

What exactly is hybrid bonding? – February 25, 2024

半导体芯片材料工艺和先进封装

[向欣电子](#) · 2024-02-21 08:09 · 391次阅读

混合键合到底是什么？

[IcbanK半导体行业观察](#)

最新更新时间：2024-02-25

阅读数：4545

????如果您希望可以时常见面，欢迎标星????收藏哦~

来源：内容由半导体行业观察（ID：icbank）编译自semianalysis，谢谢。

在之前，有人曾表示，混合键合将成为自 EUV 以来半导体制造最具变革性的创新。事实上，它将对设计流程产生比 EUV 本身更大的影响，从封装架构延伸到单元设计和布局。知识产权生态系统将发生巨大重塑，制造流程也将发生巨大重塑。2D 晶体管缩小的时代仍将继续，但步伐缓慢，但混合键合将带来芯片设计者思考 3D 的新时代。

但semianalysis却表示，随着这句充满炒作的谣言结束，我们应该注意到，将混合键合技术大规模推向市场存在许多重大的工程和技术挑战，因为如今它仅保留给少数 AMD 芯片、CMOS 图像传感器和一些供应商的 3D 芯片和NAND。这种转变将重塑供应变化和设计流程。

半导体行业观察之前的文章《混合键合，成为“芯”宠》中，就介绍了这个技术。在这里，我们编译了semianalysis的科普文章，带大家进一步了解这个技术。

在封装史上，最后一次重大范式转变是从引线键合到倒装芯片。从那时起，更先进的封装形式（例如晶圆级扇出和 TCB）一直是相同核心原理的渐进式改进。这些封装方法都使用某种带焊料的凸块作为硅与封装或板之间的互连。这些技术可以一直缩小到约 20 微米的间距。

到目前为止，我们在多部分先进封装系列中讨论的主要封装类型和工艺流程已达到 220 微米到 100 微米规模，并且主要使用焊料作为各种小芯片铜互连之间的介质。为了进一步扩展，需要另一种范式转变：采用混合键合的无凸块互连。混合键合的尺寸超出了 10 微米互连间距，路线图为 100 纳米范围，并且它不使用任何中介物（intermediary），例如具有更高电阻的焊料。

相反，不同芯片或晶圆的互连直接通过铜通孔连接。直接铜连接可以降低电阻，从而在向各种芯片发送数据时降低功耗。当与连接数量的数量级增加相结合时，需要对设计进行彻底的重新思考。

在介绍混合键合之前，我们再看一下先进封装的全部意义是什么？我们可以看到，封装技术的进步旨在实现更大的互连密度（每个区域有更多的互连），减少迹线长度（trace length）以降低每比特传输的延迟和能量。我们可以看到混合键合如何解决这两个问题：迹线长度显著缩短，因此延迟尽可能低，无需on-die，并且在某些情况下比芯片上的global rounding更短，并且互连间距可以远低于10微米以增加密度。

混合键合到底是什么？

混合键合用于芯片的垂直（或3D）堆叠。混合键合的显著特点是它是无凸块的。它从基于焊料的凸块技术转向直接铜对铜连接。这意味着顶部die和底部die彼此齐平。两个芯片都没有凸块，而是只有可缩放至超细间距的铜焊盘。没有焊料，因此避免了与焊料相关的问题。

从上图中，我们可以看到AMD 3D V-Cache的横截面，它采用台积电的SoIC-X的die-to-wafer混合键合。顶部和底部硅之间的键合界面是混合键合层，存在于硅芯片（silicon dies）的金属层的顶部。混合键合层是一种电介质（现在最常见的是SiO或SiCN），采用通常为亚10微米间距的铜焊盘和通孔进行图案化。

电介质的作用是使每个焊盘绝缘，使得焊盘之间不存在信号干扰。铜焊盘通过硅通孔（TSV）连接到芯片金属层。TSV需要向堆栈中的其他芯片传输电源和信号。当底部芯片“面朝下”（face down）放置时，需要这些通孔来连接顶部芯片上的金属层，穿过晶体管层到达底部芯片上的金属层。

信号正是通过这些铜焊盘进行芯片间通信。这种键合之所以是“混合”键合，是因为它是电介质-电介质键合（dielectric-dielectric bond）和直接铜对铜键合（direct copper-to-copper bond）的组合。键合界面之间没有使用额外的粘合剂或材料。

关键工艺条件

与以前的基于凸块的互连相比，引入了一系列全新的技术和工艺挑战。为了实现高质量的键合，对表面光滑度、清洁度和粘合对准精度有非常严格的要求。我们将首先描述其中一些挑战，因为流程是围绕缓解这些挑战而设计的。记住这些将帮助您更好地理解为什么流程是这样的，以及不同方法的优缺点。

颗粒和清洁度

在任何有关混合键合的讨论中，都会提到颗粒（Particles）。这是因为颗粒是混合键合中良率的敌人。由于混合键合涉及将两个非常光滑且平坦的表面齐平地键合在一起，因此键合界面对任何颗粒的存在都非常敏感。

高度仅为 1 微米的颗粒会导致直径为 10 毫米的粘合空隙，从而导致键合缺陷。对于基于凸块的互连，器件和基板之间始终存在间隙，因为使用了底部填充或非导电薄膜，因此可以容纳一些颗粒。

保持清洁至关重要，而且非常具有挑战性。颗粒来自晶圆切割、研磨和抛光等许多步骤。任何类型的摩擦都会产生颗粒，这是一个问题，特别是因为混合键合涉及机械拾取芯片并将其放置在其他芯片的顶部。工具中存在大量来自芯片键合头和芯片翻转器的运动。颗粒是不可避免的，但有几种技术可以减轻对良率的影响。

当然，定期进行晶圆清洗以去除污染物。然而，清洁是不完美的，并且不能一次性去除 100% 的污染物，因此最好首先避免污染物。混合键合所需的洁净室比其他形式的先进封装所需的洁净室要先进得多。

因此，混合键合一般需要 1 级/ISO 3 级或更好的洁净室和设备。例如，台积电和英特尔正在一路迈向 ISO 2 或 ISO 1 级别。这是混合键合被视为“前端”工艺的一个主要原因，即它发生在类似于晶圆厂的环境中，而不是传统封装厂商（OSAT）的环境中。鉴于清洁度要求的升级，OSAT 很难追求混合键合。如果大多数 OSAT 想要参与混合键合，则需要建造更新、更先进的洁净室，而台积电和英特尔等公司可以使用较旧的晶圆厂或按照与现有晶圆厂类似的标准进行建设。

混合键合的工艺流程还涉及许多传统上仅由晶圆厂专用的工具。ASE 和 Amkor 等外包组装和测试公司（OSAT）在化学气相沉积（CVD）、蚀刻、物理气相沉积（PVD）、电化学沉积（ECD）、化学机械平坦化（CMP）和表面处理方面经验相对较少准备/激活。

清洁度要求和工具增加相结合导致成本大幅增加。与其他形式的封装相比，混合粘合工艺并不便宜。我们将在下面介绍整个流程。

光滑度

混合键合层的表面光滑度也极其关键。HB 界面同样对任何类型的形貌都敏感，这会产生空洞和无效的键合。一般认为电介质的表面粗糙度阈值是 0.5nm，铜焊盘的表面粗糙度阈值是 1nm。为了达到这种平滑度，需要执行化学机械平坦化（CMP），这对于混合键合来说是非常关键的工艺。

抛光后，需要在整个流动过程中始终保持这种光滑度。避免任何可能损坏该表面的步骤，例如严厉的清洁。即使是用于晶圆分类的探测也需要进行调整，以免表面受到损坏。

晶圆到晶圆 (W2W) 或芯片到晶圆(D2W)

首先讨论W2W (Wafer-to-Wafer) 还是D2W (Die-to-Wafer)。混合键合可以通过晶圆到晶圆 (W2W) 或芯片到晶圆 (D2W) 工艺来完成。W2W 意味着两个制造好的晶圆直接键合在一起。W2W 提供更高的对准精度、吞吐量 and 键合良率。鉴于 W2W 相对容易，目前绝大多数混合绑定都是通过 W2W 完成的。

W2W键合良率更高的原因在于对准和键合步骤是分开的。在 W2W 工具中，有一个单独的腔室 (chamber) 来执行对齐。一旦顶部和底部晶圆对齐，它们就会被移入键合室 (bonding chamber, 处于真空中)，在那里用一点力将它们压在一起，大约 20 分钟后，形成初始预键合。

W2W 的关键在于它是一个更加干净的过程，步骤更少。在对准和键合之前，可以清洁晶圆以去除大部分颗粒。芯片分割是颗粒污染的一个来源，仅发生在键合之后。由于它是晶圆级工艺，因此还可以为对准步骤提供更多的时间，因此较长的对准时间不会像芯片级工艺那样损害良率。

腔室内也没有发生太多运动，因此腔室内本身的污染物较少。目前，W2W键合机可以实现 50nm以下的对准精度。W2W键合已经是一个成熟的工艺并不是特别昂贵。证据是我们看到它广泛应用于大众市场产品，如 3 层图像传感器和 NAND。

W2W 键合效果很好，但一个主要限制是无法执行晶圆分类来选择已知良好的芯片 (KGD)。这会导致将有缺陷的芯片粘合到良好的芯片上，从而导致良好硅的浪费。

鉴于此，W2W 用于良率高的晶圆，这通常意味着较小的设计。在下图中我们可以看到W2W和D2W的芯片面积和成本之间的关系。在较小的芯片尺寸上，W2W 更便宜，因为晶圆产量会更高。然而，当我们采用更大的芯片尺寸时，W2W 成本曲线变得更加陡峭，这主要是由丢失的好芯片的成本驱动的。随着芯片尺寸的增大，每个晶圆上的良好芯片部分会减少，从而导致有缺陷的芯片和良好芯片的接合更多。

我们可以看到，W2W 用于具有高产量的较小芯片：CMOS 图像传感器、3D NAND，并且在逻辑方面到目前为止仅用于Graphcore 的Bow IPU。

虽然 Graphcore Bow IPU 是一款较大的 HPC 芯片，但顶部芯片不是前沿逻辑，而是用于电力传输的无源电容器芯片，因此其良率应该相当高，并且硅片要便宜得多。W2W 的另一个缺点是顶部芯片和底部芯片的尺寸必须一致，因此这限制了异构集成选项的灵活性。

关于成本有多种杠杆。主要是晶圆成本、D0（缺陷密度）和建和良率。这些杠杆中的每一个都可能导致更高或更低的成本。请注意，这些是强调这一点的示例图。不要使用下面的图表，因为它没有显示实际的键合成本。

可以看出，D2W 在小芯片上更昂贵，但在大芯片上，情况就相反了。W2W 更贵。仅测试和键合已知良好芯片 (KGD) 的能力至关重要，这也是晶圆芯片 (D2W) 率先实现产品化的原因，而不是冒着缺陷堆积和浪费优质硅的风险。它可以应对较差的良率，但仍然具有商业上可行的产品。

为了克服限制，我们需要使用 D2W。D2W 键合更具挑战性。执行晶圆分类后，KGD 从顶部晶圆上分离出来，并通过拾放工具单独附着到基础晶圆上。这在键合方面更具挑战性，因为每个晶圆有更多的键合步骤。这些额外的步骤会引入更多的颗粒污染，尤其是来自芯片分割和拾放期间键合头的移动的颗粒污染。

D2W 可以是一个“collective”工艺，其中 KGD 对齐并首先临时粘合到重构的载体晶圆上。然后将重构的载体晶圆键合到基础晶圆上以进行实际的预键合。这是为了像 W2W 一样将对准和粘合分开，并允许在最终预粘合、键合之前进行清洁步骤，以清除已积累的任何污染物。缺点是涉及额外的步骤，并且额外的 W2W 键合步骤会产生更多的对准错误机会。

这实际上是一个简单的流程，因为底部芯片也可以在载体晶圆上重构。因此，顶部和底部芯片都是从原始硅晶圆上切割而成，并且对 KGD 进行分类。两组芯片都粘合到各自载体上的精确位置上。然后，通过 W2W 工艺粘合 2 个承载晶圆。这是在台积电 SOIC 中完成的。因此，每个 AMD 3D V 缓存芯片（底部 CPU 芯片到载体、3D V 缓存小芯片到载体、2x 虚拟硅到载体）和晶圆上晶圆都使用 5 个键合步骤。

重构的流程还可以用于异构集成的更极端的选择。英特尔在 IEDM 2022 上展示了“准单片芯片 (QMC : quasi-monolithic chips)”应用的一个例子是一个封装，顶部和底部各有 2 个芯片的异构集成。对于顶部和底部，每个芯片都连接到载体晶圆上。然后用厚的无机氧化物（例如 SiO₂）对晶圆进行模制。W2W 键合已完成。然后模制的芯片被分割并附着到封装基板上以完成流程。

请注意，重构区域中可能存在 TSV。

直接 D2W（Direct D2W）键合是将单个芯片直接放置到目标晶圆上进行预接合。Direct D2W 还不太成熟，但由于流程简化，未来似乎会更多地使用直接 D2W。集体 D2W（

collective D2W) 的好处之一是可以执行清洁，然后直接发送到对准室以减少污染。最近推出的 D2W 集群工具可以重新创建此流程，从而减少此集体流程的好处。此外，D2W 更适合更精细的焊盘间距，因为对准变得更具挑战性，因此消除 W2W 步骤是有好处的，这会在 W2W 步骤中引入额外的未对准风险。

鉴于 D2W 混合键合的工艺挑战和成本，当前的应用受到限制。AMD 是 2022 年的第一个采用者，并且迄今为止仍然是唯一的采用者。

需要注意的一件事是，W2W 在对准方面远远领先于 D2W，因此，如果您的设计不是异构的，并且晶圆良率足够高，那么它实际上将是一种更准确、更高良率的工艺。这种更精细的间距还将解锁许多 D2W 尚未进入的新用例。

混合键合工艺流程

接下来让我们更详细地了解 D2W 和 W2W 的流程。

1.

TSV形成

正如我们上面提到的，需要 TSV 来向封装中的所有芯片传输电源和信号。想象一下传统的倒装芯片封装。该芯片仅需要一侧互连即可接收电力并与封装基板进行数据通信。该互连层具有连接到无源布线层（也称为“金属层”或“线路后端”/BEOL）的凸块，无源布线层为开关和处理数据的晶体管层提供电源和信号。

对于 3DIC，底部的芯片需要能够与其下方的封装基板以及顶部的芯片进行通信，因此芯片的两侧都需要互连。这就是 TSV 的用武之地。TSV 有多种变体，具体取决于它们在流程中的制造时间。TSV 可以是“先通孔”（即先在晶体管层之前在硅中制造），也可以是“中间通孔”（即在晶体管层完成之后且在金属层之前制造），或者是 BEOL 之后繁荣“后通孔”（即先在晶体管层之前制造）。

3DIC 最常见的是“中间通孔”方法，因为 TSV 在金属层之间运行，一直延伸到晶体管层，并在芯片背面显露出来，因此芯片的两侧现在都有一层互连我们将对此进行描述。

晶圆上涂有光刻胶，然后使用光刻法进行图案化。然后，使用深反应离子蚀刻 (DRIE : Deep Reactive Ion Etch) 将 TSV 蚀刻到硅中，以在晶圆深处形成高深宽比沟槽，但这不会穿过整个晶圆。使用化学气相沉积 (CVD) 沉积绝缘层 (SiO₂、SiN_x) 和阻挡层 (Ti 或 Ta)。这些层是为了防止铜扩散到硅中。然后，使用物理气相沉积 (PVD) 沉积铜种子层 (copper seed layer)。该种子层沉积在沟槽中，然后使用电化学沉积 (ECD) 进行填充。这形成了 TSV。然而，该过程尚未完成，因为背面尚未显露通孔。为了显露 TSV，需要对 TSV 背面进行抛光，

并且在某些情况下进行蚀刻，以减薄背面并随后显露出 TSV。一旦完成，晶圆就可以继续形成 BEOL。

TSV 的形成并不简单，而且可能非常耗时，特别是由于需要深蚀刻。我们知道 TSV 形成是 HBM 和 CoWoS 生产的瓶颈。一些客户从硅中介层转向 CoWoS-R 的原因之一是为了避免硅中介层中昂贵的 TSV 工艺。

2.

混合键层形成

在晶圆的键合界面之后，混合键合层被制造在晶圆的 BEOL 顶部。无论 W2W 还是 D2W 都是一样的。这是一层带有细间距铜通孔图案的介电薄膜。电介质，通常是氮化硅 (SiCN)，是通过 PECVD 沉积的。然后形成焊盘。使用光刻技术对铜焊盘的孔进行图案化并蚀刻掉。沉积阻挡层和种子层，然后使用典型的铜镶嵌工艺镀铜。

然后，进行 CMP 步骤来研磨并平滑电介质表面，并获得正确的铜轮廓。铜焊盘的一个显著特征是它们凹入约 1 微米间距。如前所述，光滑的表面对于形成良好的粘合至关重要。电介质的粗糙度必须控制在 0.5nm 以内，铜焊盘的粗糙度必须控制在 1nm 以内。

HB 接口的一个特征是铜焊盘最初凹入介电层下方约 5 纳米。这是为了确保铜在退火过程中不会妨碍初始电介质-电介质键合。如果铜凹陷得太深，则可能无法正确形成铜-铜键合。

在对铜和其他金属进行 CMP 时，由于过度抛光以及金属和电介质的柔软度不同，经常会出现凹陷现象。虽然并不理想，但这种现象并不严重，并且可以解决。需要控制凹陷的精确轮廓，以防止键合过程中铜过度/生长不足。

为了获得正确的凹陷轮廓，需要结合低和高铜去除浆料的多个 CMP 步骤。CMP 是混合键合的关键工艺，可实现非常光滑的表面和最佳轮廓。

在 ECTC 上，索尼展示了当间距降至 1 微米时，最好让铜突出而不是凹陷。

3.

晶圆分类/分割

仅对于 D2W，执行晶圆分类，并对 KGD 进行分割并在载体晶圆或带框架上重组，以便可以对其进行进一步处理。如上所述，HB 给传统晶圆分类工艺带来了新的复杂性。晶圆分类涉及用探针探测晶圆凸块或焊盘以执行电气测试。

探测可能会对铜焊盘表面造成少量损坏，从而破坏 CMP 过程中表面的光滑度。虽然对焊盘的损坏很小并且在大多数情况下通常可以接受，但 HB 对少量的形貌变化更加敏感，因为这些变

化会影响粘合质量。解决此问题的一种方法是在初始 CMP 中对此进行补偿，然后执行另一轮 CMP 后探测，以消除探测造成的任何损坏。

对于分割/切割，一个问题是过程中产生的颗粒。一般不使用刀片划片，因为它最脏：导致大量颗粒和大量良率损失。激光切割和等离子切割优于刀片切割，因为它们是更清洁的工艺，但仍然会产生颗粒物质。等离子切割是最极端的方法，其机制与蚀刻掉分隔芯片的划线类似。然而，考虑到蚀刻整个晶圆所需的时间，这个吞吐量要低得多。迪斯科是这方面的领导者。

一种缓解技术是首先在晶圆上涂上一层保护层。颗粒落在保护层上，当保护层被剥离时，颗粒可以与保护层一起被去除。虽然这有助于解决分割过程中的颗粒问题，但可能会留下保护层的残留物，并且剥离过程也可能对 HB 层造成一些表面损坏，从而增加表面粗糙度。

4.

等离子激活和清洁

现在对 2 个晶圆进行处理，为粘合做好准备。它们经过 N₂ 等离子体处理以激活表面。等离子处理改变了表面的特性，以增加表面能并使其更加亲水。使两个表面更加亲水可以使表面促进氢键结合。这有助于在室温下实现下一步中发生的初始弱电介质-电介质预键合。

处理后，进行最终清洁以清除任何积累的颗粒。重要的是，在键合之前，传入的晶圆尽可能干净。清洁需要彻底但又不能造成损坏，以保持 HB 接口的完整性。

最好的方法似乎是在兆声波辅助下进行去离子水清洁。使用洗涤剂或基于等离子的清洁可能会造成太大的破坏和/或引入污染物。

5.

键合 Bonding

现在是键合步骤。更准确地说，它更像是“预键合”，因为此步骤仅形成初始电介质-电介质键合，只是弱范德华键。我们将分别介绍 W2W 和 D2W 方法的流程。

(1) W2W 键合 W2W bonding

W2W 键合良率更高的原因在于对准和键合步骤是分开的。首先是对齐步骤。W2W 对齐有多种技术。过去，红外扫描仪用于检查两块晶圆之间的对准情况。一个限制是一个晶圆必须对红外线透明。这不适用于 CMOS 晶圆，因为红外线无法透过金属层。

EVG 在 W2W 接合领域占据主导地位，拥有获得专利的 SmartView 对准技术。有 2 个相互校准的相机，一台放置在目标晶圆上方，一台放置在下方。移动固定顶部晶圆的卡盘，以便底部摄像头可以识别对准标记，并且系统记录对准标记的位置。顶部晶圆被缩回，然后底部晶圆

在相机之间移动，直到顶部相机能够识别对准标记。对准器现在可以通过计算 2 个对准标记的相对位置来对准 2 个晶圆。为了帮助保持精度和控制，晶圆彼此非常接近（50 微米以内），并且卡盘仅在 X 和 Y 平面上移动，在预键合之前没有 Z 轴（垂直）移动。

对准后，晶圆被移入键合室，在其中施加小压力约 20 分钟将它们压在一起，形成初始键合。

键合后检查可以通过声学在原位完成，如果对准不充分，则也可以重新加工键合。

在 W2W 工具中，有一个单独的室来执行对齐。一旦顶部和底部晶圆对齐，它们就会被移入键合室（处于真空中），在那里用一点力将它们压在一起，大约 20 分钟后形成初始预键合。W2W 的关键在于它是一个更加干净的过程，步骤更少。在对准和键合之前，可以清洁晶圆以去除大部分颗粒。芯片分割是颗粒污染的一个来源，仅发生在键合之后。

由于它是晶圆级工艺，因此还可以为对准步骤提供更多的时间，因此较长的对准时间不会像芯片级工艺那样损害产量。腔室中也没有发生太多运动，因此腔室本身产生的污染物较少。目前，W2W 键合机可以实现 50nm 以下的对准精度。W2W 键合已经是一个成熟的工艺并且不是特别昂贵。证据是我们看到它广泛应用于大众市场产品，如索尼、Omnivision 和三星的图像传感器，以及长江存储、西部数据和 Kioxia 的 NAND。

(2) D2W 键合 D2W bonding

D2W 接合是通过拾放工具完成的。

底部目标晶圆位于晶圆卡盘上。将待粘合的芯片面朝上放置在胶带框架上。翻转臂收集单个芯片并将其翻转，使芯片的背面朝上位于翻转器上。有一个高架键合臂，可利用键合头上的真空吸力拾取翻转的芯片。

<http://news.eeworld.com.cn/mp/lcbank/a370343.aspx>