The Impact of Semiconductor Packaging Technologies on System Integration
An Overview

Carlo Cognetti
STMicroelectronics
Corporate Package Development Dept.
Agrate Brianza (Milan) Italy
carlo.cognetti@st.com

Abstract—We are crossing the threshold of the third revolution in semiconductor packaging. In the ‘80s, Surface Mount Technology (SMT) had major impact on size reduction of all electronic systems. In the 90’s, Ball Grid Array (BGA) has been introduced, whose latest evolution allows further dramatic steps in miniaturization, with cost effective production of 3-dimensional structures and the integration of a large number of passive and active devices in the same package (System in Package – SiP). 3D-BGA platform is now well established and offers an alternative to System on Chip (SoC), i.e. the full integration at chip level. With the additional possibility of combining “heterogeneous” devices [1]. But BGA is getting close to its intrinsic limits and will not be able to serve the requirements (size, speed, thermal dissipation, cost) of next advanced systems, like future wireless applications. At present, most of R&D effort is dedicated to the development of new concepts, mixing conventional assembly and “on wafer” processes. The result is the “3D Wafer Level” platform, which will provide unprecedented levels of integration, with several breakthroughs in design, manufacturing infrastructure, supply chain.

I. INTRODUCTION

Conventional function of semiconductor package (fig.1) is to bridge active chip and external circuit (motherboard), i.e. to

Fig.1 – The four major packaging platforms

be the link between two rather different worlds, with different characteristics in terms of materials, design rules, reliability. Various attempts to eliminate the package in the interconnection chain did fail due to higher final cost of the solution. In the first 30 years of IC packaging (1970-2000), evolution has been driven by two main factors: increase of I/O count and miniaturization. The first breakthrough in the technology took place in the ‘80s, with the introduction of Surface Mount Technology (SMT) as alternative to Through Hole Technology (THT) for the assembly of components on PCB. In THT, device pitch is decided by some mechanical process drilling vias in the motherboard and is rather large: 70-100 mils (1.8-2.54mm). In SMT, device is soldered on the board surface, pitch can be reduced down to 0.3-0.4mm. The development of Quad Flat Pack, with leads on the four sides, offered further I/O count capability, up to 200-300. The concept become inadequate in early ‘90s, when Ball Grid Array was designed, using an organic carrier – instead of a metal leadframe – to interconnect the chip to the external world. The carrier, quite similar to a printed board, has two or more layers of metal, conductive through vias among the layers, and an array of contacts (solder balls) on the bottom side. This structure can serve high end processors with 00s I/Os as well as miniaturized ICs for wireless applications.

Mobile phone pushed the BGA to extreme size reduction, at the limit of “Chip Size Package”, i.e. a ratio >0.8 between die size and package footprint. This is the 2D-phase of BGA.

II. BGA: FROM 2D TO 3D

Another important characteristic of BGA is the capability of system integration, i.e. the possibility to interconnect several components, active and passive, in the same package. In early 2000, the requirement for further miniaturization in wireless pushed beyond the existing limits, in order to use the 3rd dimension of package. This was achieved by stacking two
or more chips in the same BGA, or two BGAs one on top of the other. In the same time, the concept of System in Package (SiP) was widely applied, to complete the platform of 3D BGA. This evolution of BGA, shown in fig.2, can be considered the 2nd revolution in packaging. The “stack die” process was known since several years for high end computer devices, but only after the 2000 it became widely used for high volume, cost sensitive applications. In to-day mobile phones, almost all memory devices are in 3D technology, and can achieve a memory size which would be impossible or too expensive if monolithic. In fig.3, the early roadmap is reported, with up to 10 active devices (flash memories) in one BGA (fig.3). They are now in production by several memory suppliers. Prototypes with 20 and more active dice have been published in 2007-08. As mobile phone requires a continuous reduction of package thickness, one of the most important challenges is the thickness reduction of individual chips, down to 20-30 microns.

In the same period, the need of better RF transceivers and power amplifiers pushed for the introduction of advanced passive functions with good RF response, deposited on glass or silicon carrier, and the consequent development of SiP technology platform as available to-day (fig.4).

Cost effective SiPs are obtained on multilayer BGA through advanced interconnections like chip to chip wire bonding and bumped flip chip [2]. Passive components, filters, switches are present as miniaturized SMDs (the smallest available is 125x250 micron) or in thin film technology, with a full set of high accuracy resistors, capacitors, inductors. It is known as Integrated Passive Device (IPD).

Substrate plays the most important role in BGA package, from both viewpoints of performance and cost. Technology has been in continuous evolution, especially for what refers to patterning and interconnection density. Space/line of 20/20 micron are becoming common and cost competitive, by using the so called semi-additive process. Laser based imaging are under development, for space/line in the range of 10/10 micron, yield and cost will decide their introduction in the marketplace.

BGA design and Electrical characteristics of a package must be deeply analyzed when developing a new product. Electronic Design Automation (EDA) vendors now propose integrated solutions for co-design of IC / Package / Board. At the same time several extraction tools enabling signal and power integrity studies have become available. Major advantage of BGA is the capability of several metal layers for routing with drastically increased flexibility of the
interconnection. With high speed SiP and IC, package parasitic extraction is a routine job, for validation of substrate design and generation of electrical characteristics in I/O Buffer Information Specification (IBIS) format.

BGA package platform is close to reach its intrinsic limits. In order to assist continuous improvement in miniaturization, integration and speed, two possible directions have been identified and are now under development. They take advantage of technologies “next” to packaging: the laminate board and the silicon. Therefore, two new platforms are under development: a] embedded die (in the laminate carrier, both BGA substrate or motherboard); b] wafer level packaging (WLP).

III  DIE EMBEDDED IN THE SUBSTRATE

Since 2004-05, the possibility has been demonstrated to embed active devices inside a laminate board, as part of production flow of the board itself. This requires the availability of fully tested individual chips, which are delivered from the silicon manufacturer to the board manufacturer as bare dice. Dice are interconnected to the first metallization layer of the board and embedded in an isolation layer, usually the “core”. At the end of the process, dice(s) are buried under a number of metal layers, as requested by the application. Main advantages: a] very short interconnection, for high speed; b] compatibility with 3D structures; c] compatibility with integrated shielding; d] size reduction. Main disadvantages: a] long production cycle time; b] complex supply chain; c] yield for high complexity applications.

One of the earliest (2005) “embedded die” applications is shown in fig.5. It represents a 8 layer BGA package, for a SiP with buried single chip transceiver and additional circuitry on the top surface. Active chip is fully shielded by two metal planes in z-direction and a system of metallized through vias in x-y direction. Assembly yield is in the 90% range. But for lower complexity, yield is getting close to 99%, making the technology quite attractive also at motherboard level: several suppliers in Japan, Korea, Taiwan are ready for volume production of substrates with embedded SMDs, discretes, simple ICs and Integrated Passive Devices (IPD).

IV  Fan-In and Fan-Out WAFER LEVEL PACKAGE (FI-WLP, FO-WLP)

The second possible direction beyond BGA is provided by a combination of wafer fab and conventional packaging technologies. To interconnect the active chip(s), it takes advantage of silicon manufacturing techniques, whose density can be much higher than the one of BGA substrate. To protect the chip and connect it to the motherboard, it uses an approach similar to BGA, with an array of balls in the lower surface. Earlier example of WLP is shown in fig.6a. It is called Fan-In WLP, because the footprint is smaller than the die.

Fig.5 – 6-layer Embedded Die SiP, with integrated shield

Fig.6a,b – Two concepts of WLP : Fan-In and Fan-Out

It looks like a bumped chip, but with bumps size and pitch similar to the BGA, in order to be directly mounted on the motherboard as a conventional package. It represent the extreme case of single chip package miniaturization. It is in mass production. However, due to the mismatch in coefficient of thermal expansion (CTE) with laminated board, it is exposed to failure in thermal cycling and must be limited to small dice, with side of a few millimeters. Together with relatively large ball pitch (0.3-0.5mm) needed by end user,
this limits the Fan-In WLP to relatively low pin count applications.

In order to overcome this limitation, the Fan-Out WLP has been introduced, which allows a footprint larger than the die, with higher pin count capability (fig.6b).

FO-WLP makes use of a “re-constituted” wafer, obtained after distributing tested good dice “face down” on a temporary carrier and than molding them in suitable epoxy material.

![Fig.7 – Reconstituted wafer for FO-WLP (e-WLB)](image)

Fig.7 shows the “re-constituted wafer” after removal of temporary carrier. Die-to-die distance is decided upon the needs of final package, in terms of I/O count. After reconstitution, the “wafer” follows a passivation-metallization flow typical of wafer fab, with same interconnection density. In one side, metallization is in direct contact with chip, for high speed; in the other side, it ends with solder balls, the same of BGA and is fully compatible with motherboard design rules.

In August 2008, the alliance has been announced among ST IFX and StatsChippac, for the development of 2nd generation of existing eWLB, the one-side FO-WLP designed and qualified by IFX in early 2008. This new package family (fig.8), called 3D-eWLB, represents the evolution of BGA with high density capability for stack package and SiP [3]. It has metallizations in both sides, with interconnection vias in z-direction. It is expected to be compatible with 3D chip-to-chip solutions of next paragraph and be their suitable interface to the motherboard.

![Fig.8 – FO-WLP (3D-eWLB) platform](image)

V Chip-to-Chip INTERCONNECTION (microbumps and Through Silicon Vias)

Other important wafer level technologies are contributing to the evolution of packaging: a] Through Silicon Vias (TSV); b] microbumping.

TSV is present in the literature since early ’70s, but only in the last few years it gained momentum, especially as possible solution for data storage. In fact, it is assumed that, starting from coming technology nodes at 20nm and below, single chip integration of large memory will be more expensive than multiple stacked chips of previous nodes.

The concept is expected to serve also the sensor market, both MEMS and imaging. Fig.9 shows a Wafer Level Camera developed by STMicroelectronics and Leti in Grenoble, France. It is now in production on 8” and 12” wafers. After completion of front side of the camera, wafer is fixed onto a glass plate, having double function: optical and mechanical.

![Fig.8 – Wafer Level Camera with TSV](image)
Wafer is grinded under 100 micron and through vias are etched from backside (“via last”) and then metallized, by means of an interconnection system which re-distributes the contacts under the silicon chip, in a matrix of balls. Lens can be applied at wafer level, for high throughput.

Wafer Level camera is an early example of TSV in volume production. Vias are relatively large with large pitch. But the technology is required a higher level of integration with interconnection, with smaller form factor and high speed. thousand of vias with few micron diameter. Several Companies Wafer level technologies are considered the answer to this and Institutes are moving in this direction with specific R&D need, in combination with process steps (encapsulation, programs. Target is the 3D interconnection from chip to chip, solder balling, etc) typical of semiconductor packaging. A This will provide heterogeneous systems at high speed.

Fig.9 shows a prototype of a microprocessor mounted on a different manufacturing infrastructure and supply chain [6] memory chip by means of high microbumps. Memory has TSV is envisaged. The impact on system design can be and is assembled in a BGA (bottom side). Other types of silicon remarkable. Partitioning will take advantage of “vertical” combination are possible; cost of the function can be minimized interconnections like microbumps and through vias, with higher speed and possibly lower cost.

A number of Companies and Institutes are supporting the development of the platform, with early volume production in FO-WLP (eWLB) and wafer level camera.

VI CONCLUSIONS

After the SMT in the ‘80s and the 3D-BGA in the last decade, packaging technologies are going through their third revolution (fig.10).

The new platform is intended to increase the density of interconnection, with smaller form factor and high speed. Thousand of vias with few micron diameter. Several Companies Wafer level technologies are considered the answer to this and Institutes are moving in this direction with specific R&D need, in combination with process steps (encapsulation, programs. Target is the 3D interconnection from chip to chip, solder balling, etc) typical of semiconductor packaging. A new technology (nor “front-end”, nor “back-end”), with higher speed and possibly lower cost.

A number of Companies and Institutes are supporting the development of the platform, with early volume production in FO-WLP (eWLB) and wafer level camera.

REFERENCES


