一款 22 纳米 32 兆位嵌入式 STT-MRAM 宏单元，在高达 150°C 时实现 5.9 纳秒随机读取访问和 7.4 MB/s 写入吞吐量

下位智宏 ®、松原健 ®、斋藤智也、小川智也、太藤康彦 ®、IEEE 会员、金田义信 ®、伊津政之、武田浩一、三谷英范、伊藤隆 ⊕、IEEE 会员、河野隆 ⊕

摘要- 本文介绍了一种高精度感测放大器技术、快速写入方案以及一次性可编程 (OTP) 存储器单元读取技术，这些技术应用于 22 纳米 32 兆位嵌入式 STT-MRAM(eMRAM) 宏单元，用于高端微控制器单元 (MCU)。升压交叉耦合感测放大器 (BCC-SA) 在 125°C 和 150°C 下分别实现了 5.1 纳秒和 5.9 纳秒的随机读取访问。具有快速电压设置 (VPBW-FVS) 方案的可变并行位写入 (VPBW) 和写入电压常开 (WVAO) 模式实现了 7.4 MB/s 的写入吞吐量和 73% 的写入能量减少。可变电流 (SOC-VC) 技术稳定了操作条件，使得感测放大器可以同时用于基于嵌入式磁阻随机存取存储器 (MRAM) 的 OTP 和 MRAM 单元读取模式。

关键词-嵌入式 STT-MRAM(eMRAM)、快速随机读取操作、高写入吞吐量、高端微控制器单元 (MCU)、一次性可编程 (OTP)、自旋转移矩-MRAM(STT-MRAM)。

I. 引言

人工智能 (AI) 与物联网 (IoT) 技术的融合正在强力推动智慧社会的实现。诸如家庭自动化、机器人和医疗设备等应用，在高端微控制器单元 (MCU) 与低端微处理器单元 (MPU) 性能边界交叉区域需要更高的性能，以实现复杂的控制和实时处理，如图 1 所示。在传统的 MCU 中，嵌入式闪存 (eFlash) 被用作嵌入式非易失性存储器 (eNVM)[1]，以享受更快的启动时间 (无需初始程序代码加载)、更高的安全性和较低的物料清单 (BOM) 成本等优势。然而，在 22 – nm 代及更先进的逻辑工艺中提供 eFlash 是困难的，因为在前端工艺 (FEOL) 中嵌入闪存单元和处理 10V 级擦写电压的高压晶体管会带来额外的成本，而这些工艺正是实现交叉区域所需高性能的关键。

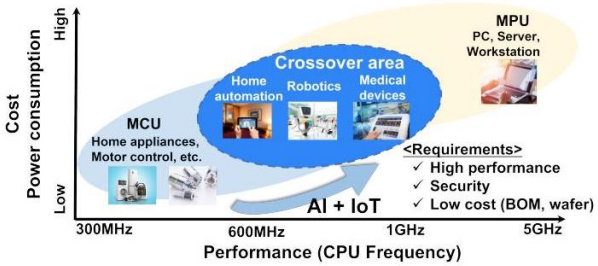


图 1. 交叉区域的应用与需求。

嵌入式自旋转移矩磁阻存储器 (eMRAM) 因其在后端工艺 (BEOL) 中需要较少的额外掩模以及较低的写入电压而被认为是 eNVM 的主要选择之一 [2], [3], [4], [5], [6], [7], [8], [9], [10], [11]。然而，与 eFlash 相比，较小的读取裕量 (RM) 是高速读取的关键挑战 [4]，特别是在汽车应用 (如符合 AECQ-100 Grade 1 标准的组件) 的高温环境下 [5]。为了解决这一问题，本文引入了一种高精度感测放大器。

在写入操作方面，eMRAM 相较于 eFlash 在重写时间上具有优势，因为它在比特单元级别具有更快的写入速度，并且不需要擦除操作。然而，这些优势在宏规格的写入吞吐量中并未得到充分利用，因为每个比特的大写入电流限制了并行写入比特的数量，并且还需要频繁的写入电压设置转换时间。考虑到这些问题，本文提出了三种方案，以增强 eMRAM 在宏规格中快速写入速度的优势，从而超越之前的 eFlash 工作 [12]。

除了对读写操作高性能的需求外，近年来确保高安全性的重要性也在增加。一次性可编程 (OTP) 存储器长期以来一直用于身份识别和加密以确保安全性。特别是在 eNVM 中的 OTP 可以提供低成本的高安全性。在这项工作中，提出了一种技术，可以在不增加重复电路或专用时序的情况下，使用相同的感测放大器读取基于 eMRAM 的 OTP 单元和磁阻随机存取存储器 (MRAM) 单元。

这些措施有助于提高客户应用程序的性能，并扩大 eMRAM 和其他 eNVM 的适用性。

稿件于 2023 年 2 月 27 日收到；2023 年 5 月 25 日和 2023 年 8 月 28 日修订；2023 年 9 月 1 日接受。发布日期为 2023 年 10 月 3 日；当前版本日期为 2024 年 3 月 28 日。本文由副编辑 Meng-Fan Chang 批准。(通讯作者:Takahiro Shimoi。)

作者隶属于日本东京 187-8588 的瑞萨电子公司 (电子邮件:takahiro.shimoi.wz@renesas.com)。

本文中一个或多个图形的彩色版本可在 https://doi.org/10.1109/JSSC.2023.3314822 获取。

数字对象标识符 10.1109/JSSC.2023.3314822

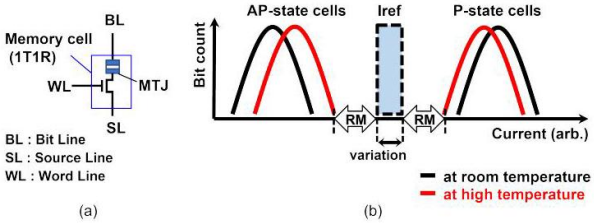
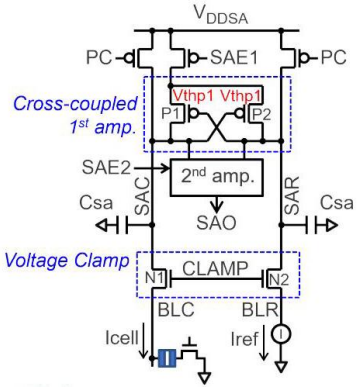


图 2. (a) STT-MRAM 单元结构. (b) 单元电流分布.



Vthp1 : P1 和 P2 的阈值电压

图 3. 传统 CC-SA.

II. 高精度传感技术

A. 读取 eMRAM 的挑战

图 2(a) 展示了简化的自旋转移矩-MRAM (STT-MRAM) 位单元结构。一个位单元由一个磁性隧道结 (MTJ) 器件和一个访问晶体管组成。STT-MRAM 有两种状态，即具有小单元电流的反平行 (AP) 状态和具有大单元电流的平行 (P) 状态，这些状态由 MTJ 的磁化方向定义。图 2(b) 展示了 STT-MRAM 单元的电流分布。在读取操作中，通常将具有一定变化的参考电流 (I_{ref}) 设置在 AP 状态和 P 状态电流的中间。AP 状态单元 (AP-cells) 和 P 状态单元 (P-cells) 之间的电流差定义为 RM，由 MTJ 的隧道磁阻 (TMR) 比率决定。通常，较小的 RM 需要更长的读取访问时间，并增加了读取失败的可能性。由于 STT-MRAM 的 RM 较小，尤其是在高温下，因此在 eMRAM 宏设计中，如何在小 RM 的情况下提高读取性能是一个挑战。

B. 传统交叉耦合感测放大器

图 3 展示了一个传统的交叉耦合读出放大器 (CC-SA)，它包含一个交叉耦合的第一级放大器和电压钳位晶体管。交叉耦合的 P1 和 P2 PMOS 对放大了 SA 内部节点 SAR 和 SAC 之间的微小电压差 ($\Delta V_{\text{SA}} = |V_{\text{SAR}} - V_{\text{SAC}}|$) .$\Delta V_{\text{SA}}$ ，该电压差源自单元电流 (I_{cell}) 和 I_{ref} ($\Delta I = |I_{\text{cell}} - I_{\text{ref}}|$) 之间的微小电流差。电压钳位晶体管 (N1 和 N2) 向位线 (BLs) 施加恒定电压以限制 I_{cell} 。由于 STT-MRAM 单元中的读取电流方向与 P 写入 (从 AP 状态写入 P 状态) 相同，较大的读取电流可能会在读取操作期间随机将位单元翻转到 P 状态。因此，需要 N1 和 N2 来防止读取干扰。V_{DDSA} 是专门为读出放大器供电的电源，其电平与 V_{DD}.C_{sa} 相同，后者是晶体管寄生电容与 SAC 和 SAR 的金属寄生电容的总和。为了读取微小的 ΔV_{SA} ，必须对 SAC 和 SAR 进行屏蔽以减少来自其他信号的串扰噪声。

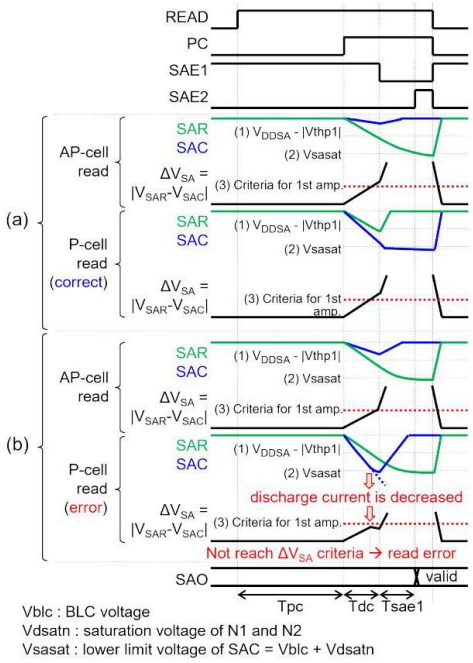


图 4. CC-SA 的时序波形。(a) 读取 P 单元的正确情况，(b) 读取 P 单元的错误情况。

图 4 展示了 CC-SA 读取操作的时序波形。首先，在 T_{pc} 期间，SAC 和 SAR 被预充电至 V_{DDSA} ，BLC 和 BLR 被预充电至 $V_{\text{clamp}} - V_{\text{thn}}$ 以防止读取干扰。在 $T_{\text{dc}}, \Delta V_{\text{SA}}$ 期间，随着 SAC 和 SAR 分别被 I_{cell} 和 I_{ref} 放电，$T_{\text{dc}}, \Delta V_{\text{SA}}$ 增加。随后，在 $T_{\text{sae 1}}$ 期间激活交叉耦合的 PMOS 对。当满足第一放大器中 P1 和 P2 之间容忍失配变化的 ΔV_{SA} 标准时，ΔV_{SA} 可以被正确放大。最后，存储的数据在 SAE2 = 1 的时序下通过锁存型第二放大器以逻辑电平信号 SAO 读出 [图 4(a)]。

为了确保 CC-SA 的正确操作，在激活交叉耦合第一放大器的时序下需要满足两个条件。在读取 AP-cell 的情况下，V_{SAR} 应低于 $V_{\text{DDSA}} - |V_{\text{thp1}}|$ 以打开 P1。另一方面，在读取具有最小 RM 的 P-cell 的情况下，如图 4(b) 所示，V_{SAC} 应高于 $V_{\text{sasat}} = V_{\text{blc}} + V_{\text{dsatn}}$ ，以使电压钳位晶体管工作在饱和区，以恒定的 I_{cell} 放电 SAC，其中 V_{blc} 是 BLC 的电压，V_{dsatn} 是 N1 和 N2 的饱和电压。当 V_{SAC} 低于 $V_{\text{sasat}}, I_{\text{cell}}$ 时，$V_{\text{sasat}}, I_{\text{cell}}$ 减小且 ΔV_{SA} 饱和。如果 ΔV_{SA} 未达到预期标准，则会发生读取错误。

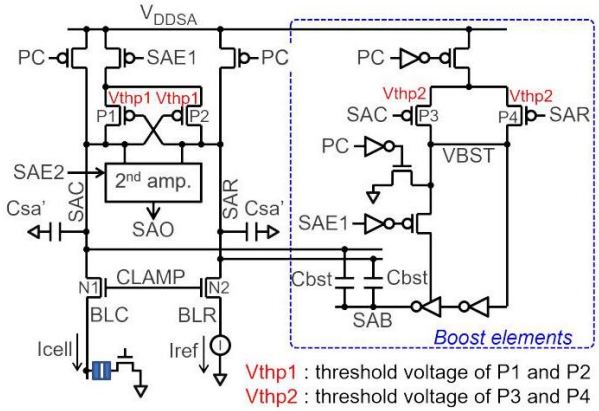


图 5. 提出的 BCC-SA。

这里，考虑了如何最大化 ΔV_{SA} 。首先，ΔV_{SA} 可以表示为

$$\Delta V_{\text{SA}} = \frac{T_{\text{dc}}}{C_{sa}} \cdot \Delta I. \tag{1}$$

此外，最大 T_{dc} $(T_{\text{dc\_max}})$ 受到以下限制

$$T_{\text{dc\_ max}} = \frac{V_{\text{DDSA}} - (V_{\text{blc}} + V_{\text{dsatn}})}{I_p} \cdot C_{\text{sa}} \tag{2}$$

其中 $V_{\text{blc}} = I_p \cdot R$($R$ 是 MTJ、访问/外围晶体管和 BL 的总电阻。
因此，最大 ΔV_{SA} ($\Delta V_{\text{SA\_max}}$) 可以表示为

$$\Delta V_{\text{SA\_max}} = \frac{\Delta I}{I_p} \cdot (V_{\text{DDSA}} - V_{\text{sasat}})$$

$$(\because V_{\text{sasat}} = V_{\text{blc}} + V_{\text{dsatn}}) \ . \tag{3}$$

公式 (3) 表明，当 ΔI 随温度升高或 MTJ 尺寸缩小而减小时，CC-SA 无法正常工作。为了通过 CC-SA 读取较小的 ΔI，一种解决方案是使用高电压 V_{DDSA}。然而，在这种情况下，从可靠性的角度来看，CC-SA 不能由核心晶体管组成，这会增加 CC-SA 的面积。此外，使用高电压晶体管会增加 SA 的失配变化，从而导致 ΔV_{SA} 的标准增加。因此，在使用核心晶体管的同时，第 II-D 节提出了另一种合适的解决方案。

C. 增强型交叉耦合感应放大器

为了扩展使用核心晶体管的 CC-SA 的工作范围，图 5 中提出了在 CC-SA 基础上增加增强元件的增强型 CC-SA(BCC-SA)。增强元件主要由两部分组成: 用于检测 SAC 和 SAR 电压降至 $V_{\text{DDSA}} - |V_{\text{thp 2}}|$ 的 PMOS 对 (P3 和 P4)，以及为 SAC 和 SAR 提供电荷的增强电容 (C_{bst})。P3 和 P4 的尺寸经过优化，使得 $|V_{\text{thp 2}}|$ 高于 $|V_{\text{thp 1}}|$，因为增强元件需要在 V_{SAC} 和/或 V_{SAR} 低于 $V_{\text{DDSA}} - |V_{\text{thp 1}}|$ 的条件下激活，以打开 P1 和/或 P2。C_{bst} 使用寄生电容，该电容最初由 CC-SA 中的屏蔽产生，如第 II-B 节所述。其他增强元件由简单的逻辑电路组成，晶体管尺寸较小，占整个宏区面积的不到 0.5%。

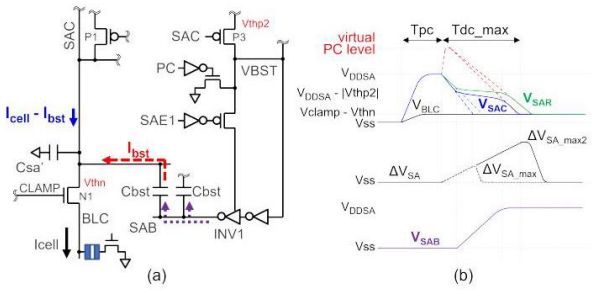


图 6. (a) 从图 5 中摘录的图示及升压激活阶段的运动。(b) BCC-SA 的概念图。

图 6(a) 展示了从图 5 的 BCC-SA 框图中摘录的记忆单元侧路径和部分升压元件，以解释 BCC-SA 的概念。当图 6(a) 中的 P3 检测到 V_{SAC} 低于 $V_{\text{DDSA}} - |V_{\text{thp2}}|$ 时，升压元件被激活。P3 的栅极连接到 SAC；因此，随着 SAC 的降低，VBST 根据 P3 的导通状态电流逐渐增加，而由第二反相器 (INV1) 提供的 SAB 电压 (V_{SAB}) 以与 VBST 相同的速率增加。因此，该方案的关键点在于，通过根据 SAC 和 SAR 的电压放电情况开启 P3 或 P4，自适应地确定向 SAC 和 SAR 供电的速率。因此，与恒定电流充电方案 [5], [13] 相比，即使存在 I_{cell} 变化，也无需调整供电时间和电荷量，即可轻松保证 SAC 的下限电压，以保持钳位晶体管处于饱和区。由于电荷通过 C_{bst} 供给 SAC 和 SAR，这相当于 I_{bst} 流动，SAC 的放电电流变为 $I_{\text{cell}} - I_{\text{bst}}$，SAR 的放电电流同样变为 $I_{\text{ref}} - I_{\text{bst}}$。SAC 和 SAR 之间的电流差 ($\Delta I$) 保持 $|I_{\text{cell}} - I_{\text{ref}}|$；因此，$\Delta V_{\text{SA}}$ 在升压元件的激活期间持续增加。换句话说，BCC-SA 可以虚拟地提升 SAC 和 SAR 的预充电电压 [图 6(b)]。因此，BCC-SA 的最大 ΔV_{SA} ($\Delta V_{\text{SA\_max 2}}$) 变为

$$\Delta V_{\text{SA\_max 2}} = \frac{\Delta I}{I_p} \cdot \left(V_{\text{DDSA}} - V_{\text{sasat}} + \frac{C_{\text{bst}}}{C_{\text{sa}'} + C_{\text{bst}}} \cdot V_{\text{DDSA}} \right)$$

(4)

其中 C_{sa} 和 C_{bst} 的总和等于 C_{sa}。
这表明与传统的 CC-SA 相比，所提出的 BCC-SA 可以扩展最大 ΔV_{SA}。

图 7 显示了 BCC-SA 的完整时序波形。首先，在 T_{pc} 期间，SAC、SAC、SAR、BLC 和 BLR 被预充电至与 CC-SA 相同的电压。在 T_{dc} 期间，SAC 和 SAR 分别通过 I_{cell} 和 I_{ref} 放电。在读取 P-cell 的情况下，当 V_{SAC} 降至 $V_{\text{DDSA}} - |V_{\text{thp 2}}|$ 以下时，SAC 和 SAR 在 T_{bst} 期间通过 C_{bst} 提升，保持 V_{SAC} 和 V_{SAR} 高于 V_{sasat}。然后，根据 BCC-SA 的概念，ΔV_{SA} 持续增加，如图 6 所示。因此，当第一个放大器激活时，V_{SAC} 和 V_{SAR} 在 $T_{\text{sae 1}}$ 期间通过具有足够 ΔV_{SA} 的交叉耦合 P1 和 P2 稳定放大。

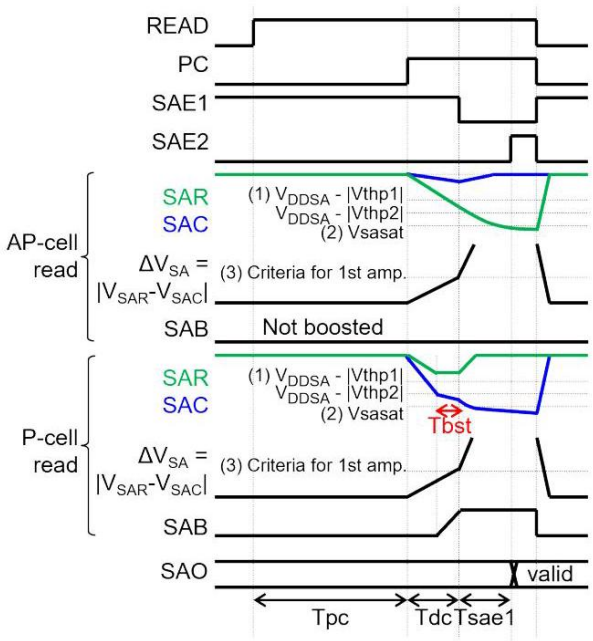


图 7. BCC-SA 的时序波形。

另一方面，在读取 AP-cell 的情况下，V_{SAR} 在 T_{dc} 期间低于 V_{SAC}。由于 I_{ref} 小于 I_{p}，V_{SAR} 不会降至 $V_{\mathrm{DDSA}} - |V_{\mathrm{thp\,2}}|$ 以下。因此，提升元件未被激活，BCC-SA 的工作方式与 CC-SA 类似。

图 8 显示了分别满足 ΔV_{SA} 标准的 T_{dc} 与 ΔI 或 V_{DDSA} 之间的关系。图 8(a) 中 ΔI 的下降相当于高温或 MTJ 尺寸缩小导致的 RM 下降。

BCC-SA 可以在 40% 比 CC-SA 更小 ΔI [图 8(a)] 和比 CC-SA 低 14% V_{DDSA} [图 8(b)] 的条件下运行。因此，与 CC-SA 相比，BCC-SA 更适合扩展操作范围，特别是在高温下更小的 ΔI 和/或先进工艺节点中更低的 V_{DDSA} 等条件下。

D. 测量结果

图 9 展示了我们宏中 BCC-SA 的访问时间 (T_{ac}) shmoo 图，其 V_{DD} 范围为 0.8 ± 0.1 V。采用所提出的 BCC-SA，分别在 -40°C 和 125°C 下实现了 T_{ac} 为 5.0 和 5.1 ns 的高速读取 [图 9(a) 和 (b)]。此外，由于 BCC-SA 在相同的 V_{DD} 范围内扩展了 ΔI 的最小限制，如图 8 中的"FP" 区域所示，因此在 150°C 下实现 T_{ac} 为 5.9 ns 的读取速度是可行的 [图 9(c)]。这一结果表明，通过采用 BCC-SA，未来可以将温度扩展至 150°C，以满足汽车应用的需求。

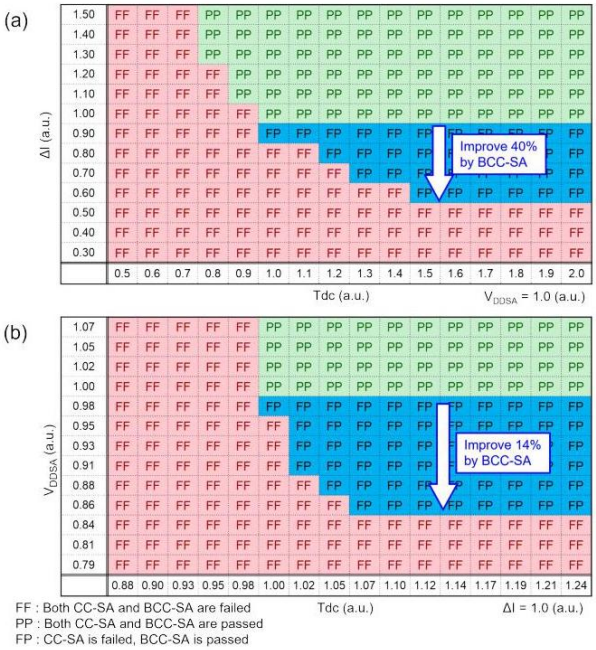


图 8. 模拟的 shmoo 图，展示了 T_{dc} 与 (a) ΔI 和 (b) V_{DDSA} 的关系，以满足 ΔV_{SA} 标准。

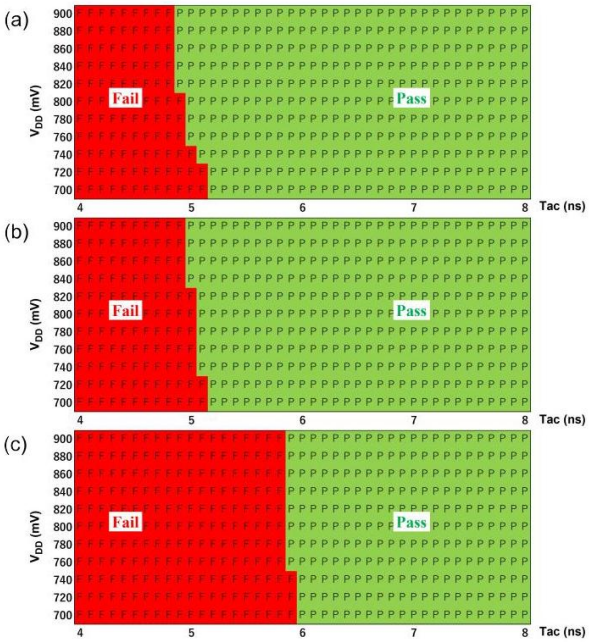


图 9. T_{ac} 在 (a) −40°C 、(b) 125°C 和 (c) 150°C 下的 shmoo 图。

III. 宽电流传感技术

A. 基于 eMRAM 的 OTP

如第一章所述，使用 eNVM 单元实现 OTP 是确保高安全性且低成本的一种候选方案。

图 10 显示了 MTJ 的 R-V 曲线。基于 eMRAM 的 OTP 利用了氧化物隧道势垒的介电击穿原理 [14]。一旦 eMRAM 单元转变为电阻低于 P 单元的 OTP 状态单元 (OTP 单元)，它将永远不会回到 P 状态或 AP 状态。

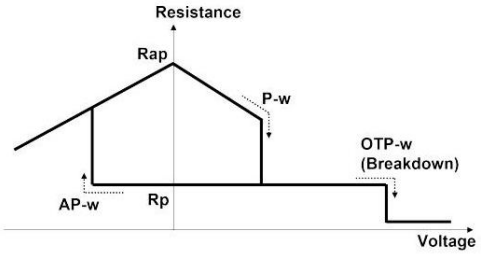


图 10. MTJ 的 R-V 磁滞曲线图像。

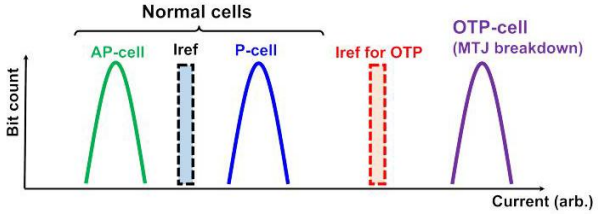


图 11. 三种状态的单元电流分布。

B. 读取基于 eMRAM 的 OTP 的传统方法

AP 单元、P 单元和 OTP 单元的电流分布如图 11 所示。OTP 单元的电流明显大于正常单元。在读取 OTP 单元时，I_{ref} 需要设置在 P 单元电流和 OTP 单元电流之间，这比读取正常单元时的 I_{ref} 要大。

通常，希望将 OTP 单元放置在由正常 AP/P 单元组成的内存阵列中，并共享外围电路 [11]。然而，使用用于正常单元读取操作的感测放大器来读取 OTP 单元是困难的。读取控制时序和感测放大器的尺寸 (如钳位晶体管) 需要单独优化。此外，还需要在正常单元读取和 OTP 单元读取之间切换 I_{ref} 的时间 (几微秒)。如果实现另一个具有更大 I_{ref} 的感测放大器专门用于读取 OTP 单元，则会导致宏面积的增加。因此，需要一种额外的感测技术来使用相同的感测放大器检测正常单元和 OTP 单元的宽电流范围。

C. 使用 SOC-VC 读取基于 eMRAM 的 OTP

在此，提出了可变电流条件下的稳定操作技术 (SOC-VC)，如图 12 所示。SOC-VC 由电流加法器 (I_{OTP}) 和电流减法器 (I_{plus}) 组成，连接到 BLC。SOC-VC 技术可以应用于传统的 CC-SA 和提出的 BCC-SA，因为它连接到 BLC。SOC-VC 所需的额外元件在电流精度足够的情况下，仅占感应放大器电路面积的几个百分点，这比添加另一个感应放大器来读取 OTP 单元的面积要小。

在读取 OTP 单元的情况下，电流源 I_{OTP} 被添加到 BLC 中 [图 13(a)]。通过 N1 (I_{N1}) 在 OTP 读取操作中输入到感应放大器的电流由于添加了 I_{OTP} 而向减少方向偏移。I_{N1} 几乎与正常单元读取操作中的电流相同，因为 I_{N1} 变为 $I_{\text{cell}} - I_{\text{OTP}}$ ，如图 13(b) 所示。由于正常单元读取的 I_{ref} 也可以通过 SOC-VC 用于 OTP 单元读取操作，因此 OTP 单元可以在与读取正常单元相同的操作条件和时序下读取，而无需 I_{ref} 的切换时间。

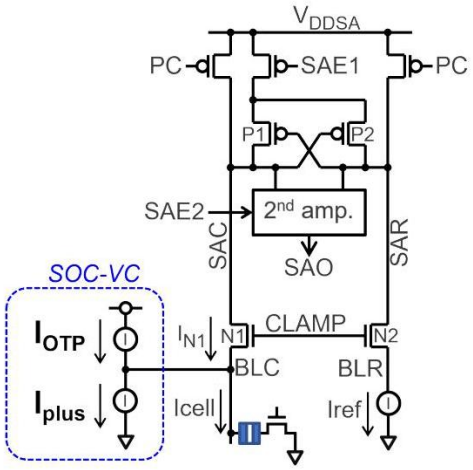


图 12. SOC-VC 技术。

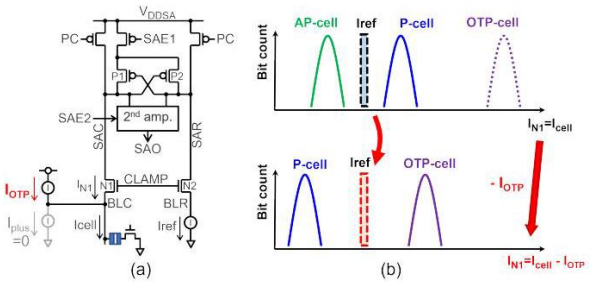


图 13. 使用 SOC-VC 读取 OTP 单元。(a) 电路图。(b) 输入到 SA 的电流分布。

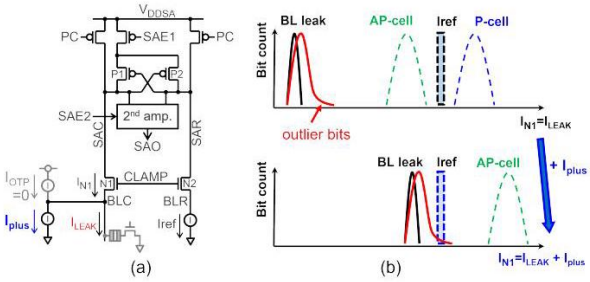


图 14. 使用 SOC-VC 检测 BL 泄漏。(a) 电路图。(b) 输入到 SA 的电流分布。

因此，SOC-VC 实现了小面积开销，并实现了基于 eMRAM 的 OTP 单元的无缝读取，增强了交叉领域应用的安全性。

D. SOC-VC 的其他用途

除了读取 OTP 单元外，SOC-VC 还可用于检测异常电流。图 14(a) 展示了激活电流源 I_{plus} 以检测位线漏电流 (I_{LEAK}) 的情况。由于异常位引起的位线漏电流 [如图 14(b) 中的红线所示] 会导致读写操作失败，因此在晶圆测试的早期阶段应进行筛选测试以检测明显缺陷。当 I_{plus} 被添加到 BLC 时，I_{N1} 变为 $I_{\text{LEAK}} + I_{\text{plus}}$ 。然后，电流分布向增加方向移动，并且通过正常单元读取操作中的相同 I_{ref} 可以正确检测到由异常位引起的位线漏电流。

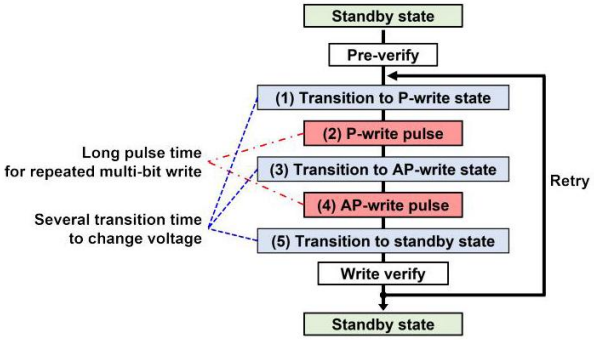


图 15. 写入流程图及写入吞吐量的瓶颈。

此外，SOC-VC 还可用于小 RM 检测。通过使用小的 I_{OTP} ，可以检测到 P 单元电流分布的下边缘。同样，小的 I_{plus} 可以检测到 AP 单元电流分布的上边缘。

IV. 写入吞吐量提升方案

A. eMRAM 中写入吞吐量的瓶颈

图 15 展示了写入流程的示意图。” 待机状态” 是接收操作命令 (如读取和写入) 之前的状态。”P 写入状态” 和”AP 写入状态” 分别是为”P 写入”(从 AP 状态切换到 P 状态) 和”AP 写入”(从 P 状态切换到 AP 状态) 设置专用写入电压作为外围电路供电电压的状态。在待机状态、P 写入状态和 AP 写入状态之间切换的时间 (1、3 和 5) 是为每个操作设置专用电压所需的。

当接受写入命令时，MRAM 单元在“(2) P 写入脉冲”和“(4) AP 写入脉冲”中被写入。随后总是执行“写入验证”操作，以读取写入的单元并确定读取数据是否与写入数据匹配，因为写入操作在一定概率下会失败。当在“写入验证”中检测到写入失败位时，仅对写入失败位执行重试写入操作。重试次数是预先确定的，以确保写入操作后的写入错误位数小于纠错码 (ECC) 可纠正的极限。因此，缩短写入脉冲时间和转换时间以提高写入吞吐量非常重要。在 IV-B-IV-D 节中提出了三种技术。

B. VPBW 方案

图 16 显示了写入误码率 (BER) 对 BL 电压 (V_{BL}) 的依赖性。由于 eMRAM 中写入特性的变化，传统上需要由内部电荷泵 (CP) 提供的高电压 V_{BL} (V_{w_all} (图 16 中) 来写入所有单元。由于 CP 电路的写入电流供应有限，写入单元通常需要分为几组 (例如，四组) N 位，并且写入脉冲依次应用于每组，如图 17(a) 所示。然而，大多数单元可以用比 V_{w_all} V_{w_1} 低得多的电压 (V_{w_1} (图 16 中) 写入，该电压可以从外部 V_{CC} (输入输出电压 1.8 V ± 10%) 提供。

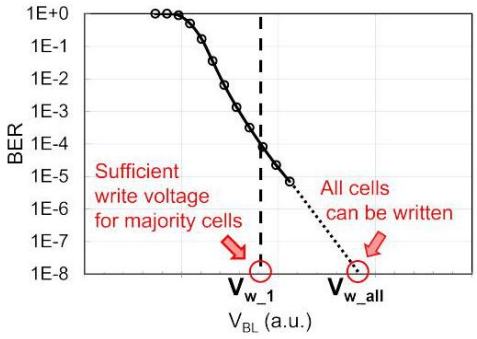


图 16. V_{BL} 与写入 BER 之间的关系。

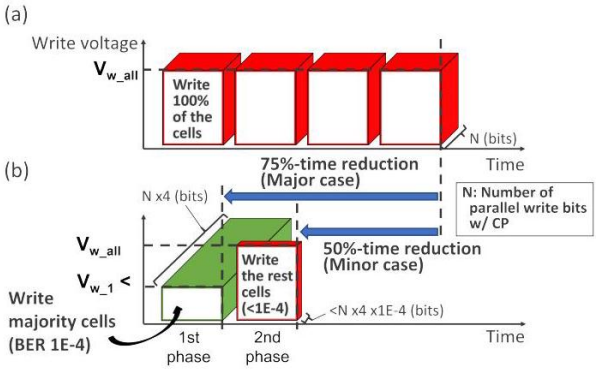


图 17. (a) 传统的恒定电压写入方案。(b) VPBW 方案。

考虑到图 16 中的写入特性，提出了可变并行位写入 (VPBW) 方案 [15] 以缩短写入脉冲时间，如图 17(b) 所示。VPBW 由两个阶段组成，每个阶段选择合适的电压源以增加并行写入位数并减少高功耗 CP 的使用。在第一写入阶段，4 N 位可以同时使用从 V_{CC} 电源降频的电压 (V_{w_1}) 进行写入。当第一写入阶段后的写入验证失败时，第二写入阶段使用 CP 写入所有剩余位，误码率 (BER) 为 1E-4 或更低。由于第二写入阶段仅在数百次中发生一次，写入操作主要在第一写入阶段完成。

因此，与传统写入方案相比，VPBW 在主要情况下可以将写入脉冲次数减少 75%，在次要情况下也可以减少 50%。此外，由于外部 V_{CC} 的高能效和 CP 使用的减少，VPBW 可以降低写入能量。

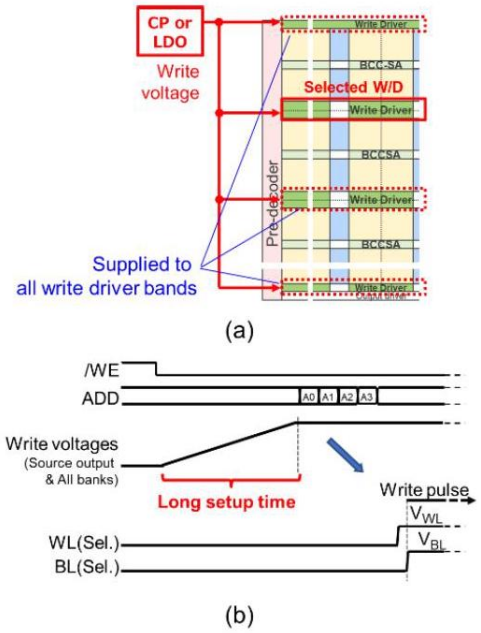


图 18. (a) 传统写入路径结构。(b) 传统时序波形。

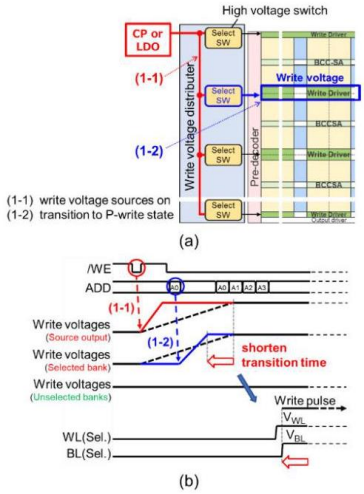


图 19. (a) 所提出的快速电压设置方案的框图。(b) FVS 的时序波形。

C. 快速电压设置方案

为了提高写入吞吐量，待机状态和写入状态之间的转换时间也应缩短。在应用写入脉冲之前，必须激活写入电压源。在传统方案中，写入电压被提供给所有写入驱动带，如图 18(a) 所示。由于所有写入驱动带中的大寄生负载需要充电，因此需要较长的电压设置时间 [图 18(b)]。

在此，图 19(a) 中提出了快速电压设置 (FVS) 方案。实现了带有高压开关的写电压分配器，仅向选定的写驱动带提供写电压。在 FVS 方案中，写电压采用两步法设置 [图 19(b)]。第一步，设置写电压以在所有开关未选中的情况下为公共节点充电 [图 19(a) 中的红线]。第二步，通过由写地址确定的选中开关，仅向一个写驱动带充电 [图 19(a) 中的蓝线]。通过上述两步限制电压供应区域，每一步的寄生负载都比传统方案更小，因此可以缩短图 15 中的过渡时间 (1)。FVS 方案还可以缩短图 15 中的其他过渡时间 (3) 和 (5)。

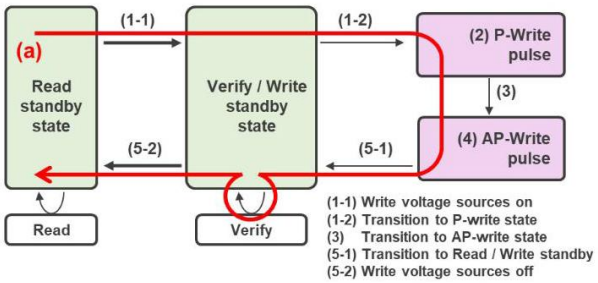


图 20. 正常写入时采用 VPBW-FVS 方案的写操作状态图。

图 20 展示了利用 VPBW 和 FVS 方案 (VPBW-FVS) 组合优势的写操作状态图。” 验证/写 (V/W) 待机” 是中间状态，用于在保持写电压的情况下启用写验证操作。通过在 V/W 待机状态下实现写电压分配器 (作为高压电源门控)，可以进行写验证操作。通过引入 V/W 待机状态，可以实现所提出的两步电压设置方法。图 20 中的红线 (a) 展示了正常的写操作。所提出的 FVS 方案缩短了写电压源的开关时间 [(1-1) 和 (5-2)]，从 V/W 待机到 P-Write 的转换时间 (1-2)，从 P-Write 到 AP-Write 的转换时间 (3)，以及从 AP-Write 到 V/W 待机的转换时间 (5-1)。

因此，与传统方案相比，所提出的 FVS 方案可以将整体转换时间缩短 46%，并提高写吞吐量。由于减少了寄生负载电容，FVS 还可以改善写能量。如第 IV-A 节所述，由于写操作总是伴随着验证操作，所提出的 FVS 方案的这些优势对于任何大小的数据都有效。

D. 写电压常开模式的突发写入

在写入大量数据 (大于写入单元) 的情况下，每个写入单元的写操作应连续执行。在这种称为突发写入操作的情况下，通过应用所提出的写电压常开 (WVAO) 模式，可以进一步提高写吞吐量。

在传统的写入操作中，写入电压源在每个写入周期中基于“读取待机”状态进行开关，以减少维持写入电压的能量。另一方面，在采用所提出的 WVAO 模式的突发写入操作中，如图 21 中的蓝线 (b) 所示，写入电压源保持激活状态，同时连续写入大量数据。由于在突发写入操作中，“(1-1) 写入电压源开启”和“(5-2) 写入电压源关闭”仅发生一次，因此采用 WVAO 模式的写入吞吐量可以得到提升。通过减少写入电压的充放电能量，突发写入操作中的写入能量消耗也有望降低。

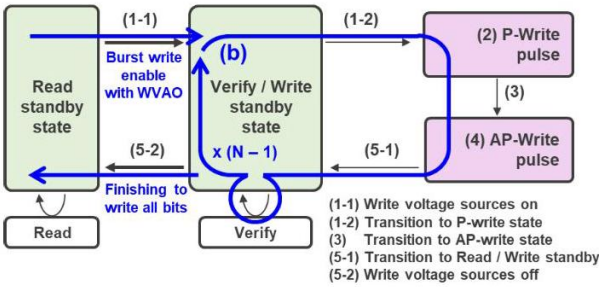


图 21. 采用所提出的 WVAO 模式在 N 倍同时写入时的突发写入操作状态图。

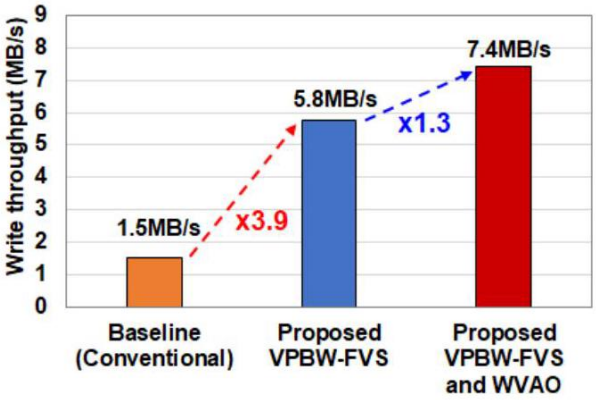


图 22. 在 $T_j = -40°\mathrm{C}$ 时的写入吞吐量提升。

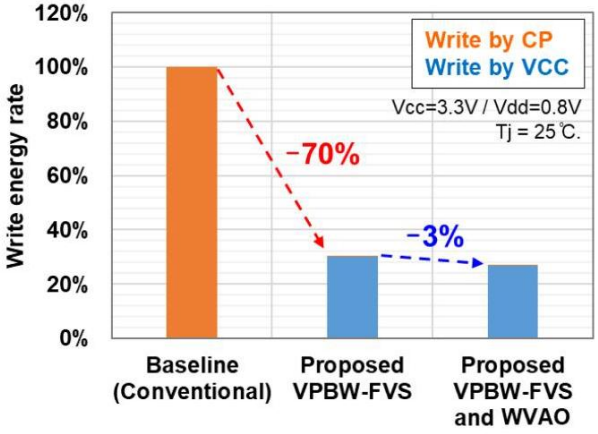


图 23. 典型条件下的写入能量减少。

图 24. 芯片显微照片。
TABLE I
宏配置

		本工作	ISSCC 2019 [6]	ISSCC 2020 [7]
技术		N22 MRAM	22FFL MRAM	N22 MRAM
单元尺寸 (μm^2)		0.0456	0.0486	0.0456
$v_{DD}\,(V)$		0.8 ± 0.1	-	0.8
$v_{cc}\,(v)$		1.8 ± 10%	-	1.8 ± 10%
工作温度 (Tj)		-40□150□	-40□105□	-40□ 125□ (功能温度 150℃)
单元编号 / BL		512	256	512
读取访问 @V $_{DD}$ =0.8V (不包括 ECC 解码)	- 40	5.0ns	-	-
	125	5.1ns	5ns (105℃)	8.5ns
	150	5.9ns	-	-
写入吞吐量 (50% 数据模式)		(a) 5.8MB/s (b) 7.4MB/s	-	> 1.5MB/s

(a) 正常写入 / (b) 采用 WVAO 的突发写入

E. 测量结果

包括设置、验证以及写入脉冲的写入吞吐量和写入能量分别如图 22 和图 23 所示。

通过采用 VPBW-FVS 方案，在最坏温度条件 $T_j = -40°\text{C}$ 下实现了 5.8MB/s 的写入吞吐量。同时，与传统写入方案相比，写入能量也减少了 70%。此外，通过结合 VPBW-FVS 和 WVAO 模式，在突发写入操作中实现了 7.4 MB/s 的写入吞吐量和 73% 的写入能量减少。

V. 结论

基于所提出的技术和方案，已在 22 纳米 STT-MRAM 工艺中开发了一款 32Mb 测试芯片。图 24 展示了芯片的显微照片，表 I 总结了宏配置。通过 BCC-SA 实现了世界上最快的随机读取访问时间，在 T_j 的 150°C 下达到 5.9 纳秒。采用 VPBW-FVS 方案和 WVAO 模式，写入吞吐量达到 7.4 MB/s。基于 eMRAM 的 OTP 与所提出的 SOC-VC 技术相结合，增强了芯片的安全性。开发的 eMRAM 宏为跨界应用领域提供了高性能、高安全性和低成本的优势。

致谢

作者感谢台积电 (TSMC) 的贡献。

参考文献

[1] M. Nakano 等，” 一款用于高端 MCU 的 40 纳米嵌入式 SG-MONOS 闪存宏，通过电荷辅助偏移消除感测放大器实现 200 – MHz 随机读取操作和 7.91Mb /mm^2 密度，”《IEEE 固态电路杂志》，第 57 卷，第 10 期，第 3094-3102 页，2022 年 10 月。

[2] M. Jefremow 等，” 用于 40 nm CMOS 中低于 80 mV 位线电压嵌入式 STT-MRAM 的时间差分感测放大器，”《IEEE 国际固态电路会议 (ISSCC) 技术论文摘要》，2013 年 2 月，第 216-217 页。

[3] C. Kim, K. Kwon, C. Park, S. Jang, 和 J. Choi, ”7.4 一种用于具有 1T1MTJ 共源线结构阵列的 STT-MRAM 的共价键交叉耦合电流模式感测放大器,” 发表于 IEEE 国际固态电路会议 (ISSCC) 技术论文摘要, 2015 年 2 月, 第 1-3 页。

[4] Y. J. Song 等, ” 高度可制造的嵌入 28 nm 逻辑的 STT-MRAM 的演示,” 发表于 IEDM 技术摘要, 2018 年 12 月, 第 18.2.1-18.2.4 页。

[5] Q. Dong 等, ” 一种 1Mb 28nm STT-MRAM，采用单电容偏移消除感测放大器和原位自写终止技术，在 1.2V VDD 下实现 2.8 ns 读取访问时间,” 发表于 IEEE 国际固态电路会议 (ISSCC) 技术论文摘要, 2018 年 2 月, 第 480-482 页。

[6] L. Wei 等, ”13.3 一种 7Mb STT-MRAM，采用 22FFL FinFET 技术，通过写-验证-写方案和偏移消除感测技术在 0.9 V 下实现 4 ns 读取感测时间,” 发表于 IEEE 国际固态电路会议 (ISSCC) 技术论文摘要, 2019 年 2 月, 第 214-215 页。

[7] Y.-D. Chih 等人, ”13.3 一款 22 nm 32Mb 嵌入式 STT-MRAM，具有 10 ns 读取速度、1M 次写入耐久性、150 下 10 年数据保持能力以及对磁场干扰的高抗性,” 发表于 IEEE 国际固态电路会议 (ISSCC) 技术论文摘要, 2020 年 2 月, 第 222-224 页。

[8] E. M. Boujamaa 等人, ” 一款 14.7Mb/m $m^2$28 nm FDSOI STT-MRAM，采用电流受限读取路径、52Ω/Sigma 偏移电压感测放大器和完全可调 CTAT 参考,” 发表于 IEEE 超大规模集成电路研讨会论文集, 2020 年 6 月, 第 1-2 页。

[9] D. Edelstein 等人, ” 一款 14 nm 嵌入式 STT-MRAM CMOS 技术,” 发表于 IEDM 技术摘要, 2020 年 12 月, 第 11.5.1-11.5.4 页。

[10] V. B. Naik 等人, ”STT-MRAM: 一种具有卓越可靠性、对外部磁场和射频源免疫性的稳健嵌入式非易失性存储器,” 发表于超大规模集成电路技术研讨会论文集, 2021 年 6 月, 第 1-2 页。

[11] P.-H. Lee 等人, ”33.1 一款 16 nm32Mb 嵌入式 STT-MRAM，具有 6 ns 读取访问时间、1M 次写入耐久性、150°C 下 20 年数据保持能力以及用于磁免疫的 MTJ-OTP 解决方案,” 发表于 IEEE 国际固态电路会议 (ISSCC) 论文集, 2023 年 2 月, 第 494-496 页。

[12] A. Kanda 等人, ” 基于 28 纳米 SG-MONOS 的 24 MB 嵌入式闪存系统，具有 240 MHz 读取操作和稳健的空中软件更新功能，适用于汽车应用,” IEEE 固态电路快报, 卷 2, 期 12, 页 273-276, 2019 年 12 月.

[13] A. Antonyan, S. Pyo, H. Jung, 和 T. Song, ” 用于替代 eFlash 的嵌入式 MRAM 宏,” 发表于 IEEE 国际电路与系统研讨会 (ISCAS), 2018 年 5 月, 页 1-4.

[14] G. Jan 等人, ” 与全功能 STT-MRAM 集成的基于 MgO 的反熔丝 OTP 设计在 Mbit 级别的演示,” 发表于 VLSI 技术研讨会 (VLSI Technol.), 2015 年, 页 164-165.

[15] T. Ito 等人, ” 在 16 nm FinFET 逻辑工艺中实现 72% 写入能量减少的 20 Mb 嵌入式 STT-MRAM 阵列，采用自终止写入方案,” 发表于 IEDM 技术文摘, 2021 年 12 月, 页 2.2.1-2.2.4.

Takahiro Shimoi 于 2010 年和 2012 年分别在日本大阪大学获得物理学学士和硕士学位.

他于 2012 年加入日本东京的瑞萨电子公司，参与了用于高端汽车 MCU 的嵌入式分栅 MONOS 闪存宏的开发。他目前致力于 MCU 的嵌入式 MRAM 电路设计.

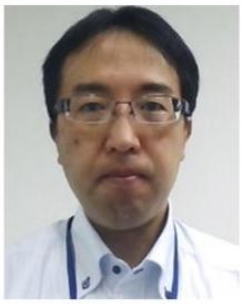
松原健于 1997 年在日本兵库县神户大学获得电气与电子工程学士学位。

2001 年, 他加入日本东京的日立公司, 从事 AND 和辅助栅极 (AG)-AND 闪存的开发工作。2003 年转入瑞萨科技公司后，他致力于嵌入式分离栅极 MONOS(SG-MONOS) 高密度闪存以及嵌入式 1T/2T-MONOS 低功耗/低成本闪存的开发。目前，他担任瑞萨电子公司东京总部存储器 IP 技术部 1 的高级经理，负责为微控制器设计和开发最先进的嵌入式非易失性存储器 (NVM) 宏单元，特别是基于 STT-MRAM 的技术。

斋藤友也于 1997 年在大阪大学获得电子工程学士学位，并于 1999 年和 2005 年分别获得电子与信息系统硕士和博士学位。

1999 年，他加入位于美国纽约的 Halo LSI 公司，从事闪存器件技术的研究。2005 年，他转入日本神奈川的 NEC 电子公司，随后转入日本东京的瑞萨电子公司。他一直致力于为消费类和汽车应用的高端微控制器开发嵌入式闪存器件。目前，他负责嵌入式 STT-MRAM 的开发，重点关注宏 IP 特性和可靠性。自 2020 年至 2025 年，他担任日本科学技术振兴机构 PRESTO 项目的研究领域顾问。

斋藤博士于 2018 年至 2022 年担任 IEEE 国际存储器研讨会 (IMW) 技术委员会成员。



小川智也于 1996 年和 1998 年分别在日本大阪大学获得物理学学士和硕士学位。

1998 年，他加入日本东京三菱电机株式会社。从 1998 年到 2002 年，他从事 32Mb 到 256Mb NOR 闪存的评估和设计工作。2003 年转入瑞萨科技株式会社东京分公司后，他致力于高密度闪存的开发。目前，他在瑞萨电子株式会社东京分公司工作，负责最先进的嵌入式非易失性存储器的设计和开发，用于微控制器 (MCU)。

太藤康彦 (IEEE 会员) 于 1991 年和 1993 年分别在日本东京大学获得应用物理学学士和硕士学位。

1993 年，他加入日本东京三菱电机株式会社，从事 NOR 和 DINOR 闪存的设计工作。他还参与了商品化和嵌入式动态随机存取存储器 (DRAM) 的设计。他开发了具有高速静态随机存取存储器 (SRAM) 接口的片上系统 (SoC) 嵌入式 DRAM 宏单元。2003 年转入瑞萨科技株式会社，2010 年转入瑞萨电子株式会社东京分公司后，他积极致力于嵌入式 NOR 和分裂栅 MONOS(SG-MONOS) 闪存宏单元的开发，主要用于高端汽车微控制器 (MCU)。目前，他是瑞萨电子株式会社存储器 IP 技术部 1 的首席工程师，负责汽车和物联网/消费类 MCU 产品的先进嵌入式非易失性存储器项目。

Taito 先生于 2018 年至 2022 年担任 IEEE 国际固态电路会议 (ISSCC) 的技术程序委员会成员。他于 2021 年担任 IEEE 固态电路期刊的客座编辑。

金田义信于 1993 年获得日本东京明治大学物理学学士学位。1993 年，他加入日本大阪三洋电机株式会社。从 1993 年到 2010 年，他从事从 8 到 128Mb 的闪存 (包括 eFlash) 的研究和设计工作。2010 年转入安森美半导体后，他致力于开发用于片上系统 (SoC) 的低功耗非易失性存储器 (NVM) 系统。他目前就职于日本东京瑞萨电子株式会社，负责为 MCU 和 SoC 设计和开发最先进的嵌入式 NVM 软宏。

伊津政之在 2001 年获得日本鸟取大学电气与电子工程学士学位。

2001 年，他加入日本东京日立 ULSI 系统有限公司。从 2001 年到 2002 年，他从事 AND 闪存的评估工作。2003 年转入瑞萨科技株式会社后，他致力于辅助栅极 (AG)-AND 闪存的评估。他目前就职于日本东京瑞萨电子株式会社，从事最先进的嵌入式非易失性存储器用于 MCU 的评估工作。

武田浩一分别于 1991 年和 1993 年获得日本仙台东北大学电子工程学士和硕士学位。

他于 1993 年加入日本神奈川的 NEC 公司微电子研究实验室。此后，他一直从事存储器电路的研究与开发，特别是静态随机存取存储器 (SRAM)。自 2015 年起，他目前在瑞萨电子公司的共享研发核心技术部门开发非易失性存储器电路。

日本东京。

武田先生是日本电子信息通信工程师学会的成员。

三谷英德于 1992 年获得日本鸟取大学电子工程学士学位。

1992 年，他加入日本东京的三菱电机公司。从 1992 年到 2002 年，他从事存储卡和闪存的研究与设计。2003 年转入瑞萨科技公司后，他致力于开发嵌入式分裂栅 MONOS(SG-MONOS) 高密度闪存和嵌入式 1T/2T-MONOS 低功耗/低成本闪存。他目前就职于日本东京的瑞萨电子公司，负责设计和开发最先进的嵌入式非易失性存储器 (NVM) 宏，如低功耗和高性能的 STT-MRAM，用于微控制器 (MCU)。

伊藤隆 (IEEE 会员) 分别于 1993 年和 1995 年获得日本福井大学电气与电子工程学士和硕士学位。

1995 年，他加入日本东京三菱电机公司，从事 64Mb 至 512Mb 动态随机存取存储器 (DRAM) 的开发工作，包括同步 DRAM、双倍数据速率 (DDR)DRAM 和低功耗伪静态随机存取存储器 (SRAM)。2003 年和 2010 年分别转入瑞萨科技公司和瑞萨电子公司后，他致力于 NOR 和辅助栅极 (AG)-AND 闪存以及多级单元闪存的开发。此后，他领导了嵌入式分裂栅极 MONOS(SG-MONOS) 从 90nm 到 28nm 工艺节点的设计团队，为汽车应用实现了高性能和汽车级 0 级可靠性。他拥有 22 项美国专利。目前，他担任瑞萨电子公司共享研发核心技术部门的杰出工程师，负责决定汽车、物联网 (IoT) 和基础设施产品中尖端嵌入式存储器 IP(如 MRAM 和 ReRAM) 的开发方向。

伊藤先生自 2022 年起担任 IEEE 国际固态电路会议 (ISSCC) 存储器分委会成员，并在 ISSCC 2023 上主持了存储器专题会议。他于 2023 年担任 IEEE 固态电路快报的客座编辑。

高野隆史于 1992 年和 1994 年分别获得日本东京大学电子工程学士和硕士学位。在研究生期间，他从理论和实验的角度研究了纳米结构化合物半导体器件 (量子线和量子点)。

1994 年，他加入日本东京三菱电机株式会社。从 1994 年到 2002 年，他在三菱电机株式会社超大规模集成电路开发实验室从事 64 至 512 兆位离散动态随机存取存储器 (DRAM) 的研究与设计，包括同步 DRAM 和 DDR/DDR2 同步 DRAM。2003 年转入瑞萨科技株式会社东京公司后，他致力于低功耗伪静态随机存取存储器 (SRAM) 和高密度 AG-AND 闪存的开发。自 2010 年起，他在瑞萨电子株式会社东京公司负责 90 至 28 纳米分离栅 MONOS 闪存技术及 IP 的开发，应用于汽车和消费/工业领域。此外，他还领导了与合作伙伴共同开发最先进 MCU 开发平台的联合项目。目前，他担任瑞萨电子株式会社共享研究与开发核心技术部门的副总裁，负责开发面向汽车和物联网/基础设施应用的先进核心技术，包括嵌入式非易失性存储器解决方案、CPU 核心及子系统、MCU 系统控制、安全、基于模型的设计解决方案等。

河野先生曾担任 IEEE 国际固态电路会议 (ISSCC) 国际技术程序委员会存储器分会成员，并在 2015 年至 2018 年期间主持了 ISSCC2015 和 ISSCC2017 的存储器分会场。