

# 关于 3D 堆叠 MRAM 热学分析方法的研究

永若雪, 姜岩峰

(江南大学物联网工程学院, 江苏无锡 214122)

**摘要:** 本文针对 3D 堆叠磁随机存储器 (Magnetic Random Access Memory, MRAM) 的热学分析问题, 在有限元法和热阻网络法的基础上, 提出了一种局部等效法, 可高精度并且快速地分析 3D 堆叠 MRAM 的热学分布. 与有限元法相比, 该方法使用直观方便, 克服了有限元法建模与求解复杂耗时的问题; 与热阻网络法相比, 局部等效法具有保持较高精度的特点, 解决了热阻网络法针对带夹层和硅通孔 (Through Silicon Via, TSV) 的复杂封装问题时存在较大误差的问题. 对比结果表明, 使用本文提出的方法得出的各叠层的上表面温度误差均小于  $0.05\text{ }^{\circ}\text{C}$ , 精度与有限元法一致, 并且更便捷高效. 同时对应的建模结构简单, 避免了热阻网络法将含铜柱的夹层和含铜柱的硅层分开考虑的不准确性. 本文的研究可为未来多层 3D 堆叠 MRAM 热学特性相关的设计与分析提供指导.

**关键词:** 磁随机存储器; 3D 堆叠; 热分析; 有限元法; 热阻网络法

**基金项目:** 国家自然科学基金 (No.61774078)

**中图分类号:** TN4

**文献标识码:** A

**文章编号:** 0372-2112(2023)10-2775-08

**电子学报 URL:** <http://www.ejournal.org.cn>

**DOI:** 10.12263/DZXB.20220275

## Research on Thermal Analysis Method of 3D-stacked MRAM

YONG Ruo-xue, JIANG Yan-feng

(School of Internet of Things Engineering, Jiangnan University, Wuxi, Jiangsu 214122, China)

**Abstract:** In this paper, a local equivalent method is proposed for thermal analysis of 3D-stacked magnetic random access memory (MRAM). The method is based on combination of two general methods of thermal analysis, including the finite element method and the thermal resistance network method. On one hand, the method can be easily applied on the thermal analysis for 3D MRAM, overcoming the complex and time-consuming problem of modeling and solving when using the finite element method. On the other hand, the method shows high accuracy on the complex package 3D MRAM with multiple interlayers and TSVs, solving the problem of inaccuracy of the thermal resistance network method. Comparing the results between the local equivalent method and the finite element method, it can be found that the temperature error of the upper surface of each layer obtained by the proposed method is less than  $0.05\text{ }^{\circ}\text{C}$ , with the benefits of higher efficiency with less simulation time. Moreover, compared with the thermal resistance network method, the proposed method shows higher accuracy. The main error of the thermal resistance network method comes from the fact that the interlayer and the silicon layer related to the copper pillar have to be modeled separately. The local equivalent method can avoid the problem and improve the accuracy. The proposed method can provide guidance for the design and analysis related to the thermal properties of multi-layer 3D-stacked MRAM.

**Key words:** magnetic random access memory (MRAM); 3D-stacked; thermal analysis; finite element method; thermal resistance network method

**Foundation Item(s):** National Natural Science Foundation of China (No.61774078)

### 1 引言

技术节点缩小引起的存储器功率密度增加, 会导致晶片的平均温度显著升高. 高温可能导致较高的功耗并降低微电子系统的性能和可靠性<sup>[1]</sup>. 例如, 功耗引起的温度升高会导致亚阈值电流呈指数上升, 从而产

生较大的泄漏功率. 载流子的迁移率对温度也很敏感, 高温会降低载流子的迁移率, 从而导致速度变慢. 高温甚至会影响晶体管的正常开关特性, 导致存储器错误率增加并且可靠性下降. 因此, 有必要在设计流程的早期阶段对存储器进行热学分析, 以便工程师能够缩短

研究周期并且提高效率。

如今,3D堆叠技术将芯片空间从二维扩展到三维,不仅减小了芯片尺寸,增加了存储容量,还提高了性能,降低了成本。尤其是对于大容量的独立存储器,需要数千甚至数百万个存储单元并行工作来执行读写操作<sup>[1]</sup>。在3D堆叠技术中,多个晶片垂直堆叠,相邻的晶片通过微凸块直接互连。互连线的总长度显著缩短,因此信号传输速度得以提高。因此,随着应用需求的提升,3D堆叠存储器成为新兴的大容量独立存储器的主流。3D堆叠存储器的示意图如图1所示。散热器位于堆叠晶片的顶部用来进行散热,基板在底层,硅中介层通过微凸块与基板连接。在3D技术中,TSV与微凸块一起使用来实现堆叠的各个晶片的物理互连,为3D堆叠存储器中的数据处理提供高带宽<sup>[2,3]</sup>。然而,由于散热能力有限,3D堆叠技术可能会导致热问题。如图1所示,散热器位于存储器的顶部。而存储层的数量可以达到16,64甚至128层,这些层产生的热量最终需要传递到散热器并散发出去。因此,3D封装内部产生的热量很难散发到外部。由于散热器到堆叠晶片的散热距离随层数不同而不同,因此散热器对各层的冷却效果也不同,离散热器较远的层会具有较高的温度。因此,热学分析对3D堆叠存储器的设计非常重要。

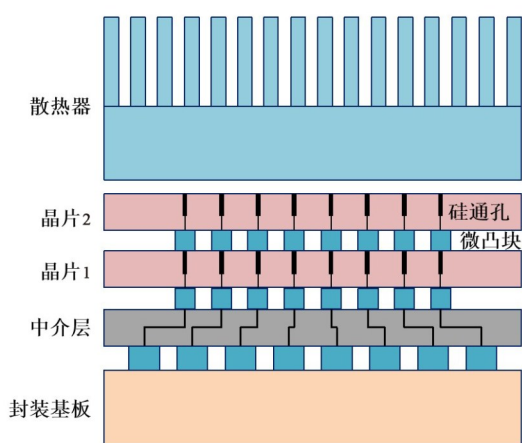


图1 3D堆叠存储器的示意图

目前,3D堆叠技术因对独立存储器的大容量需求而备受业界关注。例如,高带宽内存(High Bandwidth Memory, HBM)<sup>[4]</sup>和混合内存立方体(Hybrid Memory Cube, HMC)<sup>[2]</sup>这两类3D堆叠存储器,与传统的双数据速率同步(Double Data Rate, DDR)DRAM相比,具有更高的带宽,更适用于当今新兴的数据密集型的应用。如今,STT-MRAM是一种新兴的、很有前途的非挥发存储器(Non-Volatile Memory, NVM),被认为是下一代存储器中最有希望的候选之一。为了在保证可靠性的同时

实现更高的容量和密度,近年来科研工作者们开始在STT-MRAM上应用3D堆叠封装技术来证明其可行性。Zhu等人<sup>[5]</sup>提出的3D MRAM与基于SRAM的2D存储器相比,节省面积高达17.55%,性能提升34.74%,能耗降低13.90%。Perumkunnil等人<sup>[6]</sup>研究的晶圆对晶圆(Wafer-to-Wafer, W2W)集成3D STT-MRAM缓存可为目标片上系统(System on Chip, SoC)带来高达30%的性能提升,同时功耗降低17%,占用空间减少15%。

如今,已有几种热建模方法被提出。例如,Floworks<sup>[7]</sup>, ANSYS或Silvaco等有限元模拟器可以用于仿真封装中的热分布。使用这些商业软件,需要建立高精度的特定几何模型,包括设置所有材料的物理参数。尽管它们具有很高的准确性,但它们非常耗时。Chen等人<sup>[8]</sup>采用有限元分析方法来仿真有机3D垂直电阻随机存取存储器(Vertical Resistive Random Access Memory, VRRAM)。在封装级,Gong等人<sup>[9]</sup>建立了多芯片模块结合球栅阵列(Multi-Chips Module combined with Ball Grid Array, MCM-BGA)的3D封装模型,可用于将最高温度控制在令人满意的范围内。Huang等人<sup>[10]</sup>提出了一种非线性有限元方法(Nonlinear-Finite Element Method, N-FEM)来求解热力耦合场,并仿真了最先进的3D DRAM模型。除了单层热硅通孔(Thermal Through Silicon Via, TTSV)和TTSV阵列之外,人们还研究了多层TTSV的情况。一种紧凑热模型(Compact Thermal Model, CTM)也被提出。例如,Stan等人<sup>[11]</sup>提出了动态紧凑热模型HotSpot,使用热阻网络来构建模型,Huang等人<sup>[12]</sup>改善了其温度感知。Ayala等人<sup>[13]</sup>提出了一个完整的3D-CMP的热模型,其中纳米结构是通过使用热阻网络构建的。Wan等人<sup>[14]</sup>研究了3D堆叠集成电路的多核结构和微流控冷却的协同设计,采用紧凑热模型研究了双层结构(16核,×86多核层和L2 SRAM缓存层)在液体冷却下的热特性。Sahay等人<sup>[15]</sup>提出了3D NAND闪存静态特性的行为紧凑模型,可以模拟不同3D NAND闪存架构(具有不同的页面大小、尺寸或多个堆叠层)的电行为。

然而,现有的热学分析主要是针对处理器芯片和其他常见存储器的,很少有专门针对3D堆叠独立MRAM的热学分析。3D堆叠的独立MRAM与3D处理器芯片有很大的不同,不仅是功能和结构,还有功耗和热性能。MRAM存储器中,每个存储单元由1个MTJ器件和1个NMOS器件组成,其中NMOS器件制作在衬底有源区内,MTJ器件在金属层上。MTJ器件的尺寸远小于NMOS器件的尺寸,通孔用于外围电路的连接,用于连接MRAM阵列的字线、位线和源线等。与传统存储器相比,STT-MRAM具有更低的泄漏功率、更高的密度和更小的读写延迟。然而,为了使切换速度快,MRAM

中的自旋电子器件需要大的切换电流,从而导致在技术节点、电源电压等条件不变的情况下,其动态功耗高.之前提到的功率密度过高以及采用3D堆叠技术都有可能严重的导致热问题,从热学分析的角度来看,3D堆叠独立MRAM面临着比其他存储器更严峻的热学挑战.因此,有必要进行关于3D堆叠MRAM热学分析方法的研究,以满足工程师对存储器温度的要求.

### 2 有限元法

有限元法是一种数值分析技术,该方法把计算数学、弹性理论和计算机软件相结合.使用有限元法进行热学分析的基本原理是:首先,确定初始条件与边界条件;其次,将待分析的对象划分成有限个包含节点的子区域,这个过程被称为区域离散化;再次,根据热力学第一定律求解各个节点的热平衡方程;最后,计算得到各个节点的温度并进行进一步分析.

热量传递的方式主要有3种,分别是热传导、热对流和热辐射.热传导是指当物体内部存在温差,即存在温度梯度时,热量从物体的高温部分传递到低温部分.热对流是指固体表面和其周围接触的流体之间,由温差而引起的热量交换.热辐射是指物体发射电磁能,其他物体吸收该电磁能并将其转变为热的热交换过程.热传导和热对流是线性的,而热辐射是非线性的,有限元求解时需要考虑是否收敛.根据温度场的控制方程,瞬态情况下场变量 $T(x,y,z,t)$ 需满足以下微分方程<sup>[16]</sup>:

$$\frac{\partial}{\partial x} \left( k_x \frac{\partial T}{\partial x} \right) + \frac{\partial}{\partial y} \left( k_y \frac{\partial T}{\partial y} \right) + \frac{\partial}{\partial z} \left( k_z \frac{\partial T}{\partial z} \right) + q_{sh} = C \frac{\partial T}{\partial t} \quad (1)$$

其中,  $q_{sh}$  是单位时间、单位体积产生或者吸收的热量;  $C$  是比热容;  $\partial T/\partial t$  是单位时间的温升. 对于各向同性的材料  $k_x=k_y=k_z=k$ , 于是上式变为

$$k \left( \frac{\partial^2 T}{\partial x^2} + \frac{\partial^2 T}{\partial y^2} + \frac{\partial^2 T}{\partial z^2} \right) + q_{sh} = C \frac{\partial T}{\partial t} \quad (2)$$

在稳态情况下,  $\partial T/\partial t=0$ , 于是以上两式变为

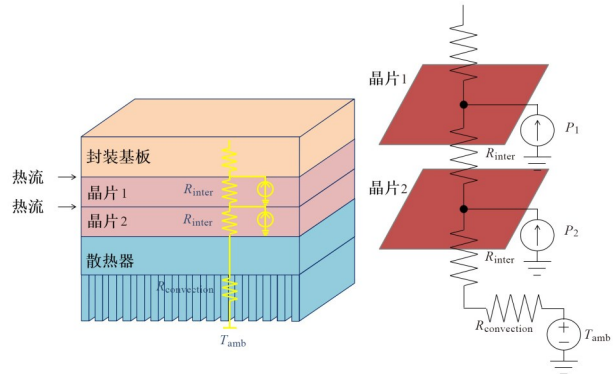
$$\frac{\partial}{\partial x} \left( k_x \frac{\partial T}{\partial x} \right) + \frac{\partial}{\partial y} \left( k_y \frac{\partial T}{\partial y} \right) + \frac{\partial}{\partial z} \left( k_z \frac{\partial T}{\partial z} \right) + q_{sh} = 0 \quad (3)$$

$$k \left( \frac{\partial^2 T}{\partial x^2} + \frac{\partial^2 T}{\partial y^2} + \frac{\partial^2 T}{\partial z^2} \right) + q_{sh} = 0 \quad (4)$$

### 3 热阻网络法

利用热和电现象之间的二元性<sup>[17]</sup>, 可以将热参数转换为等效的电参数, 从而构建热阻网络的等效电路. 一个两层3D堆叠MRAM的热模型如图2所示. 热参数和电参数之间的对应关系如表1所示, 其中温度类似于

电压, 热阻类似于电阻, 热容类似于电容, 功率类似于电流. 散热主要沿垂直方向.



(a) 两层3D MRAM封装的整体结构 (b) 对应的热阻网络模型

图2 一个两层3D堆叠MRAM的热阻网络模型

表1 热-电二元性

热参数及单位	电参数及单位
温度 $T/K$	电压 $V/V$
热阻 $R/(K \cdot W^{-1})$	电阻 $R/\Omega$
热容 $C/(J \cdot K^{-1})$	电容 $C/F$
功率 $P/W$	电流 $I/A$

在图2(a)中, 垂直相邻晶片之间的热阻为

$$R_{inter} = \frac{\Delta z}{k_{eff} A} \quad (5)$$

其中,  $\Delta z$  是晶片的厚度;  $k_{eff}$  是材料的热导率;  $A$  是晶片的横截面积. 如果考虑带通孔的情况, 那么晶片的热导率计算方式如下:

$$k_{eff} = \rho_{via} k_{via} + (1 - \rho_{via}) k_{layer} \quad (6)$$

其中, 通孔密度  $\rho_{via} = nA_{via}/(wh)$ ,  $n$  是区域内通孔的数量,  $A_{via}$  是每个通孔的截面积,  $w$  和  $h$  分别为该区域的宽度和高度;  $k_{via}$  是通孔的热导率;  $k_{layer}$  是没有任何通孔的区域的热导率. 对流热阻为

$$R_{convection} = \frac{1}{hA} \quad (7)$$

其中,  $h$  是传热系数,  $A$  是散热器的表面积.

一个两层3D堆叠MRAM的等效热模型如图2(b)所示, 红色层代表热流, 位于晶片的上表面. 对于共有  $n$  层的晶片, 将最靠近散热器的晶片的温度定义为  $T_n$ , 其计算方式为

$$T_n = T_{amb} + (R_{inter} + R_{convection}) \left( \sum_{k=1}^n P_k \right) \quad (8)$$

其中,  $T_{amb}$  是环境温度.

对于所有其他晶片定义了相应的序号  $i \in [1, 2, \dots, n-1]$ , 晶片1距离散热器最远. 这些晶片的温度可以计算为

$$T_i = T_n + R_{\text{inter}} \left( \sum_{j=1}^{n-1} \sum_{k=1}^j P_k \right) \quad (9)$$

其中,  $P_k$  是每一层的功耗.

当堆叠层数越多时, 散热问题就越严重, 特别是对于远离散热器的晶片. 通过式(9)可以从热学角度估算出可堆叠层数的最大值.

## 4 3D堆叠MRAM热分析

### 4.1 有限元法分析MRAM

第一, 需根据3D堆叠MRAM的一般结构在Solidworks中建立热学分析几何模型. 假设该MRAM为4层结构, 其横截面积为  $7.7563 \text{ mm} \times 7.7563 \text{ mm}$ , 通孔为  $130 \times 130$  个.  $R_1$  为晶片通过散热器到环境的热阻, 设为  $1 \text{ K/W}$ . 考虑到有限元法计算复杂, 取其  $1/676$ , 即  $298.32 \mu\text{m} \times 298.32 \mu\text{m}$  作为等效模型的横截面积, 通孔为  $5 \times 5$  个.  $R_1$  等效为  $R_m$ . 等效过程的示意图如图3所示. 之后提到的模型均为  $1/676$  模型. 为了采用有限元分析方法开展热学特性分析, 需要对STT-MRAM晶片进行抽象和建模. 如图4所示, STT-MRAM晶片是在CMOS工艺基础上, 在第三层金属层上面制作MTJ器件, 是典型的多层结构, 每层厚度从1纳米到几微米不等, 结构复杂, 不利于开展有限元分析. 本文从热等效原理出发, 建立了热等效模型, 将晶片等效为一个厚度  $100 \mu\text{m}$  的硅层, 与减薄后的硅片厚度相当, 并且等效功率密度是  $16 \text{ W/cm}^2 (\pm 21.7\%)$ . 这种简化可有效降低网格剖分的难度, 提高有限元分析的收敛速度. 在抽象出热等效模型的基础上, 考虑多层堆叠中的工艺具体要求, 同时为了保证多层堆叠后整体晶片温度分布的均匀性, 本文所设计的面向STT-MRAM的多层堆叠工艺, 重要的是需要设计层间的氧化层和聚合物. 考虑到温度分布的均匀性、TSV通孔的应力等影响, 仿真中提出的方案是在晶片上面叠加一层二氧化硅, 厚度为  $10 \mu\text{m}$ , 另外再淀积一层聚合物, 厚度为  $1 \mu\text{m}$ . 整体结构如图4所示, 本文基于该结构开展多层堆叠MRAM晶片的有限元分析.

第二, 将模型导入ANSYS Workbench 19.0中, 几何模型的实体图和透视图如图5所示, 图中隐藏了  $R_m$ . 之后定义各个组件的材料, 晶片材料为硅, 其热导率为  $150 \text{ W/(m}\cdot\text{K)}$ . 二氧化硅的热导率为  $1.4 \text{ W/(m}\cdot\text{K)}$ , 聚合物的热导率为  $0.3 \text{ W/(m}\cdot\text{K)}$ . 通孔材料为铜, 其热导率为  $400 \text{ W/(m}\cdot\text{K)}$ . 各个材料的尺寸和热导率如表2所示.

第三, 对模型进行有限元网格划分, FEM网格的立体图和俯视图如图6所示.

第四, 3D堆叠MRAM中的热量是由有源器件或互连中的自加热产生的, 相应的热源位于硅层顶部. 在考虑每层的功率密度时, 需要考虑由于MRAM存储器写

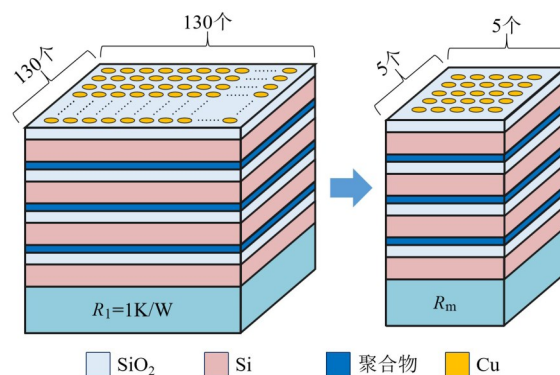


图3 3D堆叠MRAM等效过程的示意图

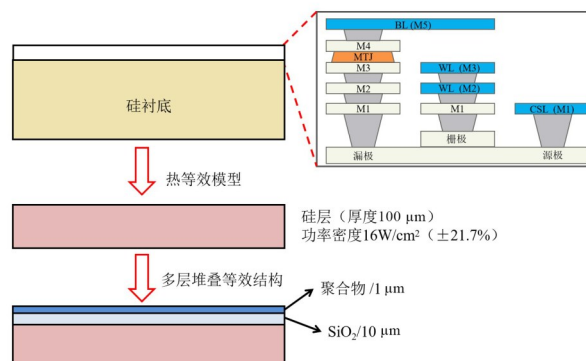


图4 单层MRAM晶片等效结构示意图

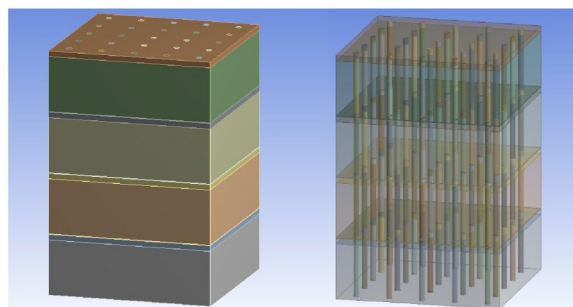


图5 四层3D堆叠MRAM几何模型的立体图和透视图

表2 各个材料的尺寸和热导率

材料名称	尺寸/mm	热导率/(W/(m·K))
硅	厚度 0.1	150
二氧化硅	厚度 0.01	1.4
聚合物	厚度 0.001	0.3
铜	直径 0.01 间距 0.05	400

入电流非对称性所带来的影响. 针对这种写入电流非对称带来的差异, 在有限元分析施加载荷时, 考虑到可能的芯片温度变化, 设计了3个芯片工作的温度值, 分别是最高温度、典型温度和最低温度, 其中最高温度是所有存储单元都进行写‘1’操作时所对应的温度, 而最低温度是所有存储单元都进行写‘0’操作时所对应的

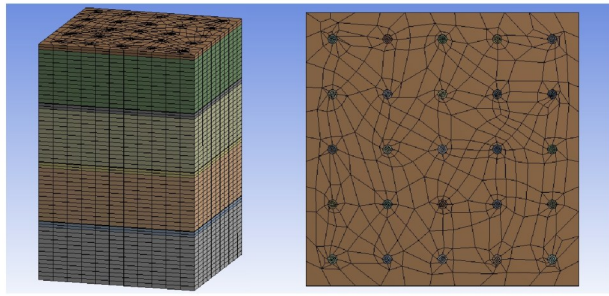


图6 四层3D堆叠MRAM有限元网格的立体图和俯视图

温度,典型温度则对应最高温度和最低温度的中间值.在有限元仿真中,通过改变施加载荷的大小,可观察到写入电流非对称带来的热学性能的影响.假设每层的功率密度为  $16 \text{ W/cm}^2 (\pm 21.7\%)^{[18]}$ ,这里的误差为  $\pm 21.7\%$ ,文献[18]中,基于RISC-V处理器仿真的2D-MRAM的功率密度,已包含最高温度值和最低温度值对功率密度的影响.在有限元仿真时,在4个晶片的表面上均施加功率为  $0.01474 \text{ W} (\pm 21.7\%)$  的载荷.散热器底面的温度设置为  $25^\circ\text{C}$ .绝热边界条件应用于其他表面.模型的剖面示意图和透视图如图7所示.

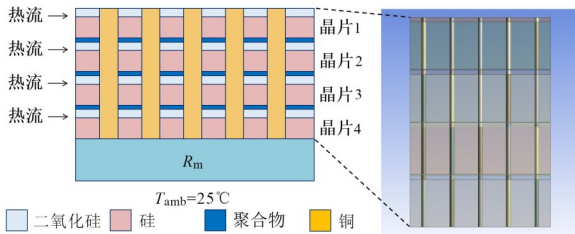


图7 四层3D堆叠MRAM的剖面示意图和透视图

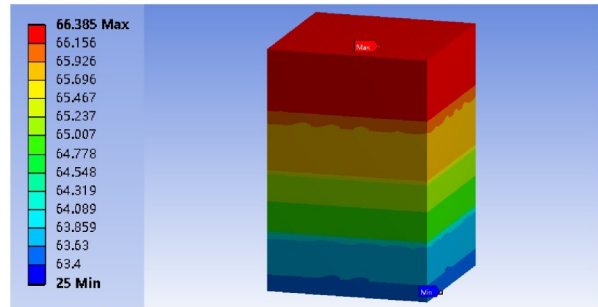
第五,仿真求解出各层晶片的上表面温度,如表3中含夹层有限元法一列所示.4层的3D堆叠MRAM有限元求解后的温度分布如图8(a)所示,距离散热器越远的晶片温度越高.距离散热器最近的晶片即晶片1的上表面温度分布如图8(b)所示,其平均温度为  $66.383^\circ\text{C}$ .沿其对角线的温度曲线如图8(c)所示.可以发现,5个铜柱中相邻两个铜柱中间温度较高,铜柱边沿温度较低,同一表面的最高温度和最低温度相差仅为  $0.006^\circ\text{C}$ .

### 4.2 局部等效法分析MRAM

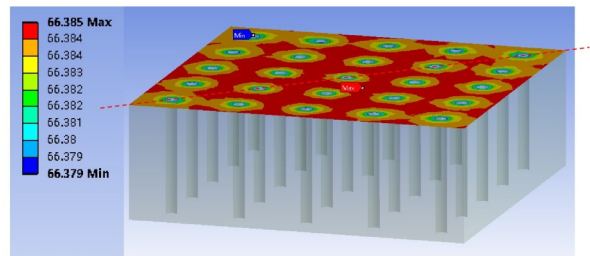
由于当考虑含有二氧化硅和聚合物的夹层时,各层晶片的上表面温度不能简单地使用热阻网络法进行计算,本文根据有限元法和热阻网络法结合提出了局部等效法.局部等效法将内含铜柱的单层晶片、聚合物层和二氧化硅层或者内含铜柱的单层晶片看作一个整体进行等效,其上下表面的温度差反映其等效热阻.

表3 不同情况下使用不同方法得出的各层晶片的上表面温度 单位:  $^\circ\text{C}$

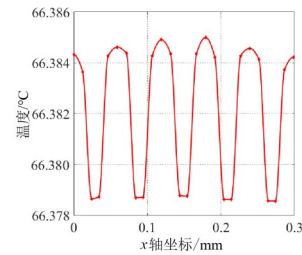
情况	含夹层有限元法	含夹层局部等效法	无夹层有限元法	无夹层热阻网络法
晶片1上表面	66.383	66.352	64.533	64.531
晶片2上表面	65.990	65.956	64.430	64.428
晶片3上表面	65.184	65.164	64.225	64.223
晶片4上表面	63.965	63.976	63.916	63.914



(a) 整体温度分布图



(b) 晶片1上表面温度分布图



(c) 沿晶片1上表面对角线的温度分布曲线

图8 四层3D堆叠MRAM有限元求解的结果

一个简单的例子如图9所示.假设对于内含铜柱的单层晶片,上表面施加均匀热流  $q$ ,下表面温度为  $T_b$ .得到上表面温度为  $T_a$ ,上下表面温度差为  $\Delta T = T_a - T_b$ ,那么该模型的等效热阻为

$$R_{eq,z} = \frac{\Delta z}{k_{eq,z} A} = \frac{\Delta z}{q \frac{\Delta z}{|\Delta T|} A} = \frac{|\Delta T|}{qA} \quad (10)$$

其中,  $\Delta z$  是模型厚度;  $k_{eq,z}$  是  $z$  方向上的等效热导率;  $A$  是截面积.所以,当  $q$  和  $A$  一定时,等效热阻和上下表面

温度差成正比。

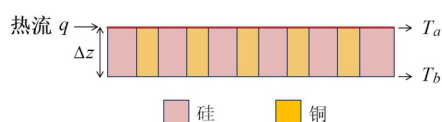


图9 内含铜柱的单层晶片剖面示意图

对于4层3D堆叠MRAM,只需先用有限元法对如图10所示的2层结构进行仿真,得到O面、P面和Q面的温度,图中隐藏了散热器的等效几何模型。同样,在2个晶片的上表面均施加功率为0.01474 W的载荷。散热器底面的温度设置为25°C,绝热边界条件应用于其他各个表面。2层的3D堆叠MRAM有限元求解后的温度分布如图11(a)所示。距离散热器最近的晶片即晶片1的上表面温度分布如图11(b)所示,其平均温度为35.140°C(O面),同一表面的最高温度和最低温度相差仅为0.006°C。并记录下P面和Q面的平均温度分别为34.744°C和34.626°C。

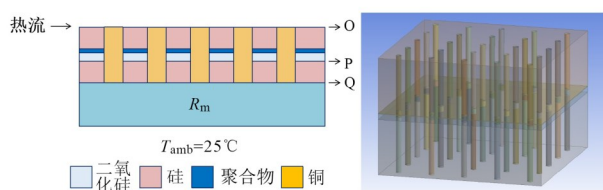
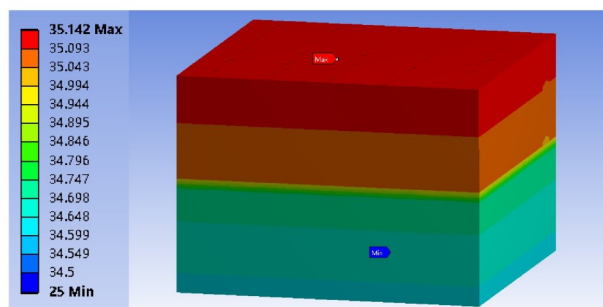
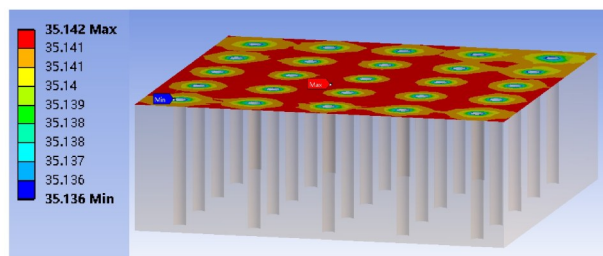


图10 两层3D堆叠MRAM的剖面示意图和透视图



(a) 整体温度分布图



(b) 晶片1上表面温度分布图

图11 两层3D堆叠MRAM有限元求解的结果

然后,根据O面与P面、P面与Q面、Q面与散热器底面之间的温度差,对等效后的模型采用热阻网络法,计算得出各层晶片的上表面温度,结果如表3所示。另外,采用有限元法和热阻网络法分析无夹层情况时的结果也列于表3中。结果表明,使用局部等效法可以有效解决含夹层的3D堆叠STT-MRAM的热学分析问题,与使用有限元法得出的结果相比,误差小于0.05°C。表中情况均含铜柱,最右侧两列用于对比无夹层情况。局部等效法不仅限于4层堆叠结构,若推广到n层( $n>1$ ),距离散热器最近的晶片即晶片1的上表面平均温度 $T_1$ 的计算方式如下:

$$T_1 = T_p + (n-1)(T_o - T_p) \quad (11)$$

其中, $T_o$ 、 $T_p$ 和 $T_q$ 分别为O面、P面和Q面的平均温度。

对于各层晶片上表面的热流不同的情况,假设共有n层晶片,最靠近散热器的晶片的上表面温度定义为 $T_n$ ,其计算方式为

$$T_n = T_{\text{amb}} + (R_{\text{eq,P-Q}} + R_m) \left( \sum_{k=1}^n P_k \right) \quad (12)$$

其中, $P_k$ 是每一层的功耗; $R_{\text{eq,P-Q}}$ 是用式(10)计算得出的P面与Q面之间的等效电阻; $R_m$ 是散热器的等效电阻。对于所有其他晶片定义了相应的序号 $i \in [1, 2, \dots, n-1]$ ,晶片1距离散热器最远。这些晶片的温度可以计算为

$$T_i = T_n + R_{\text{eq,O-P}} \left( \sum_{j=i}^{n-1} \sum_{k=1}^j P_k \right) \quad (13)$$

其中, $R_{\text{eq,O-P}}$ 是用式(10)计算得出的O面与P面之间的等效电阻。因此,当堆叠层数越多时,散热问题就越严重,特别是对于远离散热器的晶片。

本文提出的多层堆叠实现STT-MRAM大容量存储器的方法,之前尚未有相关报道,也未见相关实验数据。

由于有限元分析法是被广泛认为比较准确的热学分析方法,因此本文将提出的局部等效法与该方法进行误差比较。比较结果表明,在可接受的温度误差范围内( $<0.1^\circ\text{C}$ ),四层局部等效法与有限元分析法的结果几乎相同,并且四层局部等效法具有简单、直观、仿真时间短等特点,使用非常方便,非常适用于大面积芯片多层堆叠情况下的热学特性分析。

局部等效法由于具有上述优点,能够推广到多层使用。但是在采用有限元分析方法时,由于有限元分析法建模和求解过程极为耗时和复杂,未能获得有限元分析方法的对比数据,所以本文未列出多层使用的对比结果。这也从另一个角度证明,本文提出的局部等效法在复杂多层堆叠热学分析中具有一定优势,在相关工程问题分析中具有潜在的应用前景。

## 5 结论

本文介绍了有限元法和热阻网络法两种基本热学分析方法,并采用这两种方法分析了考虑TSV但省略夹层的3D堆叠独立MRAM的温度特性.针对带夹层和TSV的3D堆叠独立MRAM,本文将这两种方法结合提出了局部等效法并进行仿真与计算.该方法具有与有限元法相当的准确性,并且可以解决热阻网络法不能计算同时带TSV和夹层的问题.结果显示,局部等效法和有限元法相比,误差小于 $0.05^{\circ}\text{C}$ ,而且更加省时、灵活,尤其是对于层数较多的3D堆叠MRAM.因此,本文研究的热学分析方法,尤其是提出的局部等效法,对未来多层3D堆叠独立MRAM热学特性的相关设计与分析具有重要意义.

对于MRAM存储器来说,堆叠多层工艺中通过制备TSV通孔及微凸点进行多层MRAM芯片之间的互连,虽然提高了单位面积的存储密度,但该互连方式无疑会带来一定的寄生参数,尤其当堆叠层数较多时,由于寄生参数的影响,不同层之间会出现明显的时钟信号偏差,这种偏差的存在对可堆叠层数同样具有限制性.

本文提出了MRAM存储器芯片的堆叠多层技术,在具体实施中,需要考虑和优化堆叠多层的工艺,包括TSV工艺、微凸点接触电阻、硅片键合技术等,尽量减小芯片堆叠带来的寄生电阻和寄生电容等寄生参数;同时,还要开展针对堆叠多层结构的MRAM存储器的设计,除了直接将多个同样的存储器芯片进行堆叠集成之外,还可以考虑设计控制信号主-从式结构的3D-MRAM存储器,将多层MRAM存储单元的控制信号放在同一个芯片内,减少多层堆叠工艺带来的寄生参数的影响,达到改善时钟信号偏差的作用.

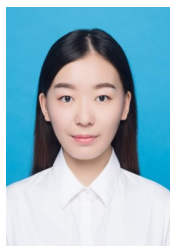
另外,堆叠多层工艺不可避免的复杂性,将会使产品的良率受到严重影响,使产品的可靠性下降、成本上升等.这些问题成为阻碍多层堆叠MRAM存储器发展的主要制约因素,也是目前堆叠多层工艺普遍具有的问题.随着相关工艺技术的逐步完善以及堆叠多层MRAM存储器设计方案的不断进步,预计这些问题会逐步得到克服,而MRAM存储器在通过多层堆叠的方法取得高存储密度和大存储容量后,会由于其优异的存储性能,逐步成为一种主流存储器.

## 参考文献

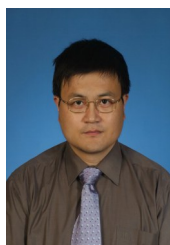
- [1] PRINCE B. 3D stacking of RAM-processor chips using TSV[M]//Vertical 3D Memory Technologies. Chichester: John Wiley & Sons Ltd., 2014: 275-344.
- [2] AYYAGIRI M. Hybrid memory cube: A revolutionary 3-D memory technology[J]. Electronic Engineering Times, 2012, 1625: 37.
- [3] AGA S, NARAYANASAMY S. InvisiMem: Smart memory defenses for memory bus side channel[C]//2017 ACM/IEEE 44th Annual International Symposium on Computer Architecture (ISCA). Piscataway: IEEE, 2017: 94-106.
- [4] LEE D U, KIM K W, KIM K W, et al. A 1.2 V8 gb 8-channel 128 GB/s high-bandwidth memory (HBM) stacked DRAM with effective I/O test circuits[J]. IEEE Journal of Solid-State Circuits, 2015, 50(1): 191-203.
- [5] ZHU L J, BAMBERG L, AGNESINA A, et al. Heterogeneous 3D integration for a RISC-V system with STT-MRAM[J]. IEEE Computer Architecture Letters, 2020, 19(1): 51-54.
- [6] PERUMKUNNIL M, YASIN F, RAO S, et al. System exploration and technology demonstration of 3D Wafer-to-Wafer integrated STT-MRAM based caches for advanced Mobile SoCs[C]//2020 IEEE International Electron Devices Meeting (IEDM). Piscataway: IEEE, 2021: 15.4.1-15.4.4.
- [7] Floworks. Fluid flow analysis for SolidWorks[EB/OL]. [2022]. <http://www.floworks.com>.
- [8] CHEN Q Y, WANG Z W, LIN M, et al. Homogeneous 3D vertical integration of parylene-C based organic flexible resistive memory on standard CMOS platform[J]. Advanced Electronic Materials, 2021, 7(2): 2000864.
- [9] GONG L, XU Y P, DING B, et al. Thermal management and structural parameters optimization of MCM-BGA 3D package model[J]. International Journal of Thermal Sciences, 2020, 147: 106120.
- [10] HUANG L J, ZHAO W S. Thermo-mechanical analysis of an improved thermal through silicon via (ttsv) structure [J]. Progress in Electromagnetics Research M, 2013, 30: 51-66.
- [11] STAN M R, SKADRON K, BARCELLA M, et al. Hot-Spot: A dynamic compact thermal model at the processor-architecture level[J]. Microelectronics Journal, 2003, 34(12): 1153-1165.
- [12] HUANG W, STAN M R, SKADRON K, et al. Compact thermal modeling for temperature-aware design[C]//Proceedings of the 41st Annual Design Automation Conference. New York: ACM, 2004: 878-883.
- [13] AYALA J L, SRIDHAR A, CUESTA D. Thermal modeling and analysis of 3D multi-processor chips[J]. Integration, 2010, 43(4): 327-341.

- [14] WAN Z M, XIAO H, JOSHI Y, et al. Co-design of multi-core architectures and microfluidic cooling for 3D stacked ICs[J]. *Microelectronics Journal*, 2014, 45(12): 1814-1821.
- [15] SAHAY S, STRUKOV D. A behavioral compact model for static characteristics of 3D NAND flash memory[J]. *IEEE Electron Device Letters*, 2019, 40(4): 558-561.
- [16] 王勔成, 邵敏. 有限单元法基本原理与数值方法[M]. 北京: 清华大学出版社, 1988.  
WANG X C, SHAO M. Basic Principle and Numerical Method of Finite Element Method[M]. Beijing: Tsinghua University Press, 1988. (in Chinese)
- [17] KREITH F. The CRC Handbook of Thermal Engineering [M]. Boca Raton: CRC Press, 2000.
- [18] ZHU L J, BAMBERG L, AGNESINA A, et al. Heterogeneous 3D integration for a RISC-V system with STT-MRAM[J]. *IEEE Computer Architecture Letters*, 2020, 19(1): 51-54.

#### 作者简介



永若雪 女, 1996年生, 新疆乌鲁木齐人. 江南大学物联网工程学院硕士研究生. 主要研究方向为集成电路设计和存储器设计等.  
E-mail: yrx18140730408@163.com



姜岩峰(通讯作者) 男, 1972年生, 甘肃白银人. 江南大学物联网工程学院教授. 主要研究方向为集成电路设计、功率半导体器件和磁性器件等.  
E-mail: jiangyf@jiangnan.edu.cn