BY GRAHAM PROPHET • EDITOR

Multi-chip packaging: tall stacks, low profiles

SOC OR SIP? AS THE DESIGN COST OF A COMPLEX SYSTEM-LEVEL CHIP ESCALATES, THE SYSTEM-IN-PACKAGE APPROACH BECOMES INCREASINGLY ATTRACTIVE. AT THE SAME TIME, NEW WAYS OF PACKING MORE CHIPS INTO A SINGLE, OUTWARDLY CONVENTIONAL PACKAGE ARE COMING ON-STREAM.

Today, they are systems-in-package, SiPs. Or multi-die packages. Before that, they were multi-chip modules: before that, they were hybrids. There's nothing fundamentally new about assembling multiple active components into what we recognise as an IC (integrated circuit) package. In fact, it goes back to the earliest implementation of an IC—in those far-off days, it was common parlance to use the terminology "monolithic IC" to differentiate the more exotic species of device in which the manufacturer had managed to get all the functionality on to a single piece of silicon. And the basic reason for building such a composite hasn't changed either. You take that route when the mix of functions you need isn't technically or economically attainable on a single chip. Over time—and between projects at the same moment in time—the balance of factors that may lead to the decision to opt for a multi-chip solution changes.

BIGGER MEMORIES, SOONER

One dimension of that decision process depends on the limitations of process technology. There is, for example, a long-established route to enhanced device density through innovative packaging in memory. As memory has followed its own route along the line of the Moore's Law graph, at any given moment there has been a corresponding maximum available size of DRAM chip. And at that same moment, there have always been some projects for which that size is not enough. By mounting multiple die within the package outline of a standard single-die part, specialist suppliers have produced parts to meet that need. By following a predictable progression in package footprints, these suppliers have been able to anticipate and emulate devices from subsequent generations of single-chip parts, many months before they come to the market. With its larger memory-cell size, SRAM always lagged one to two generations behind DRAM in density—packaging multiple SRAM die together yielded parts that provide equivalent densities in similar footprints. Today, the same is being done for flash. Suppliers such as White Electronic Designs continue to package multiple chips, and the company has recently announced a 64-MB Flash MCP (multi-chip-package), designed for use in embedded and high-reliability applications and available in commercial, industrial and military temperature ranges. It occupies a 13×22-mm, 159 plastic ball-grid array (PBGA; **Figure 1**). Organised as 8M×64, the flash features access times of 90, 100 and 120 nsec with 1 million erase/programme cycles per sector. The price of being ahead of the density curve is that the part is not cheap—its cost is $250.00 each in volumes of 500 pieces (industrial temperature).

A more subtle decision process in chip partitioning arises where a vendor has to optimise economic and technical factors—often in the light of his own particular process capabilities—to produce the best possible specification. For many years, there have been two opposing ways to produce devices that have become generically known as "smart power"—devices that combine power-control capability with some measure of intelligence, ranging from some driving and protection circuitry up to comprehensive control elements and perhaps including a microcontroller core. Some vendors choose to build a single chip; others, to build a control chip plus separate power switches into a single package. The single chip needs a much more complex

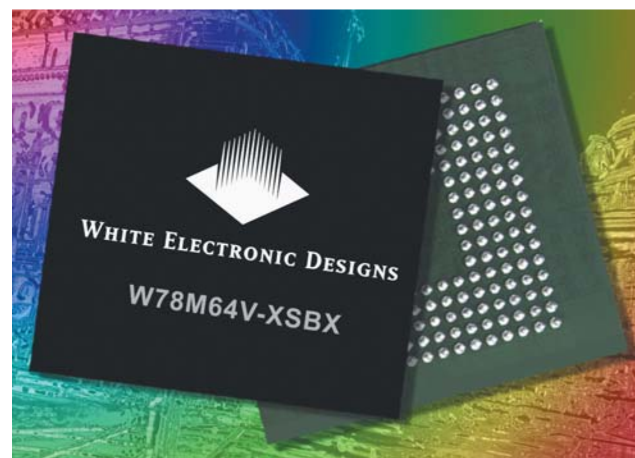

Figure 1 White Electronic Design's BGA-mounted flash chips are the latest in a long line of memory products that achieve densities not reached by equivalent single-die parts for two or three silicon process nodes.

AT A GLANCE

- Multi-die packages build on long-established principles to increase circuit density.
- The high cost of developing a system-on-chip ASIC in 90-nm geometry drives the search for a multi-chip alternative.
- Relatively simple stacked-die configurations for cell-phone memory are the forerunners of much more ambitious 3-D chip packages.
- After many years as purely a research concept, full 3-D chip-to-chip connectivity is set to become a practical reality.

fabrication process, plus more complex design to handle voltage and thermal stresses on the single die—but it allows the vendor to use a simpler package, without the need to wire together multiple chips within the package and thus arguably to improve yield. Conversely, regarding the multi-die approach, protagonists will argue that you can use optimised processes for both of the control and power elements, leading to a better overall specification and (possibly) a more reliable part. In this case, the argument appears to be finely balanced, to judge from the fact that this split has existed for more than a decade and currently vendors successfully offer both approaches in the market. As a user, you have little need to know how the manufacturer has built the smart switch, with the possible exception of attending to thermal and heatsinking considerations. The **sidebar** "Linear Technology enters power-module fray" describes a new addition to the system-in-package offerings available in the power sector.

RF designers have similar options open to them when a single package configuration for a radio function—whether RF alone or RF plus baseband—is a requirement. The various offerings for Bluetooth exemplify the single-chip approach; market leader CSR has just released the fifth generation of its Bluecore design. However, by no means all RF applications can support the massive investment in blending RF and logic CMOS technologies that vendors must invest to achieve a single-chip design. The route to an RF system-in-package was charted in a paper presented at the SAME (Sofia Antipolis Micro Electronics) Forum 2005 by Insight SiP. The company's CTO Chris Barratt notes the key role of the substrate in the design of an RF module. It can be a pure interconnection medium with IC and passive components attached to it, or it can embed passive components within the package to implement RF functions. The substrate can be of printed-circuit material, generally categorised as "laminate", or of LTCC (low-temperature co-fired ceramic), or of silicon used as a substrate. Laminate (FR4 or higher-dielectric material) substrates can support some passive RF functions such as matching networks. This, Barratt notes, is relatively mature and widely-available technology. LTCC substrates can embed more sophisticated components formed within the ceramic itself, such as capacitors, inductor, filters and baluns. Integrated passive devices represent a further step in technology—designers employ thin-film techniques to form all types of passive component on semiconductor or glass substrates.

ELECTRICAL TO PHYSICAL FLOW

Figure 2 shows the design flow that Barratt charts for an LTCC-based device, which Insight has compiled using a suite of standard EDA tools. The first step is to define the circuit function in purely electrical terms; then, the designer selects a suitable layer structure for the LTCC. Next, he chooses elements from a library of mechanical parts that maps to the chosen layers and applies electromagnetic simulation to the result to precisely model the RF behaviour of the chosen structures. He has then created a specialised library of elements appropriate to the specific device that is being designed and, via further simulation—including modelling circuit behaviour together with active devices, representing the ICs that will later be flip-chip mounted to the substrate—converges the design to a final geometric layout for all of the layers in the LTCC. A final full-system simulation checks that unintended parasitic

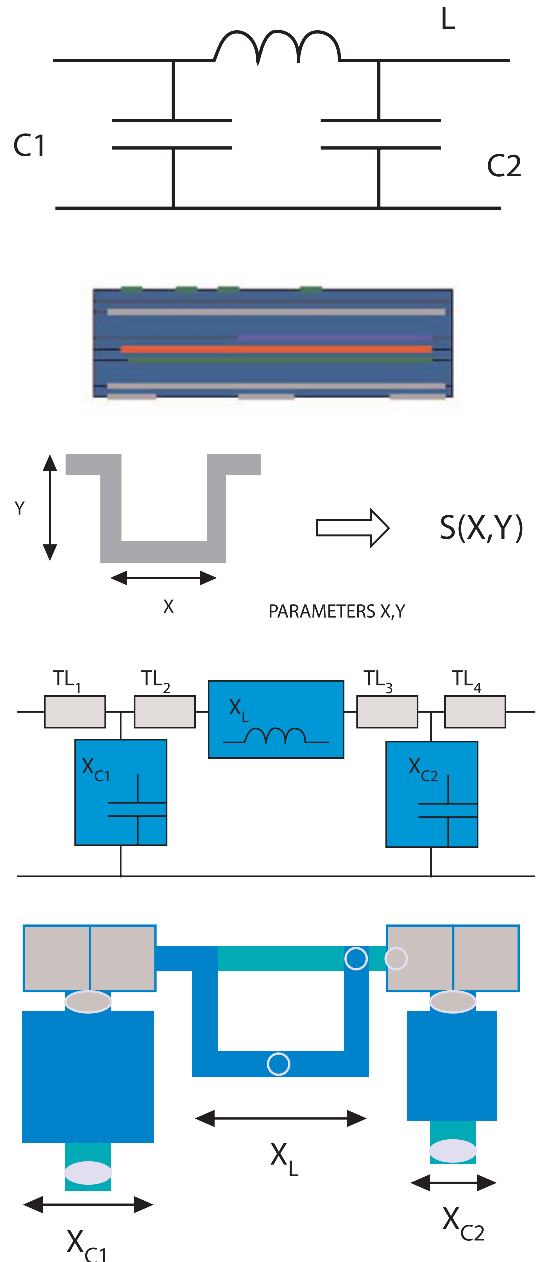

Figure 2 Design flow to create RF passive networks in a ceramic substrate: starting with the electrical circuit, the designer defines a suitable ceramic stack. Candidate component shapes are selected and simulated, and parameterised component layouts refined in simulation to converge to the original circuit.

LINEAR TECHNOLOGY ENTERS POWER-MODULE FRAY

By David Marsh, Contributing Technical Editor

Linear Technology's announcement of a new range of power modules marks a maybe surprising departure from the IC solutions supplier's traditional marketplace. Don Paulus, vice president and general manager of Linear Technology's Power Business Unit, explains: "We wanted to extend our customer base to companies that don't have the expertise or the time to invest in developing a high-performance dc-dc converter, which is significant challenge–especially at the high current levels that the LTM4600 serves." First in a new line of power μModules, the LTM4600 shrinks a 10-A step-down converter into a 15×15×2.8-mm land-grid-array package that supports automated assembly in the same way as an IC (Figure A). Paulus notes that this tiny RoHS-compliant package permits designers to pack the dc-dc conversion function into otherwise unused space, such as the back of pc boards. Its low profile and multiple-ball connections constrain junction-to-board thermal resistance to 15°C/W for a four-layer pc board, typically permitting the package to dissipate 3W with no heatsink or forced airflow while still staying safely within its 125°C junction temperature limit.

Featuring Linear Technology's own DMOS technology to optimise the FET's on-resistance versus gate-charge balance, the LTM4600 integrates the power switch, the gate driver, the buck-converter inductor, and input and output bypass capacitors. One resistor sets the output voltage within the range 0.6 to 5V, with an optional capacitor adjusting the module's soft-start characteristics. It's often unnecessary to add extra bulk capacitors to a system's normal complement. The LTM4600 accepts very wide-ranging input voltages–4.5 to 20V for the EV suffix, and 28V for the HVEV version–whereas most modules in this power range have limits of around 6V. The 20V input capability suits most intermediate-bus architectures that distribute 12V, while the 28V version suits the industrial-standard 24V level.

Capable of efficiencies as high as 92%, the current-mode architecture runs at frequencies around 800 kHz to maximise transient response. Measurements with a 12V input level reveal 80% efficiency when sourcing 1.2V, with a 10%-to-90%-to-10% load step settling within 25 μsec. Paulus notes that the current-mode design guarantees cycle-by-cycle current limiting and short-circuit protection, making fuses unnecessary. It's also easy to parallel two LTM4600s to source 20A.

Paulus stresses that reliability was a crucial design requirement: "With FIT rates indicating failure levels of around 1 event per billion operating hours, the LTM4600's reliability is 10 to 15 times better than competing dc-dc converters and approaches a monolithic IC's dependability." Available now, prices for the LTM4600EV are $16.50/1,000 with the HVEV version costing $19.50/1,000.

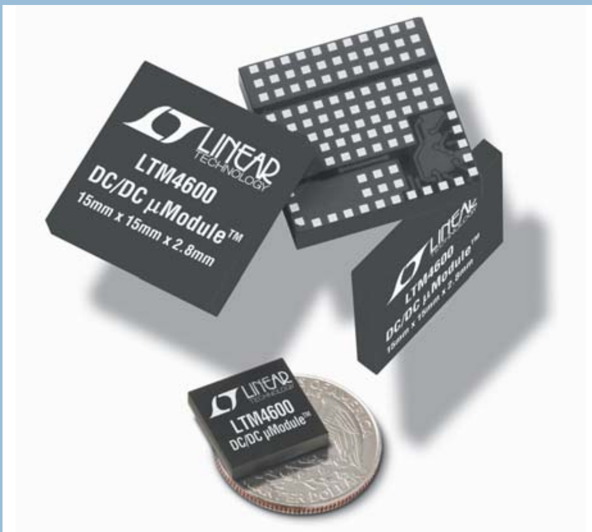

Figure A The LTM4600 shrinks a 10-A dc-dc converter into a 15-mm² LGA.

devices do not disturb the desired operating parameters. Currently, Barratt says, there is a wide choice of manufacturers for the LTCC (and laminate) substrates, although the more advanced integrated passive devices tend to be found in-house among suppliers with advanced silicon-wafer processing capability. This design flow offers a route to a highly integrated and well-characterised module where a system-on-chip (SoC) approach would be too expensive to design or might not be feasible at all.

The multi-die packages that have generated most new-product announcements in recent months have, however, been the so-called stacked-die packages. Of those, by far the most common application has been packages that stack a variety of memory die in the same package: and of those, the number-one application is in memory assemblies for mobile-phone main boards. In the mobile phone, pcb space is a at a premium and several types of memory are required: every handset design has at least one multi-chip package.

Although vendors have explored, and to some extent used, a variety of chip-to-chip and chip-to-substrate interconnection techniques, the most common practice inside packages currently on the market is for a relatively limited number of die—up to about four to six—to be mounted one on top of the other, usually

diminishing in size toward the top of the stack, in a pyramidal arrangement. The base substrate is most often a conventional laminate or ceramic micro-BGA package; and there is invariably a separation layer, possibly of adhesive alone, but more often a polymer "interposer" layer, between the semiconductor dice. The manufacturer individually wire-bonds each chip, using conventional techniques, down to the substrate. It does not take a lot of imagination to see that bond-pad space is at a premium on the substrate. This is both one of the limitations and the reason why memory stacks are attractive for multi-die offerings—a single bus may address more than one of the chips, reducing the pad-crowding effect. Previous constructions of multiple memory chips have included solid stacks of chips with solder bus connections formed in rails up and down the outer edges of the chip stack, but this is only feasible with stacks of identical dice, which is specifically what the mobile-phone application doesn't demand.

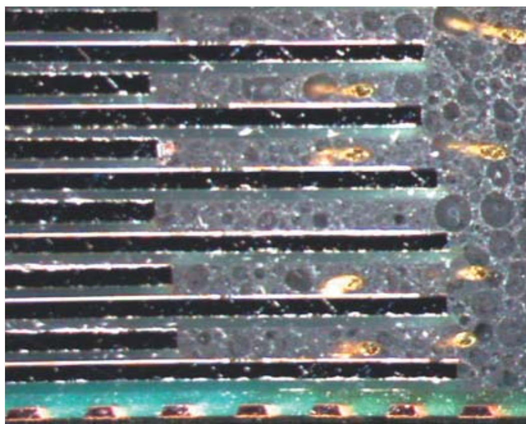

Figure 3 To demonstrate the feasibility of extended stacks of chips, ST fabricated this sandwich of eight 40-micron-thick chips with polymer separation layers. The moulding compound has filled all of the inter-chip gaps, showing that a large stack can be made robust.

One aspect the cellular-handset system board most definitely does require is low profile: component headroom above the PCB is at a premium. This poses little challenge to the multi-die package. In the "back-end of line" (BEOL) process steps, the manufacturer reduces the wafer thickness to as little as 40 microns before stacking the individual chips, so getting even a six-high stack within the profile of a standard BGA package is less of a challenge than one might think. Electrically, there's no problem with this: a chip's active layer, in which the semiconductor manufacturing process's diffusions and implants form the IC's active devices, is extremely thin, measured in Angstroms—a mere frosting of active surface on a mountain of silicon, relatively speaking. Technologies exist that enable a carrier layer of silicon to be separated from the parent wafer, so thin that it is flexible and transparent, yet still having more than enough thickness to support active circuitry. Multi-die packages don't need that level of sophistication. Nevertheless, it is remarkable that manufacturers can routinely reduce wafer thickness to no more than 40 or 50 microns, essentially by grinding and polishing away the reverse surface, before dicing the wafers into individual chips.

USING THE THIRD DIMENSION

Vendors and researchers are beginning to discuss the multi-die package as a solution to some of the problems of very large SoC designs. The very high up-front engineering cost of a large SoC in today's leading-edge technologies is well-known. One possible avenue to managing that cost might be to sub-divide the design into a series of smaller chips that mount either side-by-side (on a minature connecting substrate) or on top of each other. Aside from the economics, the functional blocks that make up an SoC design are (by definition) locked in a 2-D layout, and related blocks can end up a long way apart, relatively speaking. By working in a true 3-D fashion, designers can place functionally-related blocks in close proximity and—potentially—directly connect them in the vertical as well as the horizontal plane. Approaches under consideration for this include the interconnection of both bare and packaged die, leading to the description of package-on-package (PoP).

When does a multi-die package become a system-in-package? Different vendors use the terms somewhat loosely and sometimes interchangeably, but a working definition is that, whereas a multi-die package can act as a component that an external system can access—for example, the memory stack—an SiP performs some stand-alone processing or function. It follows that the mix of chips could therefore include a processor, or any of the other functions that are possible in silicon, such as sensors or MEMs.

At the Belgian-based Research organisation IMEC, there is a long-term project to construct an autonomous wireless sensor unit for a ubiquitous computing environment. IMEC researchers envisage it taking the form of a cube, formed of a stack of chips of identical size. Included in the stack would be environmental energy scavenging, sensing, processing, and wireless networking, all on separate dice with their own dedicated functions. Such a stack would require very advanced chip-to-chip interconnect. Working towards this aim, and also for less ambitious targets along the way, IMEC has a 3-D interconnect programme that includes bare-die and PoP research. The initial concept is to apply solder bumps around the periphery of the individual chips and reflow them together, but IMEC is also exploring much more ambitious joining techniques.

PARALLEL PROCESSING

One aim of IMEC's research is to bring to the 3-D world the same benefits of parallel processing that render conventional silicon fabrication economic. If you aim to directly connect one die to another, it is enormously more efficient if you can do so in a parallel process—however, building the die stacks one at a time, which is essentially what is done today, destroys that economic advantage. Interconnecting the chips while they are still part of a wafer would be ideal, and the organisation is researching various combinations of wafer-wafer bonding. Some of ideas involve intermediate layers between chips that incorporate through-hole interconnect; some involve direct die-to-die connection. IMEC has demonstrated concepts such as chips that researchers reduce in thickness far more than is routine today. The resulting piece of active silicon is no more than 10 microns thick. A further experimental concept will provide connections in the vertical dimension: copper plugs or "nails" pass right through the silicon, to bond to the terminal of the next chip in the stack.

More immediately, European suppliers such as ST Microelectronics and Infineon

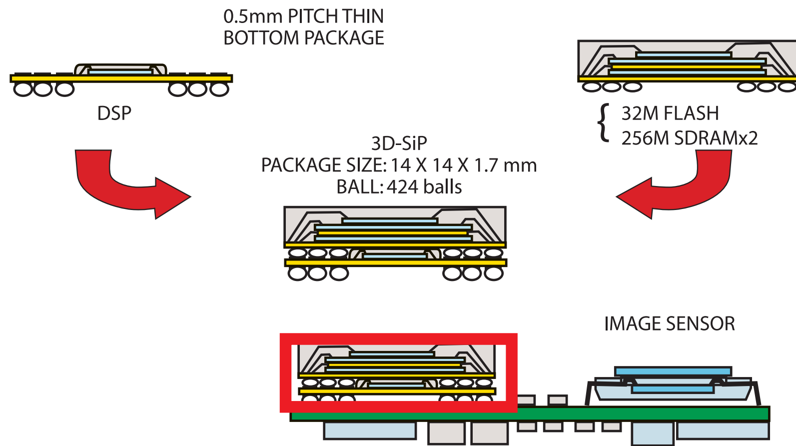

Figure 4 A substrate that carries complex routing is key to this assembly process proposed by Sharp, that potentially results in the elimination of a complete PCB.

are active in provision of multi-die packages. ST has demonstrated stacking of up to eight die in a package (**Figure 3**), although current offerings in the memory product line top out at four, with combinations such as NOR flash plus SRAM, or double NOR flash memory. ST's triple-band GSM/GPRS transceiver module is also a stacked-die product in which a Si-Ge BiCMOS RF chip is placed above an integrated active/passive device to yield a 7×7×1.4-mm package that also integrates many passives. The company further notes that the image-sensor subsystem for a camera phone is a potential area of application for the technology, saving board space by co-locating the image sensor with the image-processing DSP.

COMPETING WITH 90 NM

For the separate area of custom-chip development, ST's Group VP of new-package development, Carlo Cognetti, comments on the capability that SiP packaging offers: "There are smaller [chip] design houses that might be working one or two nodes back from the leading-edge semiconductor processes, while still doing excellent work—the investment to work with the newest processes might be beyond them. Working with a slightly more mature technology, stacking chips in the third dimension could allow such a design team to remain competitive with a high-budget, full SoC design in the most demanding process." Cognetti says that ST can support internal and external design teams with full tool suites for SiP design, just as it does for its ASIC flows.

Spansion, the AMD/Fujitsu joint-venture memory vendor, provides (as might be expected) a range of stacked-die memory products, including those for system memory and for the mobile-phone market. In the last few weeks, it has also started sampling package-on-package stacked combinations of memory and logic chips. The concept pairs the PSvfBGA package from Amkor, which carries logic and is the base element, with a packaged flash memory chip from Spansion on the upper level, all in a package 1.4 mm high. With inherently short trace connections from logic to memory, Spansion offers the concept as a solution to the high-speed signal demands of 133-MHz DDR memory.

Infineon uses stacked memory die for a range of products, including high-density system memory; the company has also demonstrated an advanced technique for direct-flip-chip interconnection of chips without intermediate "bumping" and reflow-soldering.

Sharp has recently introduced a change in the technology it employs in the construction of stacked-die packages. Previously it has used the relatively standard technique of wire-bonding all of the die in a package down to a single substrate that provided interconnect to the external environment (the system PCB). Now, it has added the capability to flip-chip mount a complex chip—a microprocessor, say—on to a substrate that provides complex routing of connections from both that chip, and from a chip stack on a separate carrier mounted above it (**Figure 4**). The technique reflow-solders the new base substrate as a component to the PCB. This seemingly small change can be significant in an application such as a digital still camera, the company says. Earlier designs would have required a separate footprint for a DSP chip, and a separate stacked memory device. The additional area needed for the image sensor would force the use of two PCBs to achieve the required surface area within the outline of the camera. When DSP and memory stack occupy the same footprint, the designer can eliminate one PCB.

The resources available to design a multi-chip package have never been greater or more varied: conversely, if you are simply purchasing a device constructed in this manner, do you need to know that it's built in that way? In many cases you don't and you can treat the component as just that: another component. Exceptions may lie in reflow-soldering—if the vendor has used multiple stages of reflow to construct the part, there may be limits on the temperature profile you can use to attach the part to your PCB. However, some of the newest chip-joining techniques overcome that issue: no longer a simple soldering process, they employ advanced metallurgy to, in effect, fuse the chips. And, if you have a very complex set of stacked chips in a single package, you may need to be careful about how much of the stack is driven at one time, for obvious thermal reasons. EDN

You can reach Editor Graham Prophet at +44-118-935-1650, fax +44-118-935-1670, and gprophet@reedbusiness.com.

FOR MORE INFORMATION

Amkor
www.amkor.com

Cambridge Silicon Radio
www.csr.com

IMEC
www.imec.be

Infineon
www.infineon.com

Insight SIP
www.insightsip.com

Linear Technology
www.ltc.com

Sharp
www.sharpsme.com

ST Microelectronics
www.st.com

White Electronic Designs
www.wedc.com