# Technology Review:

## Cooling 3D Stacked Die Packages

Qpedia continues its review of technologies developed for electronics cooling applications. We are presenting selected patents that were awarded to developers around the world to address cooling challenges. After reading the series, you will be more aware of both the historic developments and the latest breakthroughs in both product design and applications. We are specifically focusing on patented technologies to show the breadth of development in thermal management product sectors. Please note that there are many patents within these areas. Limited by article space, we are presenting a small number to offer a representation of the entire field. You are encouraged to do your own patent investigation. Further, if you have been awarded a patent and would like to have it included in these reviews, please send us your patent number or patent application.

In this issue our spotlight is on cooling 3D stacked die packages. There is much discussion about its deployment in the electronics industry, and these patents show some of the salient features that are the focus of different inventors.

### SEGMENTATION OF A DIE STACK FOR 3D PACKAGING THERMAL MANAGEMENT

US 7,928,562 B2, Arvelo, et al.

An apparatus 1000 is provided that is able to reduce a thermal penalty of a three-dimensional (3D) die stack 200 for use in a computing environment by allowing for efficient cooling of the die stack 200. The thermal penalty refers to the degree by which upper layers of the die stack 200 provide a thermal resistance to heat flow along a thermal path through the die stack 200. As noted above, where the thermal penalty of a die stack 200 is high, the thermal resistance is also high and heat flow is impeded. In this situation, a temperature of the die stack 200 may increase and performance degradation or other similar failures may occur.

The apparatus 1000 includes a substrate 300 installed within the computing environment and a first component 201 to perform operations, such as computing operations for the computing environment. The first component 201 is coupled to the substrate 300 in a stacking direction that is

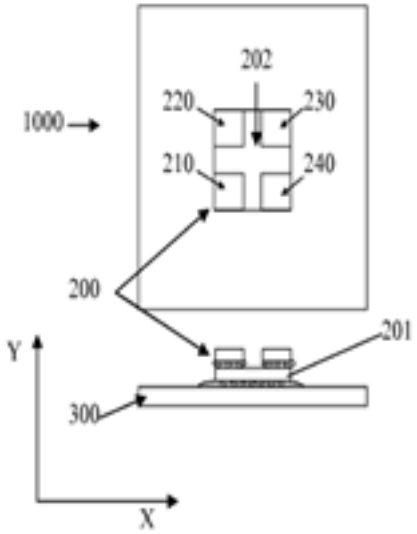
| PATENT NUMBER | TITLE | INVENTORS | DATE OF AWARD |
|---|---|---|---|
| US 7,928,562 B2 | SEGMENTATION OF A DIE STACK FOR 3D PACKAGING THERMAL MANAGEMENT | Arvelo, et al. | Apr 19, 2011 |
| US 8,030,113 B2 | THERMOELECTRIC 3D COOLING | Hsu, I., et al. | Oct 4, 2011 |
| US 8,299,608 B2 | ENHANCED THERMAL MANAGEMENT OF 3-D STACKED DIES PACKAGING | Bartley, G., et al. | Oct 30, 2012 |

substantially parallel with the thermal path. That is, the thermal path is defined to extend away from the first component 201 and is parallel with a direction that is substantially normal to a plane of the substrate 300. Conversely, a predominant path of the thermal resistance is defined to extend in the opposite direction.

One or more second components 210, 220, 230 and 240 are also configured to perform operations and are coupled to the first component 201 to thereby form an upper layer of the die stack 200. The second components 210, 220, 230 and 240 are formed by having been segmented and, as such, they cooperate to form a vacated region 202 between them and along the stacking direction. A thermal interface material (TIM) 350, through which the thermal path extends, is disposed on a top surface of the first component 201 in the vacated region 202 and on top surfaces of the second components 210, 220, 230 and 240. A lid 500, such as a cooling cap, is positioned in thermal communication with the TIM 350 and forms an exterior appearance of the die stack 200.

The lid 500 includes one or more protrusions 501 that extend into the vacated region 202 to nearly contact an exposed top surface of the first component 201. The lid 500 and the protrusion 501 further include surfaces that thermally communicate with at least the TIM 350. With this arrangement, the lid 500 and the protrusions 501 are configured to remove heat that is generated within the first and second components 201 and 210-240 during the operations from the first and second components 201 and 210-240 via the TIM 350.

In accordance with embodiments of the invention, the substrate 300 may include a printed circuit board (PCB) to which the die stack 200 is electronically coupled and which is operably installed within the computing environment. In addition, an intermediate layer may be disposed between the die stack 200 and the substrate 300.



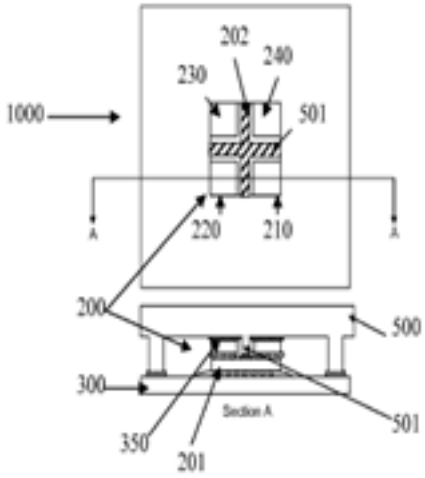

This intermediate layer may be coupled to the substrate 300 via electrical leads and may include a power regulating layer, which is configured to regulate power applied to the die stack 200, and/or a silicon carrier, which is provided with an integrated decoupling capacitance (DECAP). In accordance with further embodiments of the invention, the first component 201 may include a processor. Here, the operations performed by the first component 201 may include computational operations related to the operation of the computing environment. Concurrently, the one or more second components 210, 220, 230 and 240 may include an additional processor and/or a memory unit, including at least one of a static random access memory (SRAM) and a dynamic random access memory (DRAM).

The first component 201 is both electrically and structurally coupled to the substrate 300. The electrical coupling may be accomplished via electrical interconnects, such as controlled collapse chip connections (C4) bumps, which may be encapsulated in an underfill material, such as epoxy resin or some other suitable adhesive. The second components 210-240 are coupled to the first component 201 both electrically and structurally in generally similar manners. However, in this case, a pitch (i.e., a density) of the electrical interconnects maybe significantly higher than the coupling

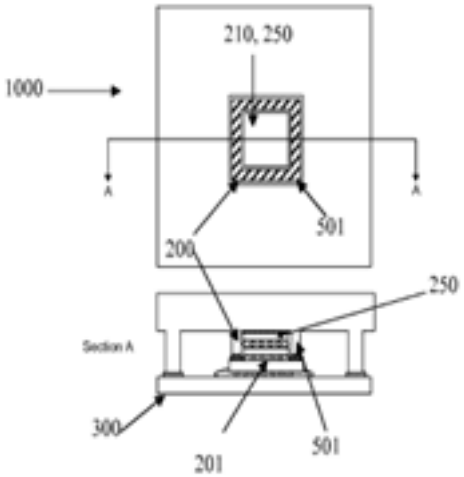
between the first component 201 and the substrate 300. An inter-chip region may, therefore, be formed between the first and second components 201 and 210-240.

During standard operations of the first component 201, which may include a testing phase and/or a normal operational phase, it may be seen that at least one hot spot may form across an X-Y plane of the die stack 200 as a result of the thermal resistance to the heat flow along the thermal path. The presence of such a hot spot may be predicted based upon the architecture of the first and second components 201 and 210-240 or may be observed during the standard operations thereof.

According to embodiments of the invention, the second components 210-240 are segmented so as to allow the lid 500 and the protrusions 501, which extend into the vacated region 202 formed by the segmented second components 210-240, to remove heat from the first and second components 201 and 210-240 via the TIM 350 to thereby limit a size and/or an intensity of the hot spot. In further embodiments of the invention, the second components 210-240 are to be segmented in a particular pattern that allows the lid 500 and the protrusions 501 to be precisely positioned proximate the hot spot. In this way, the lid 500 and the protrusions 501 may be particularly positioned to remove heat from those particular portions of the first and second components 201 and 210-240 that form the hot spot.




As an example, the second components 210-240 may be segmented into four portions. Here, each of the portions may be disposed proximate to respective corners of the first component 201 with the footprint of the vacated region 202 being defined between the respective footprints of the portions.




As another embodiment and as shown in FIGS. 4B and 5B, each of the portions may be disposed proximate to respective edges of the first component 201 with the footprint of the vacated region 202 being defined between and alongside the respective footprints of the portions.
As an yet another example the second component 210 may be singular in number and may be segmented to be narrower in at least one planar direction than the first component 201. For example, each edge of the second component 210 may be withdrawn in the X-Y plane of the die stack 200 from edges of the first component 201 by a distance T. In this way, a footprint of the vacated region 202 is defined to be above an outer region of the footprint of the first component 201 and outside of a footprint of the second component 210. In a further embodiment of the invention, the distance T may be approximately 2 mm.

In addition to the use of the segmented second components 210-240, it is understood that additional embodiments of the invention are

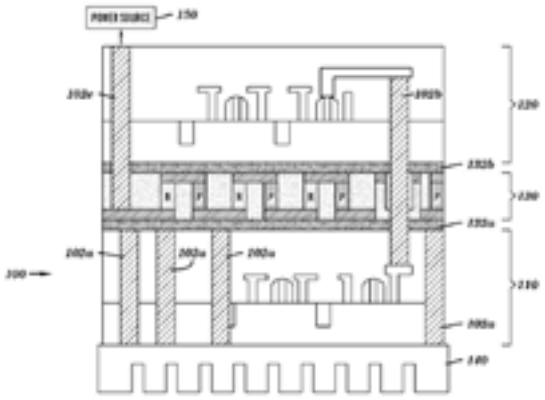
possible. For example, the apparatus 1000 may further include a cooling system to perform forced convection liquid cooling, in which coolant is pumped through the die stack 200. Similarly, a density of the C4 bumps may be increased. Of course, still further cooling options are available and may be applied to the die stack 200.

## THERMOELECTRIC 3D COOLING

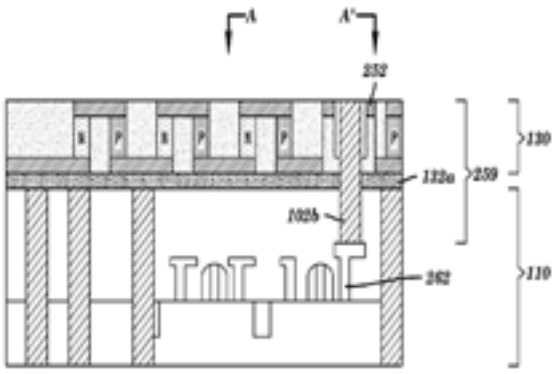
US 8,030,113 B2, Hsu, L., et al.

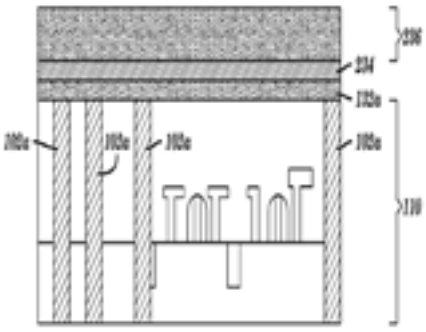
By way of overview and introduction, the embodiments of the invention are directed to thermo-electric 3D cooling. A first and second coupling layer separate a thermoelectric cooling (TEC) plate from a first and second chip. Each chip has electronic devices embedded therein. Through silicon vias (TSVs) are used for various purposes in the 3D chip stack. One TSV traverses the TEC but is electrically isolated from the TEC. Such TSV electrically connects a device in the first chip with a device in the second chip. Another TSV electrically connects the TEC with a heat sink. The TEC, sandwiched between two chips, dissipates heat generated by hot chips in the 3D stack and transfers heat to cold chips. Generally, chips located in the middle of the 3D stack are hotter than those located close to the upper and lower surfaces of the 3D stack. Such TSV then sends that heat to a heat sink or heat exchanger, which dissipates the heat from the 3D stack. Yet another TSV electronically connects the TEC with a power source. Such TSV powers the thermoelectric plate.

Cooling stacked 3D chips is not a straightforward challenge because of the various considerations the 3D chip designer must contemplate. The alignment and bonding among chips and modules is one consideration. Cooling stacked 3D chips also requires consideration of the electrical paths through the module to form connection from one chip to another. Cooling stacked 3D chips further requires the consideration of the functionality of each chip in order to properly arrange the plurality of chips and modules. For example, CPU chips inherently consume more power and thus generate more heat than the memory chips. Finally, heat dissipation is a consideration. Sometimes heat is dissipated from one to another chip, or from center core of the stack to the side of the stack, so that it can be effectively removed by means of heat sink or heat exchanger attached at the outside boundaries.



As explained the embodiment comprises two chips 110, 120 separated by a TEC plate 130. Each TSV may serve a unique purpose. TSVs 102 a dissipate heat from the TEC plate 130 to a heat sink 140. A device in chip 120 generates the heat, which the TEC plate 130 transfers to chip 110 and through TSV 102 a sends to heat sink 140. TSV 102 b, on the other hand, which traverses the TEC plate 130, remains electronically isolated from TEC plate 130, and yet connects the upper chip 120 and the lower chip 110. Lastly, TSV 102 c provides power from power source 150 to the TEC plate 130.



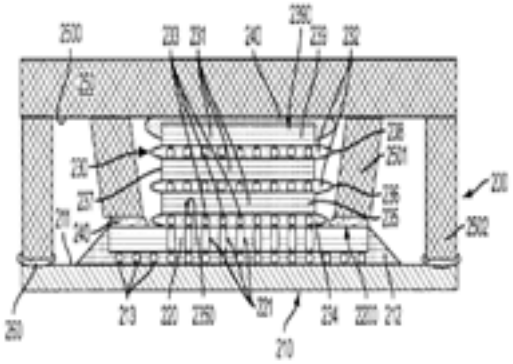


Note that the TEC plate 130 serves as a solid-state heat pump that consists of multiple pairs of N-type and P-type semiconductors as thermoelectric elements connecting electrically in series and thermally in parallel, and it absorbs heat at the cold plate and releases the heat at the hot plate. In such situation, heat is transferred downward against the temperature gradient. To facilitate further heat transfer from lower chip 110 to the heat sink 140, thermal TSV's 102 a filled with high thermal conductivity metals, e.g. copper (Cu), tungsten (W), or aluminum (Al), are placed through the lower chip 110 and in contact with the bottom of the TEC plate and the heat sink 140. Electrical TSV's 102 b, which are also filled with high conductivity metals, e.g. copper (Cu), tungsten (W), or aluminum (Al), are placed among the chip stack and through TEC plates to transmit electrical signal to achieve 3D chip integration. Coupling layers 132 a and 132 b are placed between the TEC plate 130 and the lower chip 110 and upper chip 120, which serve as electrical insulators and provide efficient heat transfer between TEC plate and chips. Dielectric materials with good thermal conductivity, such as CVD diamond or silicon carbide (SiC), are proper candidates for coupling layers 132 a, 132 b.

The bottom segment 259 of the electrical TSV 102 b is then fabricated through the entire thickness of the TEC plate 130, the coupling layer 132 a, and partially through the lower chip 110, landing on and contacting the designated metallization structure 262. Note that the electrical TSV 102 b also goes through the conducting layers of the TEC plate 130, e.g. conducting layers 234, 252 in the case shown in FIG. 2H. It is important to avoid electrical short between the TSV 102 b and layer 252.

## ENHANCED THERMAL MANAGEMENT OF 3D STACKED DIES PACKAGING

US 8,299,608 B2, Bartley, et al.



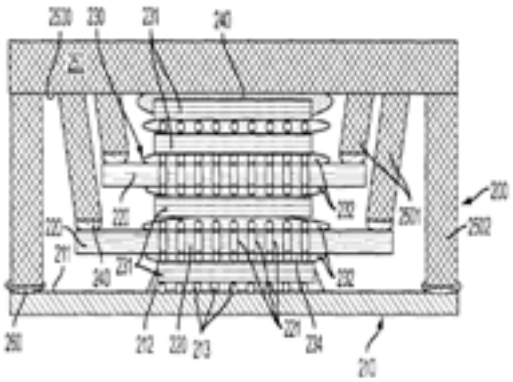
Generally, an electronic package is a hardware component in which active devices, such as logic or memory devices, and passive devices, such as resistors and capacitors, are enclosed. The electronic package performs functions of an electronic system, such as those used inside a mobile phone, a personal computer, a digital music player, etc. Common electronic packages are classified as either flip-chip or wire-bond packages.

In a typical flip-chip electronic package, a single die 10 is joined to a ceramic or organic material substrate 100 through controlled collapse chip connection (C4) bumps 130, which may be encapsulated in an underfill material 160. A lid 120 provides thermal cooling and mechanical protection for the die 10. A thermal interface material (TIM) 150, which can be an elastomer, adhesive, gel or metal, may be disposed between the chip 10 and the lid 120. A bond 170, such as an elastomer, epoxy or mechanical fasteners, attaches the lid 120 to the substrate 100. The substrate 100 may be further coupled to a printed circuit board (PCB) 110 via leads 140.

In order to increase bandwidth and function, the package is enhanced, with multiple dies 11, 12, 13 and 14 being attached to a multi-chip substrate 101. Here, the electronic package grows in dimension to accommodate the multiple dies 11, 12, 13 and 14 and presents cost, size-related and reliability trade-offs. In yet another configuration, multiple dies 15-18 are vertically stacked onto a single-chip substrate 102.

Since a set of chips, resistors, capacitors and/or memory units may be provided in a particular die stack, it may be a complete functional unit requiring few external components. As such, use of the die stack in space-constrained environments, such as mobile phones and computers, may be valuable. Also, a stacked die can provide an increased electrical interconnect density with less latency and lower power consumption, which can increase system performance. This is especially true with "multicore" chips where it is sometimes difficult to increase the bandwidth to memory adequately.

Despite its benefits, however, a problem with a vertically configured die stack exists in that the upper die provides thermal resistance along the primary heat flow path from the die stack and into the cooling lid . During normal operations, this thermal resistance causes an internal temperature of an electronic chip within a die stack to increase as compared to that of a non-stacked electronic die. As a result, performance of the electronic package with a die stack may be degraded.



The substrate 210 may be formed of ceramic and/or organic materials and may be formed as a printed circuit board (PCB) or a component that couples the die stack package 200 to a PCB for use in a computing device. The substrate 210 has a generally planar surface 211 on which underfill

212 is disposed. The underfill 212 is electrically insulating, provides for a mechanical connection between the thermal plate 220 and the substrate 210 and provides a heat bridge between the thermal plate 220 and the substrate 210. Electrical connections 213, such as solder joints and/or controlled collapse chip connection (C4) bumps, may be embedded within the underfill 212 with underfill 212 further providing structural support for the solder joints during any thermal expansion that may occur.

The thermal plate 220 is disposed on the underfill 212. The thermal plate 220 may be a substantially flat member with planar surfaces and may be formed to define a through-silicon-via (TSV) 221 by which a signal is transmittable across the thermal plate 220 and between the die stack 230 and the electrical connections 213.

The thermal plate 220 includes at least one of silicon, diamond, metallic material and/or metallic alloy. The material of the thermal plate 220 may be chosen in accordance with a material of the substrate 210 so as to provide for CTE matching and/or mechanical strength where necessary. For example, where the substrate 210 is formed of an organic material, the thermal plate may be formed of a high strength material that is CTE matched to the organic material such that thermal expansion of the die stack package 200 is permitted but warping and/or non-coplanarity is reduced due to the additional coupling through the thermal plate 220. Alternatively, for the organic substrate 210, the CTE of the thermal plate 220 can match that of the organic substrate 210 or that of the die stack 230. Here, either choice has benefits in that matching to the die stack 230 CTE improves the die stack 230 and its interconnect reliability while matching to the substrate 210 enhances the overall package coplanarity. When the substrate 210 is formed of a ceramic material, the thermal plate 220 may be formed of a relatively low CTE material that matches both the CTE of the substrate 210 and the CTE of the die stack 230.

As a whole, the die stack 230 is disposed on the thermal plate 220 or with the thermal plate 220 inserted therein and in electrical communication with the electrical connections 213. As such, during computing operations of the computing components 231, signals may be transmitted from each of the computing components 231 to the electrical connections 213 and vice versa by way of, for example, the TSV 221 of the thermal plate 220 and the additional electrical connections 233.