

快手高密数据中心散热解决之策

杨 琪

Open AI Infra 社区机房基础设施项目群机房液冷项目组联合组长
快手IDC架构师

TECHNOLOGY
FOR A MORE
VIBRANT WORLD

目录

CONTENTS

技 术 让 更 好 发 生

01 PART 01

高密散热的强输入：
AI业务爆发与芯片功耗增长

AI GROWTH AND RISING
CHIP POWER



02 PART 02

快手的解法：
风液同源架构与工程产品化

UNIFIED AIR-LIQUID ARCHITECTURE
AND ENGINEERING PRODUCTS



03 PART 03

可预见的未来与布局：
相变液冷与AI物理世界模型

PHASE-CHANGE LIQUID COOLING
AND AI WORLD MODEL




01
HIGHER
DENSITY
A COOLER
TOMORROW
PART 01
高密散热的强输入
为什么液冷是必选项
PART 01
LIQUID COOLING
A NECESSARY CHOICE
COOLING
FOR A MORE
POWERFUL TOMORROW

AI业务爆发，算力需求成倍增长

从丰富的 AI 应用场景到澎湃的算力需求，共同推动智能时代加速到来



智能驾驶

L4+ 自动驾驶加速落地
海量数据驱动算力需求



智慧医疗

AI辅助诊断加速普及
医疗大模型需求激增



短视频

海量内容生成与分发
推高实时推理算力需求



AI主播

虚拟人直播快速发展
多模态生成需求旺盛



可灵等

AIGC 创作生态繁荣
推动大规模算力消耗

智能驾驶、智慧医疗、短视频、AI主播、可灵等业务需求量大幅增长

多元AI应用全面爆发，持续拉动算力需求

单芯片算力密度持续提升，芯片功耗持续大幅度增长

更强的算力带来更高的能耗，驱动数据中心基础设施升级

有上层业务支撑，数据中心基础设施增量也会会持续扩大

AI 应用的蓬勃发展将持续带动算力基础设施的长期投资



必选项： GPU 功率密度迭代，单机柜功耗陡增

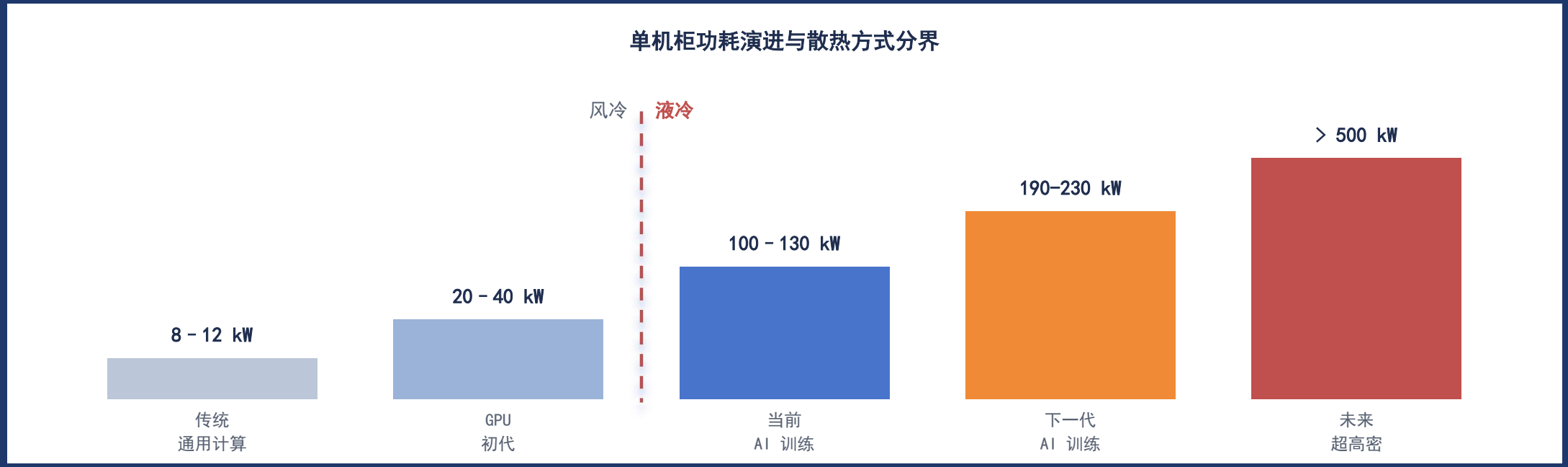
1. 芯片功耗持续攀升，风冷逼近物理极限

   ✓ 单颗 GPU：400W → 700W → 1000W→ 2000W 以上

2. 单机柜功耗从 12kW → 100kW → 200kW →500kW

   ✓ 风冷在 40kW 以上已难以经济地满足散热

   ✓ 液冷从「可选项」变为「必选项」



BUILD
A MORE EFFICIENT
SUSTAINABLE WORLD

ENGINEERING
PRODUCTS
PEOPLE
A BETTER TOMORROW

02

PART 02

快手的解法：风液同源与工程产品化

架构 · 产品 · 标准 · 验收 · 运维

AIR FLOW
LIQUID COOLING
HEAT EXCHANGE
ENERGY EFFICIENCY
OPERATION & MAINTENANCE



FROM
INFRASTRUCTURE
TO LASTING IMPACT

FLOW
PRESSURE
TEMPERATURE
EFFICIENCY
RELIABILITY

工程驱动真实改变
REAL CHANGE THROUGH ENGINEERING

ENGINEERING
FOR A BRIGHTER TOMORROW

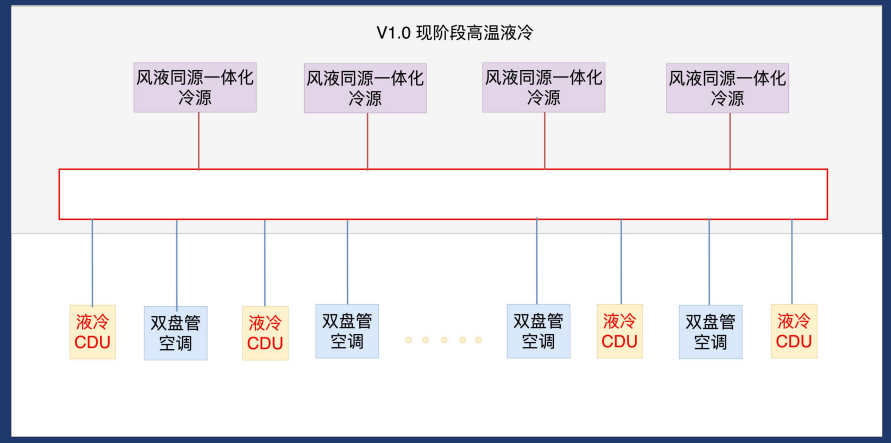
风液兼容的技术迭代

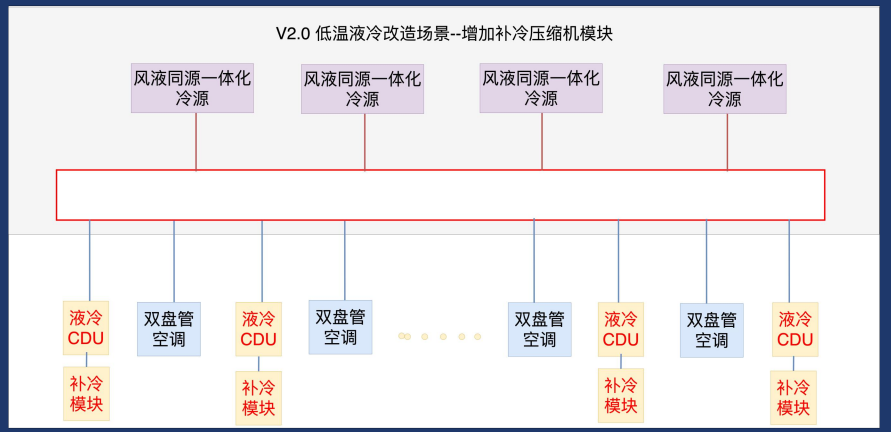
痛点

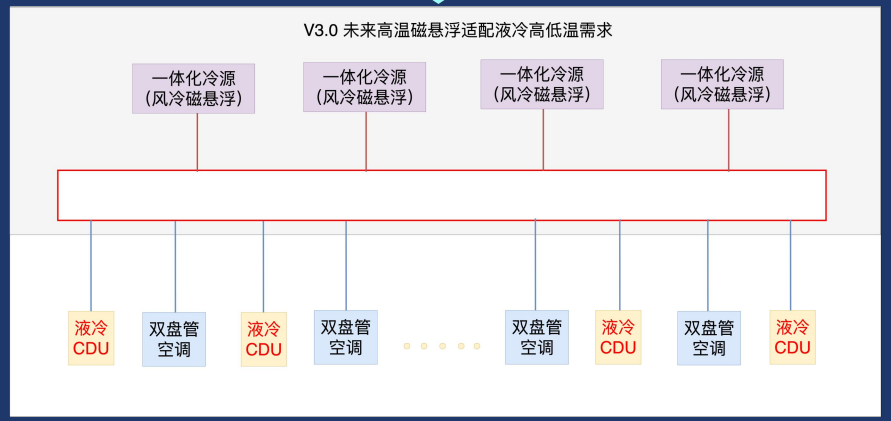
- 存在大量高温液冷机房

方案

- 改造：存在大量双盘管+CDU的高温供水液冷机房，改造为液冷机房需要补冷模块
- 新建：未来采用高温风冷磁悬浮方案，做风液同源冷源







液冷一次侧——集成智能干冷器

思路：工程产品化、减少一级供应链数量

痛点

- 乌兰察布缺水
- 风冷液冷不同源，造价高
- 一次侧工程设备多，现场工程多

方案

- 采用干冷器，解决无水问题
- 工程产品化：干冷器本体+一次侧水泵+脱气装置+控制+阀门集成，形成集成智能干冷器，车板交货快速交付
- 一次侧产品化率提升至80%。



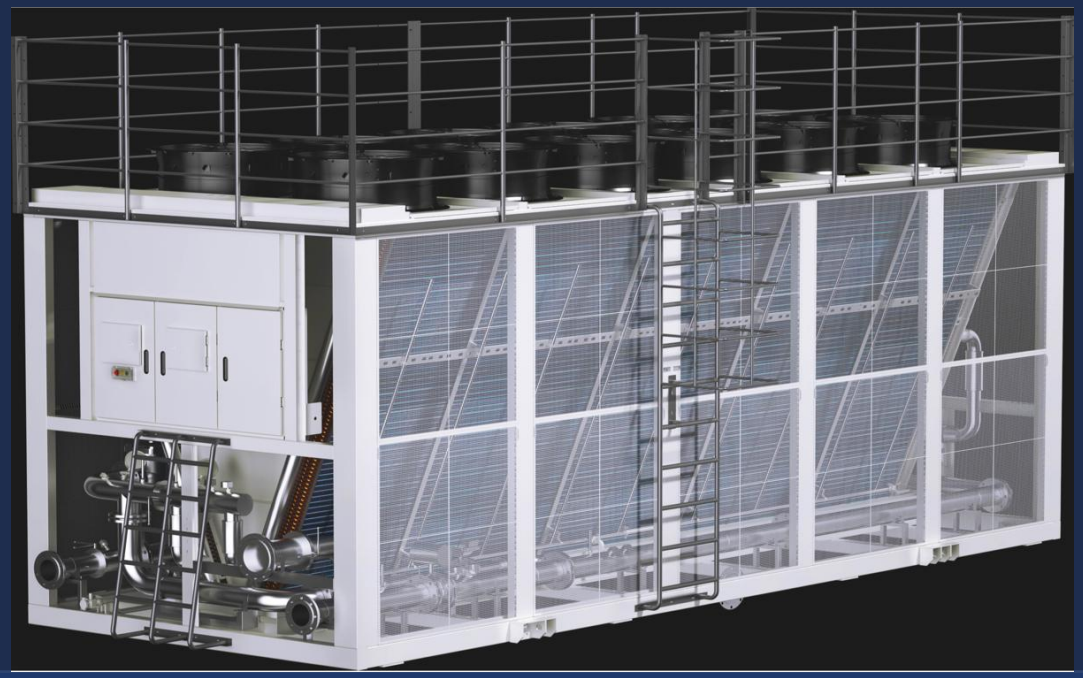












液冷二次侧——分集水器 + 上走软管架构

思路：减少隐蔽工程、工程产品化

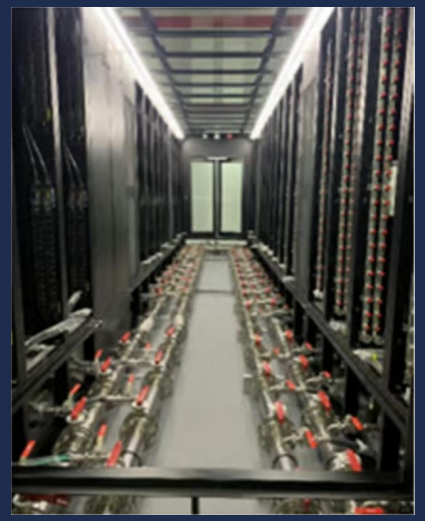
痛点

- 隐蔽工程多：传统地板下走管和上走管隐蔽工程多，运维困难
- 产品化率低：需要大量管道、阀门现场安装

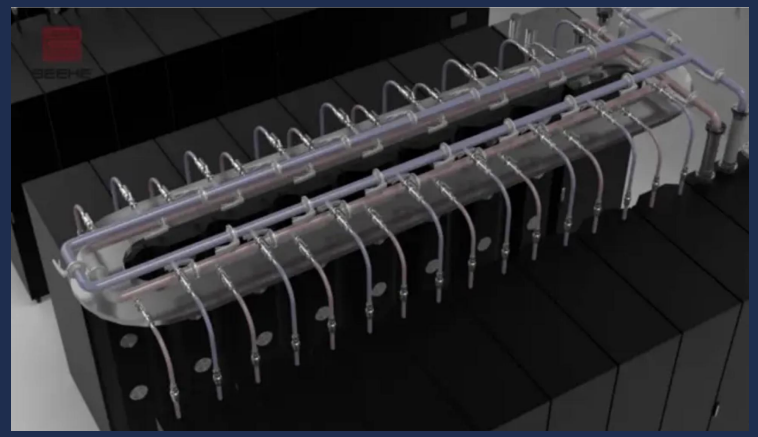
方案

- 重构架构：采用分集水器+液冷上走管方案
- 分集水器：将二次侧管道和阀门集成进分集水器里，预制化率提升至80%，开柜即检修，提高运维便利性

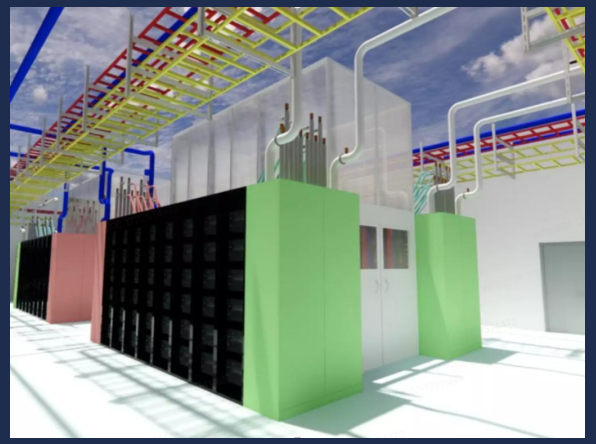
地板下走管



地板上走管



侧走管：分集水器（推荐）



补冷模块——磁悬浮冷机单元

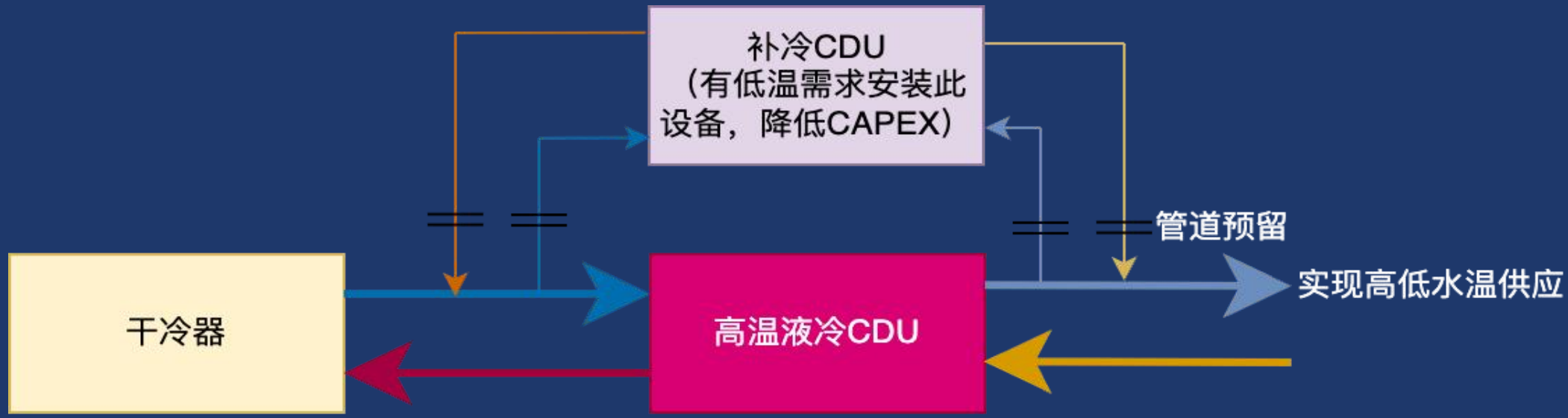
思路：今天只留接口，不装设备；有低温需求时再加装补冷 CDU，把 CAPEX 后置

痛点

- 既有高温液冷架构上需要兼容
- 场地空间有限需要架构精简

方案

- 管道预留接口、机房预留场地、水泵预留扬程
- 补冷模块仅设置：蒸发器、冷凝器、压缩机、膨胀阀、控制系统



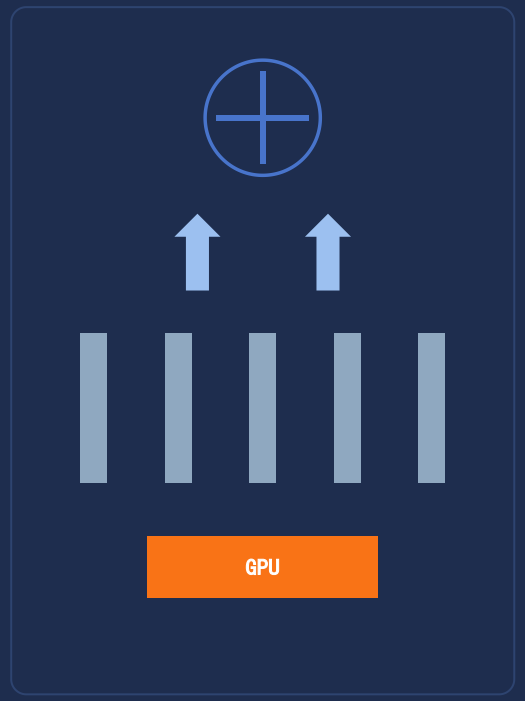

03
PART 03
可预见的未来与布局
相变液冷 · AI物理世界模型
A
MORE
SUSTAINABLE
TOMORROW
FROM
INFRASTRUCTURE
TO A BETTER TOMORROW
TECHNOLOGY
PEOPLE
A BRIGHTER TOMORROW

未来：技术演进——从风冷到单相液冷，再到相变液冷

随着功率密度再次提高，单相低温液冷解决不了，相变会成为高密散热方案必选项之一

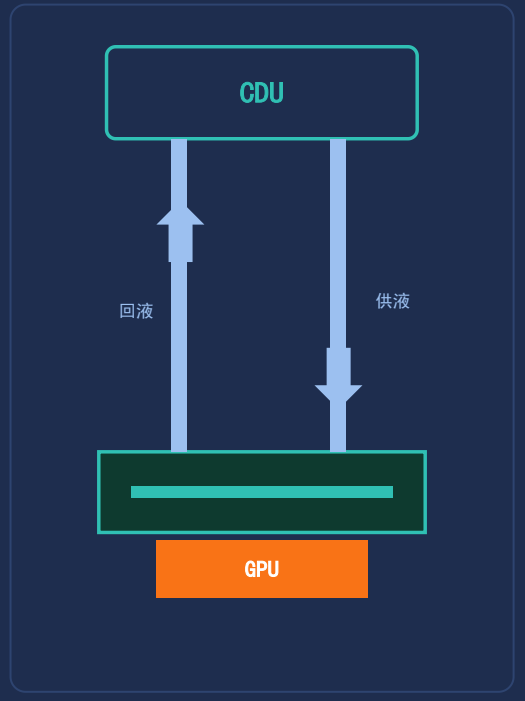
1 风冷

空气对流 · 密度受限，单机柜功率预计小于40kw



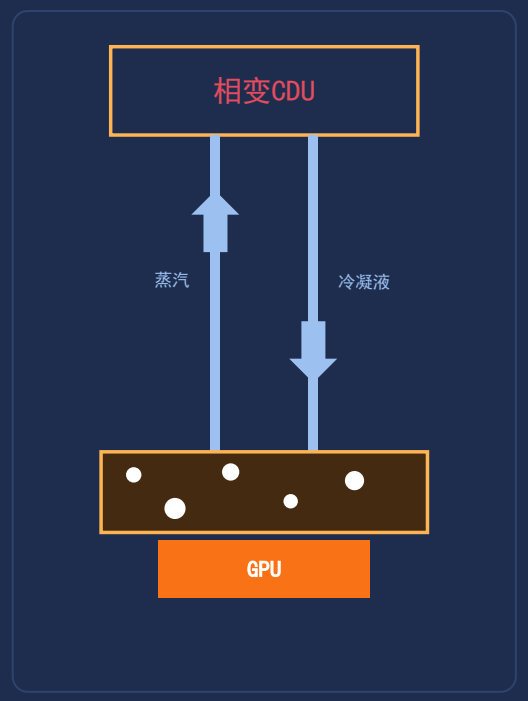
2 单相液冷

显热吸热 · 单机柜功率预计小于500kw



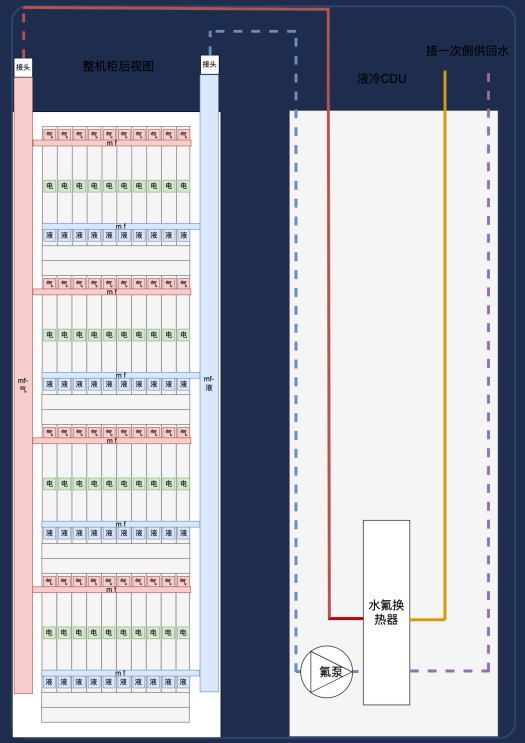
3 相变液冷

潜热吸热 · 效率更高，单机柜功率预计大于500kw



4 技术演进路径

三代解热技术全景



未来： 相变液冷--面向超高密度的下一代方案

1. 相变换热效率远超单相，芯片均温性更好

   ✓ 冷媒汽化潜热吸热，支持单机柜 500kW 以上

2. 量产前仍有两类问题待解

   ✓ 系统级：解决冷媒分配瓶颈、解决气堵风险等

   ✓ 产品级：扩展氟泵产品序列、合适的抵押冷媒、manifold、快速接头、高密假负载等

相变液冷的两类待解决问题

系统级	产品级
1. 冷媒分配：如576节点冷媒分配问题 2. 气堵问题：冷板空间小，解决蒸发排气问题 01	1. 氟泵：需要大氟泵，扩大产品序列 2. 冷媒：低压冷媒 3. manifold：保证冷媒蒸发后排气顺畅 4. 快速接头：实现气态的自密封 5. 高密假负载：模拟高密芯片散热 02

未来：AI物理世界模型

思路：物理世界模型，实现场景全覆盖，缩短数据采集时间，提升运维稳定性

痛点

传统做法：

- 需要大量以年为周期的历史数据
- 无法覆盖不同负载下全部室外温度
- 无法搭建数据中心整个系统做故障分析与溯源
- 无法做到设备故障预警

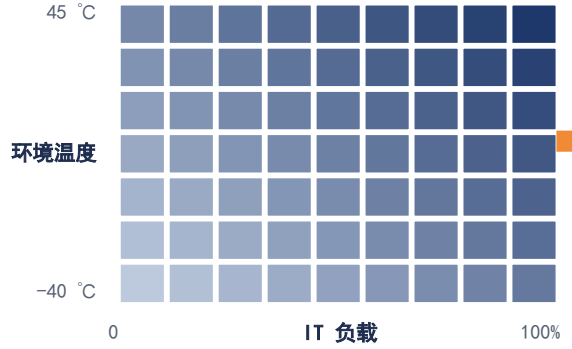
方案

搭建AI物理世界模型：

- 节能：负载--温度二维工况全覆盖，全工况数字孪生，使模型最优
- 系统故障预警+问题溯源：通过AI模拟各种故障场景，可通过AI提前预判系统故障，以及出现故障后可反向快速查找故障原因
- 设备故障预警：建立空调、柴发、UPS、电池等设备模型，根据运行参数故障前进行预告警，避免事故发生

AI 物理世界模型：二维全工况覆盖 → 数字孪生 → 三类能力输出

输入：二维全工况



全场景工况网格 · 全覆盖采样

AI 物理世界模型

数字孪生

物理机理 + 全工况仿真
替代「数学模型 + 年级历史数据」

旧：数据难全覆盖 → 模型不准

输出：三类能力

01 **节能**
0～100% 负载 × -40～45℃ 全工况寻优

02 **系统故障预警 + 问题溯源**
AI 模拟故障场景，提前预判、反向溯源

03 **设备故障预警**
空调 / 柴发 / UPS / 电池 参数级预告警

谢谢

THANK YOU

北京快手科技有限公司

快手高密散热三步走

01 **风液同源**

一套冷源双兼容 · 冷源初投资 −50%

02 **低水温预留**

今天留口 · 明天免改造

03 **相变液冷**

面向单柜 500kW 以上

共建生态　合作共赢