

A closer examination of Figure 43 will show that this integration approach, as it is the case for most of DARPA DAHI program developed integration techniques, only accomplishes a close proximity placement of circuits using heterogeneous technologies. There is no stacking of devices or impedance matching circuits. Nor, is there any substantial sharing of circuits between heterogeneous devices. The vertical integration only takes place in the transition regions to make sure impedance mismatches are minimized. It appears that the circuit transitions are at the 50-ohm impedance level, which means that each circuit is almost a stand-alone circuit. It is not difficult to imagine accomplishing this level of integration by simply flip-chip bonding separate MMIC circuits on an interposer substrate.

10.2 2.5D SiP Heterogeneous Integration

Definition: 2.5 D SiP heterogeneous integration specifically refers to the integration of chips in packages using TSV interposer layers or embedded bridge circuits. Unlike the 2D SiP integration, where chips are directly mounted on the package substrate, in 2.5D SiP an interposer or embedded bridging layers assist chip connection to the package.

Advantages: High speed connections between chips enable distribution of specialized functions to different chips without degrading performance. The modular approach offers system design flexibility and manufacturing cost reduction. This technology is nearing maturation and high volume products are are already in the market.

Disadvantages: A separately manufactured TSV interposer layer is required for each design.

10.2.1 TSV Interposer Approach

The main object of the 2.5D SiP heterogeneous integration is to increase the interconnect density and interconnect speed (bandwidth) between side-by-side placed chips in a package. This specific objective is achieved using an interposer layer with fine line interconnects and TSVs. Although it is possible to route the signal using interconnects located within the package substrate, the interconnect density is too low for wide bandwidth connections. A simple solution is to place an interposer layer under the chips with fine lines for chip-to-chip interconnects. The signal is passed on to the package substrate using TSVs. Si-based interposers are the most common type for this application for 2 reasons: 1) the thermal conductivity of Si is high enough to act as a heat spreader, 2) fine-line interconnects and vias can be fabricated on such interposers in the same Si facility where electronic components are manufactured.

The idea of making vias through the substrate, as used in TSVs, is as old as the transistor itself. A patent issued to Shockley, the co-inventor of the transistor, describes metallized holes through semiconductors [122]. A drawing of this patent claim is shown in Figure 44.

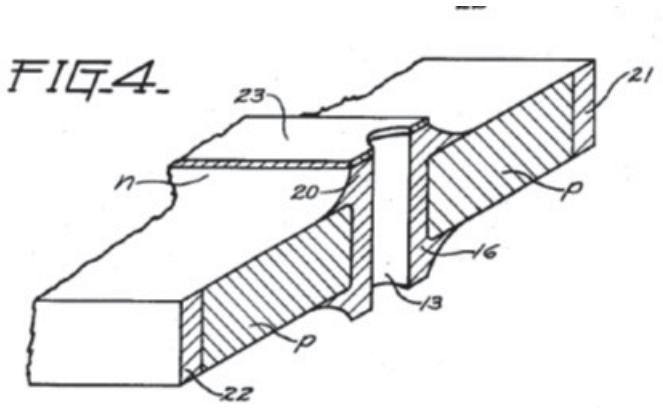


Figure 44: Metallized Vias through the Semiconductor as Described by Shockley is one of the Earliest Versions of TSV [122]

The main purpose of the interposer in 2.5D SiP integration is the fine-pitch short interconnects between flip-chip mounted chips, as illustrated in Figure 45. Chip-to-TSV contacts are accomplished using micro-bumps, which are much smaller than the bumps used in 2D SiP technology. TSV interposer has micro-bumps on the top surface and regular size bumps on the bottom surface. In effect, TSV interposer distributes the chip signal to package from fine pitch to large pitch (pitch size amplification). One of the first examples of 2.5D SiP integration by Xilinx involved the integration of 3 chips; an 8 x 28Gb/s transceiver IC and two FPGA ICs known as Super Logic Regions (SLR) on a passive silicon interposer [19].

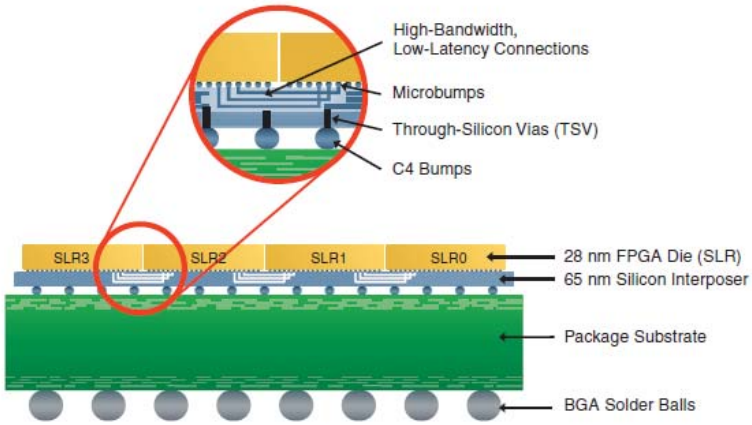


Figure 45: 2.5D SiP Integration using TSV and Micro-bumps [19]

The TSV interposer is a key enabler in SiP type heterogeneous integration. They are becoming increasingly common in more sophisticated 3D SiP integration [123, 124], as we will see in the sections below. They are deceivingly simple constructions, as depicted in cross-sectional drawings. But, in reality their fabrication requires state-of-the art fabrication facilities (e.g. GlobalFoundries, TSMC). An example of a TSV interposer is shown in Figure 46. The interconnect complexity and density are comparable to a Si CMOS BEOL process and superior to those achievable in package substrates. This is the reason why they are used for wide bandwith interconnects instead of relying on low density in-package interconnects.

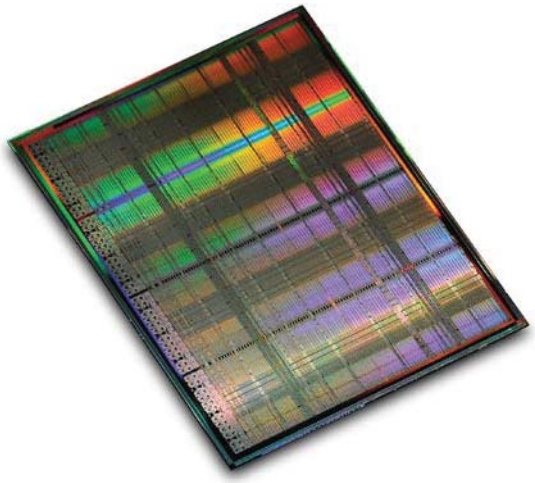


Figure 46: Example of a Passive TSV Interposer Fabricated by GlobalFoundries

Another example of a successful 2.5D SiP heterogeneous integration is the AMD Radeon™ Fury product [125]. This integration includes 4 stacks of high bandwidth memory (HBM) at 512GB/s of bandwidth and a total storage capacity of 4GB, a large graphics processing unit (GPU) (596mm^2) is in the center with 2 DRAM stacks on each die. Both the GPU and HBM connect to the interposer using 190,000 micro-bumps with 45µm pitch. The interposer is mounted to the substrate via 25,000 C4 bumps with a pitch of 165 µm. A cross-sectional view of the chip

connection to the interposer is shown in Figure 47a. Figure 47 b shows the top view of the packaged product.

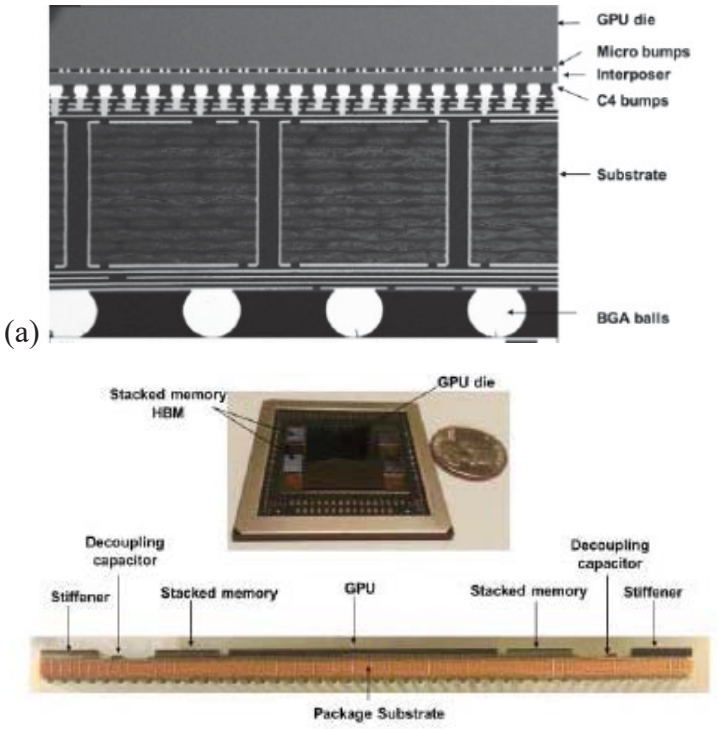


Figure 47: 2.5D SiP Integration Example (a) Cross-sectional View and (b) AMD Radeon™ Fury Product using 2.5D SiP Integration [125]

10.2.2 Embedded Multi-die Interconnect Bridge (EMIB)

One of the critical issues in 2.5D SiP heterogeneous integration is the actual size of the Si-based interposer, which can be nearly an inch square. In the example shown above, all active chips were integrated on a single interposer. Apart from the manufacturing cost, such large interposers have a problem with chip warpage. As a solution to these issues, Intel developed embedded multi-die interconnect bridge (EMIB) technology [126, 127]. Small "bridge" chips are embedded inside the package substrate such that they are only used as a bridging layer between heterogeneously integrated chips, as shown in Figure 48. This approach requires no TSVs through the bridge chip, therefore the use a large interposer is completely avoided. Integrated chip makes connection to the EMIB die using micro-bumps, whereas the same chip makes direct contact to interconnect lines in the package using larger C4 bumps. This approach has provided significant flexibility the integration process and increased production yields. It is now in high volume production, Intel also offers custom foundry service for its production. (http://www.intel.com/content/www/us/en/foundry/manufacturing-services.html)

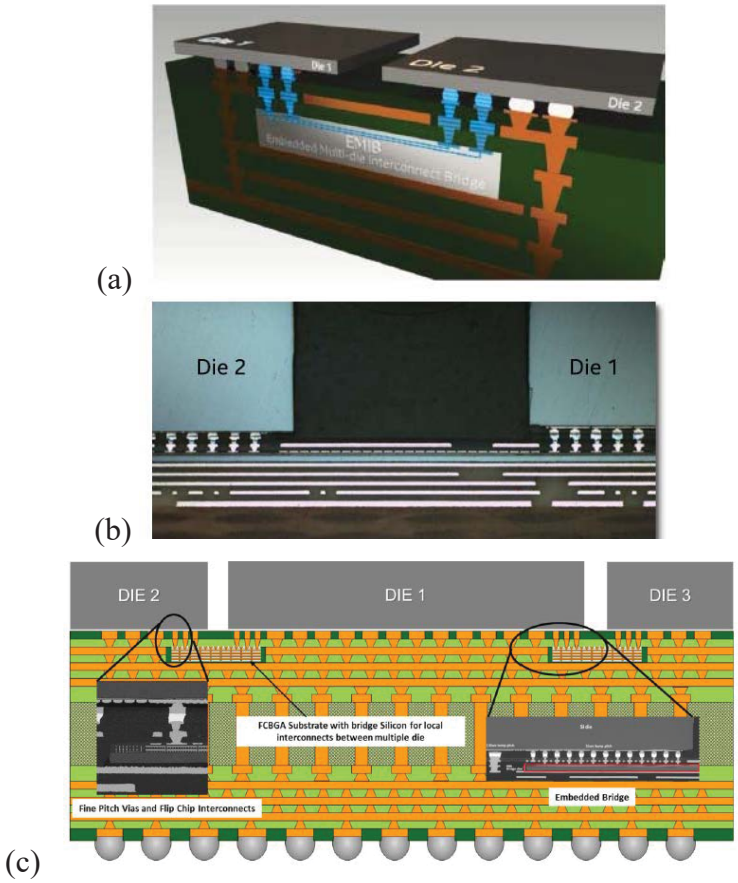


Figure 48: EMIB Technology for 2.5D SiP Heterogeneous Integration [19]

The EMIB technology is a significant breakthrough in high performance chip packaging. As a part of a SiP integration strategy, EMIB allows a great deal of flexibility in system design by allowing easy integration of heterogeneous components in the same package without giving up the performance edge. An example product that uses this technology is the Altera Stratix® 10, which integrates high performance FPGA processors with high bandwidth memory chips for wide bandwidth applications. In the schematic representation of this product, shown in Figure 49, the stacked memory chips (double date rate (DDR)) are connected to the "Core Fabric" using EMIB high speed interconnects. The memory bandwidth is increased by 10X with this type of integration compared to previous packaging types. Using 4 memory tiles with 256 Gbps aggregate bandwidth in Stratix® 10 MX product, for example, the memory bandwidth is increased to 1TBps in a single package. Also, integrated in this package are 4 transceiver chips, which are connected directly to the Core Fabric using EMIB interconnects.

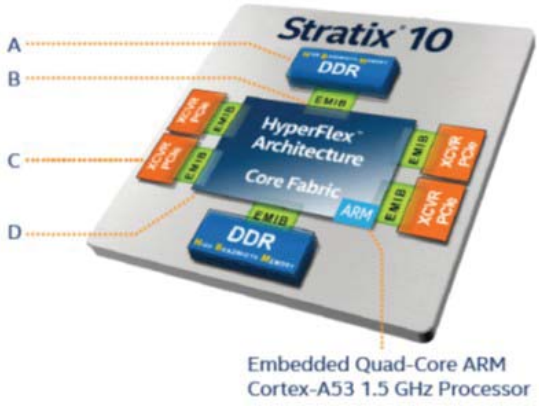


Figure 49: Intel Stratix® 10 MX Device

Compared to the TSV interposer type interconnects, EMIB approach offers several advantages. The first is that the EMIB chips are only used where they are needed in contrast to TSV interposers, which cover the entire package area. The second is the simplified fabrication, as illustrated in Figure 50. The TSV interposer fabrication involves the fabrication of TSVs, whereas EMIB does not use TSVs. A two-step assembly is needed for TSV interposer, whereas EMIB assembly is simpler.

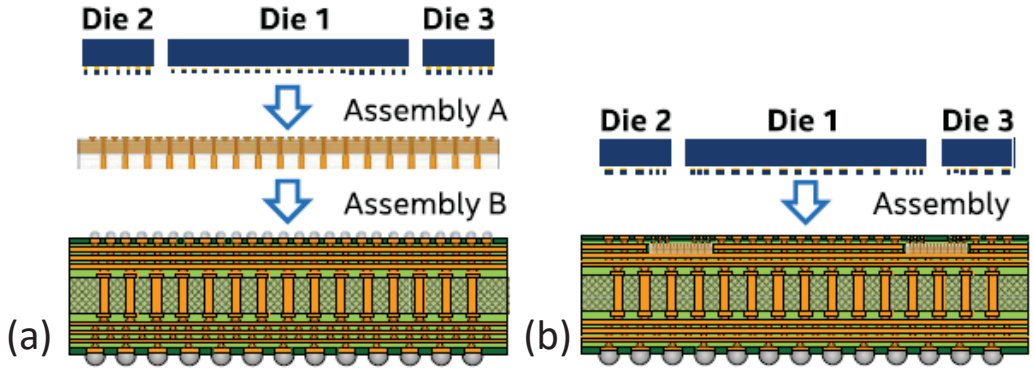


Figure 50: Comparison of 2.5D SiP Technologies: (a) Using TSV Interposer and (b) Using EMIB Technology

11. 3D INTEGRATION

11.1 3D SoC Heterogeneous Integration

Definition: 3D SoC heterogeneous integration refers to the integration of separately processed chips with different technology types on a common chip-level platform. This category also includes circuits with different device types fabricated on top of BEOL stack. It is distinguished from 2.5D SoC by the fact that footprints of devices overlap heavily and the integration density is high. The integration interface is mostly BEOL and the multi-chip integration can be facilitated using the top and the bottom surfaces of chips. This type of integration provides the highest functional density.

Advantages: The use of the 3rd dimension increases circuit density and improves communication between chips due to short length interconnects. Each chip can be optimized in performance separately before integration. Highly diverse technologies can be integrated and packaged in chip-level compact assemblies.

Disadvantages: Efficient heat dissipation is a challenge. Integrated heat sinks and cooling channels may need to be part of the integration strategy. The fabrication of diverse technology components and their integration may require close coordination between multiple foundries and add to the supply chain management issues.

The 3D SoC integration is the current technology frontier in highly compact and multi-functional chips. The final scope of this technology is not yet determined. New integration methods are being added to this category every year. Initially, 3D SoC meant the chip-level integration of diverse technology chips by connecting BEOL wiring of each chip to each other. Later, the integration was attempted at the earlier levels of the BEOL process so that the integrated final chip had global wiring levels that are common to all integrated chips. This integration technology merges with the 3D SiP at the WLP node (discussed in detail below).

It is easy to accept that the use of the 3rd dimension in chip integration reduces chip footprint, as illustrated in Figure 51. Instead of forcing the fabrication of diverse technology components on the same chip surface, as it was done in 2D and 2.5D SoC integration, and as carried out under some of the DARPA DAHI programs, diverse technology chips can be fabricated on their native substrates for optimum performance, and the final chips can be vertically integrated later. Additional functions can be simply added using other chips in the same way. Although this approach adds several levels of complexity in managing inter-chip wiring and thermal management, the final product can be very compact and can maintain the combined high performances of each chip.

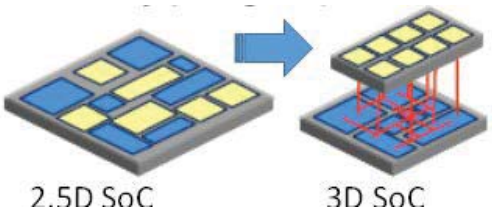


Figure 51: Illustration of Chip Footprint Reduction using 3D SoC Integration

Unlike the 2.5D SoC integration, where devices made from alternative materials are integrated at the FEOL or MEOL (middle end of line) process stage, and a common BEOL wiring is applied to all device types, 3D SoC heterogeneous integration can take place at different layers of the BEOL process (see Figure 15). The integration interface defines the type of 3D integration used. A part of Figure 15 was re-produced below (Figure 52) to guide our discussions on the types of 3D SoC integration approaches followed for military and commercial applications. For clarity, the definitions of integration types are as follows:

3D-SIC: 3-dimensional system stacked ICs.
3D-SOC: 3-dimensional system on chip.
3D-IC: 3-dimesional integrated circuit.

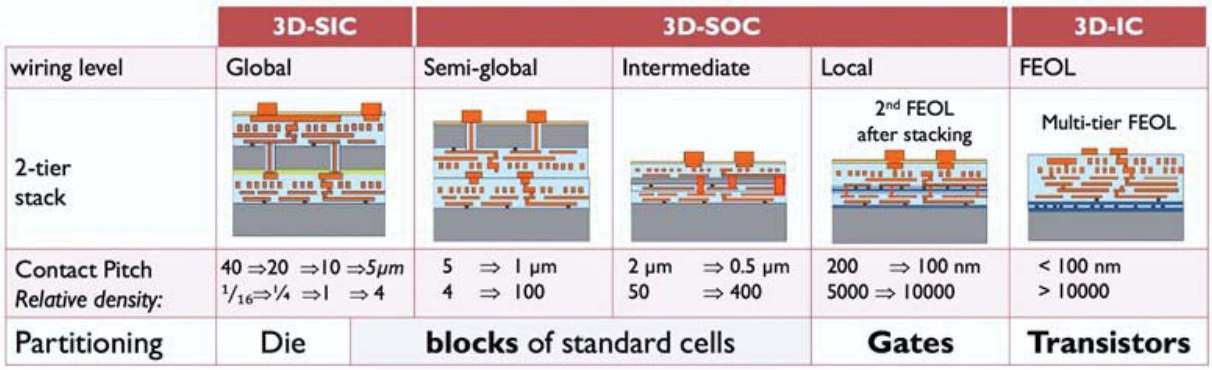


Figure 52: Classification of 3D SoC Integration Approaches based on Integration Interface [23]

11.1.1 3D-SIC

Stacking chips to achieve 3-dimensional integration is a method common to both on-chip and in-package integration. It involves the integration of circuits from fully processed wafers using micro-bumps and TSVs. There are 3 basic kinds of chip stacking, as illustrated in Figure 53. The first and the most common type is face-to-face (F2F) stacking, where chips are attached to each other using bumps at the global wiring level. TSVs are used to access the circuits, which are now sandwiched between chips. The second is face-to-back (F2B) type stacking, where chips are all oriented in the same direction and the front of one chip is connected to the back of the other. TSVs are used in this case between the chips to connect circuits of both chips. The third is back-to-back (B2B) stacking, where the backs of chips are connected. TSVs are fabricated in both chips in this case to provide connection between circuits.

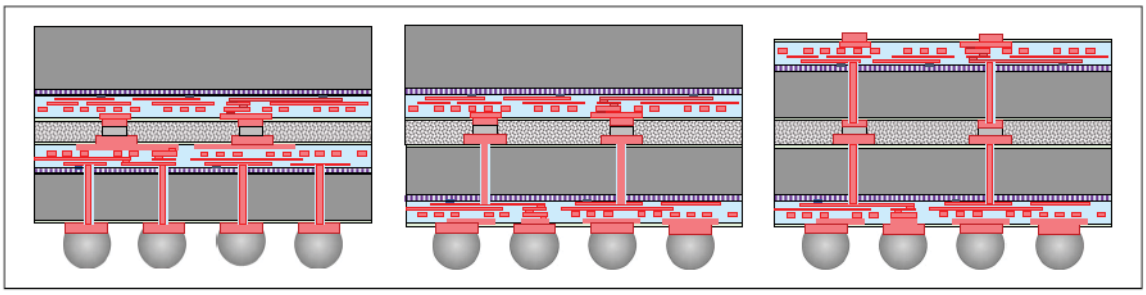


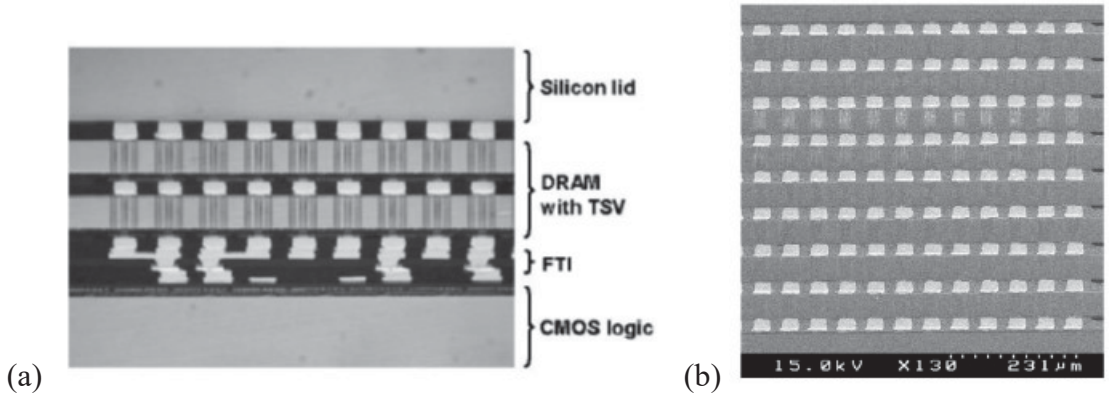
Figure 53: Basic Types of Chip Stacking using Bumps and TSVs
(Left) F2F, (middle) F2B, and (right) B2B [23]

Not only the location of TSVs but also the fabrication methods change depending on the type of stacking approach used in 3D-IC technology. There are also 3 basic types of TSV fabrication approaches. These are:

TSV first: TSVs are fabricated before IC process starts.
TSV middle: TSVs are fabricated along the IC fabrication process.
TSV last: TSVs are fabricated after the completion of wafer processing.

TSVs fabricated in each one of these processes have different properties such as minimum diameter size, aspect ratios and maximum depth.

There are several wafer stacking approaches, but in general they can be classified into "bumping" and "Cu-to-Cu" categories. The first approach involves the fabrication of micro-bumps on both sides of the thinned wafers with TSVs, and interfacing wafers with bumps aligned. The application of pressure and heat makes permanent contact between bumps. A thin adhesive layer can be used between wafers to improve mechanical strength. A cross sectional drawing of stacked ICs connected to each other with this method is shown in Figure 54 [128]. In this example, memory circuits are stacked on top of a CMOS logic circuit. Memory chips are stacked in a F2B configuration. The connection between the memory circuits and the logic circuit is facilitated with the use of a thin interposer layer called "feedthrough interposer" (FTI). Figure 54 b shows the stacking of 8 layers with this approach.



(a) (b)

Figure 54: (a) Cross-section of DRAM-COMS Logic Chip 3D-SIC Integrated Stack and (b) Bump-connected 8-layer Stack [128]

Direct bonding approaches where the Cu pads on the surface of respective chips are attached to each other directly are also popular. These approaches rely on low temperature metal-to-metal contacts and avoid the fabrication of bumps [129-131]. Figure 55 shows examples of this type of contacts. Another Cu-based direct connection between stacked wafers is the use of so-called "Cu-nails". This method, originally developed at IMEC [132, 133], uses "TSV middle" process to fabricate partial TSVs after the FEOL process but before BEOL process starts. The TSV depth is only 15 μm. After BEOL fabrication, the wafer is thinned down to 10μm to expose the TSV metal on the backside of the wafer. Selective etching of the Si wafer then allows slight protrusion of the TSV Cu metal, as shown in Figure 56. Wafers can then be stacked so that Cu-nails penetrate into the Cu pads of the other wafer. A thin layer of SiO_2 at the interface facilitates additional bonding for mechanical strength.

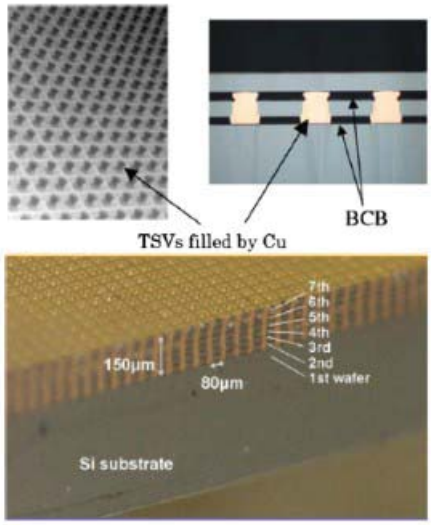


Figure 55: Seven Wafers Stacked on a Wafer with BCB Bonding and TSV Interconnection [129]

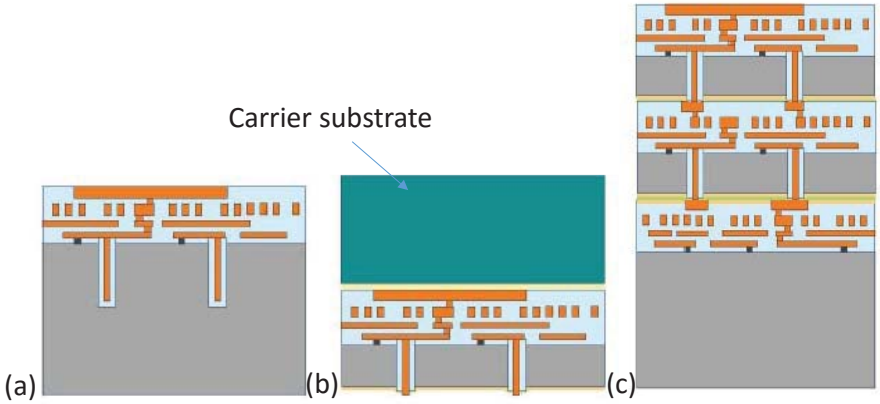


Figure 56: The "Cu-nail" Process
(a) Fabrication of Cu-filled TSVs after FEOL process, (b) wafer thinning to expose "Cu-nails", and (c) wafer stacking with Cu-Cu contacts [133].

The best example of 3D-SIC heterogeneous integration application can be found in high bandwidth memory (HBM) chips mentioned above in connection with 2.5D SiP integration. The first generation HBM products included 4 memory chips stacked on top of a microprocessor chip, as shown in Figure 57. The memory chips are stacked in a F2B configuration and TSVs are used for interconnects. The stacking between the processor chip and the lowest memory chip is B2B type.

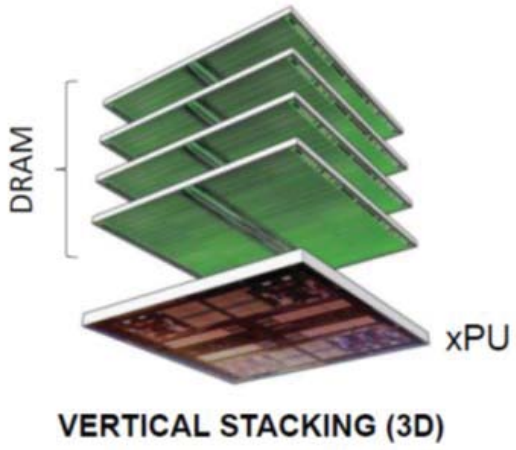


Figure 57: First Generation Stacked Chip HBM Product from AMD [134]

11.1.2 3D-SoC

This method is also called "within-die" integration since it involves integration of different types of circuits within the BEOL wiring stack itself rather than on it. A higher level of interconnects can be achieved between chips when they are interfaced at lower levels of wiring. As shown in Figure 52, the wiring density increases exponentially as lower levels are accessed. Concurrently with the increased wiring density, however, the line pitch and size also decrease exponentially. Integration at the "intermediate" and "local" levels are therefore considerably more difficult than the "semi-global" level.

In 3D-SoC integration, wafers are always integrated F2F, but the exact attachment methods fall under 2 categories. The first method is the "dielectric bonding" method with TSV-last connections, as illustrated in Figure 58. In this approach, wafers are finished with smooth top surface coatings of dielectric layers (SiO_2), the dielectric surfaces are activated for bonding, and wafers are aligned F2F. When wafers are brought together in atmospheric pressure and at room temperature, W2W takes place spontaneously [135]. The top or the bottom wafer can be thinned down and TSVs are fabricated for interface to the package. Additional BEOL processes may be applied to one or both of the wafers.

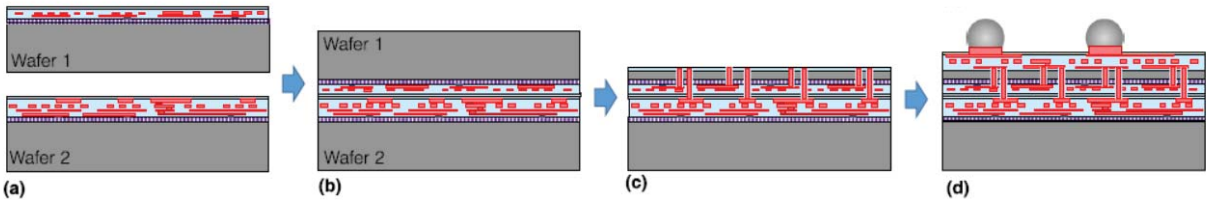


Figure 58: Dielectric W2W Bonding Approach

The second W2W bonding method is the "hybrid method" [136], as illustrated in Figure 59. In this approach, the top surface of both wafers have Cu contacts exposed in SiO_2 dielectric matrix, and very smooth surfaces. Initial bonding occurs when activated surfaces brought in contact with each other. The wafer stack is then annealed to fuse Cu contacts to each other. The success of

this approach depends on the surface finish quality as well as the alignment accuracy. One or both wafers may have partial TSVs before bonding. After bonding and wafer thinning, TSVs provide interconnects to the package. The alignment accuracy is a key parameter for this approach as the interface plane moves further into the BEOL stack. Alignment accuracies of 1.8µm have been demonstrated and 800nm accuracy is projected [23].

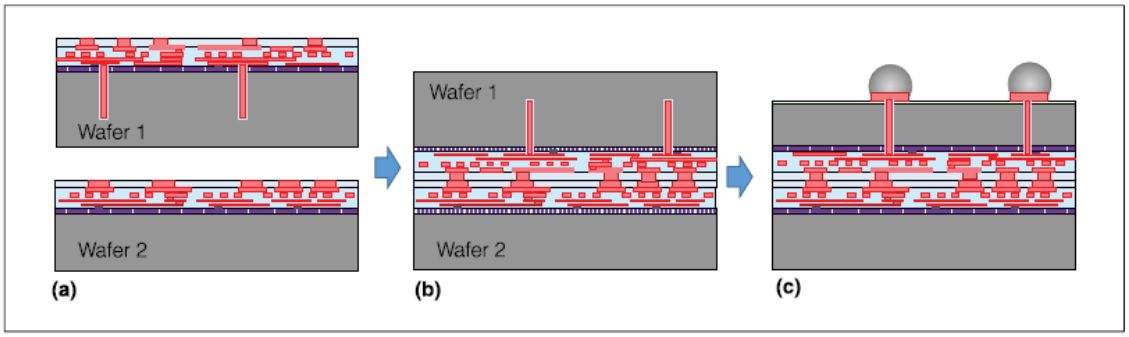


Figure 59: Hybrid W2W Bonding Method

An example the 3D-SoC heterogeneous integration using dielectric bonding approach is the CMOS-visible image sensor integrated chip that was developed at the Lincoln Laboratory [137]. This product integrated 1024x1024, 8µm pixel image sensors with a CMOS wafer containing control circuits using F2F wafer bonding approach. The integration interface layer is the top of the BEOL layers. “3D metal vias” were used to connect the imager chip metal-1 layer, control chip metal-3 layer and the back metal-1 layers, as shown in Figure 60a. A 3-tier version of this imager is shown in Figure 60b [138].

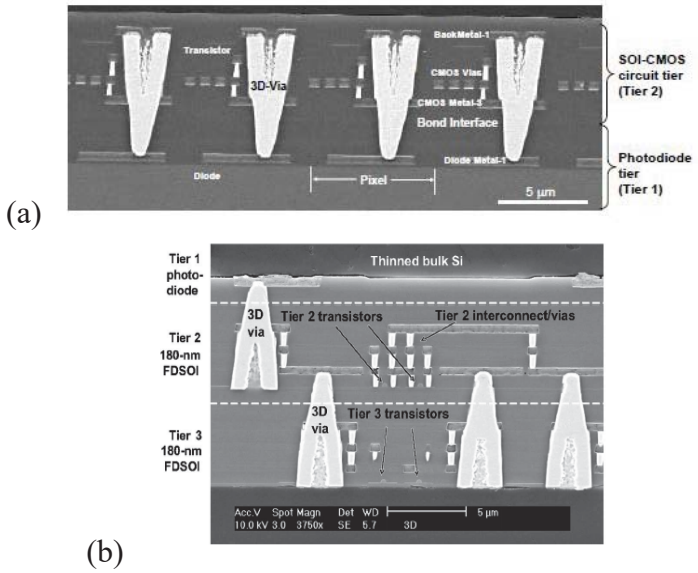


Figure 60: Cross-section of the CMOS Image Sensor [137, 139]

Another application that employs a similar approach is the 3D-laser radio detection and ranging (3D-LADAR) imager based on Geiger-mode avalanche photodiodes integrated with high speed all-digital timing circuits [140]. This integration also requires 3-tier stack and F2F bonding. The use of silicon on insulator (SOI) CMOS wafers were preferred for this technology due to the presence of an oxide layer on both sides of the wafer after substrate removal. An isometric drawing and the cross-sectional view of the integrated chip are shown in Figure 61.

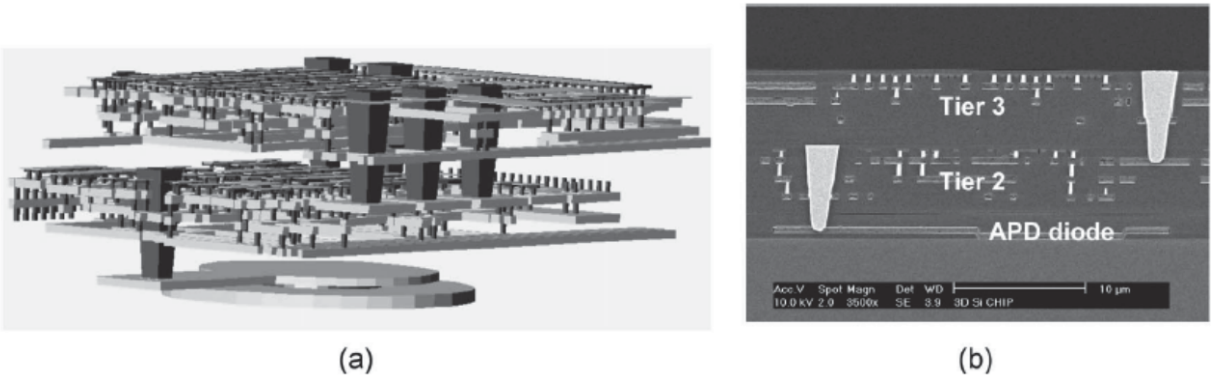


Figure 61: (a) Isometric Drawing of 3D-LADAR Pixel and (b) Cross-sectional View [139]

Lincoln Labs used the integration method previously developed for imaging application, as described above, for the fabrication of 3D RF amplifier circuits under an AFRL contract. Although, it is not strictly a heterogeneous integration, this product demonstration highlights the capabilities of the 3D SoC integration. The schematic of the circuit shown in Figure 62 indicates that the transistors are located on the tier-1 and tier-3 chips, whereas the middle chip contained inductors for input matching circuit [141]. This amplifier footprint was 40% smaller than a 2D version of the same amplifier.

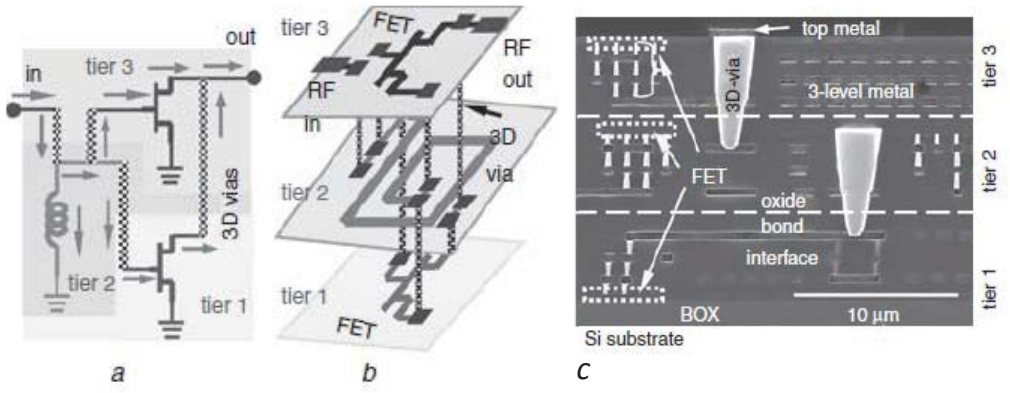


Figure 62: 3D SoC RF Amplifier with 3 Levels of Integrated Chips [141]

One of the prominent features of this technology is the use of 3D-vias, which are different than TSVs. While TSVs are fabricated mostly in substrates and contain an isolation layer between the metal and the substrate, 3D via is fabricated in the BEOL stack in the "field" (isolation) regions. As seen clearly in the cross-sectional views of the integrated ICs above, 3D via connects metal interconnects at multiple levels. It is called "3D" because not only it connects the top layer to the bottom layer, but also several other layers are connected sideways, as shown in the SEM picture in Figure 63.

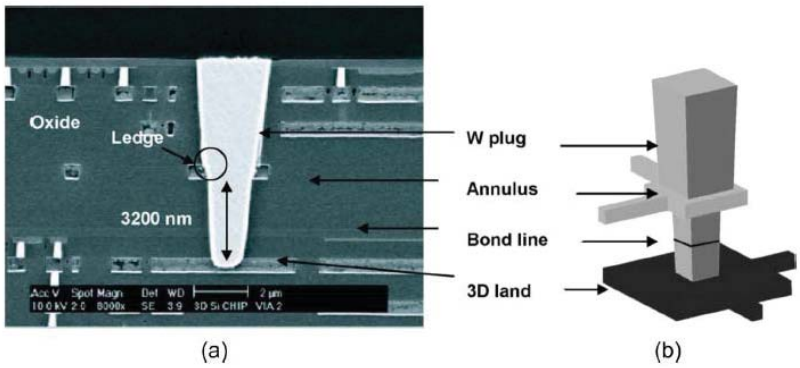


Figure 63: 3D Via Technology for Connecting Multiple Metal Interconnects [139]

While the 3D SoC integration examples examined above show great potential for compact devices, the basic technology components are still in development. Apart from the HDM applications, this technology approach has not produced high volume products yet. There are many promising new developments in 3D-SIC technology area with the introduction of stacked memory chips with controller electronics by AMD, Samsung, Micron [142] and Tezzaron. For example, Samsung announced the stacking of 32 layers of memory chips (see Figure 3) using 3D-SIC approach [143]. The maturation of this heterogeneous integration approach waits for the evolution of electronic design automation (EDA), as well as the development of standards and infrastructure [20].

11.1.3 MEMS-CMOS Integration

We have examined some of the common MEMS-CMOS integration under 2D SoC. The characteristic integration strategy for those devices was “MEMS-first” or “MEMS-middle” approaches that resulted in the fabrication of MEMS on the same chip but not occupying the same footprint. The third integration strategy involves “MEMS-last” process, where MEMS are fabricated on top of completed CMOS wafers [144]. In the example shown in Figure 64, MEMS accelerometers were fabricated on top of the BEOL stack using 10μm thick poly-Si layers [77, 78]. The CMOS circuits are visible under the MEMS structure in this figure.

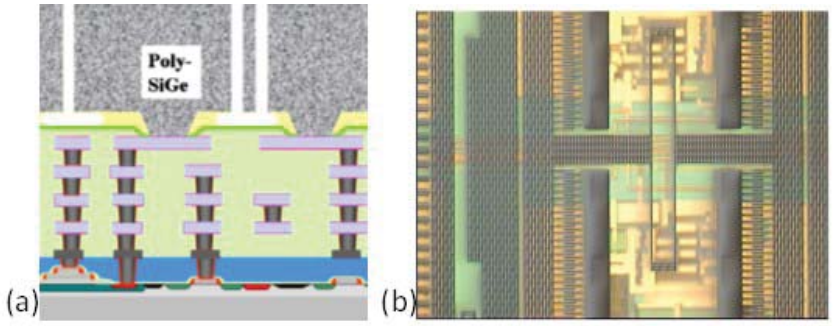


Figure 64: Cross-sectional Drawing and Top View of Accelerometer by 2.5D SoC Integration of MEMS on Top of CMOS Circuits [77, 78]

One of the most successful MEMS products, deformable mirror displays (DMD), are fabricated on the top surface of CMOS chips connected electrically to the BEOL wiring stack. The deformable mirrors are activated by the CMOS circuits laying directly below each pixel. This technology is now mature and finds applications in light projection and digital signal processing. Millions of mirrors can be fabricated on a single CMOS chip [145]. A top view of the fabricated mirrors is shown Figure 65. The cross-sectional view shows the CMOS BEOL layers and the CMOS circuits on the Si wafer.

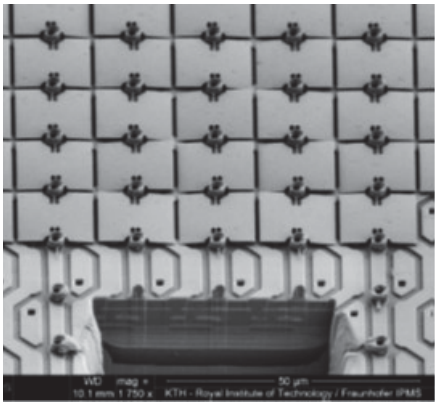


Figure 65: Deformable Micro-mirrors Fabricated on CMOS Wafers [145]

11.1.4 3D-IC

This heterogeneous integration method takes place at or near the FEOL wiring level. After the fabrication of the active devices, as second layer of active device are produced over the first layer. If the second layer devices are different than the first, this approach also becomes a heterogeneous integration. The second layer of active devices can be accommodated by layer transfer methods that involve W2W bonding and substrate removal. Alternatively, the second layer devices may be fabricated using additive techniques [146]. A common BEOL wiring is then applied to the integrated FEOL wiring. This approach is illustrated in Figure 66. What distinguishes this integration method from those under 2.5D SoC is that the functional density is higher by the substantial overlap of active device footprints.

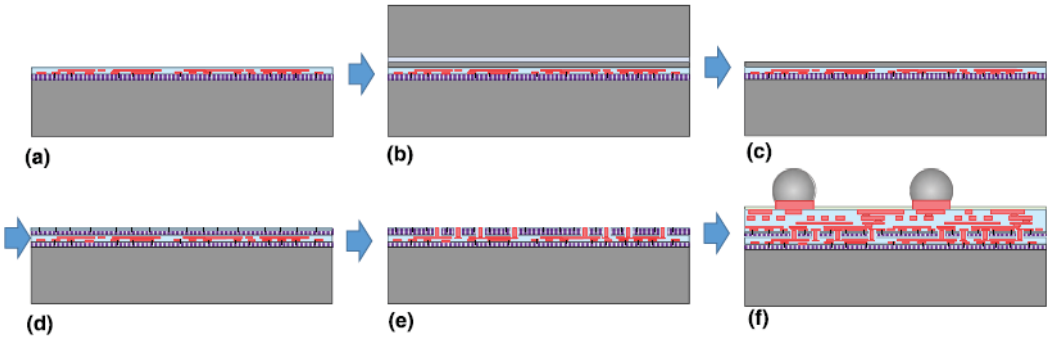


Figure 66: 3D-IC Fabrication Method

An example of 3D-IC building block is shown in Figure 67, where an nMOS transistor is fabricated directly above a pMOS transistor in such a way that a common gate electrode is used to drive both devices. Because of such intimidate integration of devices, the combined device is called “joint MOS” or JMOS [147].

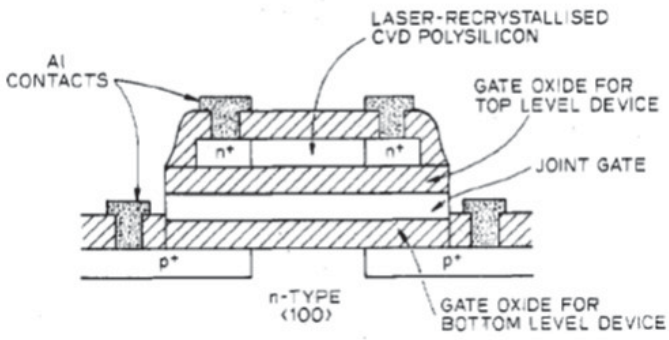


Figure 67: Cross Sectional Drawing of “Joint MOS” or JMOS Device
This type of close integration of devices are a part of 3D-IC heterogeneous integration [147]

3D-IC enables intimidate connection between heterogeneously integrated active devices, but ultimate the integration density is lower than the other 3D-SoC approaches. The reason for this is the difficulty of accessing the first layer devices through the second layer and the management of

the first metal interconnects. These difficulties cause a reduction in the integration density and limit the number of layers that can be stacked.

Additional issues with this approach are related to the quality of semiconductor devices fabricated in the second layer. Layer transfer produces the best quality devices compared to other additive technologies such as re-crystallization of thin films [148-150].

11.2 3D SiP Heterogeneous Integration

Definition: 3D SiP heterogeneous integration refers to the integration of separately processed chips with different technology types in a common package for the purpose of increasing functional density. In a broader definition, 3D SiP can include the integration of packages inside one another. The basic characteristic of this type of integration is the stacking of chips or packages to make 3D assemblies.

Advantages: The use of the 3rd dimension increases circuit density and improves communication between stacked chips due to short length interconnects. The performance of each chip or package can be optimized separately before integration. Highly diverse technologies can be integrated in compact assemblies. This approach provides the best design flexibility to system designers.

Disadvantages: Efficient heat dissipation is a challenge. Integrated heat sinks and cooling channels may need to be part of the integration strategy.

This heterogeneous integration approach is highly versatile and capable. It is also the most difficult one to classify into technology development streams. It is possible to arrive at 3D SiP heterogeneous integration following regular evolution of the BGA type packaging by going through single-chip packaging, multi-chip packaging and 2.5D SiP packaging using TSV interposers. It is also possible to arrive a similar 3D SiP integration by continuing the 3D SoC evolution and adding a wafer level package (WLP) or chip level package (CLP) to it. A schematic representation of these technology paths is shown in Figure 68. It is difficult to coherently analyze all these development paths. The packaging technology is evolving very fast and completely new types of packaging concepts are appearing every year. This dynamic technology evolution in 3D packaging is one of the reasons why heterogeneous integration roadmap is assigned to the IEEE Components, Packaging and Manufacturing Technology Society. We will examine this technology area starting with the most commonly used (in high volume production) products and progressing toward those approaches that are still in R&D stage.

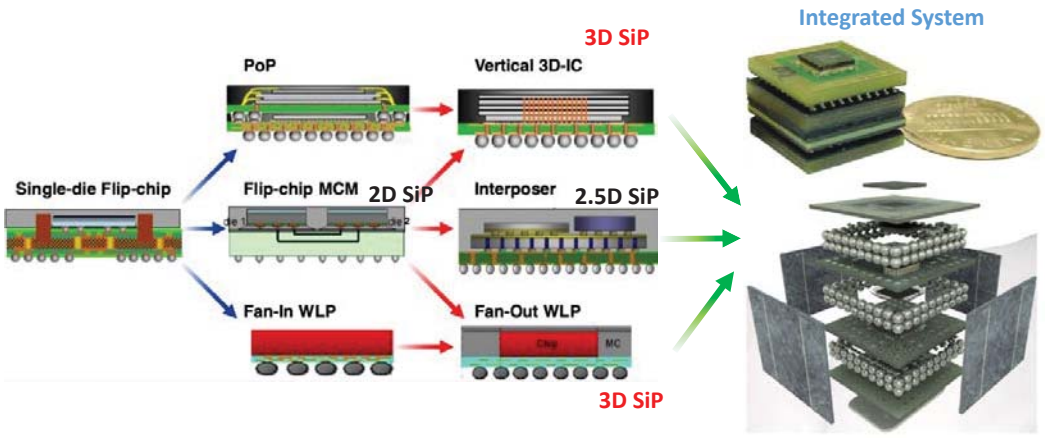


Figure 68: Alternative Paths to Arriving 3D SiP Heterogeneous Integration

11.2.1 Bond Wire Connected Stacked Chips

This technology was developed a long time ago for integrating multiple memory chips with a microprocessor chip in a package [151]. It follows is a simple yet effective concept of simply gluing several staggered chips on top of each other and bond wire connecting chips to each other or to the package. In the modern version of this concept, interposer layers are used under the chips to house RDLs [152]. Complex set of bond wires are used to connect chips to the RDL, as shown in Figure 69. The other side of the interposer layer contains BGAs for external contacts. This type of packaging is used in mobile products where because of space saving features and can be found in smart phones and tables.

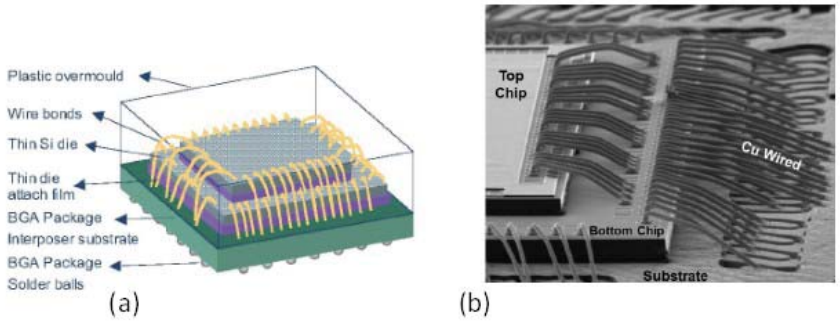


Figure 69: BGA Package with Stacked Chips and Wire Bonding
(a) Schematic drawing and (b) AMCOR's stacked chips with Cu bond wires [133, 153].

The main disadvantage of this 3D packaging approach is the use of long bond wires, which can have inductance values of several nH. Also, high density of wires is also a concern for signal cross talk. Therefore, its use will be confined to moderate level integration for medium speed circuits. However, it is an effective technique for heterogeneous integration involving electronic, acoustic and optical devices.

11.2.2 PoP Integration

PoP technique is similar to chip stacking in a package, except separate packages are stacked on top of each other. The purpose of this approach is also similar to chips stacking i.e. space saving. The added advantage is that already tested (and burned-in) packages containing diverse heterogeneous technologies and integration approaches can be stacked [154, 155]. In Figure 70 shown below, a package containing a single flip-chip mounted die is integrated with another package containing stacked chips with bond wire connections [153]. PoP integration is also used heavily in smart phones and other mobile applications due to its design flexibility and compactness.

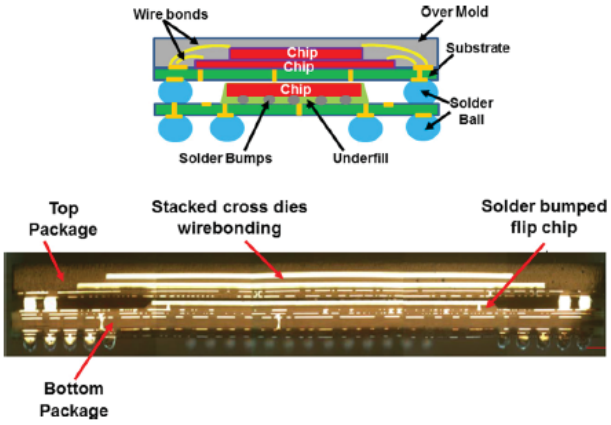


Figure 70: 3D SiP Integration using PoP Approach [153]

11.2.3 Bump-Connected Stacked Chips

Multiple chips can be stacked on top of each other inside the package to integrate multiple functions in a 3D assembly. This approach is very similar to that used for 3D-SIC SoC heterogeneous integration, except that individual chips are stacked on top of each other. This is in contrast to wafer-to-wafer bonding process used in 3D-SIC SoC, where all circuits on the wafer are stacked simultaneously. TSVs can be used for interconnects between stacked chips. The most common type of chip-to-chip bonding approach used for 3D-SiC is the bump-bonding. Cu-Cu bonding is also possible but not practical for chip stacking in a package. Chips stacked at a wafer-level process can be diced and packaged afterwards, but this is not a preferred approach either since wafer level packaging can be applied to these stacked chips to make 3D SiC. This will be discussed in the next section. Here we will only review bump-bonding approaches.

Micro-bumps used for chip stacking are typically 5µm in height and diameter. Bumps can be fabricated as a part of the BEOL process or after TSV process, depending on when the TSV process is applied (e.g TSV-first, TSV-last processes etc.). In the most common applications, each die is stacked on top of another by lining up bumps, and the dies are in F2B or F2F configurations. A large number of chips can be stacked with this approach, as seen in Figure 71.

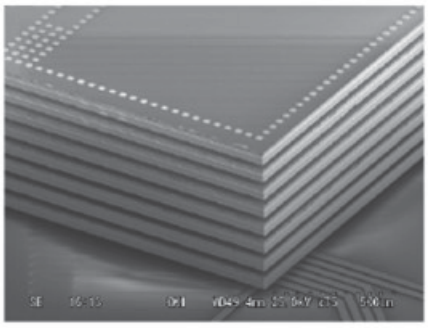


Figure 71: SEM Picture of 8-strata Stacked Chips with TSVs on Interposer Layer [128]

This type of heterogeneous integration approach was a part of the DARPA Scalable Millimeter Wave Architectures for Reconfigurable Transceivers (SMART) program, where stacked tiles of RF circuits were employed for three-dimensionally integrated active electronically steerable array (AESA) modules operating at millimeter-wave frequencies. The technical approaches used for this program, shown schematically in Figure 72 [62], make use of vertically connected Si CMOS, GaAs, InP wafers and patch antennas using Si or GaAs interposer layers. The objective of the program was to achieve 5W/cm^2 radiation power density with vertically integrated tiles whose height is not to exceed 1cm [156].

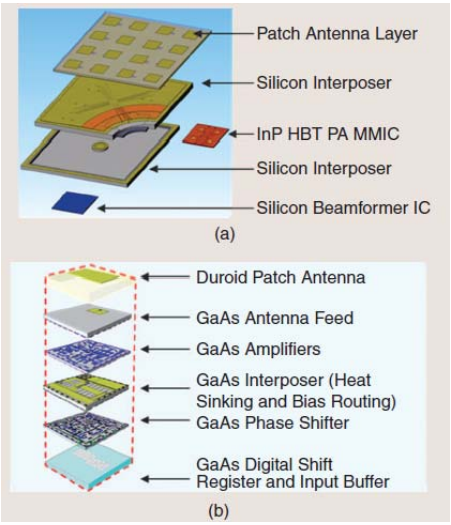


Figure 72: The DARPA SMART Program
(a) Interposer based heterogeneous integration and (b) waver-level package heterogeneous integration. MMIC: monolithic microwave IC [62].

11.2.4 Stacked Chips without TSV

The bump bonding technique is well suited to heterogeneous integration, where chip sizes may be very different and some of the integrated chips may not have TSVs. Following a concept of mother-daughter die pairing, smaller die can be attached to larger die and the larger die can be connected to the package directly, all using flip-chip bonding [157]. This concept can be extended to more than 2-chip stacking as shown in Figure 73, with the use of 4 different size bumps [158]. It can be seen that the mother-die is supporting the daughter-die and the grandma-

die is supporting the mother-die. No TSV was used for this 3D integration. One of the earlier SONY PlayStation models used this approach to connect wide I/O synchronous RAM chips to the processor chip F2F using this approach.

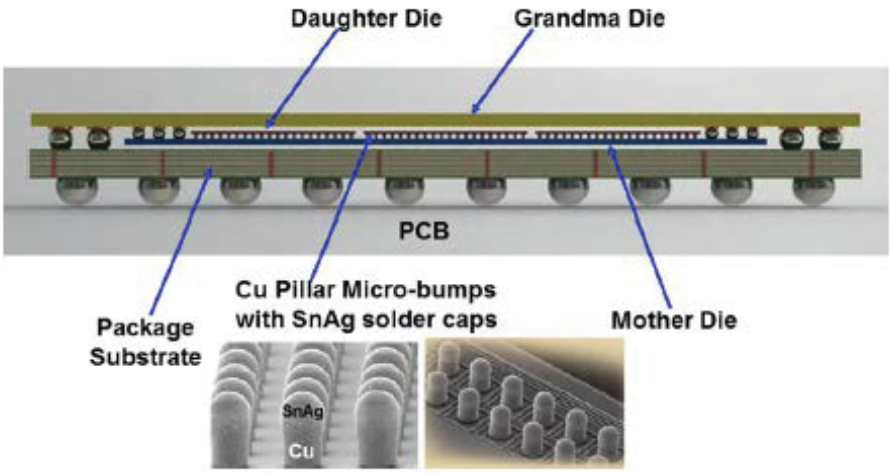


Figure 73: AMCOR Multi-chip Packaging using Different Size Bumps [158]

11.2.5 3D WLP

Wafer level packaging is already a well-established technique for single die products. After the wafer front-side processing, TSVs can be fabricated in thinned substrates and bumps can be fabricated on the wafer backside. The top of each die can be coated with plastic molding for protection. Alternatively, the bumps can be fabricated on the front-side of the wafer and TSVs are avoided. When chips are singulated, each die becomes a packaged product. This is commonly known as the WLP or CSP. We will first provide a background to WLP technology before examining its use in 3D SiP applications.

There are 2 distinct types of WLP; fan-in (FI) and fan-out (FO) types [159]. Although they look very similar in appearance, their fabrication differ significantly. Figure 74 shows the cross-sectional drawings of both packaging types. FI type is the simplest type of package for electronic chips used in low cost products and those that are size, weight and cost sensitive. The chip is made compatible with direct mounting on circuit boards using solder bumps placed directly on the front side, as shown in Figure 74a. The fabrication of FI-WLP is compatible with wafer batch processing and the incremental cost due to packaging is minimal. This approach, however, has limitations in the number of bumps that can be placed on the chip. As the chip size shrinks, this problem becomes more acute.

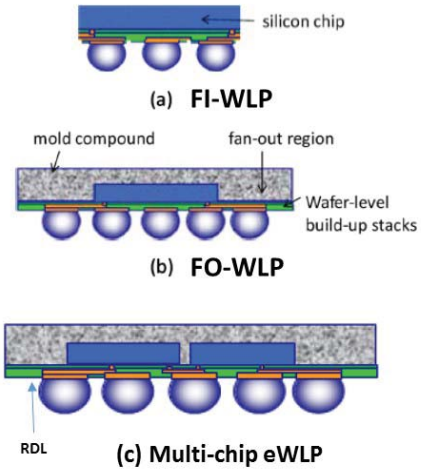


Figure 74: Comparison of FI-WLP and FO-WLP Packaging Approaches [159]

Larger number of I/O's can be accommodated if the interface surface can be expanded beyond the actual chip surface. This is the idea behind the FO-WLP concept, which is also known as the embedded WLP (eWLP), as shown in Figure 74b. The chip is fabricated with a RDL, which may contain multi-level interconnects. Larger number of bumps can be fabricated on the RDL. In its elemental form, the eWLP is similar in concept to the 2.5D SiP, except that the package is fabricated on the wafer using batch processing. The WLP is capable of integrating multiple chips for heterogeneous integration, as shown in Figure 74c. In some cases, the second chip may contain passive circuit elements that would not conveniently fit onto the first chip, as illustrated in Figure 75. This is a useful approach especially for microwave circuit applications, where high-Q passive components such as capacitors and inductors are needed but cannot be accommodated cost effectively on the integrated circuits. We will see below that eWLP is compatible with 3D heterogeneous integration using of stacked chips. It is also compatible with PoP concepts [160].

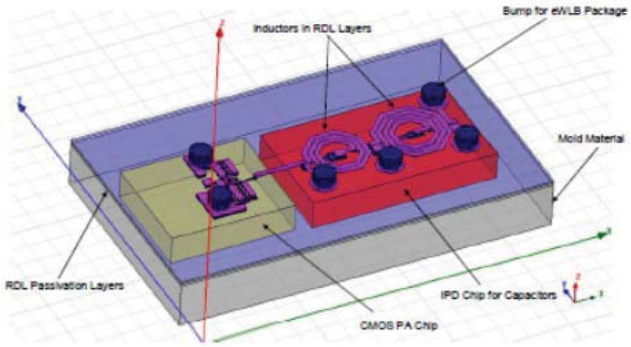


Figure 75: Schematic Drawing of eWLP to Integrate CMOS Power Amplifier Chip with an Integrated Passive Device (IPD) Chip [161]

Another way to embed passive circuit elements is to fabricate them directly in the package. The RDL provides multi-level interconnect technology for the fabrication of capacitors and inductors as well as antennas in close proximity to the IC chip. Figure 76 shows the schematic drawing of an integrated antenna concept. Also shown are inductors fabricated in the RDL layer for VCO and commercial millimeter-wave radar applications [162-165].

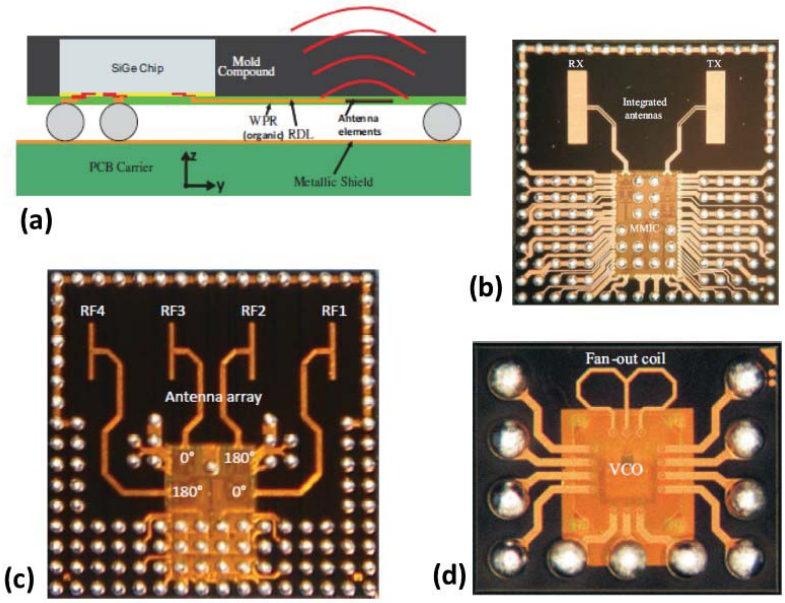


Figure 76: Integration of Antennas and Inductors with Active Semiconductor Chip in eWLP for Commercial Radar and VCO Applications [162-165]

The eWLP processing technology is not exactly the same as the regular IC wafer processing and requires post wafer processing steps. Figure 77 illustrates the industry standard "re-constructed wafer" post processing approach to FO-WLP fabrication [166-168]. The starting point for the re-constructed wafer process is a "wafer-shaped substrate" populated with singulated dies. A plastic molding layer is then applied over the chips. The original supporting substrate and the adhesive layer is then removed. At this point, the re-constructed wafer looks like that shown on the right of the picture. This wafer is processed further to fabricate organic, multi-layer RDL with bumps. Finally, packaged die is singulated.

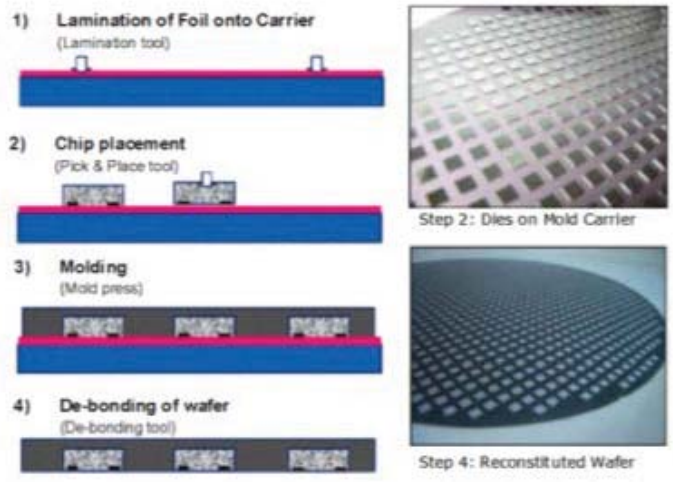


Figure 77: FO-WLP Fabrication Steps [166-168]

With the recent availability of the TSV technology, eWLP approach has been extended to the fabrication of 3D-SiP for heterogeneous integration. For example, the stacked memory and feedthrough interposer (SMAFTI) technology approach of NEC makes use of W2F of D2W stacking approaches to stack memory chips with TSVs. NEC makes use of TSV and bump technologies for chip stacking, but other direct chip stacking with Cu-Cu contacts may also work. The reconstructed wafer is populated with these singulated stacked dies, as shown in Figure 78. The reconstructed wafer is processed to fabricate the FTI layer (another term for RDL) and places logic chips on the other side of FTI layer. BGAs are then fabricated around the logic chip to complete the heterogeneous 3D SiC stack.

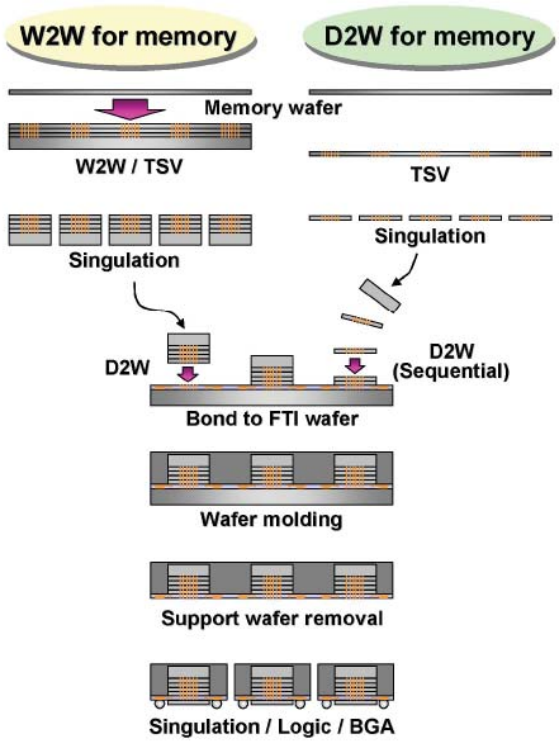


Figure 78: 3D Heterogeneous Integration on WLP [128]

As the examples show above indicate, WLP is not confined solely in the 3D SiP category, but spans the full spectrum from 2D to 3D SiP. Starting with the simple fan-in type packages, WLP extends to multi-chip integration with fan-out configuration. These integration types are confined to 2D SiP category. The use of eWLP with TSV interposer is in 2.5D SiP category. 3D SiP examples can be found in PoP, stacked chip and double-sided flip-chip integration. Figure 79 shows how the variations in the WLP technology and their categorization.

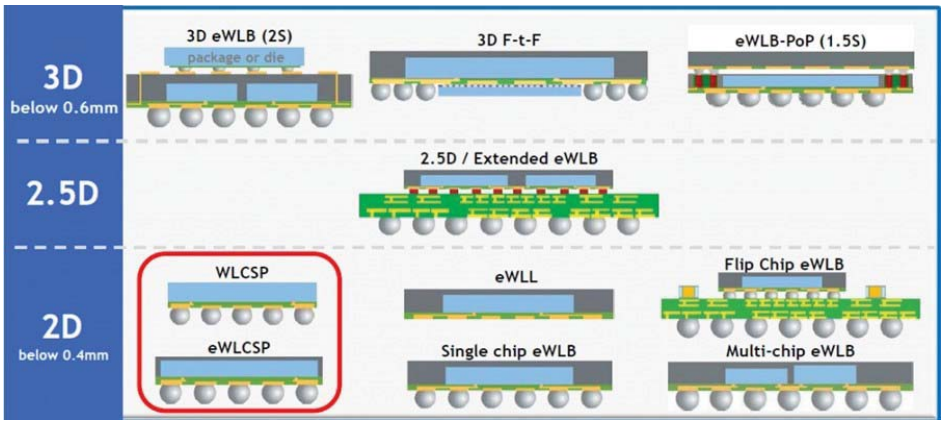


Figure 79: WLP Technology Extends over all SiP Categories
(Source: http://www.samsung.com/semiconductor/support/package-information/overview/)

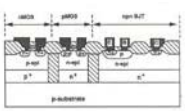
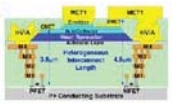
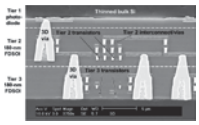
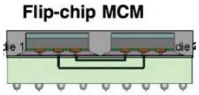
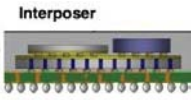
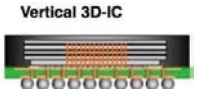
12. SUMMARY OF HETEROGENEOUS INTEGRATION TECHNIQUES

Significant progress has been made in developing technologies for heterogeneous integration of components in compact assemblies, as reviewed briefly above. This progress is not only continuing but accelerating as the advantages of these new generations of components with increased functional density become apparent. We are already seeing highly successful commercial products that integrate functions such as memory/logic or imaging/processor in 2.5D and 3D stacked chip formats. For example, the Intel/Altera Stratix 10 product, shown in Figure 49 above, is a highly successful product line that exploits 2.5D SiC heterogeneous integration technology.

The integration techniques examined in the previous chapter were fit into specific categories for reviewing convenience. Heterogeneously integrated products cannot be similarly categorized. In most cases, heterogeneous integrated products can make use multiple technologies and various integration categories depending on the application. This technology-independence (or flexibility) is the key feature that makes heterogeneous integration highly responsive to system design innovations.

Table 1 summarizes the key features of the integration techniques used for heterogeneous integration. There are many innovations involved in each technology category. The wafer-level integration approaches address the need for close proximity of different types of devices and circuits. When possible, they are fabricated on the same substrate for the highest level of integration (e.g. Si BiCMOS circuits). Otherwise, they are fabricated at different layers and combined within the same circuit, as in MEMS/GaAs pHEMT circuits. Others make use of the top surface of CMOS BEOL stack for new devices that can be controlled with the CMOS circuits below it (e.g. deformable mirrors for display applications). When the end product is the integrated chip, as in these examples, the wafer-level (SoC) integration is preferred.

Table 1. Summary of Heterogeneous Integration Techniques

Integration Location	Dimension	Example	Attributes	Advantages	Disadvantages
SoC	2D		- Integration of different device types on the same wafer surface. - Devices with alternative materials. - No overlapping device footprints.	- Increased performance and functionality. - Integration of different devices in the same circuit.	- Added process complexity. - Decreased integration density.
	2.5D		- Integration of different device types and alternative materials at multiple surface levels. - No overlapping device footprints.	- Increased performance and functionality. - Direct interface between circuits based on heterogeneous devices.	- Added process complexity. - Decreased integration density.
	3D		- W2W bonding and layer stacking. - Overlapping device footprints. - TSV and RDL based interconnects for vertical and horizontal signal distribution.	- Significantly increased functionality density. - Short interconnects between devices and circuits. - Batch-level high-volume processing.	- Compounded yield loss. - Challenging thermal management. - 3D design architecture immaturity. - High initial cost of manufacturing.
SiP	2D		- Multi-chip packaging on the same surface. - Flip-chip or bond-wire connection between chips and package. - No overlapping chip footprints.	- Compatible with miniature flat packages. - Highly suitable for microwave circuit integration.	- Low integration density.
	2.5D		- Integration on TSV interposer. - High speed interconnects with embedded bridging chips.	- High speed interconnection between integrated chips. - High interconnect density. - Use of KGD (known-good-die)	- Cost and size of TSV interposer.
	3D		- Stacked chips connected by bond wires or bumps. - Stacked packages. - Wafer level packaging of stacked chips.	- Highest functional density. - Highly flexible fabrication approaches. - Low initial cost manufacturing.	- Non-standardized and comprehensive design/test tools.

While the chip-level integration results in highly compact products, it is not always the preferred approach for heterogeneous integration. Combining two or more heterogeneous technologies on wafer require the meshing of technology IPs that are not directly compatible. The source of incompatibility may be in the material growth (lattice mismatch), thermal budget of processing, or process sequence incompatibility. These are the typical well-known problems associated with the integration of alternative material devices. There are some other serious problems even when

the devices are made from the same semiconductor material. These problems are associated with the availability of technology nodes at a given time for each technology IP (i.e. product type). In Si ICs, the transistor gate length determines the technology mode. Typically, microprocessor circuits use the most advanced node for integration density and circuit speed advantages. However, other product designs such as memory, use earlier nodes. In these case, it is advantageous to split the functions and fabricate them separately.

SiP-based heterogeneous integration techniques range from simple multi-chip packages to highly stacked 3D packages. It is also possible to stack packages on top of each other. The 2D-SiP approach has been successful for microwave modules where different types of circuits fabricated on separate chips are combined. The 2.5D SiP is also a successful approach for combining HBM chips with processor core for ultra-high bandwidth products. The 3D-SiP technology can reduce the footprint of 2.5D SiP products by vertically integrating all components.

There are mature wiring technologies to enable 3D integration in packages including wire bonding, TSVs, and solder bumps. Integration in packages is a highly flexible method that requires little or no set up cost (compared to SoC integration). Additionally, SiP integration can be undertaken cost effectively for small production batches.

A merging of SoC and SiP integration technologies can be found in the eWLP approach. This is a quasi-wafer level integration technique in a package. eWLP technology spans across all integration categories (i.e. 2D, 2.5D and 3D SiP). Stacked dies made by 3D-SoC approaches as well as individually stacked dies can be packaged in batches using this highly flexible approach.

13. HETEROGENEOUS INTEGRATION PRODUCT EXAMPLES

The technical approaches reviewed above are applicable to a variety of new product applications. In some cases, heterogeneous integration improves the performance, cost or the size of the product (e.g. high bandwidth memory). In other cases, it provides a new solution to existing problems, such as high speed connection between multi-chips in a package (e.g. Stratix 10 wide bandwidth FPGA). But, its most important contribution is for enabling the design of new compact systems that would not be possible without such integration (e.g. multi-sensor microsystems). The technology base for heterogeneous integration is already well established yet still growing rapidly. We can expect new commercial and military applications to build on this technology base now and in the future.

As mentioned before, heterogeneous integration technique is highly responsive to system-level innovations. Several integration approaches, but especially the SiP approaches, have a great deal of technology flexibility to address new system designs. Since existing component technologies are used in the integration (rather than developing new component technologies), the long time-delay associated with technology development cycle can be avoided. Also by using known good die (KGD) method, the fabrication yield can be managed for cost-effective manufacturing.

Many of the recent highly successful commercial products introduced by the industry leaders over the last 10 years employ heterogeneous integration, as shown in Figure 80.

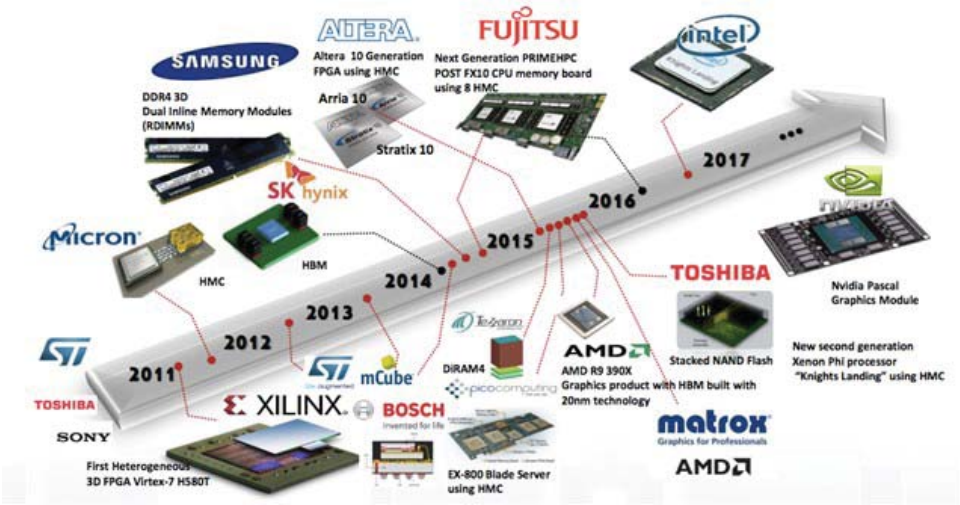


Figure 80: Many of the Recently Introduced and Highly Successful Commercial Electronic and Optoelectronic Products Employ Heterogeneous Integration
(Source: GlobalFoundries)

We have already examined the stacking technologies that enabled high bandwidth and high capacity memory components from Micron, Hynix, Tezzaron, and Toshiba. Integration of these high capacity memory chip stacks with processors enabled wide bandwidth FPGAs from Intel/Altera, AMD, and Nvidia. We have seen the first examples of heterogeneously integrated imaging arrays with processors by Xilinx. Miniaturized microwave modules integrating multiple

chips are not commonly used in cellular telephones. In this section, we will review additional examples of heterogeneously integrated electronic and optoelectronic products.

13.1 Wide Bandwidth Memories

Stacked DRAM memory chips have been providing higher capacity for several years for cellular phone applications where the data bandwidth is 10-50 Gbps. Now they are being developed for high speed computing where the bandwidth is nearly 10 times higher. HMD and hybrid memory cube (HMC) are both stacked chips with a controller chip [169] [142]. The controller chip is needed for these complicated assemblies to manage data as well as to manage failed bits and TSVs. The HMC design by Micron shown in Figure 81 is expected to increase the memory bandwidth to over 1Tbps.

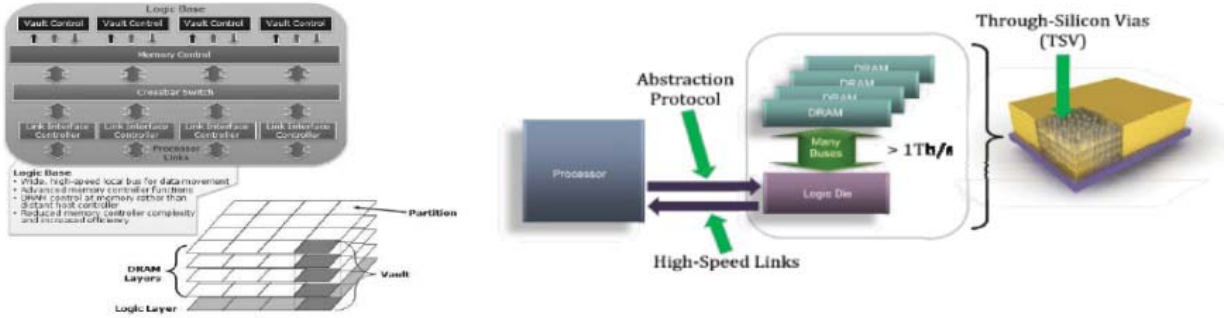


Figure 81: Block Diagram and the Structure of the Hybrid Memory Cube [142]

13.2 Integrated CMOS Image Sensors

3D heterogeneous integration is an enabler for integrated CMOS imager systems. We have seen above that both SoC and SiP approaches can be used to integrate the controller and the imaging chips in a stack. Since the imager is an array of millions of pixels and the data from each pixel must be transmitted to the processor chip with the shortest delay, the vertical integration offers a unique solution. In the example shown above (see Figure 60) SoC integration approach was used for a 1024×1024, 8µm pixel visible image sensor fabricated with oxide-to-oxide wafer bonding and 2-µm square 3D-vias in every pixel [137]. Similar integration approach was used for a high speed 8Mpixel, back illuminated imager [170]. More demanding imaging applications such as self-driving cars require stereo imaging at high frame rates of >10,000 frames/s. These requirements were met by stacking imaging chips on correlated double sampling chip (CDC) and ADC chip using chip stacking technology with TSVs [171]. Each chip was fabricated separately in different facilities using different technology nodes, therefore this product is an example of mixed node chip integration as well. Figure 82 shows the picture of the imager system, its integration approach, and cross-sectional views.

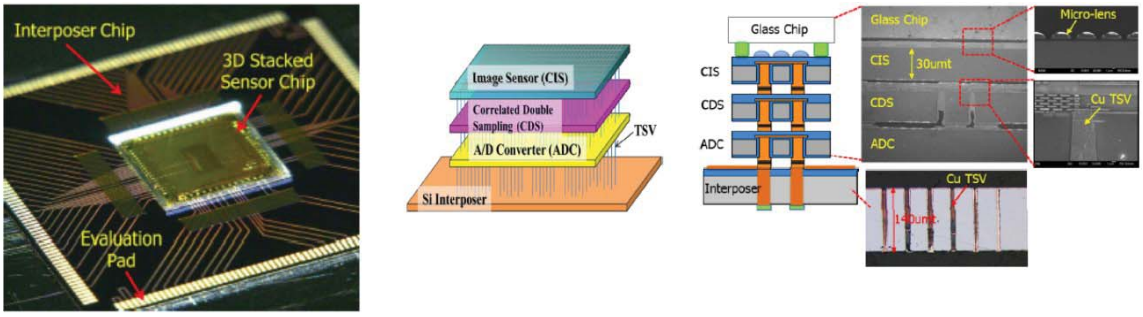


Figure 82: 3D SiP Integrated CMOS Imaging System for High Frame Rate Applications [171]

13.3 High Bandwidth Processors

The CPU-memory interface is one of the limiting factors for high speed computing. The performance of systems such as FPGAs, processors and GPUs can be more efficient by increasing the IO bandwidth rather than increasing the processing speed. As shown in Figure 83, the highest CPU-memory bandwidths are achieved using the HBM DRAM, which has a wide (1024 bit) bus running at a relatively slow 2 Gb/s compared to GDDR5 which has a 128 bit bus running at 7.2 Gb/s [127].

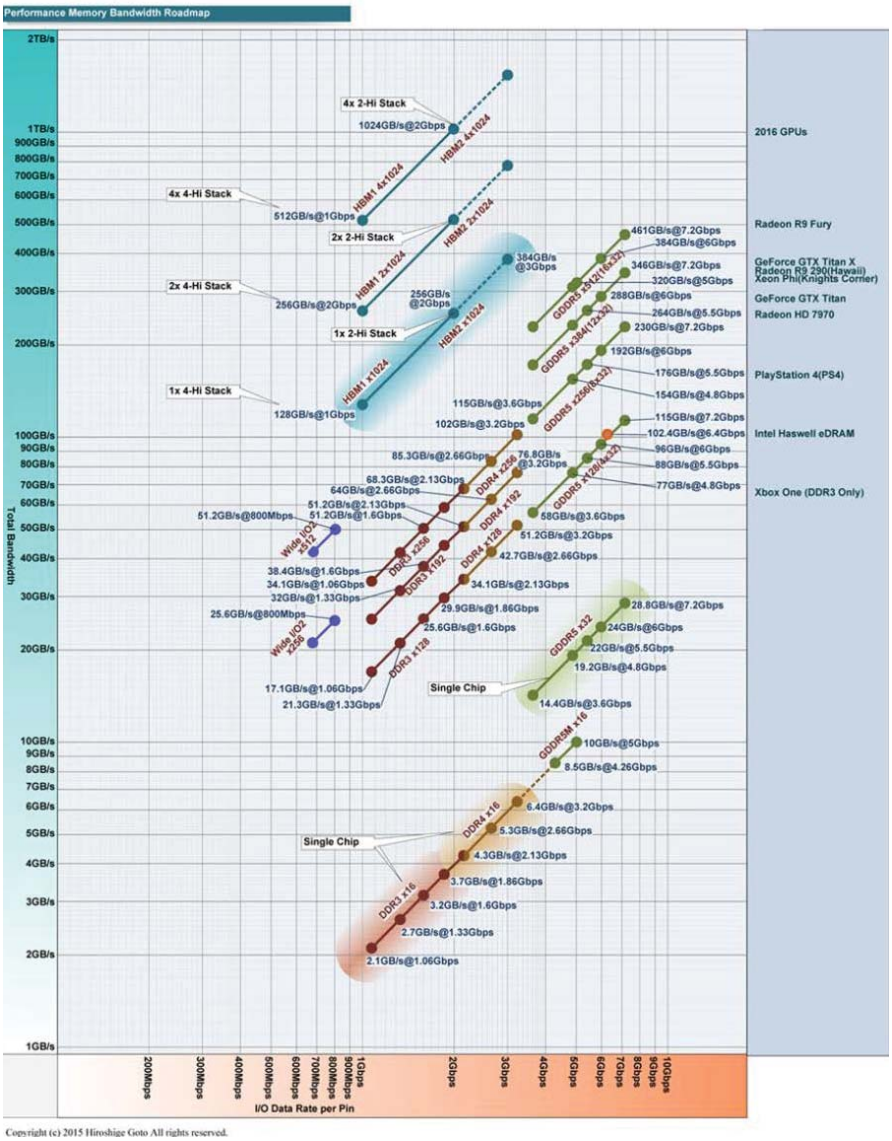


Figure 83: CPU-DRAM Memory Landscape [127]

Altera and Xilinx have developed FPGA and GPU modules that integrate central processor to several high bandwidth memory banks using 2.5D SiP technologies. The approach taken by Xilinx is to use TSVs for high density interconnects. Whereas, Altera uses Intel's EMIB technology, as shown in Figure 84. These approaches are highly popular at present and such products represent some of the best examples of the advantages gained by heterogeneous integration.

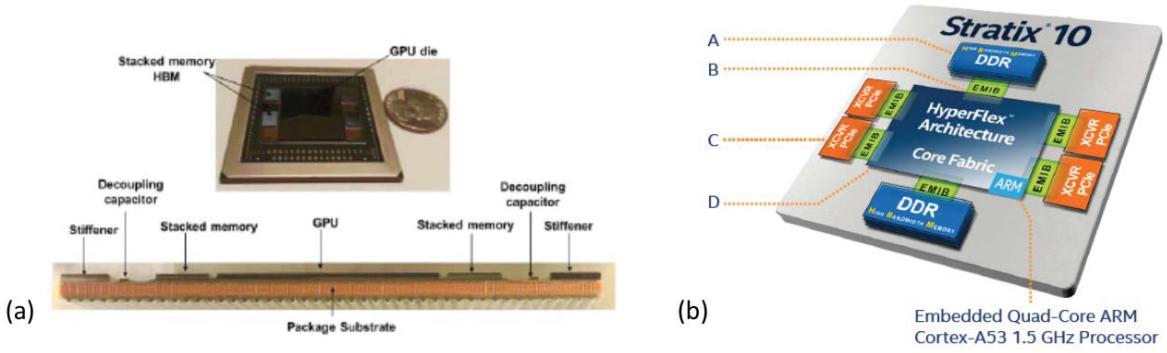


Figure 84: Examples of Products Resulting from 2.5D SiP Heterogeneous Integration
(a) AMD Radeon™ Fury and (b) Altera/Intel Stratix™ 10

13.4 Digital Light Processors (DLPs)

One of the earliest and most successful SoC integration example is the DLP developed by Texas Instruments in 1980's [172]. DLP is a display product made possible by the 3D SoC vertical integration of MEMS devices on CMOS integrated circuits. DLPs are used in projection applications such as home theaters, digital cinema, and pico-projectors in cell phones. The current display resolution is 4K using several million deformable mirrors. In industrial applications, they can be found in 3D printing, digital lithography, machine vision, and spectroscopy. They are also used in automobile heads-up display applications (Figure 85).

Figure 85: Texas instruments DLP™ and its Application in Automobile Heads-up Displays

13.5 Inertial Sensors

Another successful application of MEMS-CMOS integration is the inertial sensing products. The applications of these products range from accelerometers used in cars and industrial applications, high-g sensors in car air bags, gyroscopes, angular accelerometers used in disk read/write head assemblies. An example of a MEMS accelerometer is shown in Figure 86.

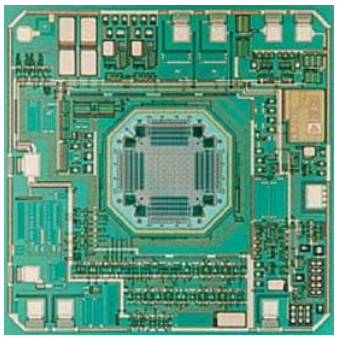


Figure 86: Analog Devices ADXL 50 Accelerometer

13.6 Integrated RF Systems

An example of an integrated electronic system produced by heterogeneous integration is the low power RF transceiver fabricated at IMEC, as shown in Figure 87 [133, 173]. This device measures 7mm x 7mm and includes 2 WLP chips. The connection between these 2 chips is made using solder balls (bump connected).

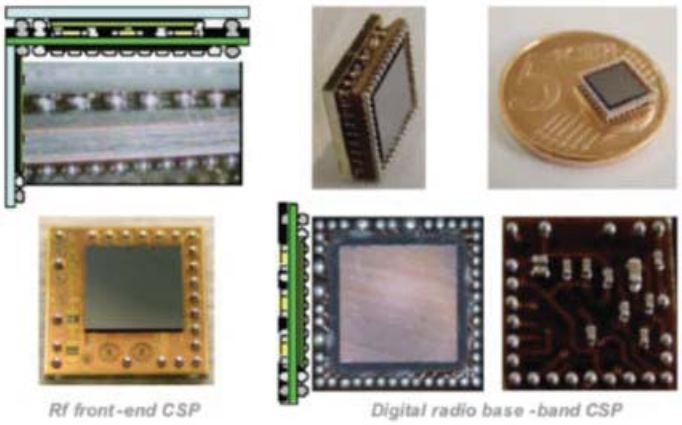


Figure 87: Fully Integrated Low Power RF Radio, Measuring 7x7x2.S mm^3, Realized by 3D Stacking of WLP (CSP) Packages [133]

The same group at IMEC is attempting to show the capabilities of SiP heterogeneous integration technologies by building an "eCube", as shown in Figure 88. This is a total sensor system complete with power sources and antennas. The design concept is to build each sub-system on a chip-level packaged component and stack all components on top of each other. Connections between sub-systems are made using micro-bumps and interposer layers. The interposer layers contained embedded passive components for RF circuits.

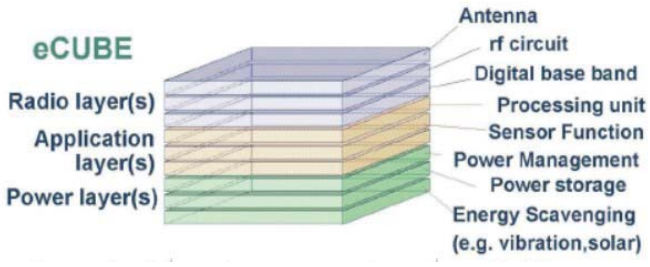


Figure 88: Schematic Representation of a 3D-SIP Concept "eCube", for the Realization of Distributed, Fully Autonomous "Ambient Intelligent" Systems
Each layer in the stack is a fully finalizing the 3D via-process [133].

Figure 89 shows the fabricated "eCube" system for medical applications measuring $1cm^3$. This system a RF subsystem with antenna, a low power digital signal processor (DSP), a 19 channel electroencephalogram/electrocardiogram (EEG/ECG) sensor die, and a power controller. Solar cells shown schematically in the lower figure can be added on the sidewalls of the cube.

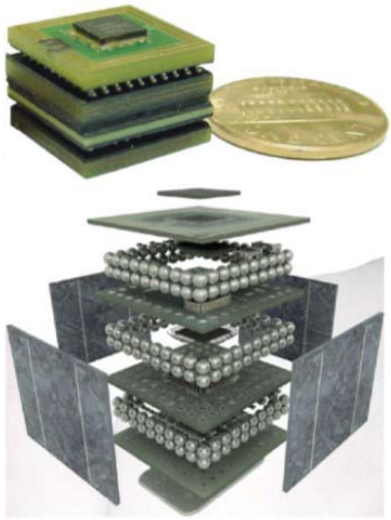


Figure 89: Photograph and a Schematic Drawing of "eCube" Autonomous Sensor System Developed at IMEC for Medical Applications [133]

A similar heterogeneous integration approach was taken at the Fraunhofer Institute for the fabrication of micro-transceivers (nodes) of a microwave localization system operating at 24GHz [174]. This system integrates a patch antenna on top a glass interposer, which also supports all RF electronics including LNA, PA, mixer, VCO, and a multiplexer. All intermediate frequency (IF) and baseband electronics are integrated on the lower substrate. Ball bonding was used to connect these substrates to each other and the whole assembly to the circuit board. This system is expected to have 50m range and 12mm location accuracy. Similar highly miniaturized wireless systems can be a part of autonomous sensor network for collecting environmental information. The role of the heterogeneous integration technology for reducing the size and the power consumption of small sensors was studied systematically [175]. A >100X reduction in overall size was shown to be possible for the same wireless system through miniaturization.

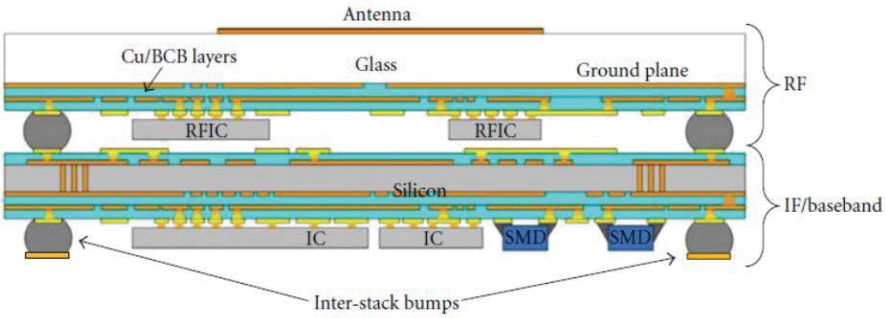


Figure 90: Schematic View of the 24 GHz 3D Sensor Node Integration Platform
The top RF module (on a glass substrate) with integrated antenna and the lower IF/baseband module on a silicon substrate [174].

A compact quad-band microwave module developed by EPCOS integrates SiP module 2.4G and 5G WLAN, Bluetooth, GPS, FM radio, and FM transmitter in a compact assembly measuring 9.5xll.9x1.2mm^3. Figure 91 shows the block diagram and a picture of the module [176]. This module is an example of 3D SiP integration featuring PoP assembly approach. The RF front-end circuits are fabricated in a separate chip, whereas digital functions are confined to another. Several filter circuits were implemented in LTCC. All IC's are integrated inside their own package (WLP) and all packaged components are further integrated in another package with 6-layer laminate interconnects and metal lids.

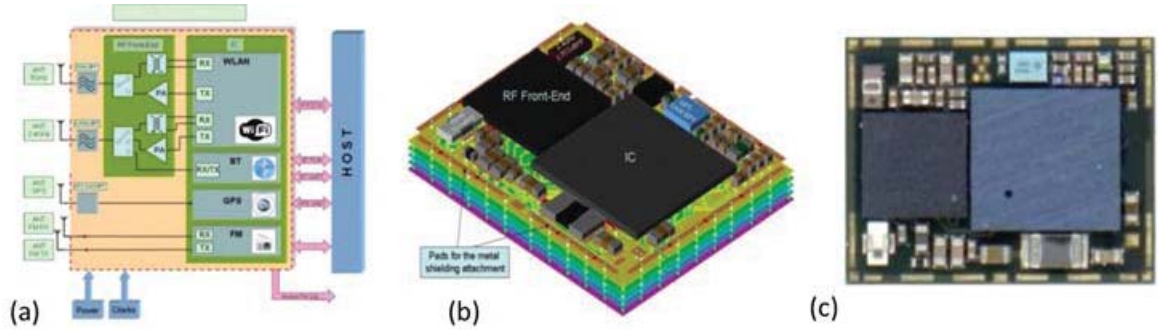


Figure 91: Compact Quad-Band Microwave Module Implemented in 3D SiP PoP Integration [176]

13.7 Integrated Radar Systems

Compact transceiver modules with integrated antennas were developed for commercial millimeter-wave radar systems using embedded wafer level ball grid array (eWLB) [164, 165], which is a slight variation of the eWLP discussed above. At mm-wave frequencies, the circuit parasitic elements such as impedance mismatches are difficult to control when circuits are assembled. Chip-level assemblies are therefore well suited for these applications where such parasitic circuit elements can be minimized. The eWLB-based SiP design makes use of metal lines on RDL layers for the fabrication of external circuit elements such as high-Q inductors and antennas. The interconnect line lengths can be kept quite small on the RDL when transceiver chips are connected together and to the antenna. Since the fabrication of eWLB package is wafer-based, a large number of modules can be fabricated simultaneously with minimal dimensional variations. Based on these technological strengths, a 77-GHz four-channel

transceiver module with four integrated antennas was fabricated in a eWLB package measuring 8mm × 8mm [177]. In comparison, the same system realized on PCB with wire bonded bare transceiver chips measured 4cm x 5cm. A picture of the integrated module is shown in Figure 92. The module integrates four half-wave dipole antennas that are realized using thin-film RDL of the eWLB. The antennas are connected to the transceiver chip using 100-Ω differential coplanar strip (CPS) lines realized in the RDL. These type of compact mm-wave modules are found in automotive radar sensor applications [178].

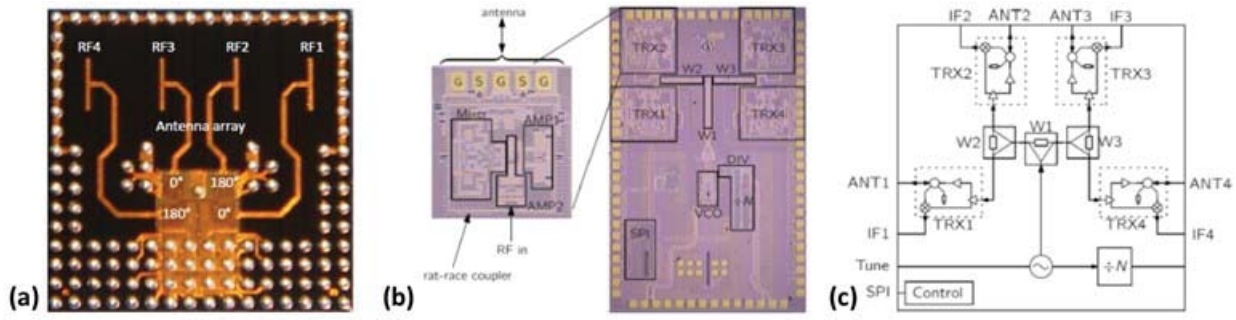


Figure 92: 77GHz, Four-channel Transceiver Module with Integrated Antennas
(a) Picture of the eWLB module, (b) the single-chip transceiver, and (c) block diagram of the four-channel transceiver [167].

Another compact radar system integrated in eWLB package is the 60 GHz frequency modulated continuous wave (FMCW) radar for industrial applications [179] and automotive radar applications [180]. The transceiver module had two dipole antennas in the package, as shown in Figure 93. The single chip transceiver contained a VCO with a buffer amplifier, mixers, power dividers, a vector modulator and IQ receiver. The maximum power output was 3dBm.

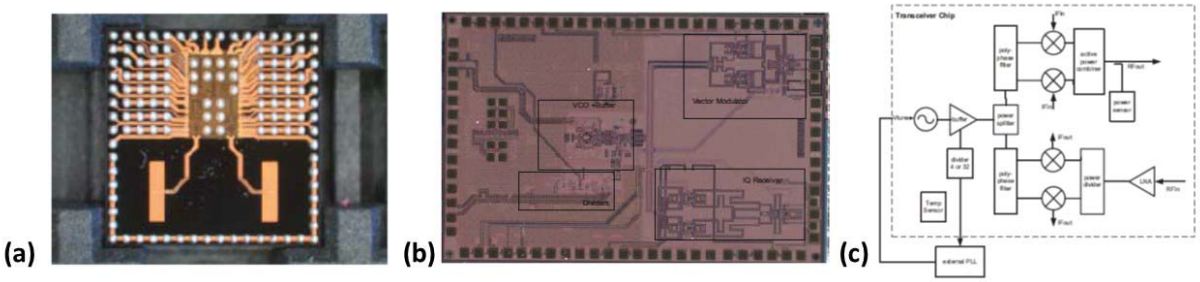


Figure 93: Compact Radar Module for 60 GHz FMCW Radar Applications
(a) The integrated module in eWLB package with 2 dipole antennas, (b) single chip transceiver chip, and (c) block diagram of the module [179]

13.8 Electro-optic Systems for Automobile Applications

Probably the most ambitious heterogeneous integration attempt was made by the Tohoku University for intelligent vehicle systems [181]. Two different types of interposer substrates were developed for this demonstration for integration of electronic functions on one and the optical functions on the other. These two substrates were then integrated to form a single module, as shown schematically in Figure 94. The electronic interposer, measuring 15mm x

8mm includes TSVs for interconnects as well as micro-channels for cooling fluid. Logic and memory chips were integrated on this substrate together with MEMS. The optical interposer, measuring 27mmx11mm, contained vertical-cavity surface-emitting laser (VCSEL) optical sources and photodiode (PD) detectors. The optical signal is routed within the interposer. The interconnect technologies used for this integration included micro-bumps and TSVs and Cu lateral interconnects [182, 183].

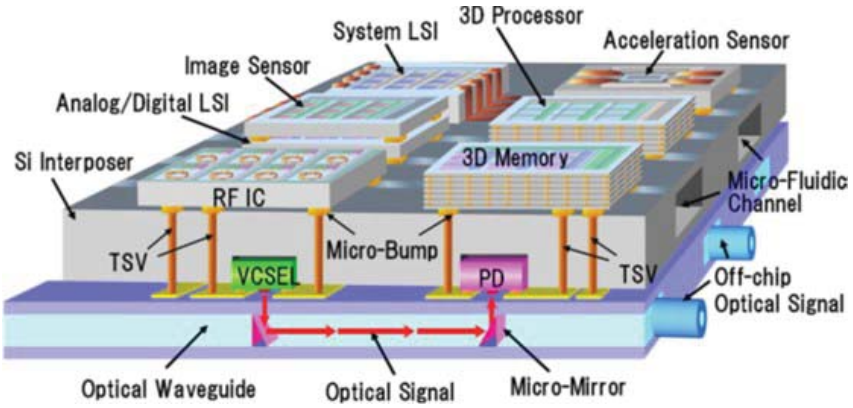


Figure 94: Conceptual 3D Integrated Sensor Module for Intelligent Vehicle Applications
This heterogeneous integration includes both electrical and optical interposer substrates [181]

13.9 Integrated Multi-Sensor Nodes

A conceptual military application of highly miniaturized sensor nodes is illustrated in Figure 95 [184]. In this application, each node contains multiple sensors heterogeneously integrated in a 3D SiP similar to those shown in Figure 89 and Figure 90. All nodes dispersed in a field of operation communicates with each other and with a base unit to download gathered information. The feasibility of such systems rely on batch fabrication of 3D integrated electronic and optical systems in small volumes [184].

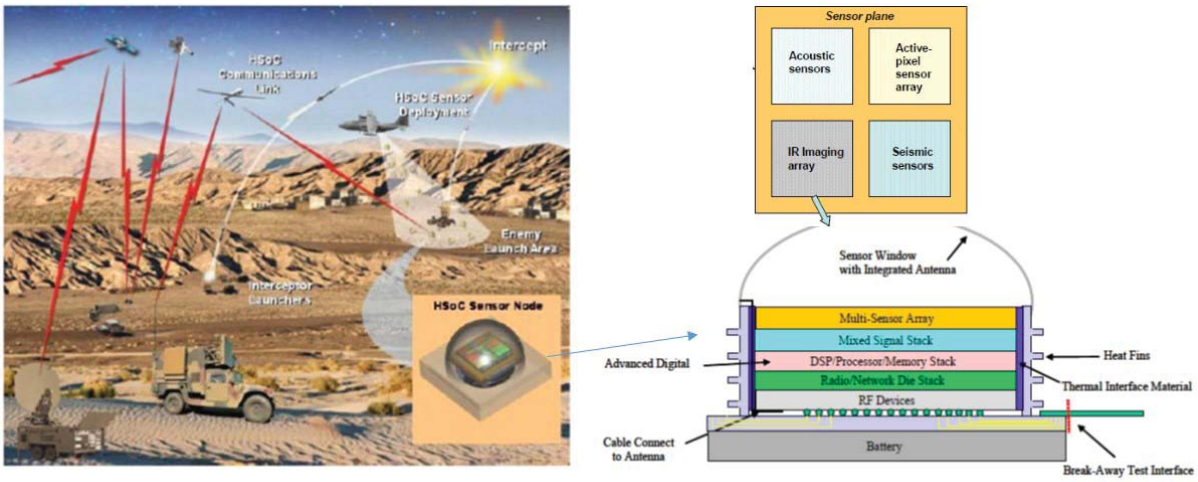


Figure 95: Highly Integrated Sensor Nodes for Integrated Classification-and-Decision-Information Extraction Capability from a Sensed Environment [184]

14. SUMMARY AND RECOMMENDATIONS

IEEE defines HI to be “the integration of separately manufactured components into a higher-level assembly that, in the aggregate, provides enhanced functionality and improved operating characteristics”. In this definition components are taken to mean any unit whether individual die, MEMS device, passive component and assembled package or sub-system that are integrated into a single package. The operating characteristics should also be taken in its broadest meaning including characteristics such as system level performance and cost of ownership.

HI is employed in the implementation of the MtM technology roadmap, which guides developments to increase the “functionality density” of electronic systems. Functional density can be increased by separately fabricating each function using the best available technology and then integrating these components in a compact unit. Functional density can be also increased by integrating different device types of devices (alternative materials, optical, mechanical etc.). The integration can take place on a chip or in a package.

HI makes use of the best available technologies and offers timely solutions to system-level innovations. A survey of the current status of HI shows that the technical approaches used for commercial products are highly specific to the applications. This is an indication that HI is more responsive to systems-pull than to technology-push. For an R&D organization such as ours with diverse technology portfolios and many system application opportunities, HI provides an important core competency platform that may act as a clearing house to match the system needs with the advanced technology solutions. In a technology development “food chain” extending from basic research to systems design, HI represents a midpoint, where it is close enough to both ends of the food chain to be responsive to innovations originating from either end of the chain. In its absence, the influence of scientific innovations take many decades to reach system designs. Similarly, system designs innovations rarely extend all the way to basic science end of the chain to influence their direction.

We have provided a general categorization of HI technologies to aid in the review. Details of integration methods under each category were examined and examples were provided. It was observed that the technology maturity levels are not linearly dependent on the increasing integration dimension, meaning that the availability of HI products seem to be relatively independent of the dimensional complexity. Some of the earliest and the most successful HI products (accelerometers and digital light processors) are a result of 3D integration. Some of the latest high volume HI products for high bandwidth memory-CPU integration involves only 2.5D. On the other hand, 2D integration of InP-CMOS circuits is still in development.

Commercial products are the main drivers behind the HI technology development. However, DoD and specifically AF has invested consistently in the development of integration technologies for microwave and high speed circuit applications. Initially, these activities were confined to integrating GaAs on Si substrates, but recently other device types (i.e. InP HBTs and GaN HEMTs) are being integrated on Si CMOS circuits. There are many direct military applications of the commercially developed HI technologies. In high speed computing applications, improvements in performance, size and cost due to the use of HI technologies will directly benefit military systems. For example, the ultra-high bandwidth FPGAs enabled by 2.5D

HI is a product that may find military applications. Other HI innovations in 3D SoC integration related to high speed imaging systems also have military applications. However, military-specific applications may require unique HI technology development effort. One of these may be the development of highly integrated microwave systems. Most commercial HI technologies address the complexity of interconnects in digital circuits, but the integration of microwave circuits is more challenging. There has been some progress in microwave multi-chip modules for commercial applications to reduce their size and cost by evolving them from 2D SiP to wafer level packaging (eWLP). This effort can be extended to full 3D SiP or 3D WLB by integrating digital and microwave circuits together with antennas [185]. This is especially important for mm-wave systems, whose design and fabrication is highly sensitive to electrical discontinuities that are inherent in hybrid solutions.

As a final note, it is worth re-emphasizing that HI requires no new basic technology development. It can be highly responsive to systems-level innovations and can be instrumental in implementing these innovations in a timely manner.

15. ACKNOWLEDGEMENTS

This technical review would not have been possible without the help and the critique of many colleagues in the AFRL Sensors Directorate Aerospace Components & Subsystems Division (RYD). The contributions of the following are gratefully acknowledged:

Paul Watson; Integrated Circuits & Microsystems Branch (RYDI)
Steve Hary; RYD
Robert Fitch; Devices for Sensing Branch (RYDD)
Tony Quach; RYDI
Kelson Chabak; RYDD
Jim Gillespie; RYDD
Brad Paul; RYDI

16. REFERENCES

[1] P. a. M. T. S. IEEE Components. (2017). *Heterogeneous Integration Roadmap* Available: http://cpmt.ieee.org/technology/heterogeneous-integration-roadmap.html

[2] J. A. Carballo, W. T. J. Chan, P. A. Gargini, A. B. Kahng, and S. Nath, "ITRS 2.0: Toward a re-framing of the Semiconductor Technology Roadmap," in *2014 IEEE 32nd International Conference on Computer Design (ICCD)*, 2014, pp. 139-146.

[3] M. M. Waldrop, " chips are down for Moore's law," *Nature,* vol. 530, pp. 145-147, 2016.

[4] R. Courtland, "Transistors could stop shrinking in 2021," *IEEE Spectrum,* vol. 53, pp. 9-11, 2016.

[5] D. Harame, "High performance BiCMOS process integration: trends, issues, and future directions," in *Proceedings of the 1997 Bipolar/BiCMOS Circuits and Technology Meeting*, 1997, pp. 36-43.

[6] S. Subbana, G. Freeman, D. Ahlgren, B. Jagannathan, D. Greenberg, J. Johnson*, et al.*, "Review of silicon-germanium BICMOS technology after 4 years of production and future directions," in *GaAs IC Symposium. IEEE Gallium Arsenide Integrated Circuits Symposium. 22nd Annual Technical Digest 2000. (Cat. No.00CH37084)*, 2000, pp. 7-10.

[7] A. Das, M. Huang, J. Mondal, D. Kaczman, C. Shurboff, and S. Cosentino, "Review of SiGe process technology and its impact on RFIC design," in *2002 IEEE Radio Frequency Integrated Circuits (RFIC) Symposium. Digest of Papers (Cat. No.02CH37280)*, 2002, pp. 325-328.

[8] T. Sakurai, "A review on low-voltage BiCMOS circuits and a BiCMOS vs. CMOS speed comparison," in *[1992] Proceedings of the 35th Midwest Symposium on Circuits and Systems*, 1992, pp. 564-567 vol.1.

[9] D. L. Plumton, F. J. Morris, and J. Y. Yang, "Method to integrate HBTs and FETs," ed: Google Patents, 1993.

[10] Y. F. Yang, C. C. Hsu, and E. S. Yang, "Integration of GalnP/GaAs heterojunction bipolar transistors and high electron mobility transistors," *IEEE Electron Device Letters,* vol. 17, pp. 363-365, 1996.

[11] M. C. F. Chang, P. M. Asbeck, and R. L. Pierson, "Planar HBT-FET Device," ed: Google Patents, 1993.

[12] K. Itakura, Y. Shimamoto, T. Ueda, S. Katsu, and D. Ueda, "A GaAs Bi-FET technology for large scale integration," in *International Technical Digest on Electron Devices Meeting*, 1989, pp. 389-392.

[13] D. C. Streit, D. K. Umemoto, K. W. Kobayashi, and A. K. Oki, "Monolithic HEMT-HBT integration by selective MBE," *IEEE Transactions on Electron Devices,* vol. 42, pp. 618-623, 1995.

[14] W. J. Ho, M. F. Chang, S. M. Beccue, P. J. Zampardi, J. Yu, A. Sailer*, et al.*, "A GaAs BiFET LSI technology," in *Proceedings of 1994 IEEE GaAs IC Symposium*, 1994, pp. 47-50.

[15] A. G. Metzger, R. Ramanathan, J. Li, H. C. Sun, C. Cismaru, H. Shao*, et al.*, "An InGaP/GaAs Merged HBT-FET (BiFET) Technology and Applications to the Design of Handset Power Amplifiers," *IEEE Journal of Solid-State Circuits,* vol. 42, pp. 2137-2148, 2007.

[16] A. Gupta, B. Peatman, M. Shokrani, W. Krystek, and T. Arell, "InGaP-Plus - A major advance in GaAs HBT Technology," in *2006 IEEE Compound Semiconductor Integrated Circuit Symposium*, 2006, pp. 179-182.
[17] B. Bayraktaroglu, "Method of integrating heterojunction bipolar transistors with heterojunction FETs and PIN diodes," ed: Google Patents, 1995.
[18] B. Bayraktaroglu, "GaAs HBT's for microwave integrated circuits," *Proceedings of the IEEE,* vol. 81, pp. 1762-1785, 1993.
[19] L. Madden, E. Wu, N. Kim, B. Banijamali, K. Abugharbieh, S. Ramalingam*, et al.*, "Advancing high performance heterogeneous integration through die stacking," in *2012 Proceedings of the ESSCIRC (ESSCIRC)*, 2012, pp. 18-24.
[20] J. H. Lau, "Evolution and outlook of TSV and 3D IC/Si integration," in *2010 12th Electronics Packaging Technology Conference*, 2010, pp. 560-570.
[21] M. Sekiguchi, H. Numata, N. Sato, T. Shirakawa, M. Matsuo, H. Yoshikawa*, et al.*, "Novel low cost integration of through chip interconnection and application to CMOS image sensor," in *56th Electronic Components and Technology Conference 2006*, 2006, p. 8 pp.
[22] D. Henry, F. Jacquet, M. Neyret, X. Baillin, T. Enot, V. Lapras*, et al.*, "Through silicon vias technology for CMOS image sensors packaging," in *2008 58th Electronic Components and Technology Conference*, 2008, pp. 556-562.
[23] E. Beyne, "The 3-D Interconnect Technology Landscape," *IEEE Design & Test,* vol. 33, pp. 8-20, 2016.
[24] M. Murugesan, J. C. Bea, T. Fukushima, T. Konno, K. Kiyoyama, W. C. Jeong*, et al.*, "Cu lateral interconnects formed between 100-um-thick self-assembled chips on flexible substrates," in *2009 59th Electronic Components and Technology Conference*, 2009, pp. 1496-1501.
[25] K. W. Lee, S. Kanno, Y. Ohara, K. Kiyoyama, J. C. Bea, T. Fukushima*, et al.*, "Heterogeneous integration technology for MEMS-LSI multi-chip module," in *2009 IEEE International Conference on 3D System Integration*, 2009, pp. 1-6.
[26] E. Beyne, "The rise of the 3rd dimension for system intergration," in *2006 International Interconnect Technology Conference*, 2006, pp. 1-5.
[27] A. R. Alvarez, "BiCMOS-has the promise been fulfilled?," in *International Electron Devices Meeting 1991 [Technical Digest]*, 1991, pp. 355-358.
[28] H. Shichijo, R. J. Matyi, and A. H. Taddiken, "Co-integration of GaAs MESFET and Si CMOS circuits," *IEEE Electron Device Letters,* vol. 9, pp. 444-446, 1988.
[29] Q. C. Streit, D. K. Umemoto, K. W. Kobayashi, and A. K. Oki, "Monolithic HEMT-HBT integration for novel microwave circuit applications," in *Proceedings of 1994 IEEE GaAs IC Symposium*, 1994, pp. 329-332.
[30] K. W. Kobayashi, A. K. Oki, D. K. Umemoto, T. R. Block, and D. C. Streit, "A novel self-oscillating HEMT-HBT cascode VCO-mixer using an active tunable inductor," *IEEE Journal of Solid-State Circuits,* vol. 33, pp. 870-876, 1998.
[31] K. W. Kobayashi, D. C. Streit, A. K. Oki, D. K. Umemoto, and T. R. Block, "A novel monolithic linearized HEMT LNA using HBT tuneable active feedback," in *1996 IEEE MTT-S International Microwave Symposium Digest*, 1996, pp. 1217-1220 vol.3.
[32] K. W. Kobayashi, D. K. Umemoto, T. R. Block, A. K. Oki, and D. C. Streit, "A monolithically integrated HEMT-HBT low noise high linearity variable gain amplifier," *IEEE Journal of Solid-State Circuits,* vol. 31, pp. 714-718, 1996.

[33] K. W. Kobayashi, H. Wang, R. Lai, L. T. Tran, T. R. Block, P. H. Liu, *et al.*, "An InP HEMT W-band amplifier with monolithically integrated HBT bias regulation," *IEEE Microwave and Guided Wave Letters,* vol. 7, pp. 222-224, 1997.

[34] H. Wang, R. Lai, L. Tran, J. Cowles, Y. C. Chen, E. W. Lin, *et al.*, "A single-chip 94-GHz frequency source using InP-based HEMT-HBT integration technology," in *1998 IEEE MTT-S International Microwave Symposium Digest (Cat. No.98CH36192)*, 1998, pp. 219-222 vol.1.

[35] S. Hyunchol, S. Jeong-Hwan, and K. Young-Se, "Monolithic integration of AlGaAs/GaAs HBT and GaAs junction-gate floated electron channel field effect transistor using selective MOCVD growth," *IEEE Microwave and Guided Wave Letters,* vol. 6, pp. 317-319, 1996.

[36] K. Kiziloglu, M. W. Yung, S. Hsiang-Chi, S. Thomas, M. B. Kardos, R. H. Walden, *et al.*, "InP-based high sensitivity pin/HEMT/HBT monolithic integrated optoelectronic receiver," in *Conference Proceedings. 1998 International Conference on Indium Phosphide and Related Materials (Cat. No.98CH36129)*, 1998, pp. 443-446.

[37] H. Y. Chang, Y. C. Liu, S. H. Weng, C. H. Lin, Y. L. Yeh, and Y. C. Wang, "Design and Analysis of a DC-43.5-GHz Fully Integrated Distributed Amplifier Using GaAs HEMT-HBT Cascode Gain Stage," *IEEE Transactions on Microwave Theory and Techniques,* vol. 59, pp. 443-455, 2011.

[38] Y. Kyounghoon, A. L. Gutierrez-Aitken, Z. Xiangkun, G. I. Haddad, and P. Bhattacharya, "Design, modeling, and characterization of monolithically integrated InP-based (1.55 μm) high-speed (24 Gb/s) p-i-n/HBT front-end photoreceivers," *Journal of Lightwave Technology,* vol. 14, pp. 1831-1839, 1996.

[39] L. M. Lunardi, S. Chandrasekhar, A. H. Gnauck, and C. A. Burrus, "20-Gb/s monolithic p-i-n/HBT photoreceiver module for 1.55-um applications," *IEEE Photonics Technology Letters,* vol. 7, pp. 1201-1203, 1995.

[40] S. Chandrasekhar, L. M. Lunardi, A. H. Gnauck, R. A. Hamm, and G. J. Qua, "High-speed monolithic p-i-n/HBT and HPT/HBT photoreceivers implemented with simple phototransistor structure," *IEEE Photonics Technology Letters,* vol. 5, pp. 1316-1318, 1993.

[41] A. L. Gutierrez-Aitken, K. Yang, X. Zhang, G. I. Haddad, P. Bhattacharya, and L. M. Lunardi, "16-GHz bandwidth InAlAs-InGaAs monolithically integrated p-i-n/HBT photoreceiver," *IEEE Photonics Technology Letters,* vol. 7, pp. 1339-1341, 1995.

[42] M. Bitter, R. Bauknecht, W. Hunziker, and H. Melchior, "Monolithic InGaAs-InP p-i-n/HBT 40-Gb/s optical receiver module," *IEEE Photonics Technology Letters,* vol. 12, pp. 74-76, 2000.

[43] E. Alekseev, D. Pavlidis, M. Dickmann, and T. Hackbarth, "W-band InGaAs/InP PIN diode monolithic integrated switches," in *GaAs IC Symposium IEEE Gallium Arsenide Integrated Circuit Symposium. 18th Annual Technical Digest 1996*, 1996, pp. 285-288.

[44] R. Fischer, H. Morkoç, D. A. Neumann, H. Zabel, C. Choi, N. Otsuka, *et al.*, "Material properties of high-quality GaAs epitaxial layers grown on Si substrates," *Journal of Applied Physics,* vol. 60, pp. 1640-1647, 1986.

[45] W. I. Wang, "Molecular beam epitaxial growth and material properties of GaAs and AlGaAs on Si (100)," *Applied Physics Letters,* vol. 44, pp. 1149-1151, 1984.

[46] R. Fischer, D. Neuman, H. Zabel, H. Morkoç, C. Choi, and N. Otsuka, "Dislocation reduction in epitaxial GaAs on Si(100)," *Applied Physics Letters,* vol. 48, pp. 1223-1225, 1986.
[47] P. N. Uppal and H. Kroemer, "Molecular beam epitaxial growth of GaAs on Si(211)," *Journal of Applied Physics,* vol. 58, pp. 2195-2203, 1985.
[48] M. Akiyama, Y. Kawarada, and K. Kaminishi, "Growth of GaAs on Si by MOVCD," *Journal of Crystal Growth,* vol. 68, pp. 21-26, 9/1/ 1984.
[49] T. Soga, S. Hattori, S. Sakai, and M. Umeno, "Epitaxial growth and material properties of GaAs on Si grown by MOCVD," *Journal of Crystal Growth,* vol. 77, pp. 498-502, 1986/09/01 1986.
[50] R. Fischer, N. Chand, W. Kopp, H. Morkoç, L. P. Erickson, and R. Youngman, "GaAs bipolar transistors grown on (100) Si substrates by molecular beam epitaxy," *Applied Physics Letters,* vol. 47, pp. 397-399, 1985.
[51] R. Fischer, W. Kopp, H. Morkoç, M. Pion, A. Specht, G. Burkhart, *et al.*, "Low threshold laser operation at room temperature in GaAs/(Al,Ga)As structures grown directly on (100)Si," *Applied Physics Letters,* vol. 48, pp. 1360-1361, 1986.
[52] R. J. Fischer, W. F. Kopp, J. S. Gedymin, and H. Morkoc, "Properties of MODFET's grown on Si substrates at DC and microwave frequencies," *IEEE Transactions on Electron Devices,* vol. 33, pp. 1407-1412, 1986.
[53] H. K. Choi, J. P. Mattia, G. W. Turner, and B. Y. Tsaur, "Monolithic integration of GaAs/AlGaAs LED and Si driver circuit," *IEEE Electron Device Letters,* vol. 9, pp. 512-514, 1988.
[54] H. K. Choi, G. W. Turner, and T. Bor-Yeu, "Monolithic integration of Si MOSFET's and GaAs MESFET's," *IEEE Electron Device Letters,* vol. 7, pp. 241-243, 1986.
[55] R. Fischer, T. Henderson, J. Klem, W. Kopp, C. K. Peng, H. Morkoç, *et al.*, "Monolithic integration of GaAs/AlGaAs modulation-doped field- effect transistors and N- m etal-oxide-semiconductor silicon circuits," *Applied Physics Letters,* vol. 47, pp. 983-985, 1985.
[56] H. K. Choi, G. W. Turner, T. H. Windhorn, and T. Bor-Yeu, "Monolithic integration of GaAs/AlGaAs double-heterostructure LED's and Si MOSFET's," *IEEE Electron Device Letters,* vol. 7, pp. 500-502, 1986.
[57] E. D. Cohen, "The MIMIC Program; A Retrospective," *IEEE Microwave Magazine,* vol. 13, pp. 77-88, 2012.
[58] R. T. Kemerley, H. B. Wallace, and M. N. Yoder, "Impact of wide bandgap microwave devices on DoD systems," *Proceedings of the IEEE,* vol. 90, pp. 1059-1064, 2002.
[59] S. Raman, T. H. Chang, C. L. Dohrman, and M. J. Rosker, "The DARPA COSMOS program: The convergence of InP and Silicon CMOS technologies for high-performance mixed-signal," in *2010 22nd International Conference on Indium Phosphide and Related Materials (IPRM)*, 2010, pp. 1-5.
[60] S. Raman, C. L. Dohrman, and T. H. Chang, "The DARPA Diverse Accessible Heterogeneous Integration (DAHI) program: Convergence of compound semiconductor devices and silicon-enabled architectures," in *2012 IEEE International Symposium on Radio-Frequency Integration Technology (RFIT)*, 2012, pp. 1-6.
[61] D. S. Green, C. L. Dohrman, A. S. Kane, and T. H. Chang, "Materials and Integration Strategies for Modern RF Integrated Circuits," in *2014 IEEE Compound Semiconductor Integrated Circuit Symposium (CSICS)*, 2014, pp. 1-4.

[62] C. L. D. Daniel S. Green, Jeffrey Demmin, Yan Zheng, and Tsu-Hsi Chang, "A revolution on the horizon from DARPA," *IEEE Microwave Magazine,* pp. 44-59, 2017.
[63] E. Johnson, "Physical limitations on frequency and power parameters of transistors," *RCA review,* vol. 26, pp. 163-177, 1965.
[64] W. K. Liu, D. Lubyshev, J. M. Fastenau, Y. Wu, M. T. Bulsara, E. A. Fitzgerald*, et al.*, "Monolithic integration of InP-based transistors on Si substrates using MBE," *Journal of Crystal Growth,* vol. 311, pp. 1979-1983, 3/15/ 2009.
[65] K. Herrick, T. Kazior, A. Liu, D. I. Loubychev, J. M. Fastenau, M. Urteaga*, et al.*, "Direct Growth of III-V Devices on Silicon," *MRS Proceedings,* vol. 1068, 2008/001/001 2008.
[66] T. E. Kazior, "Beyond CMOS: heterogeneous integration of III–V devices, RF MEMS and other dissimilar materials/devices with Si CMOS to create intelligent microsystems," *Philosophical transactions. Series A, Mathematical, physical, and engineering sciences,* vol. 372, p. 20130105, 2014.
[67] T. E. Kazior, J. R. LaRoche, D. Lubyshev, J. M. Fastenau, W. K. Liu, M. Urteaga*, et al.*, "A high performance differential amplifier through the direct monolithic integration of InP HBTs and Si CMOS on silicon substrates," in *2009 IEEE MTT-S International Microwave Symposium Digest*, 2009, pp. 1113-1116.
[68] T. E. Kazior, R. Chelakara, W. Hoke, J. Bettencourt, T. Palacios, and H. S. Lee, "High Performance Mixed Signal and RF Circuits Enabled by the Direct Monolithic Heterogeneous Integration of GaN HEMTs and Si CMOS on a Silicon Substrate," in *2011 IEEE Compound Semiconductor Integrated Circuit Symposium (CSICS)*, 2011, pp. 1-4.
[69] W. E. Hoke, R. V. Chelakara, J. P. Bettencourt, T. E. Kazior, J. R. LaRoche, T. D. Kennedy*, et al.*, "Monolithic integration of silicon CMOS and GaN transistors in a current mirror circuit," *Journal of Vacuum Science & Technology B, Nanotechnology and Microelectronics: Materials, Processing, Measurement, and Phenomena,* vol. 30, p. 02B101, 2012.
[70] J. W. Chung, J. k. Lee, E. L. Piner, and T. Palacios, "Seamless On-Wafer Integration of Si(100) MOSFETs and GaN HEMTs," *IEEE Electron Device Letters,* vol. 30, pp. 1015-1017, 2009.
[71] C. Kamesh, J. M. Michael, L. D. Carl, and A. F. Eugene, "Monolithic CMOS-compatible AlGaInP visible LED arrays on silicon on lattice-engineered substrates (SOLES)," *Semiconductor Science and Technology,* vol. 22, p. 29, 2007.
[72] M. L. Schuette, A. J. Green, K. D. Leedy, J. P. McCandless, and G. H. Jessen, "Heterogeneous integration of low-temperature metal-oxide TFTs," 2017, pp. 1010512-1010512-7.
[73] H. Baltes, O. Brand, A. Hierlemann, D. Lange, and C. Hagleitner, "CMOS MEMS - present and future," in *Technical Digest. MEMS 2002 IEEE International Conference. Fifteenth IEEE International Conference on Micro Electro Mechanical Systems (Cat. No.02CH37266)*, 2002, pp. 459-466.
[74] A. C. T. Harrie, R. Walter De, and B. Eric, "MEMS for wireless communications: 'from RF-MEMS components to RF-MEMS-SiP'," *Journal of Micromechanics and Microengineering,* vol. 13, p. S139, 2003.
[75] H. C. Kim and K. Chun, "RF MEMS technology," *IEEJ Transactions on Electrical and Electronic Engineering,* vol. 2, pp. 249-261, 2007.

[76] G. M. Rebeiz and J. B. Muldavin, "RF MEMS switches and switch circuits," *IEEE Microwave Magazine,* vol. 2, pp. 59-71, 2001.
[77] A. Witvrouw, "CMOS–MEMS integration today and tomorrow," *Scripta Materialia,* vol. 59, pp. 945-949, 11// 2008.
[78] G. K. Fedder, R. T. Howe, T. J. K. Liu, and E. P. Quevy, "Technologies for Cofabricating MEMS and Electronics," *Proceedings of the IEEE,* vol. 96, pp. 306-322, 2008.
[79] J. H. Smith, S. Montague, J. J. Sniegowski, J. R. Murray, and P. J. McWhorter, "Embedded micromechanical devices for the monolithic integration of MEMS with CMOS," in *Proceedings of International Electron Devices Meeting*, 1995, pp. 609-612.
[80] J. H. Smith, S. Montague, J. J. Sniegowski, J. R. Murray, R. P. Manginell, P. J. McWhorter*, et al.*, "Characterization of the embedded micromechanical device approach to the monolithic integration of MEMS with CMOS," 1996, pp. 306-314.
[81] J. A. Yasaitis, M. Judy, T. Brosnihan, P. M. Garone, N. Pokrovskiy, D. Sniderman*, et al.*, "A modular process for integrating thick polysilicon MEMS devices with sub-micron CMOS," 2003, pp. 145-154.
[82] J. M. Tsai and G. K. Fedder, "Mechanical noise-limited CMOS-MEMS accelerometers," in *18th IEEE International Conference on Micro Electro Mechanical Systems, 2005. MEMS 2005.*, 2005, pp. 630-633.
[83] M. Kaynak, K. E. Ehwald, J. Drews, R. Scholz, F. Korndörfer, D. Knoll*, et al.*, "BEOL embedded RF-MEMS switch for mm-wave applications," in *2009 IEEE International Electron Devices Meeting (IEDM)*, 2009, pp. 1-4.
[84] H. Lakdawala, X. Zhu, H. Luo, S. Santhanam, R. Carley, and G. Fedder, "Micromachined High-Q Inductors in 0.18μm Cu Interconnect Low-K CMOS Process."
[85] G. M. Rebeiz, T. Guan-Leng, and J. S. Hayden, "RF MEMS phase shifters: design and applications," *IEEE Microwave Magazine,* vol. 3, pp. 72-81, 2002.
[86] E. R. Brown, "RF-MEMS switches for reconfigurable integrated circuits," *IEEE Transactions on Microwave Theory and Techniques,* vol. 46, pp. 1868-1880, 1998.
[87] K. J. Herrick, G. Jerinic, R. P. Molfino, S. M. Lardizabal, and B. Pillans, "S-Ku band intelligent amplifier microsystem."
[88] E. A. Sovero, R. Mihailovich, D. S. Deakin, J. A. Higgins, J. J. Yao, J. F. DeNatale*, et al.*, "Monolithic GaAs PHEMT MMICs integrated with high performance MEMS microrelays," in *1999 SBMO/IEEE MTT-S International Microwave and Optoelectronics Conference*, 1999, pp. 257-260 vol. 1.
[89] A. Ansari, V. J. Gokhale, J. Roberts, and M. Rais-Zadeh, "Monolithic integration of GaN-based micromechanical resonators and HEMTs for timing applications," in *2012 International Electron Devices Meeting*, 2012, pp. 15.5.1-15.5.4.
[90] L. Binet and D. Gourier, "Origin of the blue luminescence of β-Ga2O3," *Journal of Physics and Chemistry of Solids,* vol. 59, pp. 1241-1249, 8// 1998.
[91] B. W. Bahr, L. C. Popa, and D. Weinstein, "1GHz GaN-MMIC monolithically integrated MEMS-based oscillators," in *2015 IEEE International Solid-State Circuits Conference - (ISSCC) Digest of Technical Papers*, 2015, pp. 1-3.
[92] R. R. Tummala, "Multichip packaging-a tutorial," *Proceedings of the IEEE,* vol. 80, pp. 1924-1941, 1992.
[93] C. Bartlett, J. Segelken, and N. Teneketges, "Multichip Packaging Design for VLSI-Based Systems," *IEEE Transactions on Components, Hybrids, and Manufacturing Technology,* vol. 10, pp. 647-653, 1987.

[94] T. A. Midford, J. J. Wooldridge, and R. L. Sturdivant, "The evolution of packages for monolithic microwave and millimeter-wave circuits," *IEEE Transactions on Antennas and Propagation,* vol. 43, pp. 983-991, 1995.
[95] B. A. Kopp, M. Borkowski, and G. Jerinic, "Transmit/receive modules," *IEEE Transactions on Microwave Theory and Techniques,* vol. 50, pp. 827-834, 2002.
[96] Y. Mancuso, P. Gremillet, and P. Lacomme, "T/R- Modules Technological and Technical Trends for Phased Array Antennas," in *2006 IEEE MTT-S International Microwave Symposium Digest*, 2006, pp. 614-617.
[97] K. Fujii, Y. Hara, Y. Shibuya, T. Sakai, and Y. Takano, "A highly integrated T/R module for active phased array antennas," in *1998 IEEE Radio Frequency Integrated Circuits (RFIC) Symposium. Digest of Papers (Cat. No.98CH36182)*, 1998, pp. 77-80.
[98] P. Schuh, H. Sledzik, R. Reber, K. Widmer, A. Fleckenstein, B. Schweizer*, et al.*, "T/R-module technologies today and future trends," in *The 40th European Microwave Conference*, 2010, pp. 1540-1543.
[99] A. Bettidi, A. Cetronio, M. Cicolani, C. Costrini, C. Lanzieri, S. Maccaroni*, et al.*, "X-band T/R module in state-of-the-art GaN technology," in *2009 European Radar Conference (EuRAD)*, 2009, pp. 258-261.
[100] P. Schuh, H. Sledzik, R. Reber, A. Fleckenstein, R. Leberer, M. Oppermann*, et al.*, "GaN MMIC based T/R-Module Front-End for X-Band Applications," in *2008 European Microwave Integrated Circuit Conference*, 2008, pp. 274-277.
[101] T. Vähä-Heikkilä, "RF to millimeter wave integration and module technologies," 2015, pp. 94340S-94340S-8.
[102] J. Wu, D. Coller, M. J. Anderson, and G. Guth, "RF SiP technology: integration and innovation," in *International conference on compound semiconductor manufacturing*, 2004.
[103] M. J. Franco, "Mobile handset power amplifiers," *IEEE Microwave Magazine,* vol. 10, pp. 16-19, 2009.
[104] F. v. Straten, B. Smolders, A. v. Zuijlen, and R. Ooijman, "Multiband cellular RF solutions," *IEEE Journal of Solid-State Circuits,* vol. 39, pp. 1598-1604, 2004.
[105] G. Zhang, S. Chang, and A. Wang, "WCDMA PCS Handset Front End Module," in *2006 IEEE MTT-S International Microwave Symposium Digest*, 2006, pp. 304-307.
[106] F. Roozeboom, A. L. A. M. Kemmeren, J. F. C. Verhoeven, F. C. van den Heuvel, J. Klootwijk, H. Kretschman*, et al.*, "Passive and heterogeneous integration towards a Si-based System-in-Package concept," *Thin Solid Films,* vol. 504, pp. 391-396, 5/10/ 2006.
[107] A. Gutierrez-Aitken, P. Chang-Chien, D. Scott, K. Hennig, E. Kaneshiro, P. Nam*, et al.*, "Advanced Heterogeneous Integration of InP HBT and CMOS Si Technologies," in *2010 IEEE Compound Semiconductor Integrated Circuit Symposium (CSICS)*, 2010, pp. 1-4.
[108] B. Oyama, D. Ching, K. Thai, A. Gutierrez-Aitken, and V. J. Patel, "InP HBT/Si CMOS-Based 13-b 1.33-Gsps Digital-to-Analog Converter With > 70-dB SFDR," *IEEE Journal of Solid-State Circuits,* vol. 48, pp. 2265-2272, 2013.
[109] Y. C. Wu, M. Watanabe, and T. LaRocca, "InP HBT/GaN HEMT/Si CMOS heterogeneous integrated Q-band VCO-amplifier chain," in *2015 IEEE Radio Frequency Integrated Circuits Symposium (RFIC)*, 2015, pp. 39-42.

[110] J. C. Li, Y. Royter, P. R. Patterson, T. Hussain, J. R. Duvall, M. C. Montes, *et al.*, "Heterogeneous wafer-scale integration of 250nm, 300GHz InP DHBTs with a 130nm RF-CMOS technology," in *2008 IEEE International Electron Devices Meeting*, 2008, pp. 1-3.
[111] P. Watson, A. Mattamana, R. Gilbert, Y. Royter, M. Lau, I. Valles, *et al.*, "A wide-bandwidth W-band LNA in InP/Si BiCMOS technology," in *2014 IEEE MTT-S International Microwave Symposium (IMS2014)*, 2014, pp. 1-4.
[112] A. Carter, M. Urteaga, P. Rowell, S. Hong, R. Patti, C. Petteway, *et al.*, "Wafer-Scale InP/Si CMOS 3D Integration Using Low-Temperature Oxide Bonding," *Compound Semiconductor Week (CSW, Santa Barbara, CA,* pp. 493-494, 2015.
[113] M. Urteaga, A. Carter, Z. Griffith, R. Pierson, J. Bergman, A. Arias, *et al.*, "THz bandwidth InP HBT technologies and heterogeneous integration with Si CMOS," in *2016 IEEE Bipolar/BiCMOS Circuits and Technology Meeting (BCTM)*, 2016, pp. 35-41.
[114] M. Lisker, A. Trusch, A. Krüger, M. Fraschke, P. Kulse, Y. Borokhovych, *et al.*, "InP-Si BiCMOS Heterointegration Using a Substrate Transfer Process," *ECS Transactions,* vol. 53, pp. 245-254, May 3, 2013 2013.
[115] M. Lisker, A. Trusch, A. Krüger, M. Fraschke, P. Kulse, Y. Borokhovych, *et al.*, "InP-Si BiCMOS Heterointegration Using a Substrate Transfer Process," *ECS Journal of Solid State Science and Technology,* vol. 3, pp. P17-P20, January 1, 2014 2014.
[116] M. Lisker, A. Trusch, A. Krüger, M. Fraschke, P. Kulse, S. Marschmeyer, *et al.*, "(Invited) Combining SiGe BiCMOS and InP Processing in an on-top of Chip Integration Approach," *ECS Transactions,* vol. 64, pp. 177-194, August 12, 2014 2014.
[117] N. G. Weimann, D. Stoppel, M. I. Schukfeh, M. Hossain, T. Al-Sawaf, B. Janke, *et al.*, "SciFab –a wafer-level heterointegrated InP DHBT/SiGe BiCMOS foundry process for mm-wave applications," *physica status solidi (a),* vol. 213, pp. 909-916, 2016.
[118] T. Kraemer, I. Ostermay, T. Jensen, T. K. Johansen, F. J. Schmueckle, A. Thies, *et al.*, "InP-DHBT-on-BiCMOS Technology With <formula formulatype="inline"> <img src="/images/tex/20715.gif" alt="f_{T}/f_{\max }"> </formula> of 400/350 GHz for Heterogeneous Integrated Millimeter-Wave Sources," *IEEE Transactions on Electron Devices,* vol. 60, pp. 2209-2216, 2013.
[119] T. Jensen, T. Al-Sawaf, M. Lisker, S. Glisic, M. Elkhouly, T. Kraemer, *et al.*, "A 164 GHz hetero-integrated source in InP-on-BiCMOS technology," in *2013 European Microwave Integrated Circuit Conference*, 2013, pp. 244-247.
[120] M. Hossain, T. Kraemer, I. Ostermay, T. Jensen, B. Janke, Y. Borokhovych, *et al.*, "A 246 GHz Hetero-Integrated Frequency Source in InP-on-BiCMOS Technology," *IEEE Microwave and Wireless Components Letters,* vol. 24, pp. 469-471, 2014.
[121] M. Hossain, N. Weimann, M. Lisker, C. Meliani, B. Tillack, V. Krozer, *et al.*, "A 330 GHz hetero-integrated source in InP-on-BiCMOS technology," in *2015 IEEE MTT-S International Microwave Symposium*, 2015, pp. 1-4.
[122] S. William, "Semiconductive wafer and method of making the same," ed: Google Patents, 1962.
[123] J. U. Knickerbocker, P. S. Andry, E. Colgan, B. Dang, T. Dickson, X. Gu, *et al.*, "2.5D and 3D technology challenges and test vehicle demonstrations," in *2012 IEEE 62nd Electronic Components and Technology Conference*, 2012, pp. 1068-1076.

[124] J. U. Knickerbocker, P. S. Andry, B. Dang, R. R. Horton, C. S. Patel, R. J. Polastre, *et al.*, "3D silicon integration," in *2008 58th Electronic Components and Technology Conference*, 2008, pp. 538-543.
[125] C. C. Lee, C. Hung, C. Cheung, P. F. Yang, C. L. Kao, D. L. Chen, *et al.*, "An Overview of the Development of a GPU with Integrated HBM on Silicon Interposer," in *2016 IEEE 66th Electronic Components and Technology Conference (ECTC)*, 2016, pp. 1439-1444.
[126] M. Hutton, "Stratix- 10: 14nm FPGA delivering 1GHz," in *2015 IEEE Hot Chips 27 Symposium (HCS)*, 2015, pp. 1-24.
[127] R. Mahajan, R. Sankman, N. Patel, D. W. Kim, K. Aygun, Z. Qian, *et al.*, "Embedded Multi-die Interconnect Bridge (EMIB) -- A High Density, High Bandwidth Packaging Interconnect," in *2016 IEEE 66th Electronic Components and Technology Conference (ECTC)*, 2016, pp. 557-565.
[128] Y. Kurita, S. Matsui, N. Takahashi, K. Soejima, M. Komuro, M. Itou, *et al.*, "A 3D Stacked Memory Integrated on a Logic Device Using SMAFTI Technology," in *2007 Proceedings 57th Electronic Components and Technology Conference*, 2007, pp. 821-829.
[129] C.-T. Ko and K.-N. Chen, "Low temperature bonding technology for 3D integration," *Microelectronics Reliability,* vol. 52, pp. 302-311, 2// 2012.
[130] J. J. McMahon, J. Q. Lu, and R. J. Gutmann, "Wafer bonding of damascene-patterned metal/adhesive redistribution layers for via-first three-dimensional (3D) interconnect," in *Proceedings Electronic Components and Technology, 2005. ECTC '05.*, 2005, pp. 331-336 Vol. 1.
[131] Y.-S. Tang, Y.-J. Chang, and K.-N. Chen, "Wafer-level Cu–Cu bonding technology," *Microelectronics Reliability,* vol. 52, pp. 312-320, 2// 2012.
[132] A. Jourdain, S. Stoukatch, P. D. Moor, W. Ruythooren, S. Pargfrieder, B. Swinnen, *et al.*, "Simultaneous Cu-Cu and Compliant Dielectric Bonding for 3D Stacking of ICs," in *2007 IEEE International Interconnect Technology Conferencee*, 2007, pp. 207-209.
[133] E. Beyne, "3D System Integration Technologies," in *2006 International Symposium on VLSI Technology, Systems, and Applications*, 2006, pp. 1-9.
[134] B. Black, "Die stacking is happening," in *Intl. Symp. on Microarchitecture, Davis, CA*, 2013.
[135] K. Warner, J. Burns, C. Keast, R. Kunz, D. Lennon, A. Loomis, *et al.*, "Low-temperature oxide-bonded three-dimensional integrated circuits," in *2002 IEEE International SOI Conference*, 2002, pp. 123-124.
[136] P. Enquist, "Scalable direct bond technology and applications driving adoption," in *2011 IEEE International 3D Systems Integration Conference (3DIC), 2011 IEEE International*, 2012, pp. 1-5.
[137] V. Suntharalingam, R. Berger, J. A. Burns, C. K. Chen, C. L. Keast, J. M. Knecht, *et al.*, "Megapixel CMOS image sensor fabricated in three-dimensional integrated circuit technology," in *ISSCC. 2005 IEEE International Digest of Technical Papers. Solid-State Circuits Conference, 2005.*, 2005, pp. 356-357 Vol. 1.
[138] C. L. Keast, B. Aull, J. Burns, C. Chen, J. Knecht, B. Tyrrell, *et al.*, "Three-dimensional integration technology for advanced focal planes," in *MRS Proceedings*, 2008, pp. 1112-E01-02.

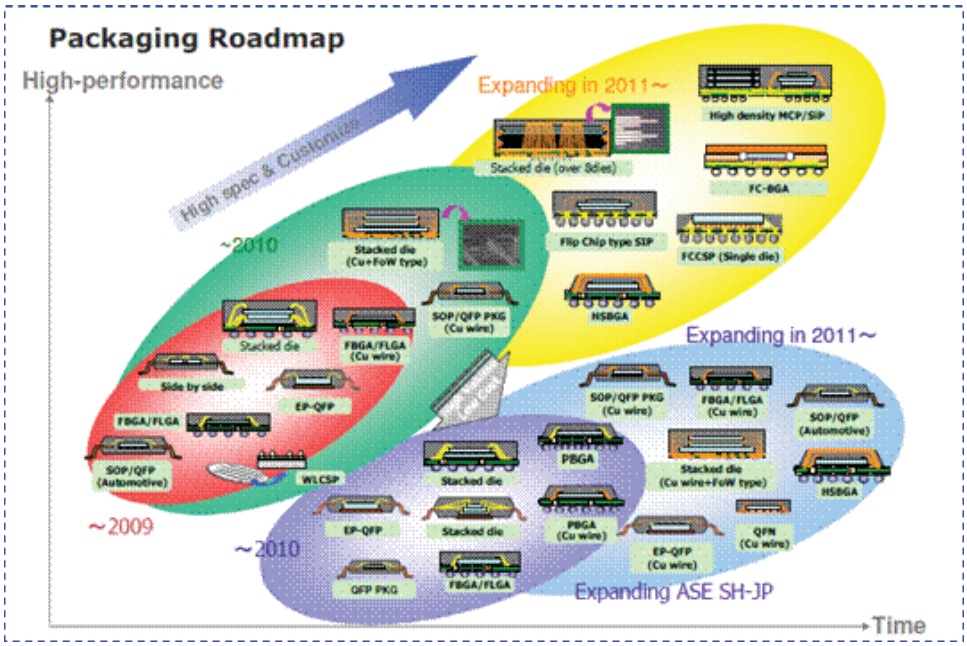
[139] J. A. Burns, B. F. Aull, C. K. Chen, C. Chang-Lee, C. L. Keast, J. M. Knecht, *et al.*, "A wafer-scale 3-D circuit integration technology," *IEEE Transactions on Electron Devices,* vol. 53, pp. 2507-2516, 2006.
[140] B. Aull, J. Burns, C. Chen, B. Felton, H. Hanson, C. Keast, *et al.*, "Laser Radar Imager Based on 3D Integration of Geiger-Mode Avalanche Photodiodes with Two SOI Timing Circuit Layers," in *2006 IEEE International Solid State Circuits Conference - Digest of Technical Papers*, 2006, pp. 1179-1188.
[141] C. L. Chen, C. K. Chen, D. r. Yost, J. M. Knecht, P. W. Wyatt, J. A. Burns, *et al.*, "Three-dimensional integration of silicon-on-insulator RF amplifier," *Electronics Letters,* vol. 44, pp. 746-747, 2008.
[142] J. Jeddeloh and B. Keeth, "Hybrid memory cube new DRAM architecture increases density and performance," in *2012 Symposium on VLSI Technology (VLSIT)*, 2012, pp. 87-88.
[143] S. Kim, E. J. Yun, J. S. Seo, D. E. Kim, B. H. Cho, and B. S. Kim, "Semiconductor memory device with stacked memory cell and method of manufacturing the stacked memory cell," ed: Google Patents, 2012.
[144] A. E. Franke, J. M. Heck, K. Tsu-Jae, and R. T. Howe, "Polycrystalline silicon-germanium films for integrated microsystems," *Journal of Microelectromechanical Systems,* vol. 12, pp. 160-171, 2003.
[145] F. Zimmer, M. Lapisa, T. Bakke, M. Bring, G. Stemme, and F. Niklaus, "One-Megapixel Monocrystalline-Silicon Micromirror Array on CMOS Driving Electronics Manufactured With Very Large-Scale Heterogeneous Integration," *Journal of Microelectromechanical Systems,* vol. 20, pp. 564-572, 2011.
[146] Y. Hayashi, K. Oyama, S. Takahashi, S. Wada, K. Kajiyana, R. Koh, *et al.*, "A new three dimensional IC fabrication technology, stacking thin film DUAL-CMOS layers," in *International Electron Devices Meeting 1991 [Technical Digest]*, 1991, pp. 657-660.
[147] J. F. Gibbons and K. F. Lee, "One-gate-wide CMOS Inverter on laser-recrystallized polysilicon," *IEEE Electron Device Letters,* vol. 1, pp. 117-118, 1980.
[148] M. Koyanagi, T. Nakamura, Y. Yamada, H. Kikuchi, T. Fukushima, T. Tanaka, *et al.*, "Three-Dimensional Integration Technology Based on Wafer Bonding With Vertical Buried Interconnections," *IEEE Transactions on Electron Devices,* vol. 53, pp. 2799-2808, 2006.
[149] P. Ramm, D. Bonfert, H. Gieser, J. Haufe, F. Iberl, A. Klumpp, *et al.*, "InterChip via technology for vertical system integration," in *Proceedings of the IEEE 2001 International Interconnect Technology Conference (Cat. No.01EX461)*, 2001, pp. 160-162.
[150] M. Koyanagi, H. Kurino, L. Kang Wook, K. Sakuma, N. Miyakawa, and H. Itani, "Future system-on-silicon LSI chips," *IEEE Micro,* vol. 18, pp. 17-22, 1998.
[151] D. B. Tuckerman, L. O. Bauer, N. E. Brathwaite, J. Demmin, K. Flatow, R. Hsu, *et al.*, "Laminated memory: a new 3-dimensional packaging technology for MCMs," in *Proceedings of IEEE Multi-Chip Module Conference (MCMC-94)*, 1994, pp. 58-63.
[152] M. Karnezos, "3D packaging: where all technologies come together," in *IEEE/CPMT/SEMI 29th International Electronics Manufacturing Technology Symposium (IEEE Cat. No.04CH37585)*, 2004, pp. 64-67.
[153] J. H. Lau, "Overview and outlook of through-silicon via (TSV) and 3D integrations," *Microelectronics International,* vol. 28, pp. 8-22, 2011.

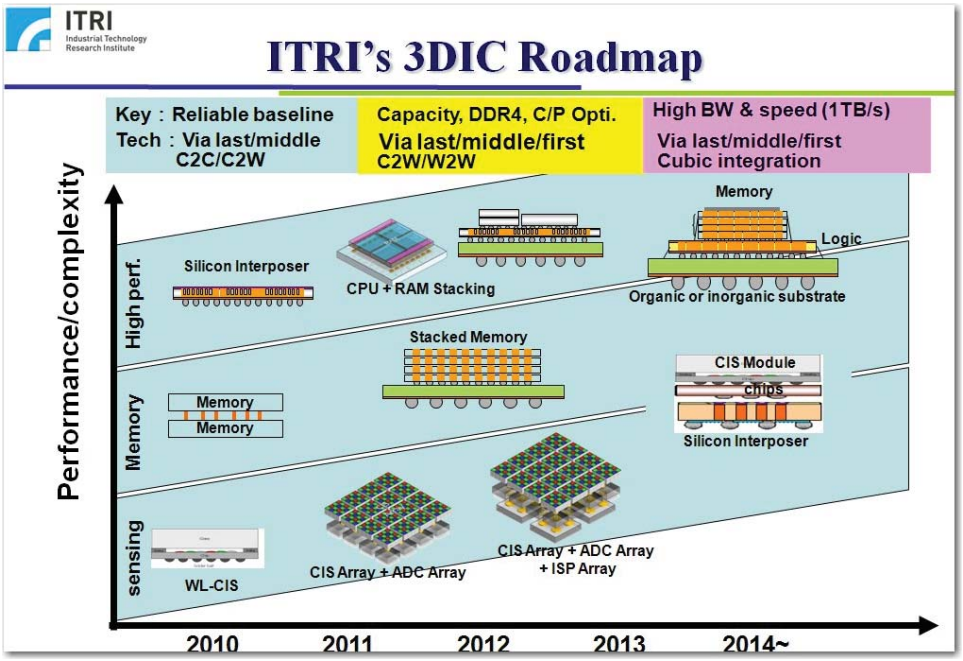
[154] A. Yoshida, J. Taniguchi, K. Murata, M. Kada, Y. Yamamoto, Y. Takagi, *et al.*, "A study on package stacking process for package-on-package (PoP)," in *56th Electronic Components and Technology Conference 2006*, 2006, p. 6 pp.
[155] M. Dreiza, A. Yoshida, K. Ishibashi, and T. Maeda, "High Density PoP (Package-on-Package) and Package Stacking Development," in *2007 Proceedings 57th Electronic Components and Technology Conference*, 2007, pp. 1397-1402.
[156] M. J. Rosker, "Technologies for Next Generation T/R Modules," in *2007 IEEE Radar Conference*, 2007, pp. 944-947.
[157] S. P. S. Lim, V. S. Rao, W. Y. Hnin, W. L. Ching, V. Kripesh, C. Lee, *et al.*, "Process Development and Reliability of Microbumps," *IEEE Transactions on Components and Packaging Technologies,* vol. 33, pp. 747-753, 2010.
[158] J. Sutanto, "POSSUMTM Die Design as a Low Cost 3D Packaging Alternative," *3D Packaging,* vol. 25, pp. 16-18, 2012.
[159] X. Fan, "Wafer level packaging (WLP): Fan-in, fan-out and three-dimensional integration," in *2010 11th International Thermal, Mechanical & Multi-Physics Simulation, and Experiments in Microelectronics and Microsystems (EuroSimE)*, 2010, pp. 1-7.
[160] K. Pressel, G. Beer, T. Meyer, M. Wojnowski, M. Fink, G. Ofner, *et al.*, "Embedded wafer level ball grid array (eWLB) technology for system integration," in *CPMT Symposium Japan, 2010 IEEE*, 2010, pp. 1-4.
[161] M. Prashant, K. Liu, and S. W. Yoon, "Integrated passive devices (IPD) integration with eWLB (embedded wafer level BGA) for high performance RF applications," in *2010 12th Electronics Packaging Technology Conference*, 2010, pp. 677-680.
[162] M. Wojnowski, V. Issakov, G. Knoblinger, K. Pressel, G. Sommer, and R. Weigel, "High-Q embedded inductors in fan-out eWLB for 6 GHz CMOS VCO," in *2011 IEEE 61st Electronic Components and Technology Conference (ECTC)*, 2011, pp. 1363-1370.
[163] M. Wojnowski, V. Issakov, G. Knoblinger, K. Pressel, G. Sommer, and R. Weigel, "High-Q Inductors Embedded in the Fan-Out Area of an eWLB," *IEEE Transactions on Components, Packaging and Manufacturing Technology,* vol. 2, pp. 1280-1292, 2012.
[164] M. Wojnowski, R. Lachner, J. Böck, C. Wagner, F. Starzer, G. Sommer, *et al.*, "Embedded wafer level ball grid array (eWLB) technology for millimeter-wave applications," in *2011 IEEE 13th Electronics Packaging Technology Conference*, 2011, pp. 423-429.
[165] M. Wojnowski and K. Pressel, "Embedded wafer level ball grid array (eWLB) technology for high-frequency system-in-package applications," in *2013 IEEE MTT-S International Microwave Symposium Digest (MTT)*, 2013, pp. 1-4.
[166] M. Brunnbauer, T. Meyer, G. Ofner, K. Mueller, and R. Hagen, "Embedded Wafer Level Ball Grid Array (eWLB)," in *2008 33rd IEEE/CPMT International Electronics Manufacturing Technology Conference (IEMT)*, 2008, pp. 1-6.
[167] M. Brunnbauer, E. Fürgut, G. Beer, and T. Meyer, "Embedded wafer level ball grid array (eWLB)," in *2006 8th Electronics Packaging Technology Conference*, 2006, pp. 1-5.
[168] S. W. Yoon, A. Bahr, X. Baraton, P. C. Marimuthu, and F. Carson, "3D eWLB (embedded wafer level BGA) technology for 3D-packaging/3D-SiP (Systems-in-Package) applications," in *2009 11th Electronics Packaging Technology Conference*, 2009, pp. 915-919.

[169] D. U. Lee, K. W. Kim, K. W. Kim, H. Kim, J. Y. Kim, Y. J. Park, *et al.*, "25.2 A 1.2V 8Gb 8-channel 128GB/s high-bandwidth memory (HBM) stacked DRAM with effective microbump I/O test methods using 29nm process and TSV," in *2014 IEEE International Solid-State Circuits Conference Digest of Technical Papers (ISSCC)*, 2014, pp. 432-433.
[170] S. Sukegawa, T. Umebayashi, T. Nakajima, H. Kawanobe, K. Koseki, I. Hirota, *et al.*, "A 1/4-inch 8Mpixel back-illuminated stacked CMOS image sensor," in *2013 IEEE International Solid-State Circuits Conference Digest of Technical Papers*, 2013, pp. 484-485.
[171] K. W. Lee, Y. Ohara, K. Kiyoyama, S. Konno, Y. Sato, S. Watanabe, *et al.*, "Characterization of chip-level hetero-integration technology for high-speed, highly parallel 3D-stacked image processing system," in *2012 International Electron Devices Meeting*, 2012, pp. 33.2.1-33.2.4.
[172] L. J. Hornbeck, "Combining Digital Optical MEMS, CMOS and Algorithms for Unique Display Solutions," in *2007 IEEE International Electron Devices Meeting*, 2007, pp. 17-24.
[173] E. Beyne, "Multilayer thin Film technology as an Enabling technology for System-in-Package (SIP) and" Above-IC" Processing," in *Proceedings of the International Microelectronics And Packaging Society Conference*, 2004.
[174] F. Ohnimus, U. Maaß, G. Fotheringham, B. Curran, I. Ndip, T. Fritzsch, *et al.*, "Design and Comparison of 24 GHz Patch Antennas on Glass Substrates for Compact Wireless Sensor Nodes," *International Journal of Microwave Science and Technology,* vol. 2010, 2011.
[175] M. Niedermayer, S. Guttowski, R. Thomasius, D. Polityko, K. Schrank, and H. Reichl, "Miniaturization platform for wireless sensor nodes based on 3D-packaging technologies," in *Proceedings of the 5th international conference on Information processing in sensor networks*, 2006, pp. 391-398.
[176] M. Shevelov, V. Sieroshtan, O. Ruban, A. Burygin, P. Komakha, O. Aleksieiev, *et al.*, "Small form factor and low profile quad-band System-in-Package (SiP) module," in *2011 IEEE MTT-S International Microwave Symposium*, 2011, pp. 1-4.
[177] M. Wojnowski, C. Wagner, R. Lachner, J. Böck, G. Sommer, and K. Pressel, "A 77-GHz SiGe single-chip four-channel transceiver module with integrated antennas in embedded wafer-level BGA package," in *2012 IEEE 62nd Electronic Components and Technology Conference*, 2012, pp. 1027-1032.
[178] J. Hasch, E. Topak, R. Schnabel, T. Zwick, R. Weigel, and C. Waldschmidt, "Millimeter-Wave Technology for Automotive Radar Sensors in the 77 GHz Frequency Band," *IEEE Transactions on Microwave Theory and Techniques,* vol. 60, pp. 845-860, 2012.
[179] R. Agethen, M. PourMousavi, H. Forstner, M. Wojnowski, K. Pressel, R. Weigel, *et al.*, "60 GHz industrial radar systems in silicon-germanium technology," in *Microwave Symposium Digest (IMS), 2013 IEEE MTT-S International*, 2013, pp. 1-3.
[180] M. Al Henawy and M. Schneider, "Integrated antennas in eWLB packages for 77 GHz and 79 GHz automotive radar sensors," in *Microwave Conference (EuMC), 2011 41st European*, 2011, pp. 1312-1315.
[181] K. W. Lee, A. Noriki, K. Kiyoyama, T. Fukushima, T. Tanaka, and M. Koyanagi, "Three-Dimensional Hybrid Integration Technology of CMOS, MEMS, and Photonics Circuits for Optoelectronic Heterogeneous Integrated Systems," *IEEE Transactions on Electron Devices,* vol. 58, pp. 748-757, 2011.

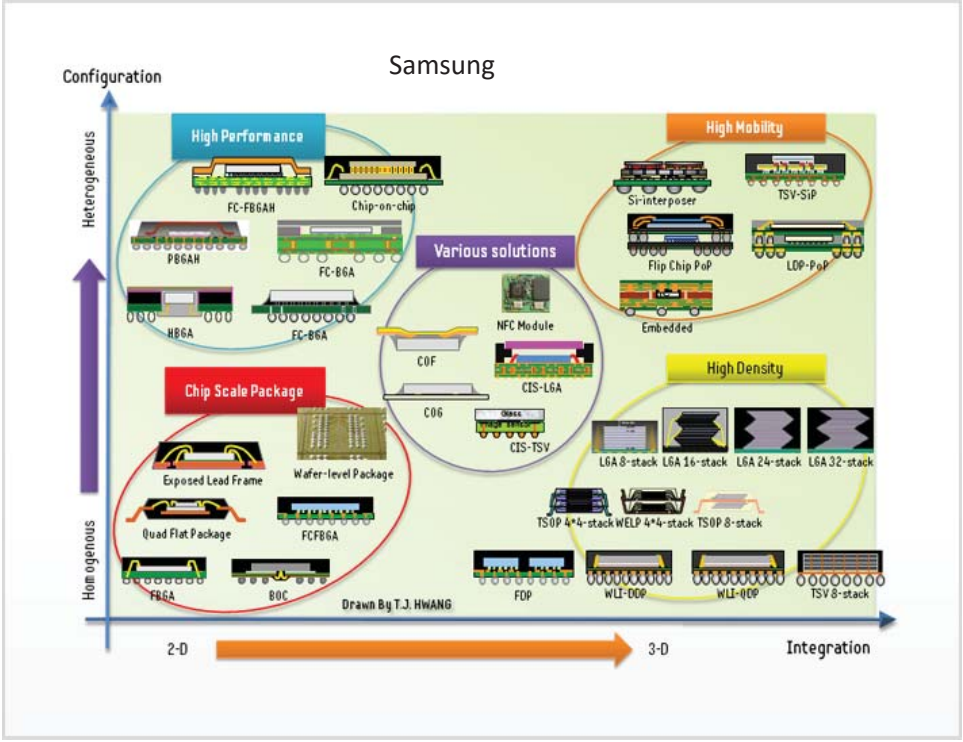
[182] K. W. Lee, A. Noriki, K. Kiyoyama, S. Kanno, R. Kobayashi, W. C. Jeong, *et al.*, "3D heterogeneous opto-electronic integration technology for system-on-silicon (SOS)," in *2009 IEEE International Electron Devices Meeting (IEDM)*, 2009, pp. 1-4.
[183] M. Fujiwara, S. Terada, Y. Shirato, H. Owari, K. Watanabe, M. Matsuyama, *et al.*, "Low-loss optical interposer with recessed vertical-cavity surface-emitting laser diode and photodiode chips into Si substrate," *Japanese Journal of Applied Physics,* vol. 47, p. 2936, 2008.
[184] S. Bhansali, G. H. Chapman, E. G. Friedman, Y. Ismail, P. Mukund, D. Tebbe, *et al.*, "3D heterogeneous sensor system on a chip for defense and security applications," in *Defense and Security*, 2004, pp. 413-424.
[185] J. Q. Lu, "3-D Hyperintegration and Packaging Technologies for Micro-Nano Systems," *Proceedings of the IEEE,* vol. 97, pp. 18-30, 2009.

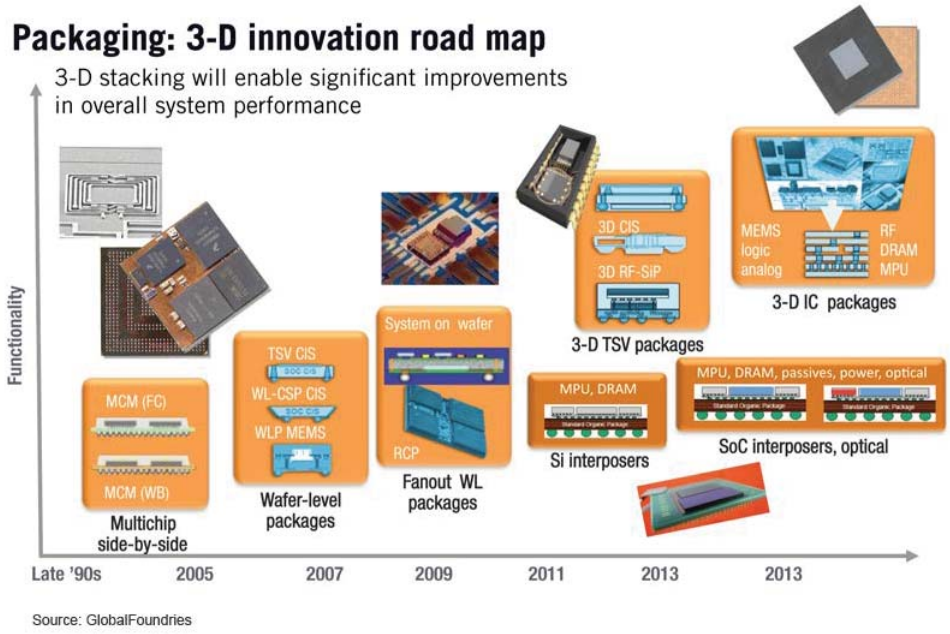
APPENDIX: INTEGRATION ROADMAPS OF LEADING ELECTRONICS COMPANIES

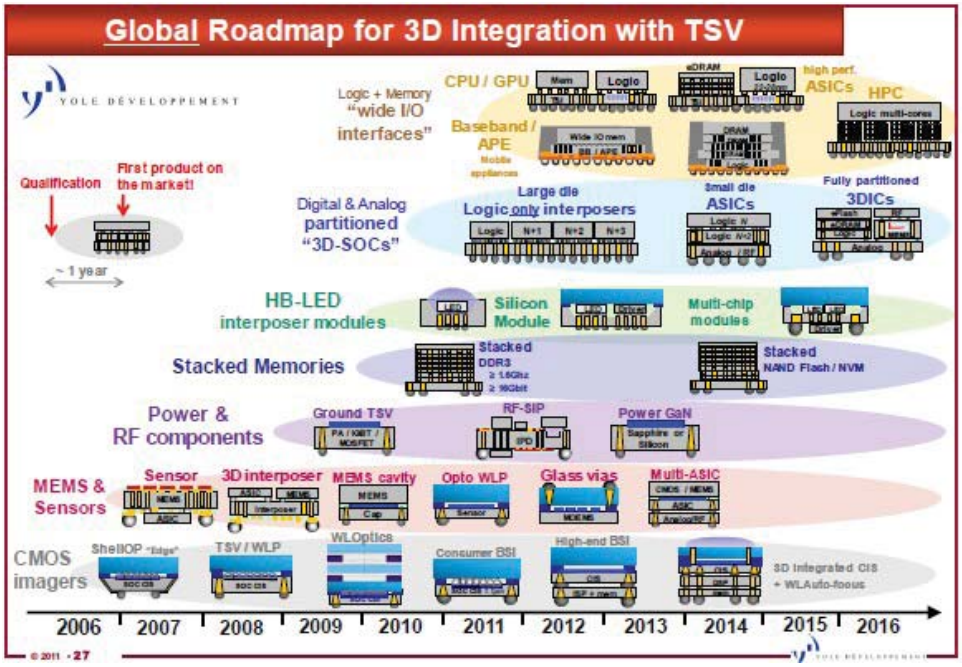


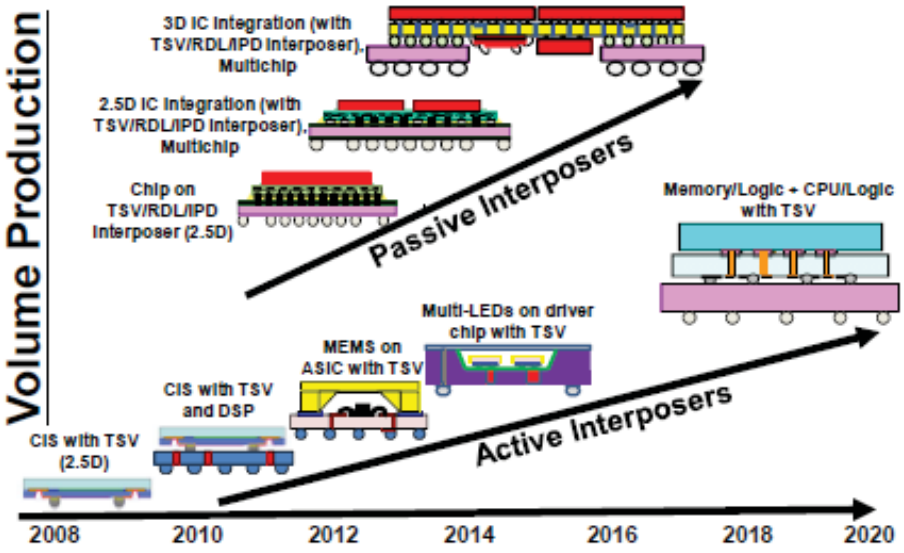


Samsung

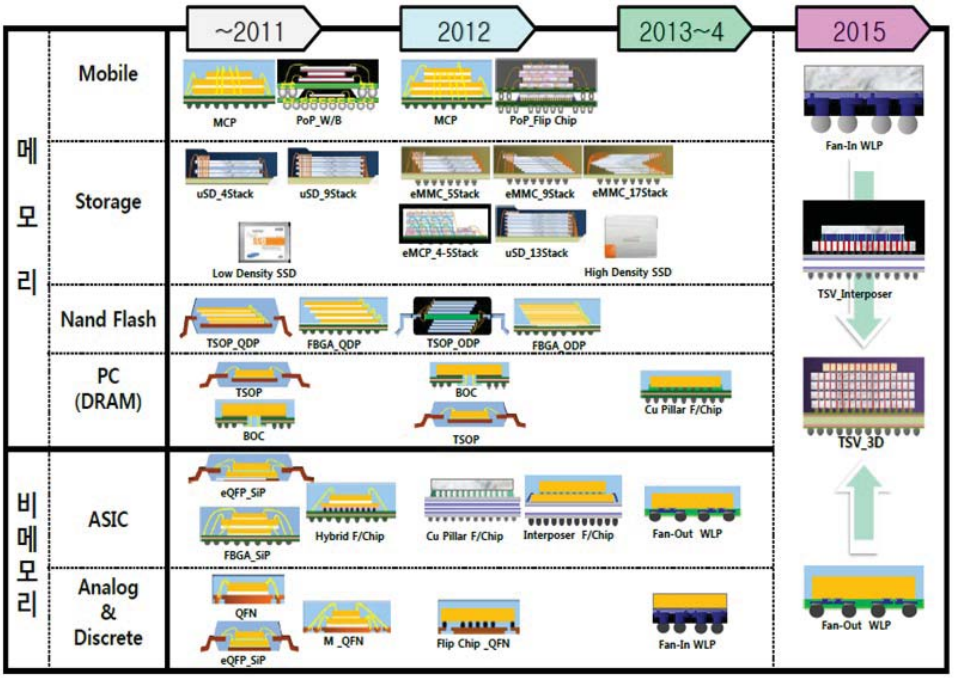








Reference [20]



Samsung

LIST OF ACRONYMS, ABBREVIATIONS, AND SYMBOLS

ACRONYM	DESCRIPTION
AESA	active electronically steerable array
AFOSR	Air Force Office of Scientific Research
AFRL	Air Force Research Laboratory
ARCHIPELA	Adaptable Radio Frequency Criteria-High Performance Cell Array
ARO	Army Research Office
B2B	back-to-back
BEOL	back end of line
BGA	ball grid array
BT	bipolar transistor
CDC	correlated double chip
CLP	chip level package
CMOS	complementary metal-oxide-semiconductor
COSMOS	Compound Semiconductor Materials on Silicon
CPMT	Components, Packaging and Manufacturing Technology Society (IEEE)
CPU	central processing unit
CSP	chip scale packaging
DAC	digital-to-analog converter
DAHI	Diverse Accessible Heterogeneous Integration
DARPA	Defense Advanced Research Projects Agency
DDR	double date rate
DFN	dual flat no-lead
DMD	deformable mirror display
DoD	Department of Defense
DRAM	dynamic random-access memory
EDA	electronic design automation
EDS	Electron Devices Society (IEEE)
EMIB	embedded multi-die interconnect bridge
eWLP	embedded wafer level packaging
F2B	face-to-back
F2F	face-to-face
FA CSP	flip chip scale packaging
FBH	Ferdinand-Braun-Institut
FC	flip chip
FEM	front end module
FEOL	front end of line
FET	field effect transistor
FI	fan-in
FO	fan-out
FOM	figure-of-merit
FPGA	field-programmable gate array
FTI	feedthrough interposer
GPU	graphics processing unit
HBM	high bandwidth memory

ACRONYM	DESCRIPTION
HBT	heterojunction bipolar transistor
HEMT	high electron mobility transistor
HI	heterogeneous integration
HMC	hybrid memory cube
IC	integrated circuit
IEEE	Institute of Electrical and Electronics Engineers
IGZO	indium gallium zinc oxide
IPD	integrated passive device
IRFFE	Intelligent Radio Frequency Front End
ITRS	International Technology Roadmap for Semiconductors
JMOS	joint metal-oxide-semiconductor
KGD	known good die
LADAR	laser radio detection and ranging
LED	light emitting diode
LF	lead frame
LNA	low noise amplifier
LTCC	low temperature co-fired ceramic
MAU	math accelerator unit
MBE	molecular beam epitaxy
MEMS	micro-electromechanical systems
MESFET	metal-semiconductor field-effect transistor
MM	More Moore (roadmap)
MMIC	monolithic microwave integrated circuit
MMU	memory management unit
MOCVD	metal-organic chemical vapor deposition
MtM	More than Moore (roadmap)
PA	power amplifier
PCB	printed circuit board
PoP	package on package
QFN	quad flat no-lead
R&D	research and development
RDL	redistribution line
RF	radio frequency
RYD	Sensors Directorate Aerospace Components & Subsystems Division
RYDD	Devices for Sensing Branch
RYDI	Integrated Circuits & Microsystems Branch
S&T	science and technology
SAW	surface acoustic device
SEM	scanning electron microscopy
SEMI	Semiconductor Equipment and Materials International
SIC	stacked integrated circuit
SiP	system in package
SLR	Super Logic Regions
SMAFTI	stacked memory and feedthrough interposer
SMART	Scalable Millimeter Wave Architectures for Reconfigurable Transceivers

ACRONYM	DESCRIPTION
SoC	system on chip
SOI	silicon on insulator
SWAP-C	size, weight, power, and cost
TFT	thin film transistor
TSV	through Si vias
VCO	voltage controlled oscillator
W2W	wafer-to-wafer
WB	wire bonding
WLP	wafer level packaging