

SUPRO-SPEC-001

HMGO SuPro 1 产品规格书

五款超级计算产品。一套架构，从兼容迁移到下一代预览。

HMGO SILICON · SUPRO PROGRAM

Revision 1.0 · Public Concept Edition

Portfolio · Specifications · Platforms · Qualification

· 2026-07

DOCUMENT MAP

文档导航

00

阅读路径

本文件把可追溯的概念参数、架构假设和工程边界放在同一条叙事中。数字用于建立产品层级与系统设计空间，不构成性能保证、销售承诺或第三方比较结论。

01	系列总览	产品命名、共性能力与选择逻辑
02	SuPro 1 Lite	AI2 Ultra 等级的迁移入口
03	SuPro 1	标准超算节点
04	SuPro 1 Pro	服务器、训练与持续高负载
05	SuPro 1 Ultra	800 倍投影的极限封装
06	SuPro 2 Preview	下一代技术预览
07	平台与合规	主机、机架、供电、冷却与资格验证

概念声明：HMGO SuPro 1 为原创概念芯片系列。所有制程、性能、功耗、接口和软件名称均为设计设定。Apple 与 M5 Max 为其各自权利人的商标；本文的 800 倍表述仅对应 HMGO 定义的内部 FP4 合成投影。

PORTFOLIO OVERVIEW

系列总览

01

SuPro 1 系列不是同一颗芯片的简单频率分档，而是围绕不同通信半径、内存容量、可靠性目标和冷却边界形成的五种计算封装。

61.44 POPS

Lite FP4

12.8 EFLOPS

Ultra FP4

48 TB

Ultra HBM4E

65,536

单域节点

命名与产品角色

型号	角色	一句话定义	生命周期
SuPro 1 Lite	Compatibility	把 AI2 Ultra 级能力带入 SuPro 系统	标准支持 5 年
SuPro 1	Standard	面向实验室与企业的标准超级计算芯片	标准支持 7 年
SuPro 1 Pro	Professional	大型服务器、模型训练与关键业务	长期支持 10 年
SuPro 1 Ultra	Frontier	单封装极限和超大规模机群	精选合作 7 年
SuPro 2 Preview	Early Access	下一代软件与系统的前置验证	滚动预览 18 个月

所有型号共享

- HM-Ark V6 64 位执行模型、Aurora CPU、Matrix Ocean、Nebula GPU 和 Orion DPU。
- HM-Mesh X1 集合通信、统一遥测、链路加密与拓扑感知调度。
- Secure Supercompute Enclave、内存加密、度量启动和远程证明。
- HM-Forge SDK、HMCC 编译器、OceanDNN、MeshCollective 和 RackScope。

选型建议：先按模型或网格是否能驻留封装内存选择容量，再按通信比例选择链路，最后按作业 SLA 和恢复目标选择 RAS 等级。不要仅按峰值算力选型。

PRODUCT DEFINITION

HMGO SuPro 1 Lite

02

与 HMGO AI2 Ultra 处于同一处理能力等级，但拥有完整的服务器内存、互联与可靠性能力。

兼容迁移型 · 单芯粒组合 · 空冷或液冷

项目	概念规格
计算峰值	61.44 POPS FP4 · 7.68 POPS FP8 · 0.48 PFLOPS FP64
计算配置	96 Aurora cores · 8 Matrix Ocean clusters · 32 Nebula partitions
内存	2 TB HBM4 · 8 TB/s · 32 GB stacked SRAM
互联	4 × HM-Link X1 · 3.2 Tb/s aggregate scale-out
I/O	PCIe 7.0 ×32 · CXL 4.0 ×16 · 2 × 800G Ethernet
功率	450 W sustained · 600 W boost · 30-45°C coolant
RAS	ECC · link retry · package checkpoint · secure boot
封装	110 × 110 mm module · 2.6 kg with cold plate

目标工作负载

- 将 AI2 Ultra 工作站代码迁移到 HM-Forge 超算软件栈。
- 33B-400B 量化模型推理、微调和向量检索服务。
- 小规模 CAE、基因分析、遥感预处理和教学集群。

设计边界：处理能力相当指 FP4 峰值等级，不保证所有 AI2 Ultra 应用等速；内存层级、时钟与软件调度均不同。

PRODUCT DEFINITION

HMGO SuPro 1

03

面向研究机构、企业 AI 平台和部门级超级计算的标准配置。

标准型 · 双计算芯粒 · 液冷优先

项目	概念规格
计算峰值	0.42 EFLOPS FP4 · 52.5 POPS FP8 · 3.2 PFLOPS FP64
计算配置	192 Aurora cores · 32 Matrix Ocean clusters · 96 Nebula partitions
内存	4 TB HBM4E · 18 TB/s · 96 GB stacked SRAM
互联	8 × HM-Link X1 · 12.8 Tb/s aggregate scale-out
I/O	PCIe 7.0 ×64 · CXL 4.0 ×32 · 4 × 800G Ethernet
功率	700 W sustained · 950 W boost · direct liquid cooling
RAS	Lite 全部能力 + 芯粒迁移 + 内存页隔离
封装	140 × 140 mm module · 4.8 kg with cold plate

目标工作负载

- 通用大模型推理、7B-400B 训练与 1T 级混合专家模型研究。
- FP64 数值模拟、渲染、材料计算与气象区域模型。
- 8-72 节点实验室集群或部门级私有算力池。

设计边界：0.42 EFLOPS 是 FP4 稠密矩阵理论峰值；FP64、稀疏和端到端工作负载具有不同吞吐。

PRODUCT DEFINITION

HMGO SuPro 1 Pro

04

为大型服务器、模型训练和持续数周的关键任务提供更大的内存、更强的通信与完整 RAS。

专业型 · 六计算芯粒 · 双故障域

项目	概念规格
计算峰值	2.8 EFLOPS FP4 · 350 POPS FP8 · 18 PFLOPS FP64
计算配置	512 Aurora cores · 128 Matrix Ocean clusters · 256 Nebula partitions
内存	12 TB HBM4E · 64 TB/s · 512 GB stacked SRAM
互联	16 × HM-Link X1 · 51.2 Tb/s aggregate scale-out
I/O	PCIe 7.0 ×96 · CXL 4.0 ×64 · 8 × 1.6T Ethernet
功率	1.8 kW sustained · 2.2 kW boost · dual-loop liquid cooling
RAS	双控制域 · SDC 哨兵 · 细粒度检查点 · 热备 DPU
封装	2 × compute baseboards · 11.5 kg service unit

目标工作负载

- 400B-10T 参数模型预训练、后训练与多租户推理。
- 大型数据库加速、图计算、风险模拟和数字孪生。
- 72-4,608 节点生产集群与关键 SLA 服务。

设计边界：Pro 需要双路电源、受监控液冷和 HMGO 认证基板；不支持桌面 PCIe 插卡形态。

PRODUCT DEFINITION

HMGO SuPro 1 Ultra

05

世界最强大的概念芯片：在 HMGO 定义的 FP4 Transformer 投影中，相较单个 Apple M5 Max 系统最高 800 倍。

极限型 · 多基板融合 · 光电共封装

项目	概念规格
计算峰值	12.8 EFLOPS FP4 · 1.6 EFLOPS FP8 · 64 PFLOPS FP64
计算配置	1,536 Aurora cores · 512 Matrix Ocean clusters · 768 Nebula partitions
内存	48 TB HBM4E · 192 TB/s · 2 TB stacked SRAM
互联	32 × HM-Link X1 · 102.4 Tb/s aggregate scale-out
I/O	PCIe 7.0 ×128 · CXL 4.0 ×128 · 16 × 1.6T Ethernet
功率	6 kW sustained · 7.5 kW boost · 2-phase direct liquid cooling
RAS	三重控制域 · 光链路旁路 · predictive evacuation · N+2 DPU
封装	8 compute tiles + 4 I/O tiles · 28 kg service unit

目标工作负载

- 10T-100T 参数模型训练、极长上下文推理和大规模专家并行。
- 全球气候、聚变、分子动力学和高分辨率地球数字孪生。
- 4,608-65,536 节点前沿超级计算中心。

设计边界：“800 倍”是跨类别概念投影，不是第三方实测；Ultra 的 6 kW 持续功率与 M5 Max 的整机功耗边界完全不同。

PRODUCT DEFINITION

HMGO SuPro 2 Preview

06

在 SuPro 2 正式系列发布前，让软件、网络、供电和冷却团队提前验证下一代接口。

早期预览型 · HM-Ark V7 Preview · HBM5-P

项目	概念规格
计算峰值	20.0 EFLOPS FP4 · 2.8 EFLOPS FP8 · 96 PFLOPS FP64
计算配置	1,792 Aurora-V7 preview cores · 640 Matrix Ocean Gen5 clusters
内存	64 TB HBM5-P · 256 TB/s · 4 TB stacked SRAM
互联	32 × HM-Link X2 preview · 204.8 Tb/s aggregate scale-out
I/O	PCIe 8.0 preview · CXL 5.0 preview · optical memory ports
功率	5.5 kW target · adaptive voltage islands
RAS	Ultra 基线 + self-healing routing preview
封装	non-production evaluation module

目标工作负载

- 编译器、框架、数值库和通信库的 SuPro 2 前置适配。
- HBM5-P 数据布局、光内存和 HM-Link X2 拓扑研究。
- 下一代数据中心电力密度、冷却回路与运维流程验证。

设计边界：Preview 不提供生产 SLA、长期 ABI 稳定或最终频率承诺；固件更新可能要求重新认证。

PLATFORM REQUIREMENTS

平台与合规

07

SuPro 系统从供电、冷却、基板、固件到软件版本形成一条资格链。任何一环脱离认证范围都会改变可用功率和可靠性。

7.1 参考平台

平台	支持型号	节点数	机架功率	冷却
SuPro Node 2	Lite / Standard	2	2.4 kW	空冷或液冷
SuPro Tray 8	全系列	8	6-52 kW	直接液冷
SuPro Rack 72	Standard / Pro / Ultra	72	68-480 kW	双回路液冷
SuPro Pod 4608	Pro / Ultra	4,608	4.5-31 MW	设施级热回收

7.2 环境要求

- 冷却入口 30-45°C；水质、电导率、颗粒物和腐蚀抑制剂按 HMGO CL-4 规范。
- 双路 415 V 三相供电；托盘级隔离与机架级直流母线支持 20 ms ride-through。
- 每个机架至少两条独立管理网络和两条独立 HM-Link 光平面。
- 地板承重、软管弯曲半径、光纤清洁与漏液检测必须进入站点验收清单。

7.3 资格验证阶段

- EVT：验证电气、信号、冷却和启动路径。
- DVT：验证持续负载、故障注入、性能稳定与安全边界。
- SVT：验证整机架拓扑、服务流程、固件回滚与远程证明。
- Site Acceptance：以客户设施中的 24-72 小时混合负载完成最终验收。

所有型号、数量、性能与供应计划均为概念设定。正式产品若存在，将以最终订单、合规和支持文件为准。