美光推出全球首款高容量 256GB LPDRAM SOCAMM2，为数据中心基础架构树立新标杆

March 5, 2026 at 10:00 AM CST

新闻亮点：

- 采用业界首款单晶粒 32Gb LPDDR5X，**相较标准 RDIMM，功耗降至其 1/3，尺寸亦缩小至其 1/3**
- 长上下文 LLM 推理的**首个 token 生成时间加速 2.3 倍**，在独立 CPU 应用中，**每瓦性能提升 3 倍**
- **单一模组容量提升 1.33 倍** — 每颗 8 通道服务器 CPU 可配置 2TB LPDRAM，适用于 AI 及高性能计算（HPC）场景

2026 年 3 月 5 日，爱达荷州博伊西市 — 美光科技股份有限公司（纳斯达克股票代码：MU）近日宣布开始向客户送样业界容量领先的 LPDRAM 模块 256GB SOCAMM2，进一步巩固其在低功耗服务器内存领域的领导地位。依托业界首款单晶粒 32Gb LPDDR5X 设计，这一里程碑式成就为 AI 数据中心带来变革性突破，提供足以实现全新系统架构的低功耗内存容量。

AI 训练、推理、代理式 AI 和通用计算的融合，正推动更严苛的内存需求，并重塑数据中心的系统架构。现代 AI 工作负载催生了大模型参数、扩展的上下文窗口及持久性键值（KV）缓存的需求，而核心计算则在数据强度、并发性和内存空间方面持续扩展。

面对上述工作负载，内存容量、带宽效率、延迟和能效已成为系统层面的主要瓶颈，直接影响性能、可扩展性和总体拥有成本。LPDRAM 融合上述特性的独特优势，在功耗与散热限制日益严苛的数据中心环境中，成为 AI 及核心计算服务器的关键解决方案。美光正与 NVIDIA 携手合作，共同设计高性能内存解决方案，以满足先进 AI 基础架构的需求。

美光高级副总裁暨云端存储事业部总经理 Raj Narasimhan 表示：“美光 256GB SOCAMM2 为 AI 及高性能计算（HPC）提供更具能效的 CPU 附加内存解决方案。此次产品发布充分展现出美光在技术与封装领域的突破，打造业界容量领先、低功耗、小尺寸的模块化内存解决方案。美光在数据中心低功耗内存解决方案领域持续保持领先地位，这一独特优势使我们率先推出单晶粒 32Gb LPDRAM，协助推动业界加速采用更节能、更高容量的系统架构。”

专为容量、能效和工作负载性能优化而设计
美光的 256GB SOCAMM2 为各种 AI 和通用计算工作负载提供更高的内存容量、更低的功耗，以及更快的性能。

- **为 AI 服务器扩展内存容量：**256GB SOCAMM2 容量较前代最高规格 192GB SOCAMM2 提升三分之一，可为每颗 8 通道 CPU 提供 2TB LPDRAM 容量，从而支持更大的上下文窗口及更复杂的推理工作负载。
- **功耗更低、尺寸更小：**与相同容量的 RDIMM 相比，SOCAMM2 的功耗仅为其三分之一，尺寸亦缩减至三分之一，有效提升机架密度并降低总体拥有成本。[1]
- **提升推理与核心计算性能：**在统一内存架构中，与现有解决方案相比，256GB SOCAMM2 用于 KV 缓存卸载时，可将长上下文、实时 LLM 推理的首个 token 生成时间加速 2.3 倍。[2]在独立 CPU 应用中，针对高性能计算工作负载，LPDRAM 的每瓦性能较主流内存模块提升超 3 倍。[3]
- **易维护、可扩展的模块化设计：**模块化 SOCAMM2 设计可提升设备可维护性、支持液冷服务器架构，并能随着 AI 与核心计算内存需求的持续增长，实现未来容量扩充。

NVIDIA 数据中心 CPU 产品部门主管 Ian Finder 表示：“先进 AI 基础架构需要在各个层面进行极致优化，才能有效应对严苛的 AI 推理工作负载对性能与能效的需求。美光通过 256GB SOCAMM2，以低于传统服务器内存的功耗，实现超大内存容量与带宽的突破，为下一代 AI CPU 提供关键助力。”

推动行业标准制定，加速低功耗内存普及
美光在 JEDEC SOCAMM2 规范制定过程中持续发挥领导作用，并维持与系统设计人员的深度技术合作，以推动下一代数据中心平台在能效与性能方面实现全行业性提升。

美光现已面向客户送样 256GB SOCAMM2 产品，并提供业界最全面的数据中心 LPDRAM 产品组合，涵盖 8GB 至 64GB 组件及 48GB 至 256GB 的 SOCAMM2 模块。

更多资源：

- 美光白皮书：LPDDR at Scale: Enabling Efficient LLM Inference Through High-Capacity Memory
- 美光博客：每一瓦特都很重要：低功耗内存如何改变数据中心
- SOCAMM2 产品网页
- 数据中心内存产品网页

关于 Micron Technology Inc.（美光科技股份有限公司）

美光科技是创新内存和存储解决方案的业界领导厂商，通过改变世界使用信息的方式来丰富全人类生活。我们始终以客户为中心，专注引领技术创新，追求卓越制造与运营，向客户交付丰富的高性能内存和存储产品组合——包括 DRAM、NAND 及 NOR。美光团队打造的创新产品，每一天都助力数据经济的发展，推动人工智能（AI）和计算密集型应用的突破，释放从数据中心到本地智能设备的无限机遇。如需了解 Micron Technology Inc.（美光科技股份有限公司，纳斯达克股票代码：MU）的更多信息，请访问 micron.cn

美光媒体联络人
高诚公关
潘平 / 美光服务团队
电话：+86 188 8388 2632
E-mail：ppan@golin.com

---

[1] 三分之一的功耗依据单个 128GB、128 位总线宽度 SOCAMM2 模块与两个 64GB、64 位总线宽度 DDR5 RDIMM 的功耗瓦数对比计算。三分之一的尺寸依据 SOCAMM2 的面积（ 14x90 mm）与标准服务器 RDIMM 的面积之比。

[2] 结果基于美光内部测试，使用 Llama3 70B 模型（FP16 量化）进行实时推理测试，测试配置为：上下文长度 500K，并发用户数 16。首 token 响应时延（TTFT）的预期提升，基于每 CPU 配置 2TB LPDRAM 时延 0.12 秒，对比每 CPU 配置 1.5TB LPDRAM 时延 0.28 秒测算。有关测试条件详情，请参阅本月稍早发布的白皮书：LPDDR at Scale: Enabling Efficient LLM Inference Through High-Capacity Memory。

[3] 美光内部测试使用相同容量的 LPDDR5X 和 DDR5 进行 Pot3D 太阳物理 HPC 代码性能测评。