



服务器风冷 实践 & 思考



ZOMI

Content

- 服务器风冷散热效率如何，是否能满足高性能服务器的需求？ AI 场景下多挂 8 张 NPU/GPU，功耗可不是逆上天，赶紧放海底 ~~~



Content github.com/Infrasys-AI/AIInfra

AI 系统 + 大模型全栈架构图

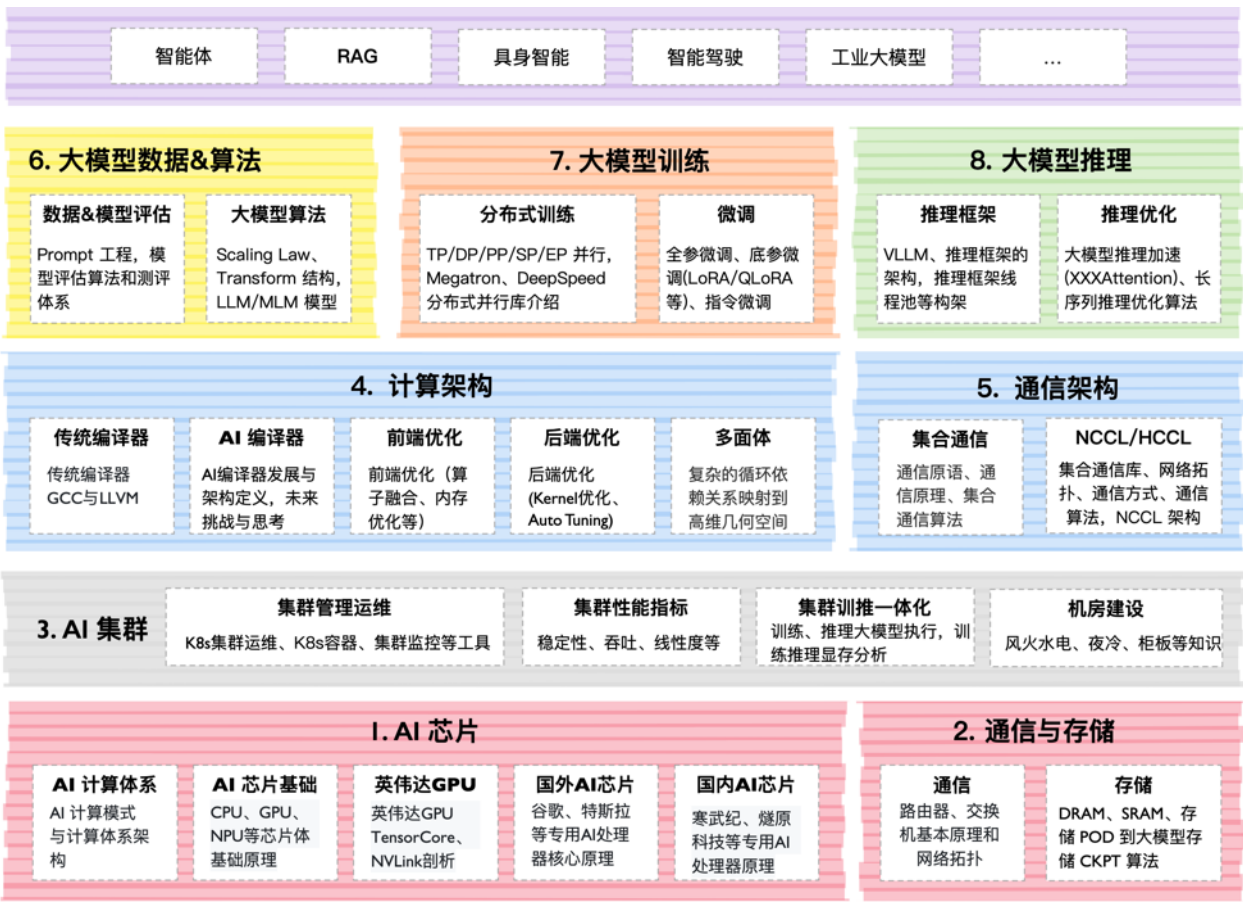


时事
热点

大模型
训推

编译
计算
架构

硬件
体系
结构



Content

L0 集群基础设施建设

服务器节点

通用服务器、AI 服务器 (GPU/NPU)

风冷：实践与思考

挑战、设计原则、实践与思考

液冷：技术发展趋势

风冷挑战、液冷发展、风液混合

液冷：实践与思考

液冷挑战、实践与思考

服务器机柜

风冷柜、板冷柜、全液冷柜

集群机房供电

供电技术、供电分类



Content

1. 服务器风冷挑战与设计原则
2. 服务器风冷实践



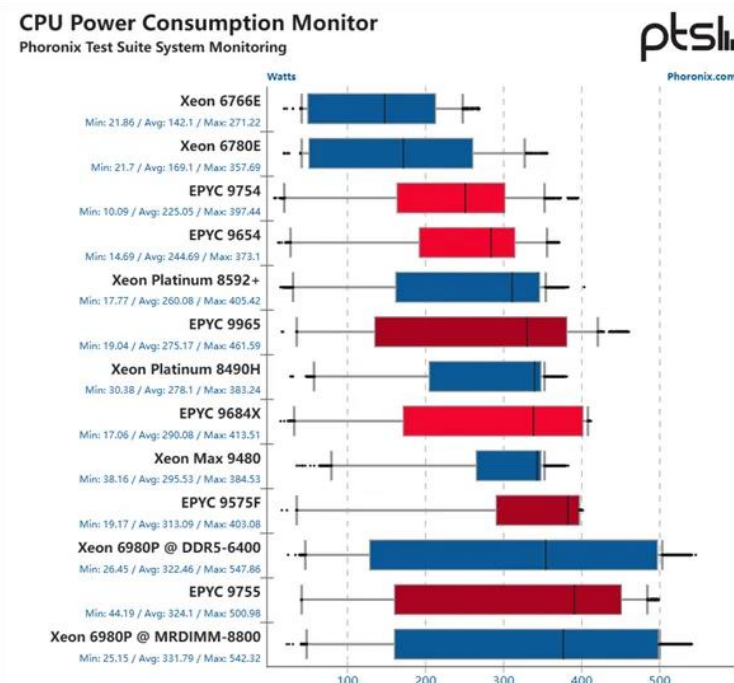
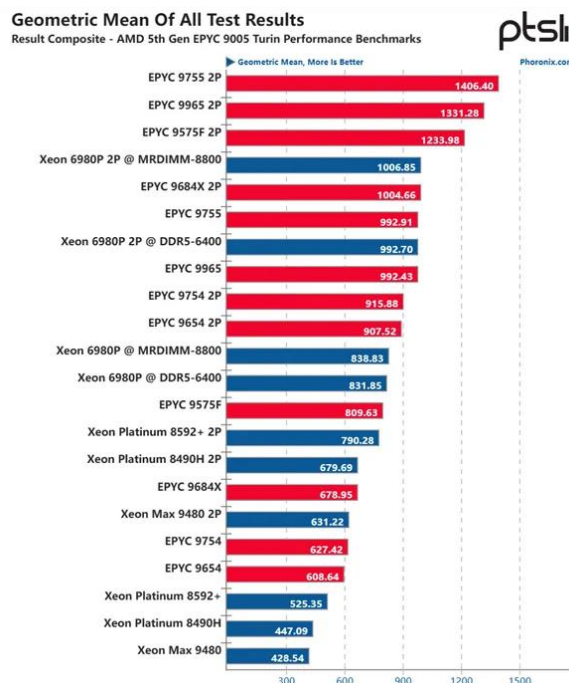
01

服务器风冷 挑战与设计原则



风冷场景挑战

- **峰值算力**: 释放芯片&服务器的峰值算力, 在缺乏先进制程下提升频率, 对标友商 $xx\% +$ 性能
- **性能达标**: 系统突破 $xxx\text{FLOPS}/xxx\text{W}$ 算力, 功率密度突破 $xxx\text{W}/\text{cm}^2$ 散热
- **系统能效**: 缺乏先进制程, 系统存在 $xx\%$ 功耗 GAP, 能效存在 $xx\%$ GAP



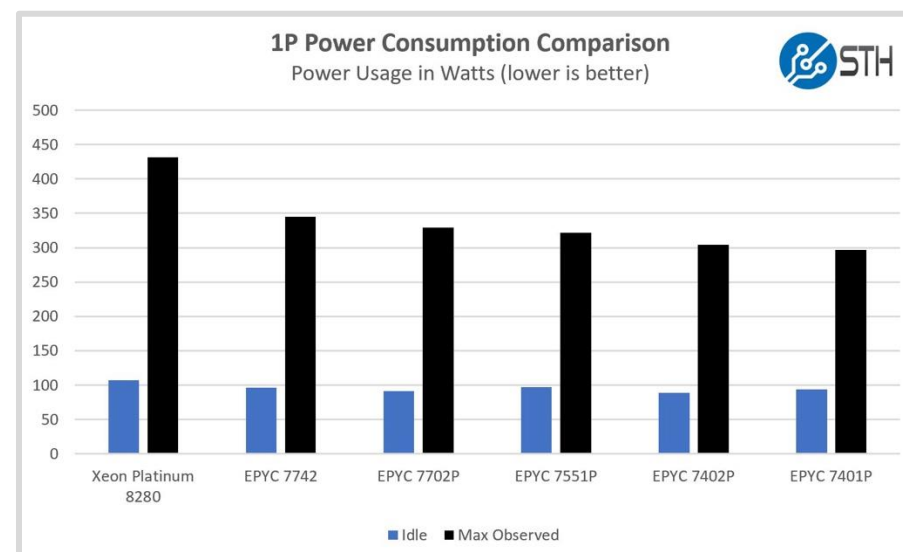
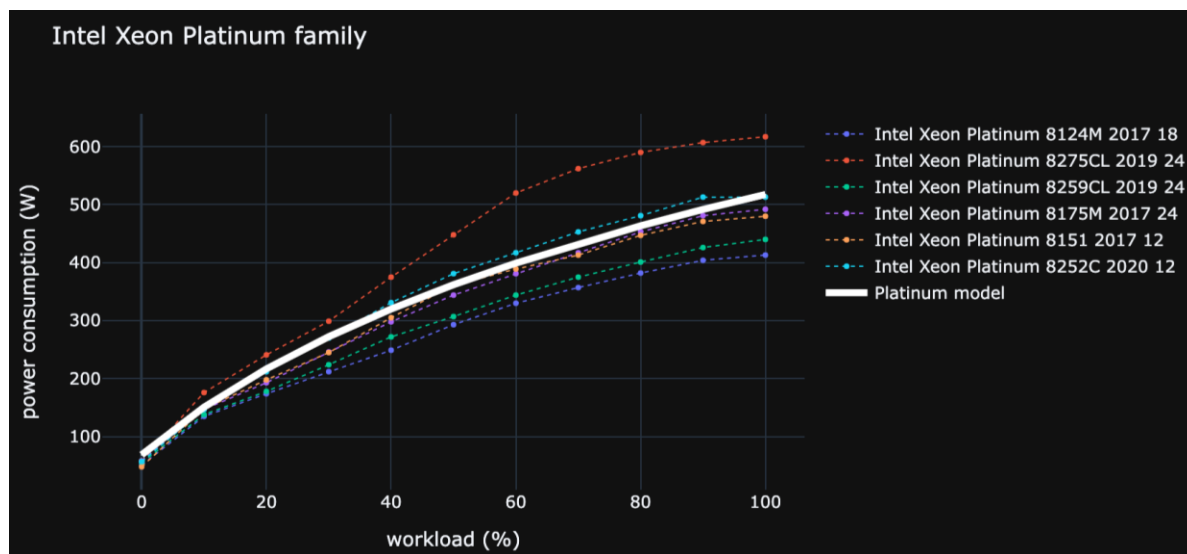
风冷场景设计原则

- **设计原则：**
 - 峰值性能场景以性能为中心，释放 CPU 极致算力；
 - 主流业务场景以能效/性价比为中心，平衡性能/功耗/成本/规格诉求，确保性价最优；
- **设计思路：** 基于峰值性能、系统能效及性价比构建商业场景竞争力



风冷场景设计思路

- **设计思路：** 基于峰值性能、系统能效及性价比构建商业场景竞争力
 - **峰值性能：** 风冷场景通过散热组件（风扇、LAAC/VCE散热器、导热垫）来提升效率，进行风道优化；
 - **系统能效：** 结合动态 CPU 节温和风扇调节，降低系统动态功耗，达到系统能效最优；



02

服务器 风冷实践



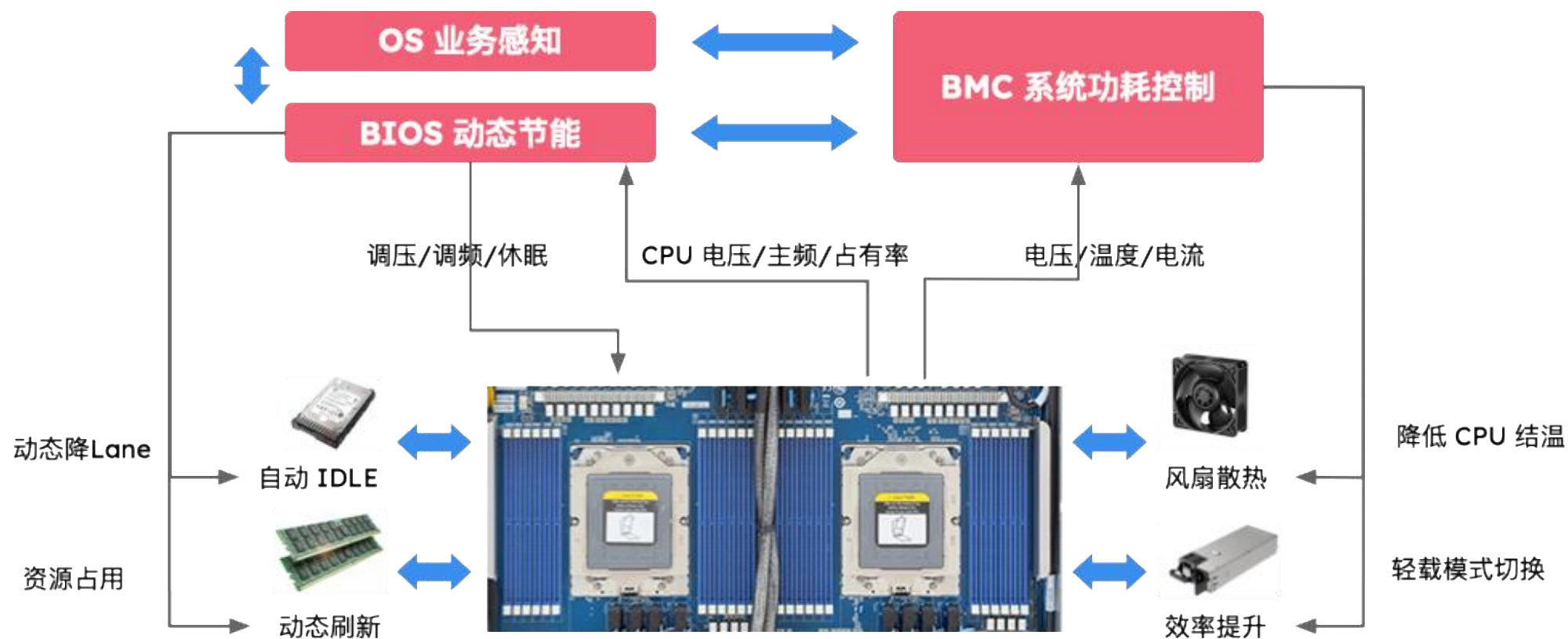
场景负载分析

- **设计目标：**分级能效管理，达成系统级最优
- **系统能效：**场景化能效最优
- **业务负载场景：**基于实际业务建模进行能效寻优（包括散热效率提升，系统风道优化，动态 CPU 节温，风扇调节），实现典型场景智能化节能
- **低负载场景：**通过动态休眠、动态调频/调压、降Lane降速、动态关断等技术，达成零负载低功耗（<5%）；



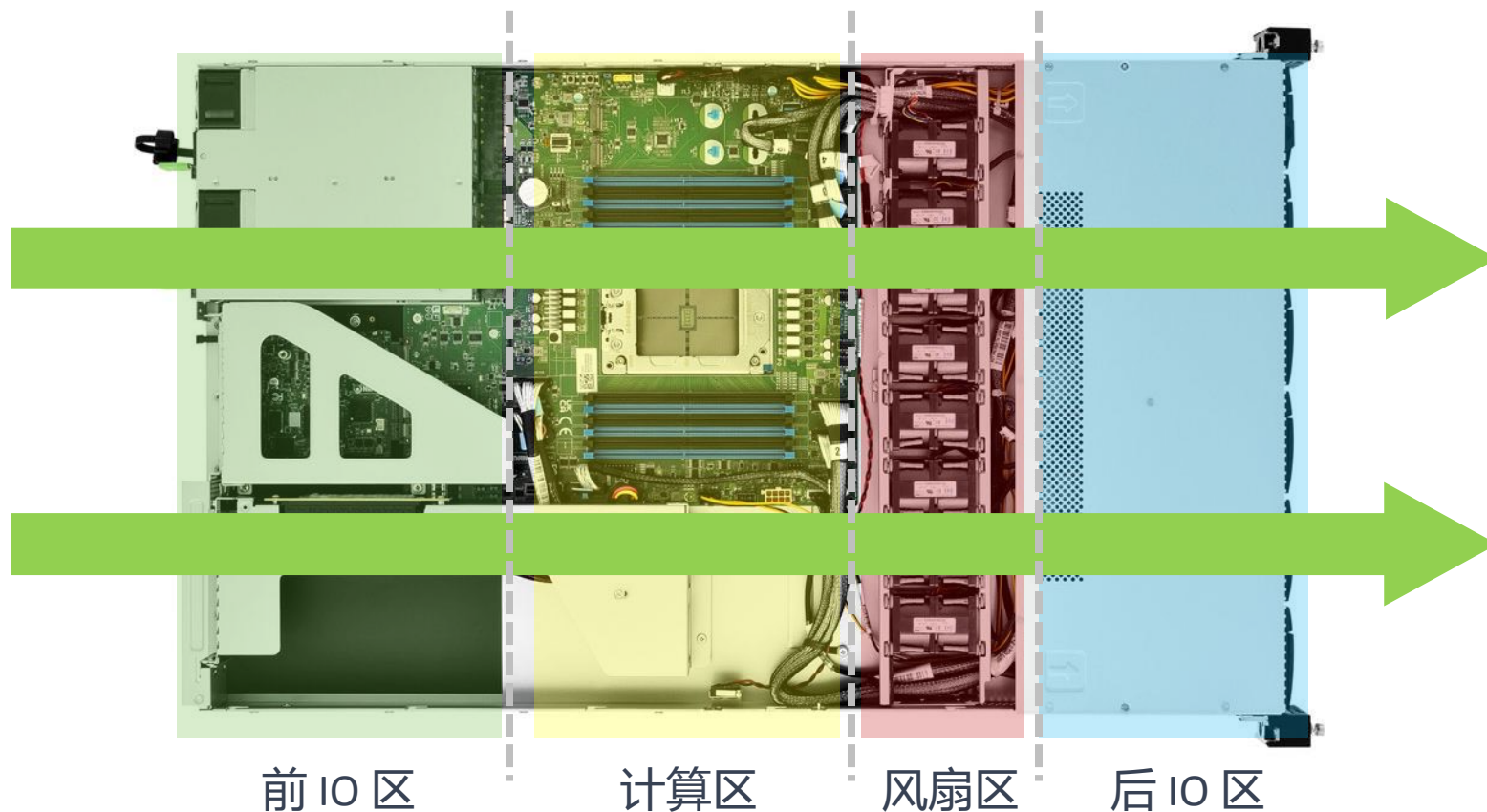
场景负载能效管理

- 低负载场景\业务负载场景：业务、管理、硬件多级联动进行能效管理



系统级风道优化

- **系统级优化：**增加开孔率、风道优化/管理等风阻设计，多段式优化设计，2/4U 整机 XXXXW 散热。
- **提升峰值性能：**通过风道/风量优化，修改散热器/换热器，热从芯片带到系统，支持 CPU/NPU/GPU 算力释放



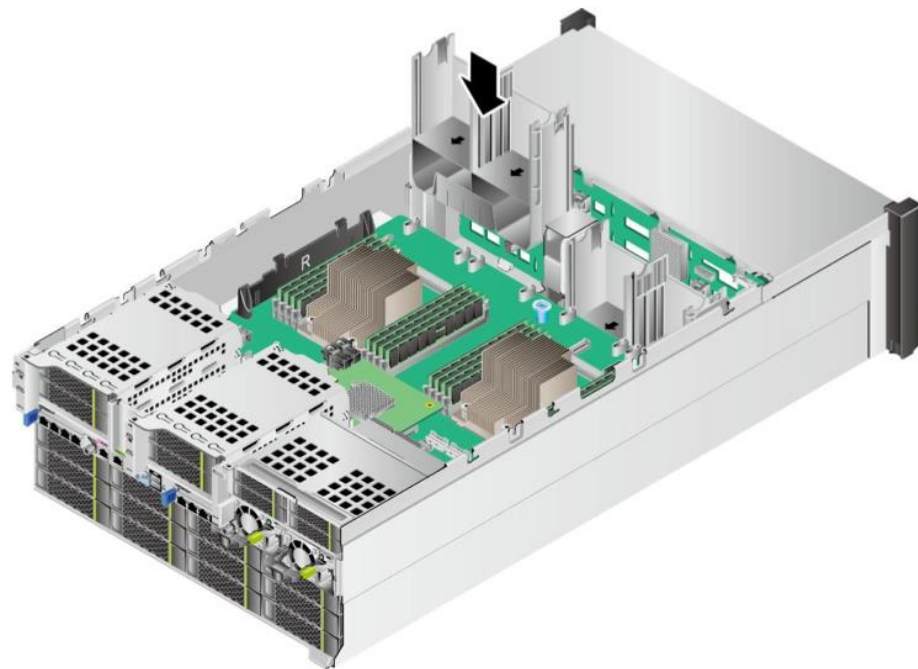
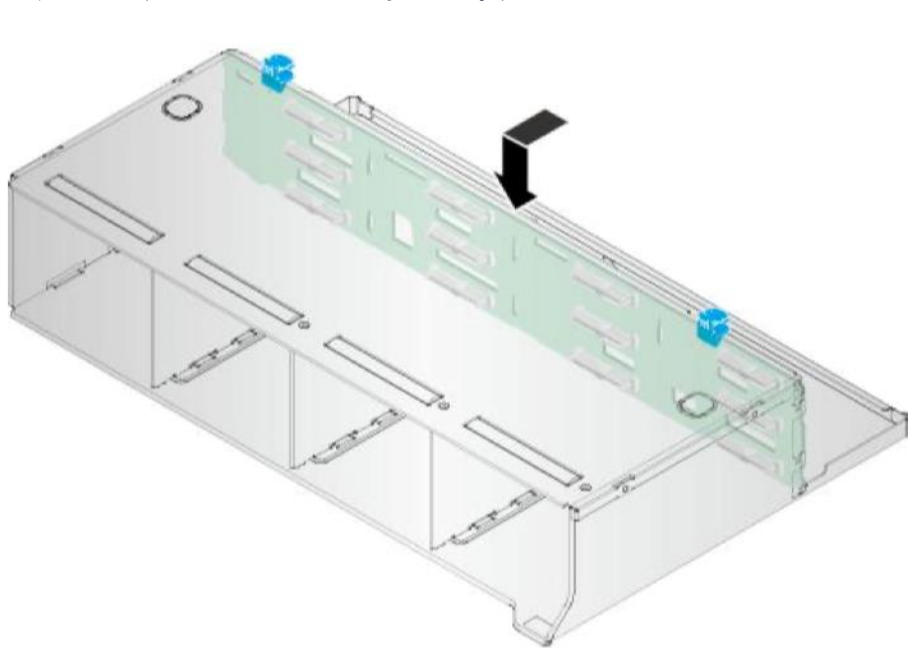
背板架构优化

- **挑战:**
 - CPU/MEM/SSD/DPU/GPU 往大功耗演进;
 - 传统背板架构无法提供足够风量支撑大功耗散热;
 - 系统功耗大量浪费在散热上, 散热效率低, 工艺受限;
- **策略: 重构背板优化系统风量/风道, 使风进得来出得去, 提升系统散热能力和能效**
 - 后置盘入风口温度高散热差, 导致系统风扇高转速, 系统功耗增加, 入风口修改
 - 后置背板竖直背板->水平背板, 最大化开孔率, 提升等效开孔率, 开孔率 XX% +
 - 风扇功耗降 XXW, 盘散热能力由 XXW->XXW, 噪声降低 X.X dBA, 系统散热能力提升 XX%



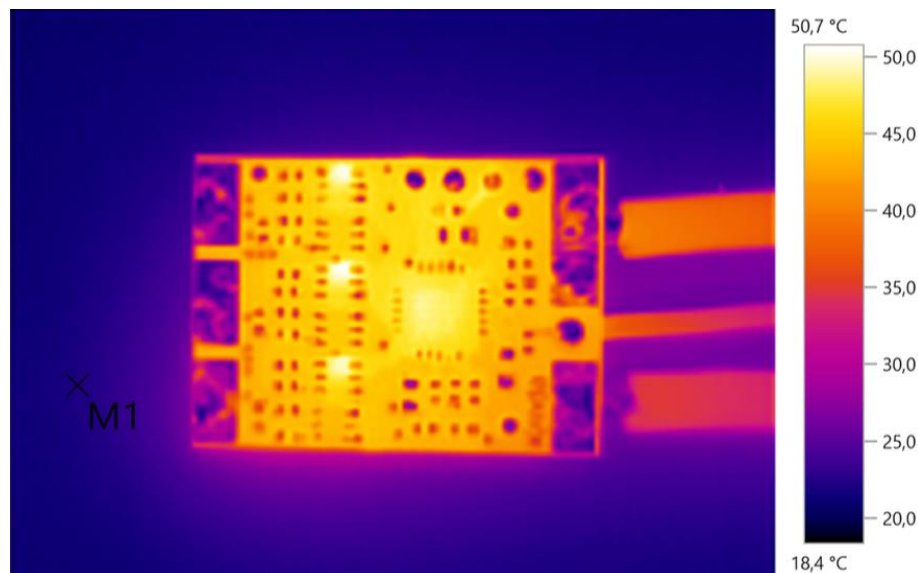
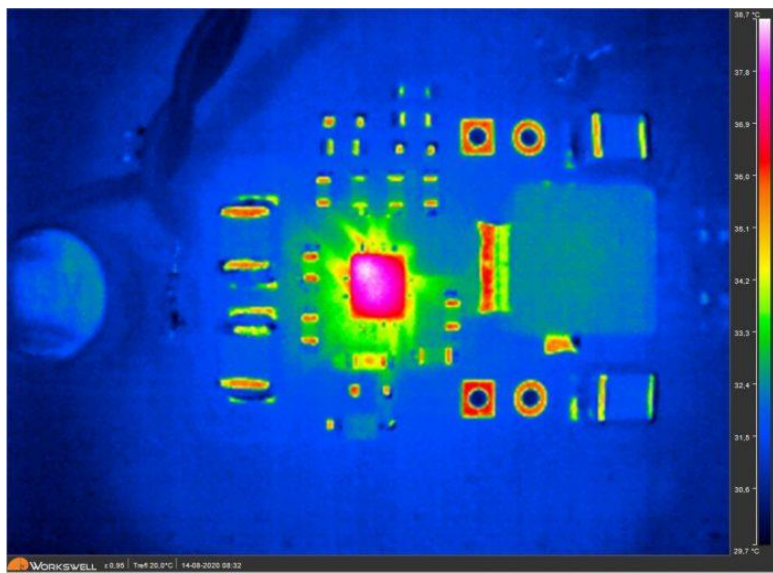
背板架构优化

- **策略：** 重构背板优化系统风量/风道，使风进得来出得去，提升系统散热能力和能效；
 - 后置盘入风口温度高散热差，导致系统风扇高转速，系统功耗增加，入风口修改
 - 后置背板竖直背板->水平背板，最大化开孔率，提升等效开孔率，开孔率 XX% +
 - 风扇功耗降 XXW，盘散热能力由 XXW->XXW，噪声降低 X.X dBA，系统散热能力提升 XX%



Question

- 什么是芯片结温?
- 芯片结温 (Junction Temperature) 芯片内部半导体材料温度，特别是晶体管PN结处温度。是衡量芯片内部热量积累和散热效果关键参数，对于确保芯片的稳定运行和可靠性至关重要。



散热组件动态结温控制

- 挑战:

- 工艺受限, 芯片 & 服务器友商整机存在 XX%+ 能效 GAP

- 策略: 工程补工艺, 整机散热、供电优化, 缩小能效 GAP, 做到能效持平/领先

- 分级散热组件提升系统散热效率, 采用高效散热方案
- 减少系统风扇功耗、提升散热效率, 效降低芯片结温

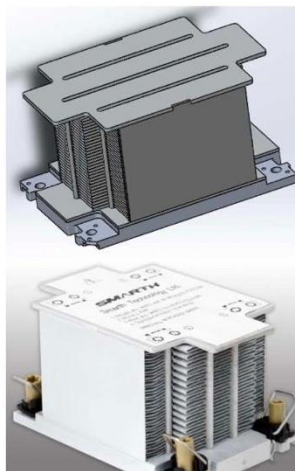
Heatpipe HS



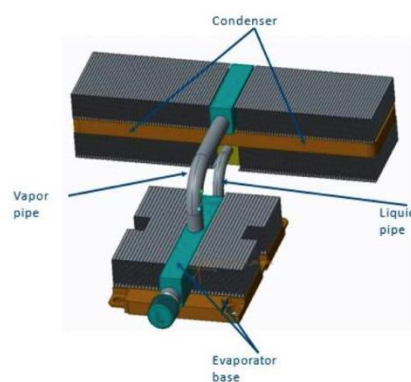
VC HS



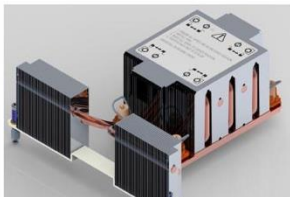
3DVC



Thermosyphon



EVAC



散热组件动态结温控制

率密度	部件种类	散热方案
高密度发热器件	CPU/GPU/NPU/DPU	Heat pipe 热管、Vapor Chamber 均热板 EVAC 增强型蒸汽腔空气冷却器 Thermosyphon 热虹吸管 3DVC三维真空腔均热板
低密度发热器件	CPU/DIMM VR, PCH, NIC, Memory chipset(HBM)	铝挤HS
<3W 发热器件	BMC	N/A

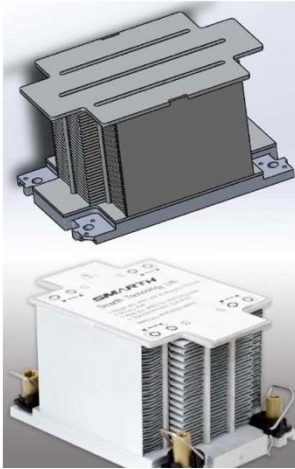
Heatpipe HS



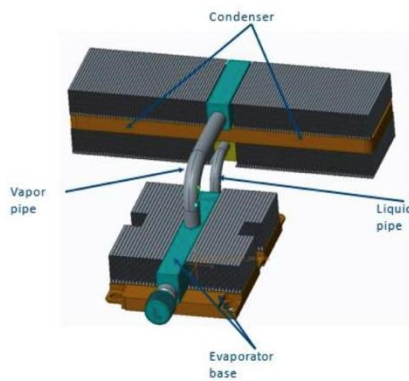
VC HS



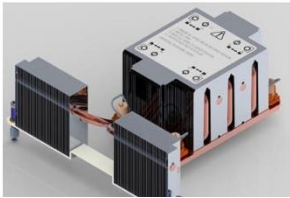
3DVC



Thermosyphon



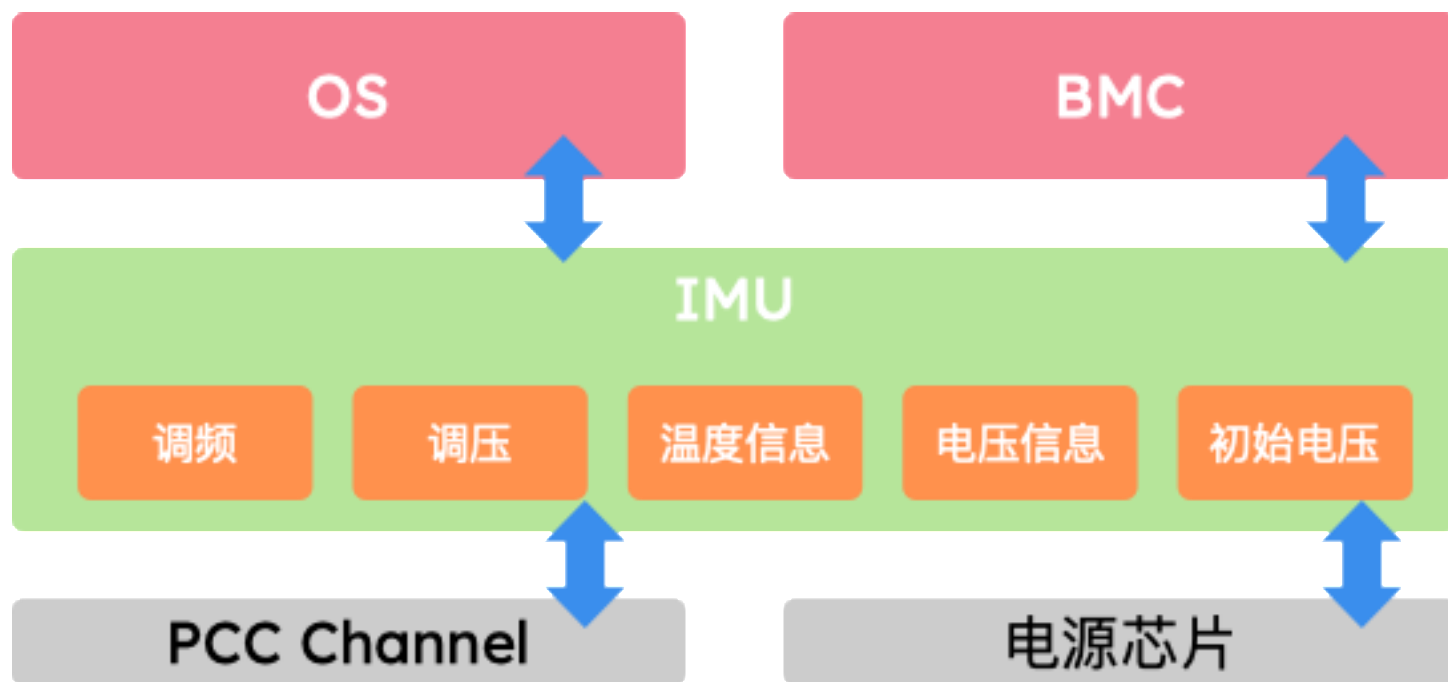
EVAC



软硬件协同智能节能

- 板级节能:

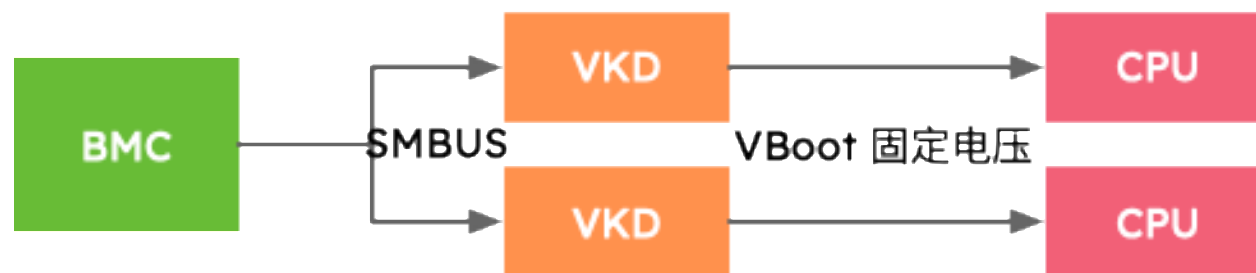
- AVS/Load line: IMU 获取 CPU HPM 信息计算芯片工作电压, 一芯一压;
- 板级DVFS: IMU 根据 CPU 实际工作主频, 动态调整 CPU 电压, 实现动态调压;



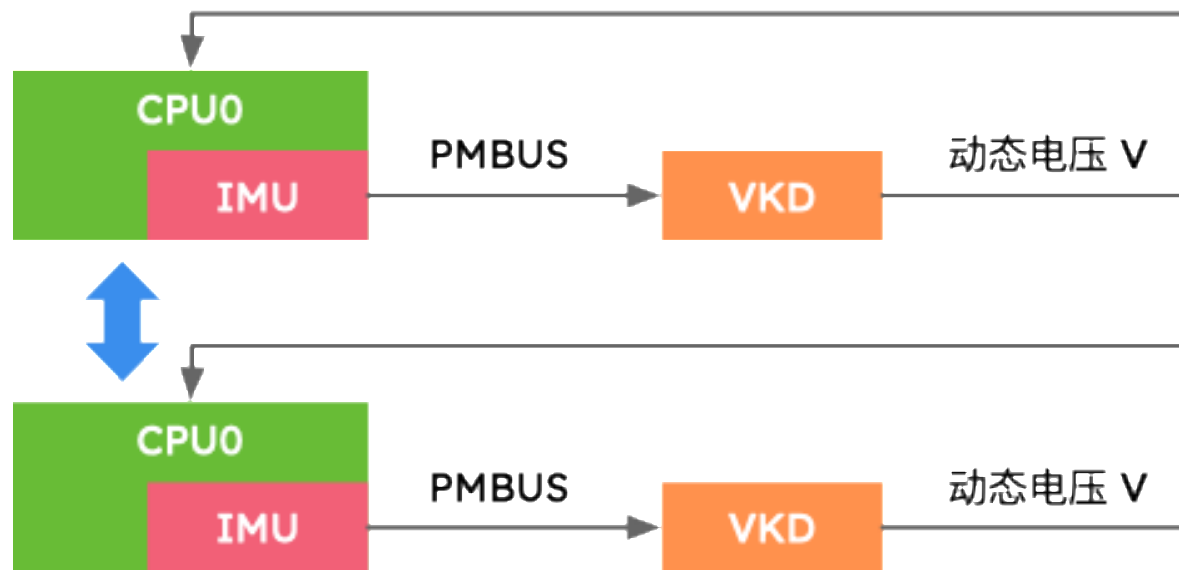
软硬件协同智能节能

- 系统节能:

- 联合 OS 优化调频算法, 业务感知优化调度算法;
- 系统风扇调速算法优化, MPC 取代传统 PID;
- 软件控制框圈, 硬件根据负载调整电压&主频;



传统方案: 固定 Vboot 电压供电



优化方案: 按工作主频动态调整电压



总结与思考



Question

- 大功率芯片 + 高密机柜，未来散热走向何方？
- 深挖动态调频/调压、MPC 风扇调速及 OS 调频算法优化等多路径功耗寻优；
- 联合业务进行场景化分析（分析业务负载），软硬协调识别关键功耗点；
- 结合应用精准设计各负载下的调频和调压甜点；
- 说白了，**针对特殊业务进行 hardcode**！！！！！



Question

- 纳尼? GPU/NPU 等 AI 场景怎么办?
- 传统 CPU 散热已经玩烂了!





Thank you

把 AllInfra 带入每个开发者、每个家庭、
每个组织，构建万物互联的智能世界

Bring AI Infra to every person, home and
organization for a fully connected,
intelligent world.

Copyright © 2025 [Infrasys-AI](#) org. All Rights Reserved.

The information in this document may contain predictive statements including, without limitation, statements regarding the future financial and operating results, future product portfolio, new technology, etc. There are a number of factors that could cause actual results and developments to differ materially from those expressed or implied in the predictive statements. Therefore, such information is provided for reference purpose only and constitutes neither an offer nor an acceptance. [Infrasys-AI](#) org. may change the information at any time without notice.



ZOMI

GitHub github.com/Infrasys-AI/AllInfra

Book infrasys-ai.github.io



引用与参考

1. <https://navitassemi.com/nvidias-grace-hopper-runs-at-700-w-blackwell-will-be-1-kw-how-is-the-power-supply-industry-enabling-data-centers-to-run-these-advanced-ai-processors/>
2. <https://www.eet-china.com/mp/a357098.html>
3. <https://www.ctyun.cn/developer/article/401197533323333>
4. <https://www.dell.com/zh-hk/shop/dell-xe8640-laac-%E6%A8%A1%E5%9D%97/apd/412-bbhs/wi-fi-%E5%92%8C%E7%B6%B2%E8%B7%AF%E5%8A%9F%E8%83%BD>

PPT 开源在: <https://github.com/Infrasys-AI/AllInfra>

