

用面对面叠加法实现芯片的 3D 集成

谢苑林¹, Deborah Patterson²

(1. 拓朗半导体, 美国; 2. 安靠封装测试, 美国)

本文原载于美国 Chip Scale Review 杂志 2013 年 5 月 -6 月期, 并已获准翻译成中文在“中国集成电路”杂志发表。

摘要: 面对面芯片叠加使一颗芯片置于另一颗芯片的倒装凸块阵列中, 从而极大减小了封装厚度。POSSUM™ 封装是指两颗或多颗芯片用面对面的方式叠加, 其中较小的芯片置于较大芯片上没有互联倒装凸块的区域。较小的芯片在薄化后通过铜柱微块组装到较大的芯片上。因此, 薄化了的较小芯片和它的铜柱微块的总体高度就比其附着的较大芯片的倒装凸块经回流后的高度要低很多。一旦组装完毕, 较小芯片就被有效地夹在上面的较大芯片和下面的基板中, 同时被一环或多环倒装锡凸块包围。底部填充剂的使用同时保证了铜柱和无铅锡凸块在测试和使用中的可靠性要求。因为信号在两颗芯片的表面运行, 所以互连线极短, 可以实现近距离信号匹配, 和小电感的信号传输。这种面对面芯片叠加方式保证了很好的芯片间信号传输的完整性, 是穿硅孔 (TSV) 封装技术的低成本的有效替代。

关键词: 芯片对芯片封装; 袋式封装; 穿硅孔; 现场可编程门阵列; 铜柱

Realizing 3D IC Integration with Face-to-Face Stacking

John Xie¹, Deborah Patterson²

(1. Altera Corporation, USA; 2. Amkor Technology, Inc., USA)

Abstract: Face to face die stacking can offer clever approaches to minimize overall stacked height, especially when one die can be “hidden” within the bump array of another. The POSSUM™ stacked die configuration describes two or more devices assembled face-to-face where a smaller die is nested within the I/O-free areas of the larger die. The smaller die is thinned significantly and joined to the larger die through copper pillar micro-bumps. The combined height of the thinned die and its copper pillar bumps are much less than the collapsed flip chip bumps that mount the larger die to the substrate or PCB. Once mounted, the smaller die is effectively sandwiched between the larger top die, the bottom substrate, and surrounded by one or more rings of solder bumps. Underfill ensures that both the copper pillar and lead free solder bumps meet package and board level reliability requirements. Because the signals are routed between the top sides of the two devices, the I/Os can be closely matched up, ensuring extremely short interconnect

lengths and very low inductance copper pillar joints. The resulting structure ensures high die-to-die signal integrity and provides a cost effective alternative to a through silicon via (TSV) solution.

Key words: CoC package; Possum™ package; TSV; FPGA; Copper pillar

1 概述

摩尔定律驱动的半导体制程的进步是当今高科技工业的发展动力。但是,传统的制程换代已逐渐不再满足新一代系统性能、产出和功耗的要求。虽然芯片并排(2D结构)、封装叠加(PoP)和封装套封装(Package-in-Package, PiP)等平台可以集成两个或多个芯片,但它们不能提供足够的密度、带宽和功耗来满足新一代产品路线图的要求。

SoC(System-on-Chip)是将更多的功能集成在一颗芯片上,而系统封装(System-in-Package, SiP)或多芯片封装(Multi-Chip Package, MCP)是将多个芯片集成在一个封装体中。如何平衡SoC或SiP的应用,需要很好地考量。优化系统性能和控制整体成本同等重要。而最重要的考量之一是加速进入市场的步伐,因为对最终用户来说,这是战略性的成功因素。

为了正确评估上述平衡,业界正积极研究2.5D和3D封装。这些研究用于考量芯片的使用效率对比功能集成(制程优化),考量从工程到大批量生产的芯片良率,和封装策略对信号完整性的影响等。理想的封装在提供保护和散热的功能外,对信号最好是透明的。

通过使用穿硅孔(TSV)的芯片叠加是3D集成的一个例子。在3D芯片叠加中,焊锡凸块(solder bump)被用于从上到下芯片间的连接,使信号得以传输。3D芯片叠加体可以单独封装,或通过并排的方式在基板上和其它芯片互联。相比之下,2.5D芯片集成是使用带有TSV的多层硅基板,使排列于其上端的多个主动芯片或芯片叠加体实现高密度互联。穿硅孔(TSV)将硅基板上部的经过互联后的信号传到基板下部的倒装焊锡凸块上(flip chip bump)。芯片通常使用更小的铜柱微块(Cu pillar

micro-bump)组装到硅基板上。

在信号完整性上,使用硅基板实现主动芯片的互联相比使用传统基板材料,性能优越许多。事实上,硅基板让我们可以换一种方式去思考SoC和SiP的平衡问题。2.5D封装提供硅级别的互联密度,并且由于主动硅芯片之间连接线路缩短,信号延迟极大减小。此外,由于I/O驱动可以设计成硅对硅(Si-Si)的直接连通,从而减少了焦耳热损耗,芯片功耗也随之降低。2.5D封装还具有快速进入市场的优越性,这是因为3D芯片在设计、生产和封装上都需要较长的周期。图1是叠加芯片性能优越性在互联密度指数、热阻和功耗上的比较,包括独立封装、2.5D封装和3D封装三种形式。

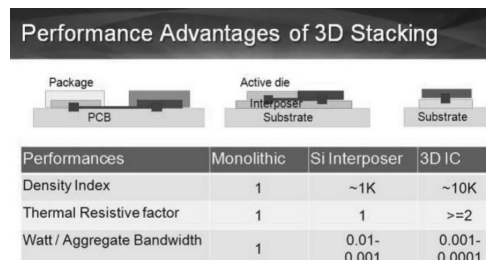


图1 三种多芯片叠加方式在连点密度,热阻和功耗上的比较

与传统的多芯片封装(MCP)相比,2.5D叠加封装的优越性为趋于SoC目标提供了一个新的视角,也让我们从另一个角度审视2.5D封装对芯片和封装功能集成的近期和远期路线图的影响。除了2.5D和3D半导体封装方式外,还有其它方式用于提高功能和性能,并同时保持小尺寸封装,支持产品快速进入市场。

优化芯片叠加方式,并实施合理的设计和生产流程可以实现低成本集成。意识到这一点,Altera和Amkor合作制定了功能集成的路线图:例如,是否使用SoC,或重新构建芯片,将其设计成多芯片体,以提高芯片的良率。

2 什么是“面对面”封装

芯片叠加创新是众多半导体公司和封测公司的开发重点。Altera 公司的研发部门制定了未来 5 年的芯片叠加技术路线图。类似地, Amkor 的先进封装技术开发部支持各式芯片叠加的封装技术, 包括打线和倒装解决方案。

“面对面”封装技术是裸片叠加 (CoC) 的一种。无需使用穿硅孔 (TSV), 两颗裸芯片通过铜柱微凸块, 面对面地将相应电路紧密相连。作为母体, 其中一颗芯片必须足够大, 使得它与基板的倒装连接点可以置于芯片周边, 而作为子体的小芯片可以置于中间。子体芯片和它的铜柱微凸块的总体高度必须小于母体倒装凸块回流后的高度, 并使子芯片和基板表面有足够的间隙。图 2 是这种封装结构的一个例子, 即 Amkor 的 POSSUM™ (袋式) 封装。

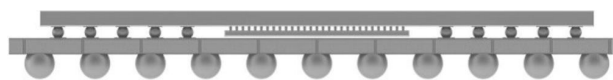


图 2 无需 TSV 的面对面裸片叠加封装 (POSSUM™), 母芯片通过倒装焊锡凸块或铜柱与 BGA 基板相连, 而子芯片通过铜柱微凸块与母芯片相连

这种封装形式的关键是无需 TSV 或硅基板, 而只用铜柱微凸块将两颗主动芯片直接互联。事实上, 这种封装形式应用现有成熟的倒装基板及组装技术。它提供了几乎所有 3D IC 叠加的优势: 堆叠芯片直接互联, 减少信号延迟, 提高信号质量, 并比 2.5D 的结构更省功耗。POSSUM™ 封装尺寸减小而 Z 方向上的厚度并不增加。集成成本远少于 2.5D 叠加方式。面对面的倒装结构还消除了因在主动芯片中使用 TSV 而导致的复杂设计和相应的较高生产成本。

面对面叠加有多种封装形式。除了面对面倒装芯片尺寸封装 (CSP) 或球阵列 (BGA) 外, 也可用于支架类封装 (图 3)。面对面塑封球阵列 (PBGA) 代表着此类封装最先进的形式, 因其包含了一个 POSSUM™ 裸片安装到较大芯片表面, 并实施 PBGA

F-2-F Integration
Including FCBGA, WB PBGA and Lead-frame Packages

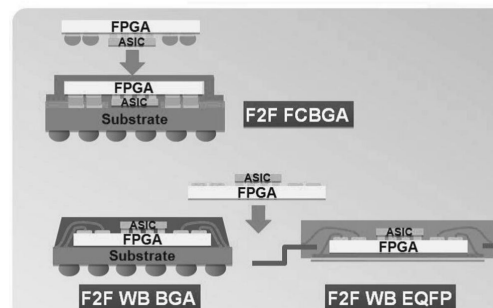


图 3 三种面对面芯片堆叠的封装形式

封装工艺。

3 POSSUM™ 封装偏离成本和互联密度关系曲线

随着 I/O 互联密度的增加, 封装成本也随之增加。图 4 用左下角的典型 2D 倒装 BGA (FCBGA) 作为基准点, 其互联间距为 $100\mu\text{m}$, 互联线宽为 $10\mu\text{m}$ (只有在硅片上才能实现)。随着互联间距和线宽的减小, 封装成本增加。当使用超高互联密度 (UHDI-O) 的封装时, 互联间距减至 $50\mu\text{m}$; 而使用 2.5D 硅基板封装更可以使互联间距小于 $50\mu\text{m}$ 。这个互联密度增加的趋势也带来了成本的增加。与这个趋势相背离的是面对面的封装形式。无论是倒装 (flip chip) 的还是引线 (wire bonding) 的封装方式, 在同样互联密度的情况下, 成本却得到降低。

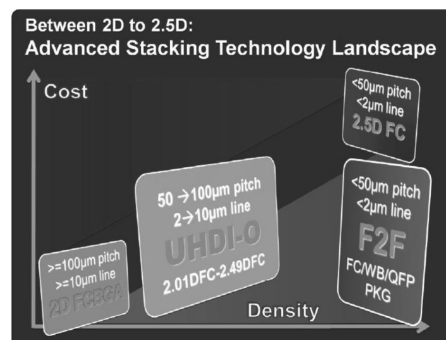


图 4 随着互联密度增加的封装成本趋势

4 面对面叠加技术的应用

很多半导体芯片的组合都可以应用面对面叠加技术。例如，一个高密度的现场可编程门阵列（FPGA）和一个专用集成电路（ASIC）的组合可以在特殊应用环境中将 ASIC 的高性能和 FPGA 的灵活性都能很好发挥。也许最明显的应用是将用不同半导体技术生产的芯片进行组合，如一颗 RF（射频）前端芯片和一颗 FPGA 芯片的组合。面对面联接使两颗芯片间有足够带宽来允许与 RF 电路中数字信号直接互联。类似地，FPGA 和外部存储芯片的直接互联，大大拓展了 FPGA 与低延迟存储芯片的互联能力，从而减小了使用高频芯片带来的能量损耗。

因为系统吞吐量需求和工作存储带宽的增加，存储能耗问题开始变大。图 5 比较了不同的存储方案下每个 I/O 互联信道的转换能耗。它显示了面对面 SRAM 的叠加形式在能耗和带宽上与 3DIC 叠加性能相近。

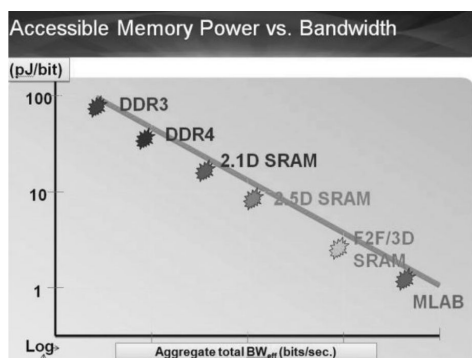


图 5 模拟数据显示 SRAM 到 FPGA 的面对面叠加与用 TSV 的 3D 叠加具有相同的能耗和带宽（MLAB 是指嵌入式存储或 SOC 设计）

5 培育 3D 封装的替代品

在芯片开发中，同时考量芯片设计、生产和封装的相互作用是十分重要的。这种多方面的考量将极大影响后续产品的技术方向，特别是，当芯片设计必

须同时优化总体成本、产品性能和市场推入时间等因素时，更是如此。这种理念促成了芯片供应商和外包半导体组装和测试服务商（OSAT）在芯片开发的早期开始合作。为了支持产品路线图上更具进取的目标，在确认更有效的芯片封装技术上，对各类专家、资源和经验的组合借鉴是必要的。

面对面 POSSUM™ 封装正是这样一个组合借鉴的案例。这种封装结构的选取是因为它可以用更短、更快的互联方式将两颗或多颗芯片组装。更少的电感、串扰和寄生电阻意味着这种方法可以用于高频高带宽芯片。这种封装基于已经量产的生产工艺，如铜柱微凸块、倒装焊锡凸块、芯片对芯片的组装回流/热压键合、芯片对芯片的引线键合，或芯片对基板的倒装或引线键合等等。此外，这种封装无需引进复杂的晶圆工艺设备，能很快进入大规模量产。

6 面对面的芯片与芯片的组装

这种面对面的芯片与芯片组合封装平台是铜柱晶圆植凸块（copper pillar wafer bumping）技术和先进的细间距组装（fine pitch assembly）技术的直接结果。植凸块工艺经修改后用于产品的设计和组装。

在母芯片晶圆和子芯片晶圆经过晶圆铜柱植凸块工艺后，要进行晶圆背磨和芯片分割步骤。对于母芯片晶圆来说，考虑到工艺优化的影响，可以选择芯片分割或保持晶圆状态。接着，子芯片用铜柱微凸块的联接方式组装到母芯片上。大量的研究对铜柱凸块的结构，材料和工艺进行了优化，以控制铜柱凸块的高度，保证更好的可靠性、电性能和传热性能。40 μm 间距周边分布和 80 μm 间距交叉阶梯分布的铜柱微凸块已经可以大量生产，而 30 μm 间距周边分布和 60 μm 间距交叉阶梯分布的铜柱微凸块也已经可以小批量生产。

7 多种凸块高度的倒装工艺实现

在目前的面对面的芯片组装工艺中，子芯片拥

有较高的铜柱微块,而母芯片的铜柱微块较低。图6可以看见子芯片和母芯片互联时,两个铜柱微块的细节。

还可以注意到,较大的母芯片上安植了先进的两种高度的铜柱凸块。较低的铜柱微块用于芯片与芯片的互联,而周边较高的铜柱凸块则用于母芯片与其它载体,如基板的互联。

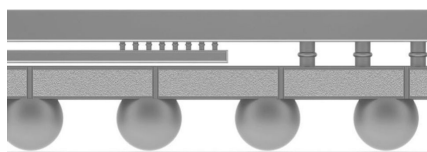


图6 不同尺寸的母芯片和子芯片的铜柱微块互联

8 无需 TSV 的 3D 封装

利用面对面铜柱键合方式实施两颗芯片组合的封装方法有两种。第一种是焊锡回流加虹吸底部填料的方式(MR+CUF)。这种方式常用于凸块间距大于 $80\mu\text{m}$ 的情形,且正朝着用于凸块间距大于 $50\mu\text{m}$ 的情形发展。

第二种芯片组装工艺利用非导电粘合剂作为底部填料进行热压键合(TC+NCP)。这种工艺目前用于芯片到基板的倒装封装的大规模量产,铜柱间距为 $40/80\mu\text{m}$,具有很高的良率和可靠性。

TC+NCP的组装工艺可以用于母芯片和子芯片的互联。在面对面的芯片组装完成之后,芯片组合体将通过焊锡回流键合方式组装到基板上。填料工艺需要保证填料流动合理,并均匀地填充母芯片和子芯片之间的缝隙,以避免底部填料成型后气泡的形成。这种工艺被进一步优化来保证无孔填料的形成和母芯片与子芯片之间的合理缝隙。芯片和基板的翘曲变化都需放在考量之列。芯片和基板组合之后,还要进行金属盖粘结、BGA球粘结、激光标注和测试等。图7是一个面对面封装体的截面图和几个放大区域图。

正在进行的开发工作推进了这种芯片对芯片封装结构的影响力。在母芯片处于晶圆状态下,更小

的子芯片对母芯片的键合,即所谓的芯片对晶圆(CoW)的封装工艺,就是正在开发的技术。CoW也可以在POSSUM™平台上支持晶圆级芯片尺寸封装(WLCSP)。3D技术平台中应用的先进材料,类似地也可以用于CoC封装技术中。

我们对使用铜柱加锡银帽(Cu+SnAg)作为凸块的封装结构进行了温度循环试验。在 -55°C 到 125°C 的测试条件下,这些封装体通过了2000个温度循环(JESD22-A104D, B级标准)。还有一些封装体中,母芯片和子芯片分别使用镍柱加锡银帽(Ni+SnAg)和铜柱加锡银帽(Cu+SnAg),它们可以在 -55°C 到 125°C 的测试条件下,通过3000个温度循环。

9 总结

芯片供应商和封装厂商在芯片开发之初进行合作,更易于实现较高性能价格比的封装解决方案,以支持终端客户要求的多样性和积极进取性。

在很短的时间里,这支共同开发团队在晶圆的凸块设计和生产、基板结构和工艺,及封装材料组合(BOM)和生产流程上都取得了巨大成效。这个团队还研究了其它几种芯片叠加方法和它们到现有封装平台的组装(如:从引线键合和倒装到BGA和框架类封装形式)。

POSSUM™结构展示了如何在不使用昂贵的2.5D和3D TSV技术的条件下,优化面对面的封装形式,以实现高性能的多芯片叠加。

虽然使用TSV的3D IC集成仍然是工业界的目标,但近期我们可以用更聪明的方法实现低成本的芯片叠加。面对面的倒装封装技术可以用于开发复杂的、高集成度的和高性能的芯片产品,这些芯片适用于众多的计算、通信、汽车和消费类终端产品中。这项技术对具有编程灵活性的FPGAs来说尤其重要,因为FPGAs对芯片功能提高、尺寸减小和重量减轻的要求更为强烈。

封装

中国集成电路
China Integrated Circuit

CIC

感谢

作者希望感谢来自 Altera 公司 MJ Lee, Yuan Li, Zhe Li 和 Altera 在 San Jose 和 Penang 的封装研发团队。作者还要感谢来自 Amkor 的 Jemmy Sutanto, DongHe Kang, Michael Oh, KwangSeok Oh, KyungRok Park, Sa Yun Ma, Rober Lanzone, Dave Hiner, Ron Huemoeller 和他们的先进封装技术开发团队, 以及 Amkor 韩国对此项目的贡献。CIC

作者简介

谢苑林, 中国科学院物理学博士, 加州伯克利分校和 Lawrence 伯克利实验室博士后, 目前是拓朗半导体封装技术研发部总监, 公司位于美国加州圣荷西。jxie@altera.com。

Deborah S. Patterson, 产品和技术市场部高级总监, 就职于安靠技术公司, 公司位于美国利桑那。Deborah.Patterson@amkor.com。