

2021 年 VLSI 技术与电路研讨会 技术要点

2021 年 VLSI 技术与电路研讨会是首个记录了微/纳米集成电子的步伐、进展和演变的国际会议，定于 2021 年 6 月 13 日至 19 日举行。两次研讨会将使用线上的形式举行，包括现场论文宣讲和预先录制的材料。

研讨会的主题“**改变生活方式的 VLSI 系统**”旨在将先进的技术开发、创新的电路设计以及它们所能实现的应用整合在一起，努力促进全球社会向智能、互联设备和系统所代表的新时代过渡，从而改变人类相互交流的方式。

以下是针对此主题的一些重点论文：

技术要点

先进的 CMOS 技术

论文 (T2-1) T0030

imec 的研究人员报道了叉片 N-和 PFET 与环栅纳米片 FET 的共集成情况。叉片短通道控制与低至 22nm 栅极长度（饱和亚阈值斜率=66-68mV/dec）的纳米片相当。通道释放后的湿法清洁优化可改善叉片的 I_{ON} 和 I_{OFF} 特性，这归因于栅极堆叠界面陷阱密度的降低。双功函数金属栅极在 17nm N-P 空间集成，突出了叉片用于 CMOS 面积缩放的关键优势。

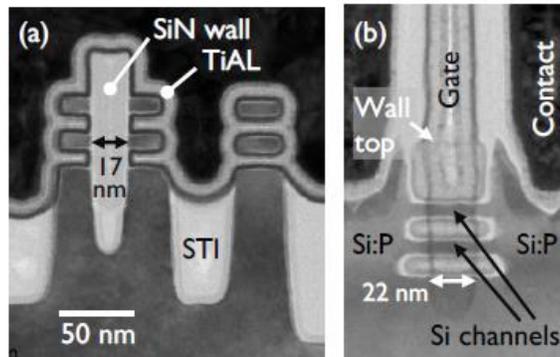


Fig. 5. End-of-process TEM images (a) across co-integrated fork- and nanosheet structures and (b) along a forksheet channel (NMOS).

Fig. 5. End-of-process TEM images (a) across co-integrated fork- and nanosheet structures and (b) along a forksheet channel (NMOS).

图 5.过程结束时的 TEM 图像(a)跨共集成的叉片和纳米片结构以及(b)沿着叉片子通道 (NMOS)。

“用于先进 CMOS 缩放的叉片 FET: 17nm N-P 区域的叉片纳米片集成和双功函数金属栅”, H. Mertens 等人, imec

论文 (T2-3) T0035

英特尔报告称, 二维材料沟道可实现 MOSFET 晶体管的最终缩放, 并将有助于摩尔定律数十年的延续。它们使用厚度小于 1nm 的单层 TMD (过渡金属二卤化物) 沟道, 结合可制造的 CVD、MBE 或晶种生长来展示 n 和 p-MOSFET 的状态。使用新型金属接触的转移 MBE MoS₂ 上的 NMOS 器件显示低离散性 (最低接触电阻(R_c) 0.4 kΩ·μm)、低磁滞和 77 mV/dec 的良好亚阈值斜率(SS)。使用 CVD WSe₂ 的 PMOS 器件的 SS 为 89 mV/dec, 在生长膜上的 PMOS 中表现最好, 但导通电流仍低于 NMOS。无转移、区域选择性的 WS₂ 晶体管的导通电流达到 10 μA/μm, 这是使用晶种生长法的 WS₂ 的最高报道值。

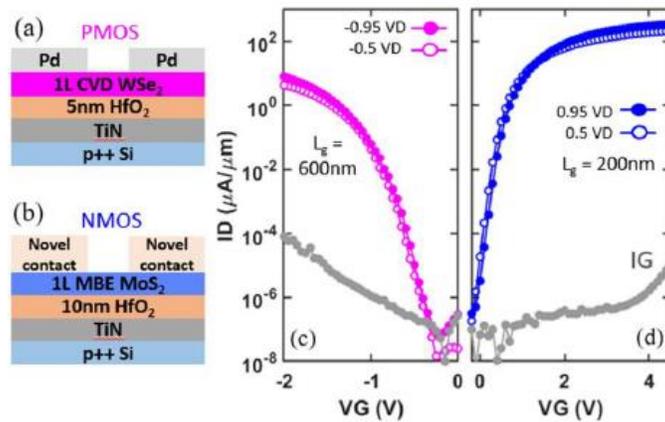


Fig. 6. Cross-section of PMOS (a) and NMOS (b) device. (c) Id-Vg characteristics of NMOS and PMOS transistors.

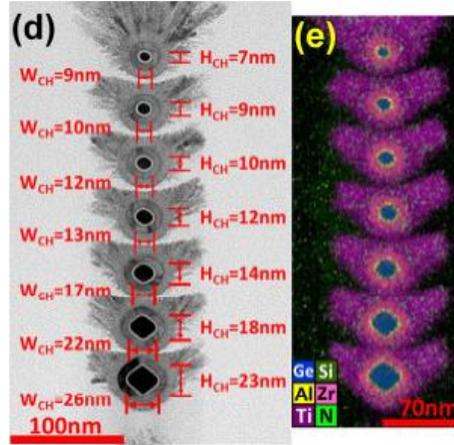
Fig. 6. Cross-section of PMOS (a) and NMOS (b) device. (c) Id-Vg characteristics of NMOS and PMOS transistors.

图 6. PMOS (a) 和 NMOS (b) 器件的横截面。(c) NMOS 和 PMOS 晶体管的 Id-Vg 特性。

“用于终极 CMOS 微缩的从生长到范德华接口工程的先进单层 2D NMOS 和 PMOS 晶体管集成”, C. J. Dorow 等人, 英特尔公司

论文 (T15-2) T0039

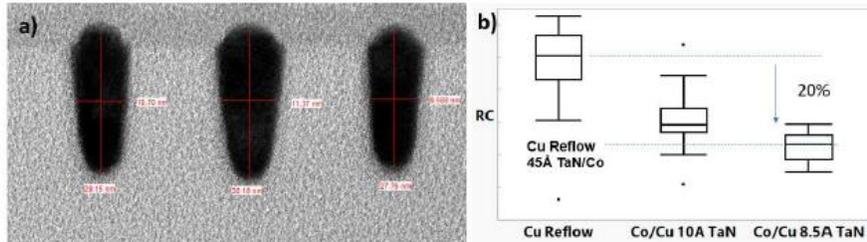
国立台湾大学的研究小组报告称, 通过 H₂O₂ 湿法刻蚀可以实现 8 层堆叠的 Ge_{0.75}Si_{0.25} 纳米片和 7 层堆叠的 Ge_{0.95}Si_{0.05} 纳米线。8 层堆叠的 Ge_{0.75}Si_{0.25} 具有很高的通道间均匀性。由于 L4 谷的输运有效质量(m_t)小, 态密度(m_{DOS})有效质量大, R_{S/D}/R_{total} 低, 七层堆叠的 Ge_{0.95}Si_{0.05} 表现出很高的性能。在报道的 Ge/GeSi 3D nFET 中, 在 V_{OV}=V_{DS}=0.5V 时达到 I_{ON}=110μA/堆栈 (每通道足迹为 4100μA/μm) 最高值, 在 V_{DS}=0.5V 时达到了高 G_{m,max}=340μS (13000μS/μm) 较高值。



“第一个高度堆积的 $Ge_{0.95}Si_{0.05}$ nGAAFET，通过湿法蚀刻在 $V_{OV}=V_{DS}=0.5V$ 时实现 $I_{ON}=110 \mu A$ ($4100 \mu A/\mu m$) 最高值，在 $V_{DS}=0.5V$ 时实现 $G_{m,max} = 340 \mu S$ ($13000 \mu S/\mu m$) 较高值”，Yi-Chun Liu 等，国立台湾大学

论文 (T5-2) T0107

IBM 的研究表明，对于低于 28 nm 间距的 BEOL 互连，双大马士革(DD)集成方案具有可扩展性。他们开发了两种全新的工艺流程：(1)扩展具有选择性沉积的 TaN 阻挡层的铜基大马士革互连，该阻挡层可降低通孔电阻，而不会损害可靠性，(2)创新的冶金双金属化方案，旨在提高细线和宽线的性能。这些工艺创新可显著改善通孔、信号和电源线的电阻。对于高性能计算(HPC)应用，具有高长宽比电源轨的双冶金技术可提供最佳性能。



Fig's. 6. (a) TEM image of 10nm wide Co interconnects (b) RC comparison between Co (10 \AA / 8.5 \AA TaN) and Cu with 4.5 nm TaN/Co.

Fig's. 6. (a) TEM image of 10nm wide Co interconnects (b) RC comparison between Co (10 \AA / 8.5 \AA TaN) and Cu with 4.5 nm TaN/Co.

图 6.(a) 10nm 宽 Co 互连线的 TEM 图像(b) Co (10 \AA / 8.5 \AA TaN)和具有 4.5 nm TaN/Co 的 Cu 之间的 RC 比较。

具有铜回流/选择性 TaN 和 Co/Cu 复合材料的双大马士革 BEOL 扩展性”，P. Bhosale 等人，IBM 研究

量子计算

论文 (JFS5-5) T0150

受埋线技术启发, AIST 首次提出了一种可实现高速/低离散性硅自旋量子比特操作的埋线纳米磁铁(BNM)。高速量子门操作源自于 BNM 在自旋量子位附近产生的大倾斜磁场, 以及归因于自对准制造工艺的低保真度偏差。利用基于 TCAD 的仿真, 它们表明 BNM 实现了比以前工作快 10 倍的 Rabi 振荡 (更快的自旋翻转), 并且在某些工艺变化下保真度大于 99%。而且, 所提出的 BNM 布置对于采用 2D 格状量子位布局的可纠错大规模量子计算机是可实现的。这项技术为实际的大规模硅量子计算机铺平了道路。

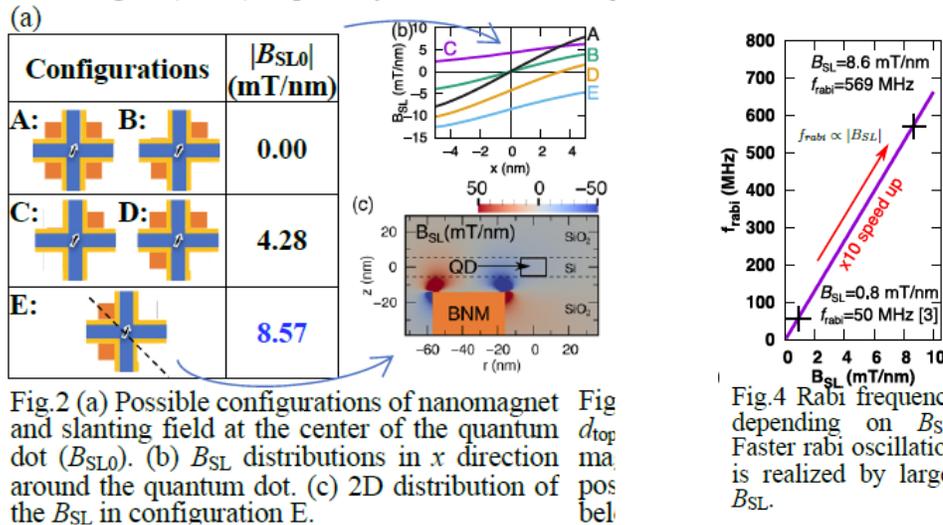


Fig. 2 (a) Possible configurations of nanomagnet and slanting field at the center of the quantum dot (B_{SL0}). (b) B_{SL} distributions in x direction around the quantum dot. (c) 2D distribution of the B_{SL} in configuration E.	图 2 (a) 纳米磁体和量子点(B_{SL0})中心斜场的可能配置。(b) B_{SL} 沿量子点在 x 方向上的分布。(c) 配置 E 中 B_{SL} 的 2D 分布。
Fig.4 Rabi frequency depending on B_{SL} . Faster rabi oscillation is realized by larger B_{SL} .	图 4 取决于 B_{SL} 的 rabi 频率。较大的 B_{SL} 实现了更快的 rabi 振荡。

“可实现高速/低离散性硅自旋量子位的埋入式纳米磁体: 可用于可纠错的大规模量子计算机中”, S. Iizuka 等人, AIST

人工智能/机器学习设备

论文 (T13-3) T0077

IBM 研究报告了 ARES——一种基于 14nm 相变存储器(PCM)的测试芯片, 包括多个交叉阵列单元, 每个阵列单元都可以在 512x512 唯一权重上进行并行乘累加(MAC)推理。大规模并行的 2D 网格以持续时间格式在芯片上传输深度神经网络

(DNN)激励，在阵列单元与集成的着陆垫(LP)之间传输，数字数据由此输入和输出芯片。为了进行精确的权重编程 (<3%的权重误差)，他们采用了一种逐行编程方案，以最小的超调量对每个模拟权重中的 4 个 PCM 器件进行有效编程。他们以接近软件的等效精度验证了两个 DNN，用完全片上 2 层网络展示了阵列单元到阵列单元的传输，在循环回到下一个片上 MAC 之前使用片外激活函数，并使用循环 LSTM（长期短期存储）网络测试了误差累积的恢复能力。

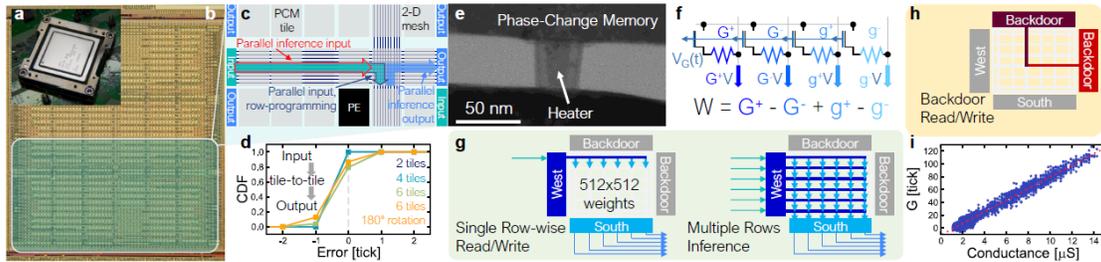


Fig. 1 (a) ARES chip, (b) micrograph and (c) components: input LP, output LP, PCM tiles, and duration transport across tiles using a 2-D parallel-signal mesh. (d) CDF of bit-errors in duration transport, for various travel distances. (e) TEM of a PCM device integrated in 14-nm back end. (f) Each MACC weight is encoded using 4 PCM devices. (g) Parallel read/write (left) is achieved by programming and then inferencing one row at a time, using the same circuits (and any non-idealities) encountered during full inference (right). (h) Single-device ("backdoor") read/write circuitry can measure device conductance in μS . (i) Correlation between measurements using (g) and (h).

Fig. 1 (a) ARES chip, (b) micrograph and (c) components: input LP, output LP, PCM tiles, and duration transport across tiles using a 2-D parallel-signal mesh. (d) CDF of bit-errors in duration transport, for various travel distances. (e) TEM of a PCM device integrated in 14-nm back end. (f) Each MACC weight is encoded using 4 PCM devices. (g) Parallel read/write (left) is achieved by programming and then inferencing one row at a time, using the same circuits (and any non-idealities) encountered during full inference (right). (h) Single-device ("backdoor") read/write circuitry can measure device conductance in μS . (i) Correlation between measurements using (g) and (h).

图1(a) ARES 芯片, (b)显微照片和(c)组件: 输入 LP, 输出 LP, PCM 阵列单元以及使用 2-D 并行信号网格跨阵列单元的持续时间传输。(d)各种传输距离的持续时间传输中的位错误 CDF。(e)集成在 14 nm 后端中的 PCM 器件的 TEM。(f)每个 MACC 权重使用 4 个 PCM 器件进行编码。(g)并行读取/写入(左)通过编程实现,然后在完整推理(右)中使用相同的电路(和任何非理想状况)一次计算一行。(h)单器件("后门")读取/写入电路可以测量器件电导(单位: μS)。(i)基于(g)和(h)测量数据之间获得的相关性。

“通过对基于 PCM 的权重进行精确的逐行编程并以持续时间格式进行并行矢量传输实现 14nm 的完全片上 MAC”，P. Narayanan 等人，IBM

3D 异构集成，非硅衬底/材料和设备

论文 (T2-4) T0011

普渡大学的研究人员首次展示了采用后端(BEOL)兼容的原子层沉积(ALD)工艺在鳍片结构上涂覆 In_2O_3 形成三维晶体管，并进而实现集成电路。通过通道厚度工程和沉积后退火技术，可以实现 $113 \text{ cm}^2/\text{V}\cdot\text{s}$ 的高迁移率和最大漏极电流达 $2.5 \text{ mA}/\mu\text{m}$ 最高值的高性能平面 In_2O_3 晶体管。展示了基于高性能 ALD In_2O_3 的零 VGS 负载反相器，最大电压增益为 38 V/V，最小电源电压(V_{DD})低至 0.5 V。受益于 ALD 的保形沉积能力，还实现了 ALD In_2O_3 三维鳍片晶体管。这些结果表明，ALD 氧化物半导体和器件具有独特的优势，并有望实现 BEOL 兼容的单片三维集成电路。

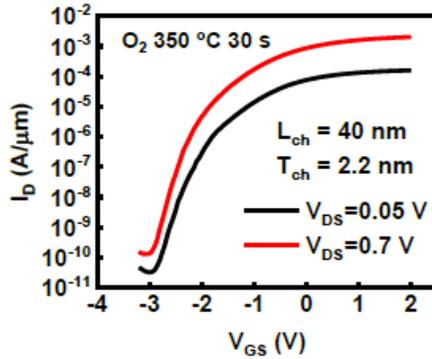


Fig. 10. I_D - V_{GS} characteristics of an In_2O_3 transistor with L_{ch} of 40 nm and T_{ch} of 2.2 nm with O_2 annealing at 350 °C.

Fig. 10. I_D - V_{GS} characteristics of an In_2O_3 transistor with L_{ch} of 40 nm and T_{ch} of 2.2 nm with O_2 annealing at 350 °C.

图 10. 350 °C O_2 退火时, L_{ch} 为 40 nm, T_{ch} 为 2.2 nm 的 In_2O_3 晶体管的 I_D - V_{GS} 特性。

“原子层沉积 BEOL 兼容的 In_2O_3 三维鳍片晶体管和集成电路的首次演示: In_2O_3 反相器具有 $113^2/\text{V}\cdot\text{s}$ 的高迁移率, $2.5 \text{ mA}/\mu\text{m}$ 的最大漏极电流和 38 V/V 的最大电压增益”, Mengwei Si 等人, 普渡大学

论文 (T2-5) T0142

索尼公司的一个研究小组展示了世界上第一个用于缓存的 40nm 嵌入式 STT-MRAM (自旋转矩转移磁随机存取存储器), 它与三维堆叠 CMOS 图像传感器 (CIS) 工艺兼容。他们优化了一种基于 CoFeB 的垂直磁性隧道结 (p-MTJ), 以抑制三维堆叠晶圆工艺造成的磁性能退化。通过改进工艺, 他们可以在 -30°C 的典型工作电压条件下实现 40ns 以下的高速写入操作, 在 105°C 的环境下可承受高达 $1\text{E}+10$ 个循环, 在缓存所需的 85°C 的条件下实现了 1s 的数据保留。此外, 为了拓宽嵌入式 MRAM (eMRAM) 的应用, 他们提出了一种全新的融合技术, 该技术将嵌入式非易失性存储器 (eNVM) 和缓存型的嵌入式 MRAM 集成在同一芯片中。他们使用融合技术实现了 $1 \text{ s} \sim 10$ 年的数据保留, 并具有足够的写入余量。

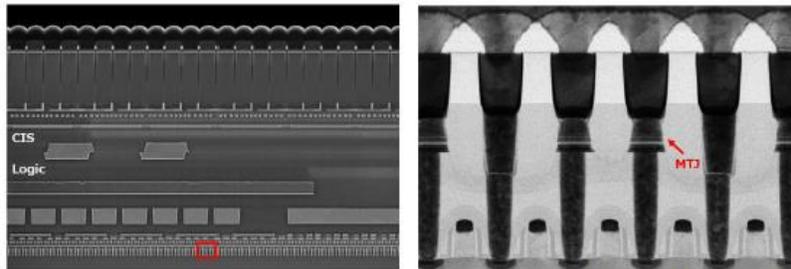


Fig. 1: Cross section image of eMRAM integrated into 3D stacked CIS (a) SEM image of 3D stacked CIS (b) TEM image of eMRAM arrays.

Fig. 1: Cross section image of eMRAM integrated into 3D stacked CIS (a) SEM image of 3D stacked CIS (b) TEM image of eMRAM arrays.

图 1: 集成到三维堆叠 CIS 中的 eMRAM 截面图像 (a) 三维堆叠 CIS 的 SEM 图像 (b) eMRAM 阵列的 TEM 图像。

“用于缓存的三维堆叠且 CIS 工艺兼容的 40nm 嵌入式 STT-MRAM”，M. Oka 等人，索尼半导体

论文 (JFS2-6) T0046

imec 的研究人员报告称，通过以下方式实现缩放硅通道 FinFET ($L_{gate} > 20nm$, 45nm 鳍片距) 的背面连接：采用低温(LT)、晶圆到晶圆(W2W)介电键合后，使得晶圆极度变薄 (STI 氧化物下硅薄膜的几种目标厚度：约 370nm 到约 20nm) 并且形成不同高度 (与晶圆变薄值有关) 的 W 填充纳米直通硅通孔(n-TSV)。该方案的目标是实现信号和电源网络的解耦，并通过将后者移到晶圆背面来预期实现 IRdrop 效应的降低。在此基础上，全面评估了三维工艺对器件特性的影响，显示：1) nmos 迁移率和驱动电流增强了 (高达 15%)；2) 对于 pmos, I_{ON} 损失小 (约 3% 至 10%)， R_{ext} 较大，通过 NBD (纳米束衍射) 评估了各种布局的通道应变；3) $\Delta V_T \sim 130mV$ ，这一变化可以通过在末端进行额外退火来恢复，同时保持较小的离散性和有效的匹配控制。没有观察到 BTI 退化，进一步表明最终退火选择有利于静电和可靠性的提高。

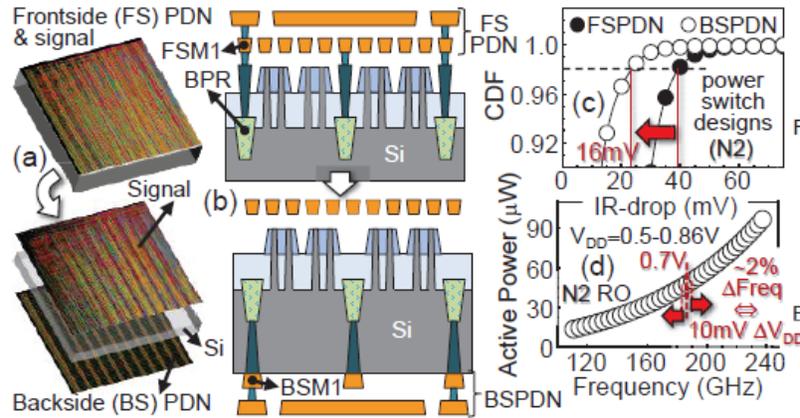


Fig.1 – a,b) Illustration of the concept to move power delivery networks (PDN) to the backside of thinned wafers using n-TSVs and BPR technology. Smaller IR-drop values are predicted for this configuration (c) leading to enhanced system performance (d).

Fig.1 - a,b) Illustration of the concept to move power delivery networks (PDN) to the backside of thinned wafers using n-TSVs and BPR technology. Smaller IR-drop values are predicted for this configuration (C) leading to enhanced system performance (d).

图 1- a,b) 使用 n-TSVs 和 BPR 技术将输电网络 (PDN) 移动到薄晶圆背面的概念说明。对于这种构造 (C)，预期较小的 IR 降会有助于系统性能的提高 (d)。

“通过 n-TSV 实现背面连接的逻辑及其作为扩展助推器的潜力”，A. Veloso 等人，imec

存储技术

论文 (T11-3) T0071

TSRI 首次报道用界面增强的合成反铁磁体 (SAF) 和改进的离子束刻蚀 (IBE) 技术实现了 CMOS 兼容并且 400°C 热稳定的 42 nm 垂直自旋轨道转矩磁隧道结 (p-SOT-MTJ) 器件，具有 130% 的隧道磁阻 (TMR) 比，并达到了 SAF 最高的 440°C 热稳定性。磁耦合 fcc 织构多层膜 (MCFTM) 缓冲层增强了 Co/Pt 多层膜 (ML) 和参考层之间

的 SAF 场(HSAF)和磁耦合。通过基于 W(3Å)的织构解耦扩散多壁垒(TDDMB)，可有效降低热应力过程中的 Pt-Fe 互扩散，从而获得磁场抗扰性。TaN/W 和 Ta/W 复合 SOT 通道突破了 β -W (<约 5 nm) 的厚度限制，并扩大了 MTJ 蚀刻窗口。TaN/W 通道具有~ -0.27 的较大有效自旋霍尔角。实现了具有自旋转移转矩(STT)辅助的确定性无场 SOT 写入。

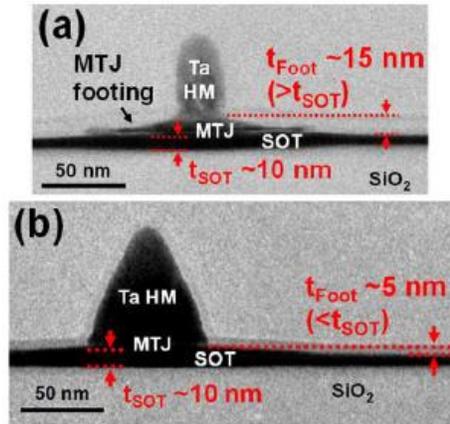


Fig. 10 TEM of p-SOT-MTJ after IBE (a) process I and (b) process II. Process II including 20° Ar etch (angle between out-of-plane axis) can reduce t_{Foot} less than t_{SOT} .

Fig. 10 TEM of p-SOT-MTJ after IBE (a) process I and (b) process II. Process II including 20° Ar etch (angle between out-of-plane axis) can reduce t_{Foot} less than t_{SOT} .

图 10 IBE 后 p-SOT-MTJ 的 TEM (a) 过程 I 和 (b) 过程 II。过程 II 包括 20° Ar 刻蚀（平面外轴之间的夹角）可以将 t_{Foot} 减小到小于 t_{SOT} 。

“首次展示通过界面增强 SAF，使 400°C 热稳定的 42 nm p-SOT-MTJ 单元具有 STT 辅助的无场开关和复合通道”，Ya-Jui Tsou 等人，TSRI

论文 (T6-3) T0052

TSMC 研究了用 SILC（应力诱导漏电流）测量方法分析不同电场应力下 HfZrO 铁电体的极化疲劳。在高场下，他们观察到极化唤醒和 SILC 增加之间有很强的相关性，这归因于氧空位的重新分布和渗流路径的形成，尤其是在高频循环下。相比之下，发现在低场发生的极化疲劳没有伴随 SILC 增加。通过 TEM/PED（进动电子衍射）进行的 P-E 环测量和材料分析表明，在低偏压下，电荷俘获是主要原因。他们提出，低场应力引起的疲劳可以通过在高场散布的周期性短期循环来有效地恢复，以管理电荷俘获和氧空位再分布，从而将循环耐久性延长至 $>1\text{E}12$ ，而不会在室温下发生 SILC 退化。

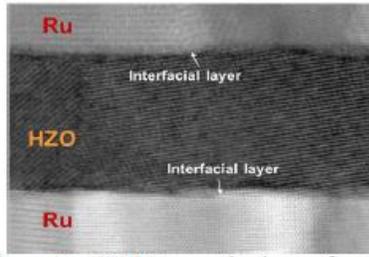


Fig. 6 TEM analysis of the Ru/HZO/Ru MFM capacitor shows a very thin interfacial layer (<1nm) formation between the 10nm HZO and metal electrodes at the top and bottom interface.

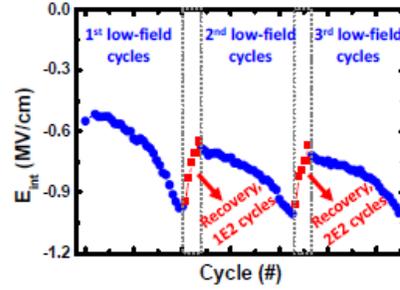


Fig. 17 Internal field E_{int} as a function of cycling shows a recoverable imprint due to de-trapping process by high-field recovery pulse after each low-field cycling loop.

<p>Fig. 6 TEM analysis of the Ru/HZO/Ru MFM capacitor shows a very thin interfacial layer (<1nm) formation between the 10nm HZO and metal electrodes at the top and bottom interface.</p>	<p>图 6 对 Ru/HZO/Ru MFM 电容器的 TEM 分析表明，在 10nm HZO 与顶部和底部界面的金属电极之间形成了非常薄的界面层(<1nm)。</p>
<p>Fig. 17 Internal field E_{int} as a function of cycling shows a recoverable imprint due to de-trapping process by high-field recovery pulse after each low-field cycling loop.</p>	<p>图 17 作为循环函数的内场 E_{int} 在每个低场循环回路之后，由于高场恢复脉冲的释放过程而显示出可恢复的印记。</p>

“HfZrO 铁电体的疲劳特性及其恢复行为”，P.J.Liao 等人，TSMC

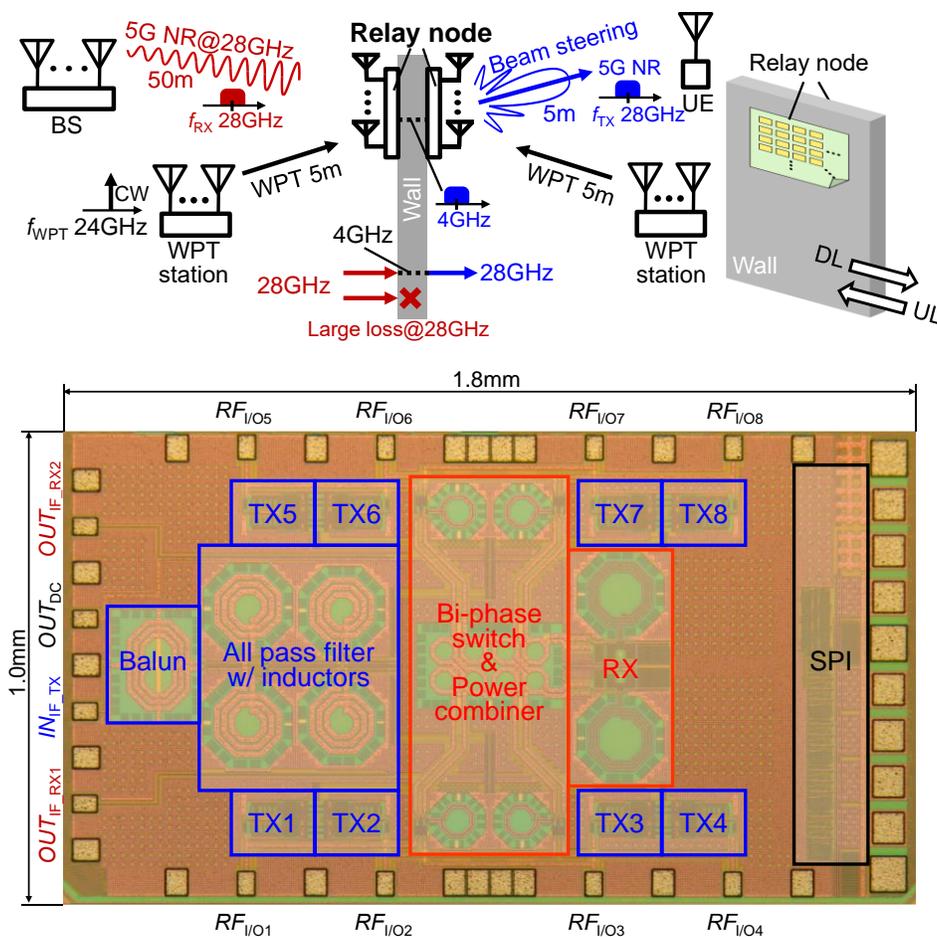
电路亮点

5G 收发器

论文(C11-1) 面向第五代移动通信的 28GHz 无电池中继收发机

面向 5G 应用的毫米波具有很强的方向性和较短的通信距离，限制了 5G 网络的服务范围。东京工业大学的研究人员提出了一个大规模的中继网络以扩大 5G 和 5G 以外系统的空间覆盖和容量。这篇文章提出了一种无线供电的 28 GHz 相控阵中继收发机，并没有使用有线电源的场景下展示了一个基于 5G 毫米波标准的 400 MHz 64QAM 调制的无线通信。

“使用具有 24 GHz 无线功率和 LO 传输的反向散射矢量和的 5G 网络 28 GHz 相控阵中继收发机”Michihiro Ide 等人，东京工业大学



[图1和5]使用 24 GHz 无线功率传输的 28 GHz 相控阵中继收发机

光子互连与计算

论文 JFS3-4: 基于硅光子微环调制器的 4 x 112 Gb/s WDM 发射器

来自英特尔的研究者提出了一种用于 400G 以太网模块和共封装光学器件的混合集成 $4\lambda \times 112 \text{ Gb/s}/\lambda$ 波分复用(WDM)发射器。光子 IC (PIC)包含一个带有集成加热器的微环调制器(MRM)阵列，用于高效 WDM。基于 28nm CMOS 工艺实现的电子 IC (EIC)具有带非线性 FFE 和控制电路的 PAM4 MRM 驱动器，以稳定 MRM 性能使其免受工艺和温度变化的影响。作者声称这是针对基于 O 波段环的 WDM 发射器所报告的最高每- λ 数据速率。

“基于硅光子微环调制器的 4 x 112 Gb/s O 波段 WDM

28nm CMOS 中基于环形光电流的热控制发射器”Jahnvi Sharma 等人，英特尔公司

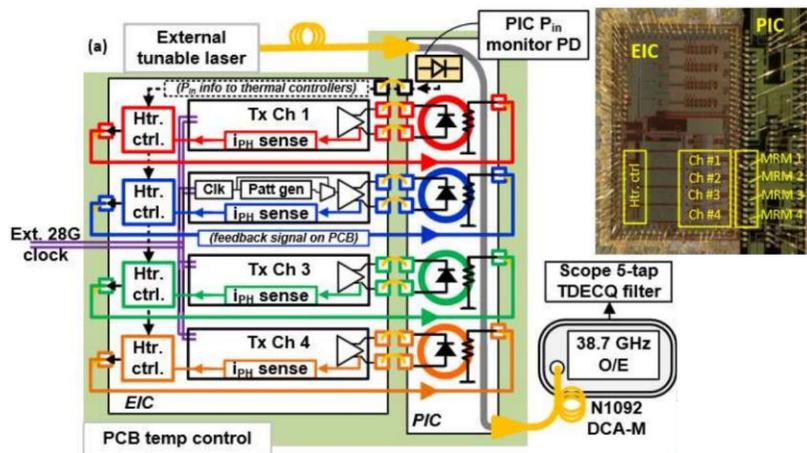
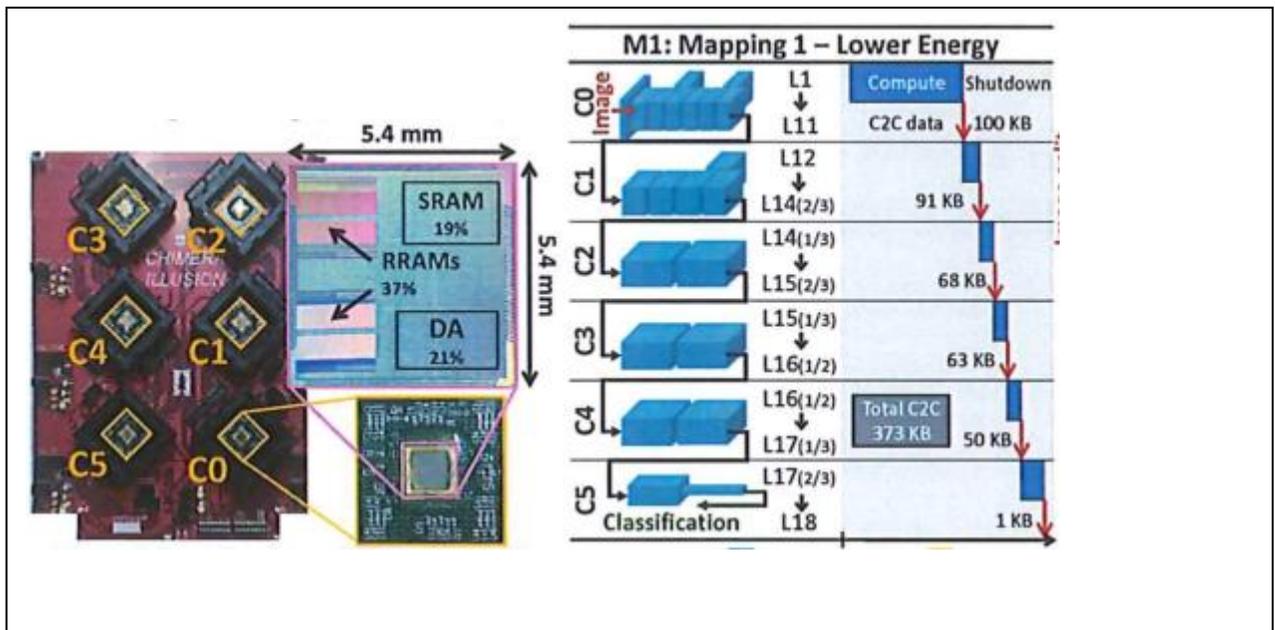


图1(a) O 波段 WDM Si-Ph TX 的架构和芯片照片。

人工智能，机器学习

论文(CFS1-2) 在多芯片可扩展处理器中边缘利用 RRAM 的深度神经网络加速器 斯坦福大学和 TSMC 的研究则会介绍了一款边缘 AI 加速器 CHIMERA。这个加速器以推理和增量学习为目标。它集成了一种非易失性电阻存储器(RRAM)，并通过具有芯片间链接的多芯片可扩展方法来利用其特性。作者将权重深度分布在六个芯片 (C0 至 C5) 上，然后将计算结果从一个芯片传播到下一个芯片。一旦每个芯片完成了它的计算部分，它就会断电。由于重量记忆的非易失性，唤醒时间仅为 33 μ s，在不使用时允许完全断电。作者还提出了一种将权重更新降至最低的训练算法，以解决 RRAM 磨损以及高写入能量和延迟问题。该芯片采用 40nm 工艺制造，可实现每个芯片 0.92TOPS，和 2.2TOPS/W 的能量效率。

“CHIMERA: 一个用于高效训练和推理的带有 2 兆字节的片上铸造电阻 RAM 的 0.92 TOPS、2.2 TOPS/W 边缘 AI 加速器”Massimo Giordano 等人，斯坦福大学和 TSMC

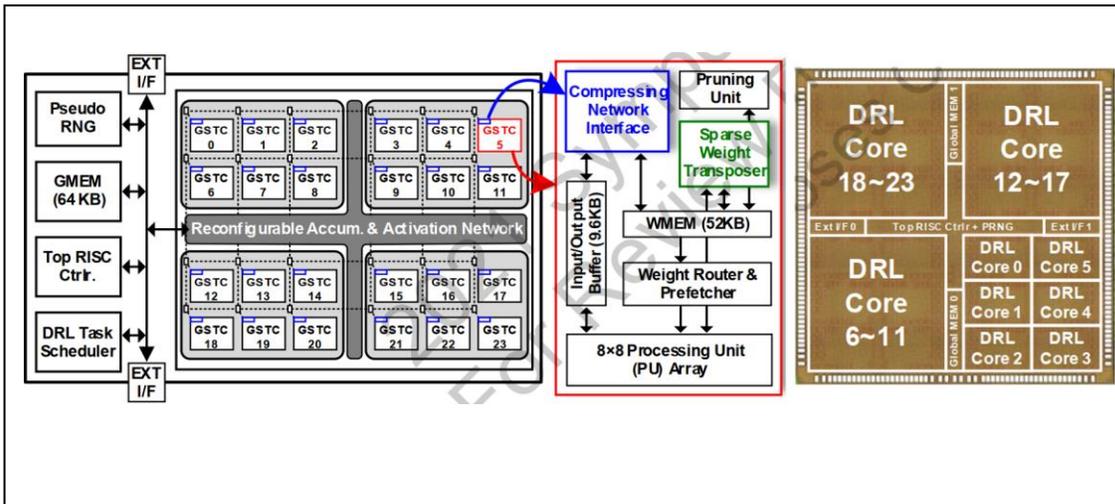


[图 14 和图 5 左。]DNN 推理应用的内核照片和示例性多芯片映射系统 PCB 概述

论文(CFS1-3) 用于加速训练的高性能节能深度强化学习(DRL)处理器

KAIST 的研究人员提出了一种高效的深度强化学习(DRL)处理器 OmniDRL，用于加速训练任务。OmniDRL 具有多达 24 组稀疏训练核(GSTC)，利用精简和基于块循环的权重分组，使训练速度提高 2 倍。此外，DRL 处理器在压缩网络接口(CNI)中采用指数均值增量编码(EMDE)，以提高指数压缩率 (1.6 倍) 并降低内存访问能力 (23.3%降幅)。最后，集成了一个稀疏的权重换位器(SWT)，以实现片上压缩权重的转置，以减少对外部存储器的访问。该处理器采用 28nm CMOS 设计，可实现 4.18TFLOPS 的峰值性能和 29.3TFLOPS/W 的峰值能量效率。

“OmniDRL: 具有双模式权重压缩和片上稀疏权重换位器的 29.3 TFLOPS/W 深度强化学习处理器, ”Juhyoung Lee 等人, KAIST



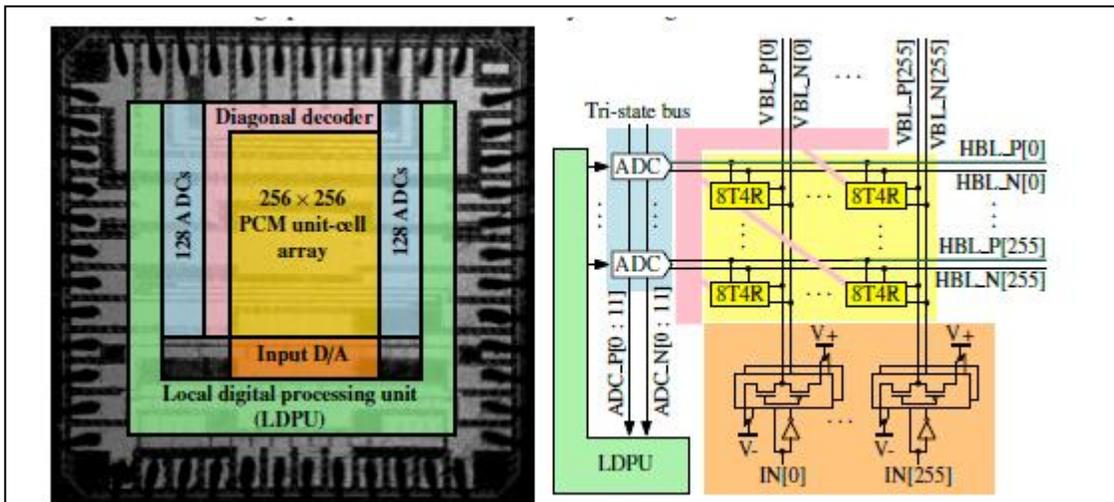
[图2 和图6。]具有 24 组稀疏训练核心(GSTC)、压缩网络接口(CNI)和稀疏权重换位器(SWT)模块的 DRL 处理器架构和芯片显微照片的概述

存算一体

论文(JFS2-5) 一种使用基于线性化 CCO 的 ADC 阵列和本地数字处理的基于 CMOS 和 PCM 的内存计算内核

IBM 展示采用一款基于 14nm CMOS 设计和制造的带后端集成多级相变存储器 (PCM) 的 256×256 内存计算内核。它包括 256 个紧凑型 $4\mu\text{m}$ 间距的基于线性化电流控制振荡器 (CCO) 的 ADC，以及一个执行仿射缩放和 ReLU 操作的本地数字处理单元。本文引入了一种新的 CCO 频率线性化技术，在 1 GHz 以上的频率工作时，可产生精确的片上矩阵矢量乘法 (MVM)。这篇文章给出了采用两个内核进行深度学习 (DL) 推理时在 MNIST 和 CIFAR-10 数据集上测得的分类精度。在 $1.59 \text{ TOPS}/\text{mm}^2$ 的性能密度下，测得的能量效率为 $10.5 \text{ TOPS}/\text{W}$ 。

“HERMES 内核 - 一种使用基于 $300\text{ps}/\text{LSB}$ 线性化 CCO 的 ADC 阵列和本地数字处理的基于 14nm CMOS 和 PCM 的内存计算内核，”R. Khaddam-Aljameh 等人，IBM Research Europe & IBM Research, Albany & IBM Systems and Technology & IBM T. J. Watson 研究中心



[图 1] 芯片照片和系统概述

存储技术

论文(C16-1) 在 5nm 高 K 金属栅极 FinFET CMOS 中的 16Kb 反熔丝一次性可编程存储器

TSMC 首次展示了在 5nm 高 K 金属栅 FinFET CMOS 中制造的 16Kb 一次性可编程 (OTP) 反熔丝存储器。采用自举高电压方案 (BHVS)、读端点检测 (REPD) 和伪差分检测 (PDS) 以在 5nm SoC 中进行现场编程时，可实现低于 1ppb 的固有误码率 (BER)，并在 125°C 时可保留 10 年的数据。

“具有自举高压方案、读取端点检测和伪差分传感的 5nm 高 K 金属栅极 FinFET CMOS 中的 16Kb 反熔丝一次性可编程存储器，”Shaun Chou 等人，TSMC。



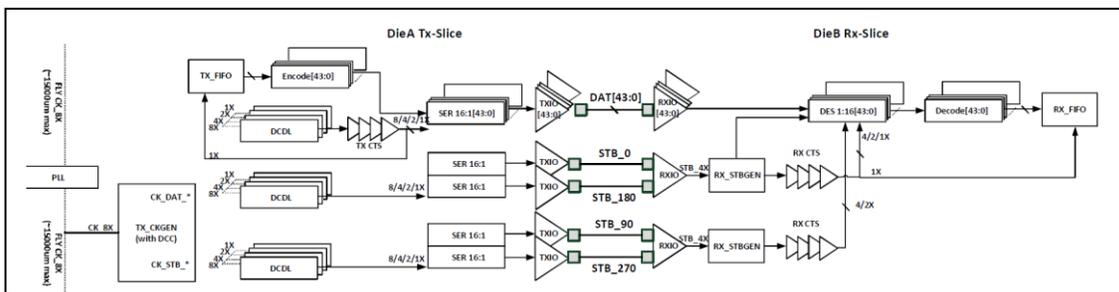
[图14] 芯片照片。

高级异构/3D 集成

论文(JFS1-3) 用于 7nm CMOS 技术中高级异构集成的高密度和低功耗芯片间通信:

这篇文章提出了一种用于超短距离(USR)芯片间通信的高密度低误码率和低功耗 Mlink (MediaTek 链路) PHY。所提出的 Mlink 采用 TSMC 7nm FinFET 1P15M CMOS 技术制作完成。互连通过台积电晶圆级封装(CoWoS)和台积电集成扇出(InFO)封装技术进行演示。Mlink PHY 采用了一系列高效节能的方案,包括无负载单端、基于 quarter rate strobe 和不平衡方案的收发机技术、最小固有自动对准和全新的抗噪编码方法。在 1 mm 超短距离平台下实现 20Gb/s/wire 和 0.46pJ/bit 的(目标 BER 为 1E-25)。带宽密度分别用 shoreline 5.31Tb/s/mm 和面积 2.25Tb/s/mm² 归一化。

“一个使用最小固有自动对准和抗噪编码、具有 BER 1E-25 芯片间链路的 7nm 0.46pJ/位 20Gbps, ”Ying-Yu Hsu 等人, 联发科公司



[图1] 单向一对 Tx/Rx 片的多链路 PHY 架构。

电源转换电路和时钟电路

论文(C3-1) 封装集成的 GaN 电源模块，用于服务器应用的 5 V 至 1 V DC – DC 转换器

英特尔提出了一种高频、高密度(9 A/mm^2)降压变换器，该变换器将一个低压 GaN 功率晶体管（FoM 比 Si 好 5-10 倍）与 CMOS 驱动器 IC 集成在一起，并实现了 $4 \text{ mm} \times 4 \text{ mm}$ 封装。在 3 MHz 的开关频率下，该转换器在 5 V 至 1 V 电压转换中实现了高达 94.2% 的峰值效率。

“一个具有封装集成式低压 GaN NMOS 功率晶体管的 32 A 5 V 输入、 94.2% 峰值效率的高频功率转换器模块，”Nachiket Desai 等人，英特尔公司

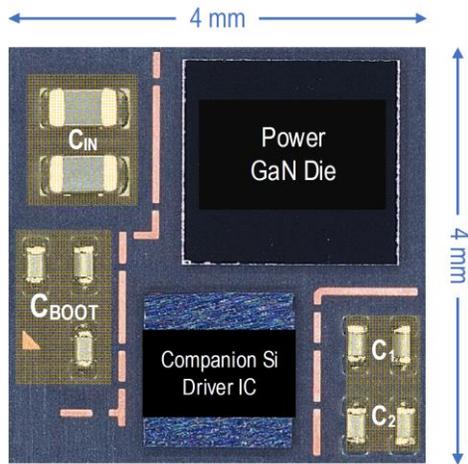
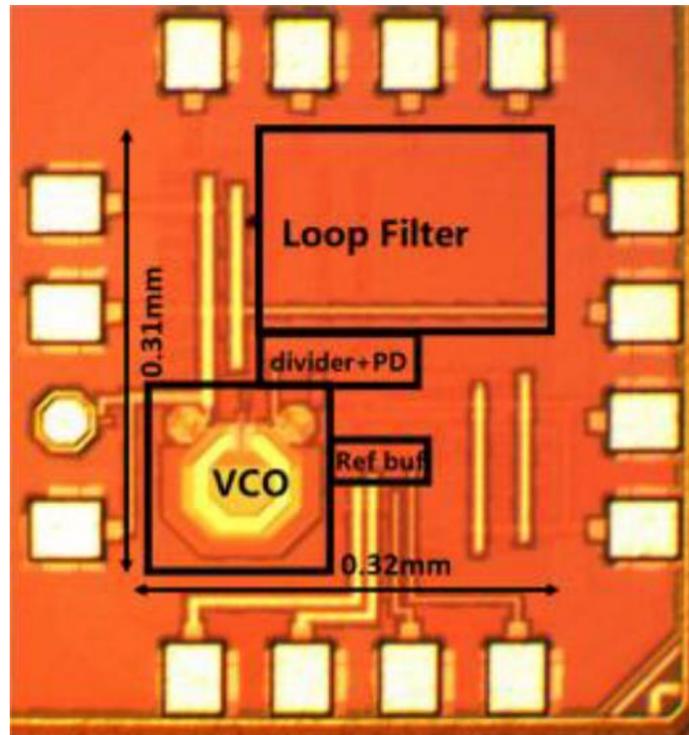


图 4.2 芯片封装集成 GaN 电源模块。

论文(C18-1) [C0165] 一个具有 20.3-fs 抖动的 19-GHz PLL

加州大学洛杉矶分校(UCLA)的研究人员将展示一种基于 28 nm CMOS 工艺设计的超低抖动 PLL。通过基于新型重定时的双边沿采样技术，这颗芯片在 250 MHz 的基准时钟和 12 mW 的功耗下实现了 20.3 fs 的抖动(rms)。

“具有 20.3 fs 抖动的 19 GHz PLL，”Yu Zhao 和 Behzad Razavi, UCLA



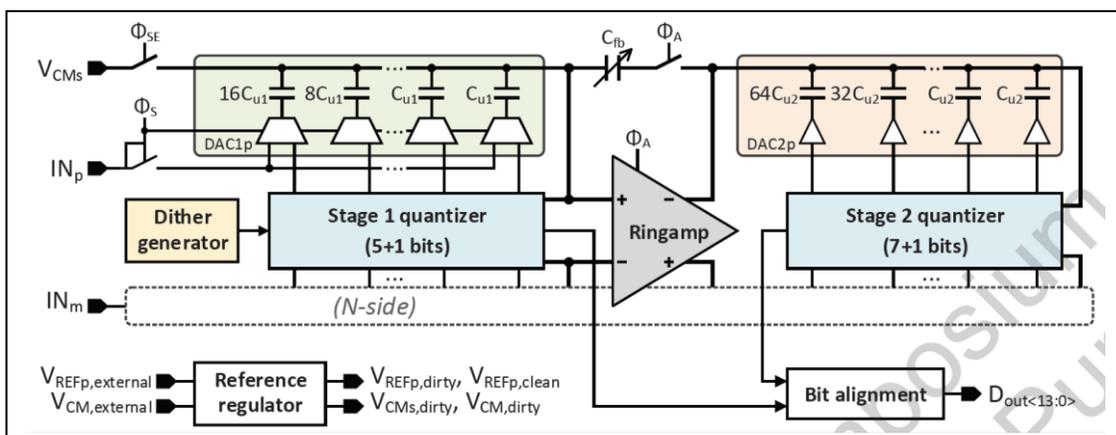
[图5].超低抖动PLL的芯片显微照片

传感器和数据转换器

论文(C15-1) 具有背景校准的全动态流水线 SAR ADC

全动态电路（如环形放大器、动态量化器和稳压器）促进了 ADC 最近的能效提升。今年，Imec 将推出一款单通道全动态流水线 SAR ADC，该 ADC 利用环形放大器，通过背景校准确保鲁棒性和优化性能，从而同时实现高带宽和高能效。本文提出了一种全新的动态量化器和窄带扰动注入，实现了对 DAC 失配、级间增益和环形放大器偏置最优性的快速、全面的背景校正。这个 ADC 还包括一个片上宽量程、全动态参考调节系统。ADC 在 500 MS/s 的功耗下为 3.3 mW，达到 10.0ENOB 和 75.5 dB SFDR，实现 6.2 fJ/c.s 的 Walden FoM。

“一个基于 10.0 ENOB，6.2 fJ/转换步长，500 MS/s 环形放大器的流水线 SAR ADC，16nm CMOS 中具有背景校准和动态参考调节功能，”J. Lagos 等人，imec

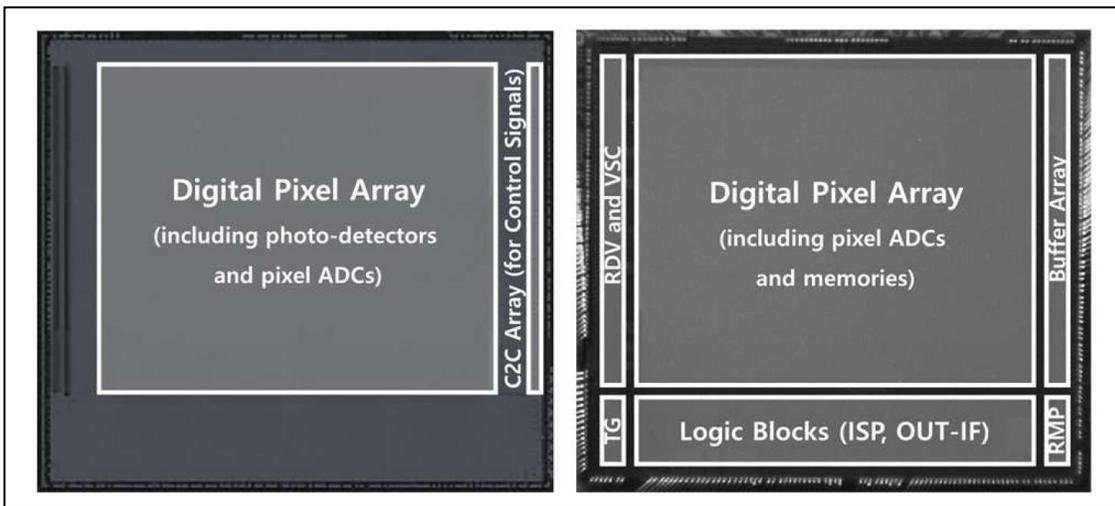


[图1] ADC 架构

论文(JFS4-4) 具有像素级 ADC 和 Cu 间集成的 CMOS 图像传感器

本文提出了一种先进的 CMOS 图像传感器，该传感器由基于 65 nm 工艺的集成传感器和使用 Cu 间(C2C)互连器的 28 nm 工艺集成逻辑芯片实现。它显示了 2.6 e-rms 的低随机噪声，116.2 mW 的低功耗（在视频速率下），以及高达 960 fps 的 200 万像素高速全局快门式 CMOS 图像传感器(CIS)并采用先进的 DRAM 技术。为了实现高性能的全局快门 CIS，他们为数字像素传感器提出了一种全新的架构，该架构包含采用像素级 ADC 和数字存储器的出色全局快门操作 CIS。每个像素具有两个小间距 Cu 间互连器用于晶圆级堆叠，每个单位像素的间距小于 5 μm ，这是世界上同时嵌入像素级 ADC 和 22 位存储器的最小像素。

“采用像素级 ADC 和像素内存的 2.6 e-rms 低随机噪声、116.2 mW 低功率 200 万像素全局快门 CMOS 图像传感器，”Min-Woong Seo 等人，三星电子

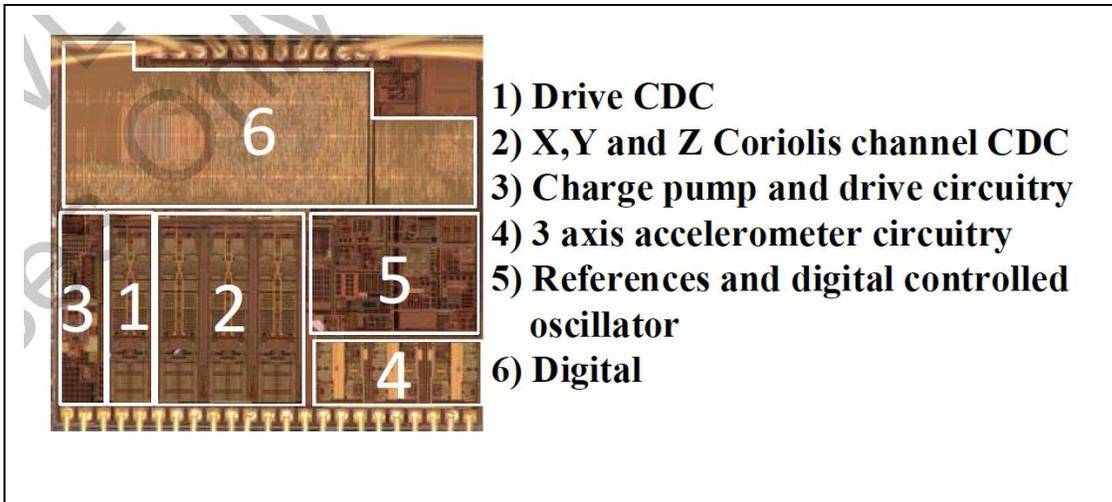


[图5] 芯片显微照片（左：顶部芯片，右：底部芯片）

论文(C19-1) 宽量程陀螺仪前端电路

Robert Bosch & Bosch Sensortec 提出了一种陀螺仪前端电路，用于感应、放大和数字化来自三轴电容感应型陀螺仪传感器设备的信号。这篇文章的显著特点是其高达 $\pm 8000 \text{ deg/s}$ 的宽输入范围，是现有技术的 4 倍。该芯片在其他方面也表现出相当的性能，例如 $0.0047 \text{ deg/s}/\sqrt{\text{Hz}}$ 的噪底。

“直接数字化开环陀螺仪前端具有 $\pm 8000^\circ/\text{s}$ 的满量程范围和 $0.0047^\circ/\text{s}/\sqrt{\text{Hz}}$ 的本底噪声，”Chinwuba Ezekwe 等人，Robert Bosch & Bosch Sensortec



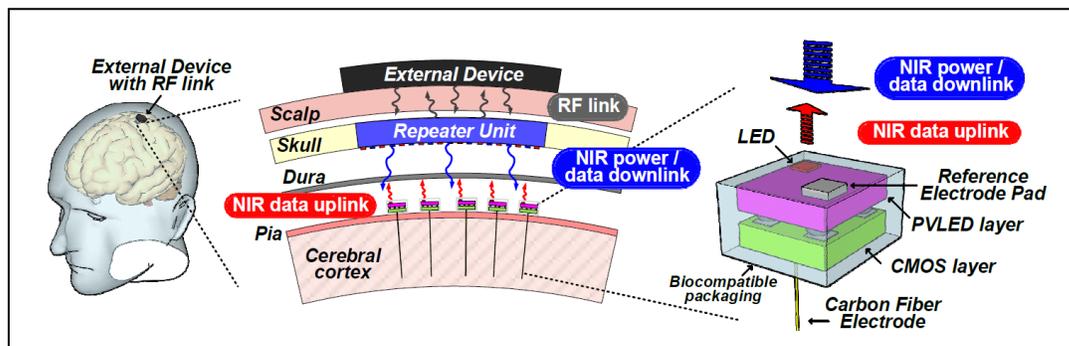
[图5] 芯片显微照片(1.94 mm × 1.8 mm)

生物医学电路

论文(C2-2) 用于无线超小型大脑传感器的神经记录芯片

本文提出了一种用于近红外超小型大脑传感设备的电源管理、神经传感和光信号传输芯片。本文的主要特点是提高了芯片的耐光性。作者实施了一种电路设计方法，即使 CMOS 芯片受到来自封装的意外光线照射，也可以保持其功能。作者实现了高达 $300 \mu\text{W}/\text{mm}^2$ 曝光的足够耐光度（高于组织极限），并在 38°C 时消耗 $0.57\mu\text{W}$ ，使之在独立微粒中功耗最低的同时集成了片上特征提取和独立增益控制。

“一种用于近红外自由浮动微粒的耐光神经记录IC，”Jongyup Lim 等人，密歇根大学、苏黎世联邦理工学院和特拉华大学



[图1] 近红外无线神经记录微粒概念图