

# 下一代XPU模组项目技术规划分享

汪新新 单宝灯

Open AI Infra 社区XPU模组项目孵化小组

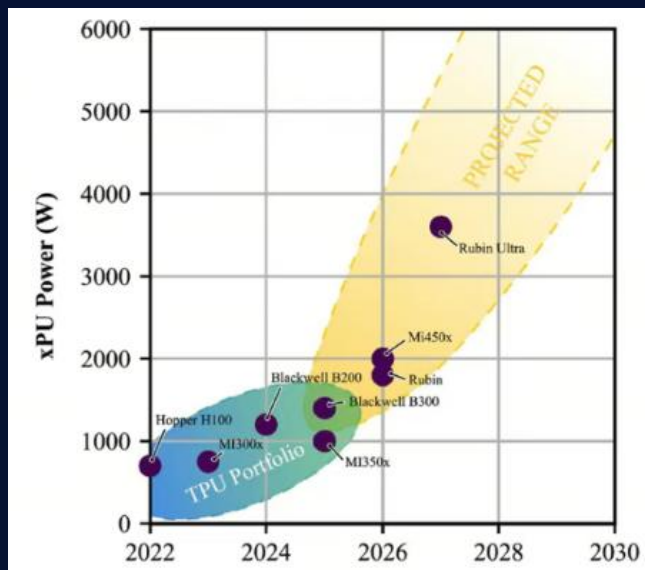


# 目录

XPU芯片&模组现状分析

XPU模组标准化思路

- 27-28年预计单GPU芯片会到3000W;
- 2029年, 单GPU功耗预计到6000W;
- Scales from 100kW to 1MW+。



## Power and Grid: Challenges

### Power and Grid

- GPU clusters create correlated load oscillations
- Causes load swings at Grid and Power Source
- Law of large numbers no longer smooths usage
- Can cause Grid Instability and power generation equipment fatigue & malfunction

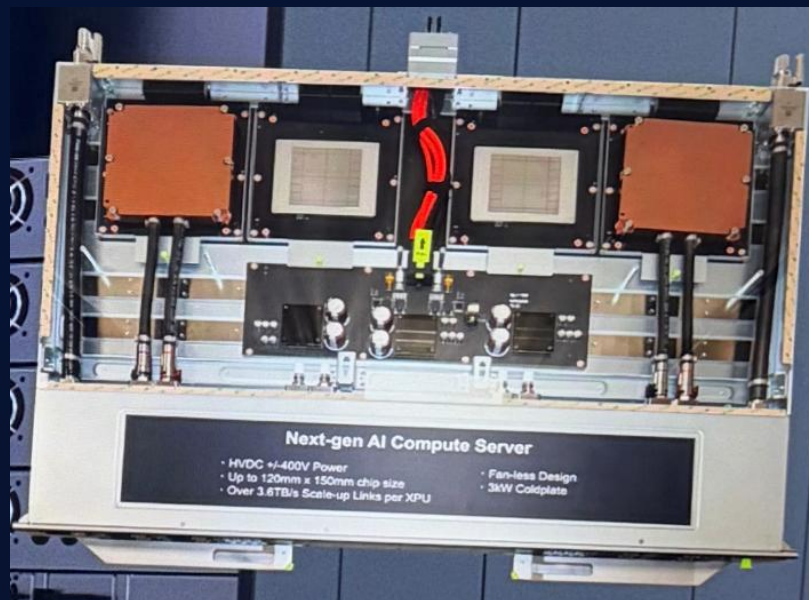
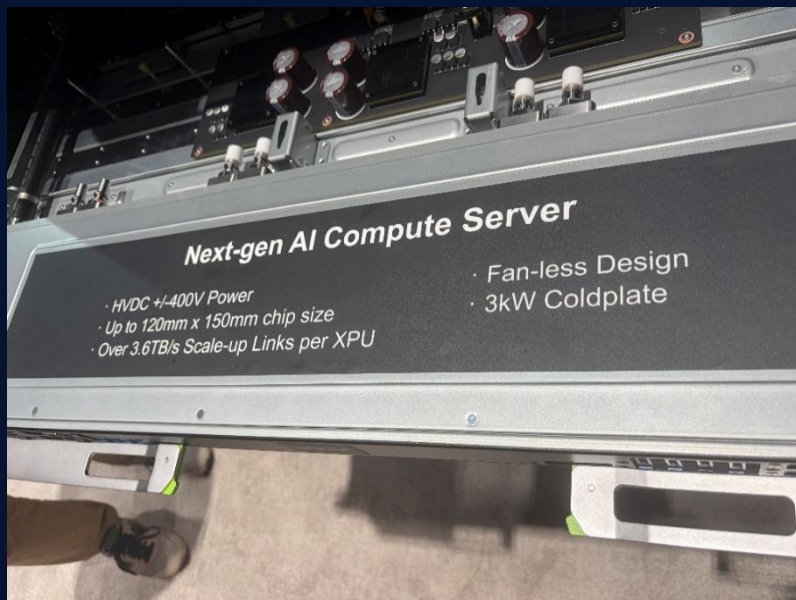
### Future-proofing

- Rack Power Evolving Fast:
  - 10KW -> 100KW -> 500KW -> 1MW
- DC Lifetime Higher than Server Rack Lifetime



## ■ 封装:

- 基于国外某GPU单GPU芯片尺寸 $120 \times 150\text{mm} > \text{OAM2.0}$
- 国产GPU国内制程受限，非先进制程，尺寸较大，功率较大
- 部分国产推理GPU和高端训练GPU共用封装，芯片尺寸 $\sim 120\text{mm}$ 。



- 国外GPU生态已基本收敛，国内处于蓬勃发展阶段，目前国内GPU厂家至少超过20+家；
- 亟需可以快速落地的整机柜全生态的标准化方案，加速国产GPU厂家的快速产业化落地。

国 外



国 内

20+



## CEM 限制

## 升级难度

|         |                |      |
|---------|----------------|------|
| 功率(TBP) | ~600W          | 高    |
| 散热      | 风冷优先<br>液冷难以实现 | 高    |
| PCB 厚度  | 1.57mm OR 阶梯板  | 高    |
| PCIE接口  | X16 lane       | 高    |
| 高速接口    | 无              | 无法升级 |



## CEM 2.0限制

## 升级难度

|         |           |   |
|---------|-----------|---|
| 功率(TBP) | ~1400W    | 高 |
| 散热      | ~1400W    | 高 |
| 尺寸      | 102*170   | 高 |
| 垂直供电    | 不支持       | 高 |
| 高速接口    | 112G/224G | 高 |
| 最新版本    | 2023年     | 高 |

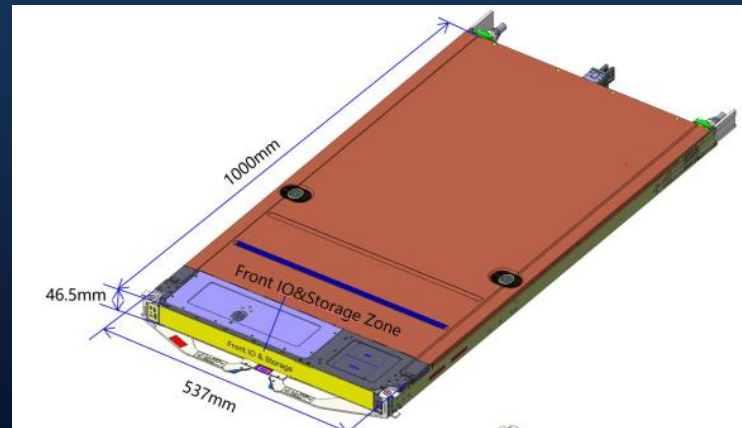
现有模组规格限制大



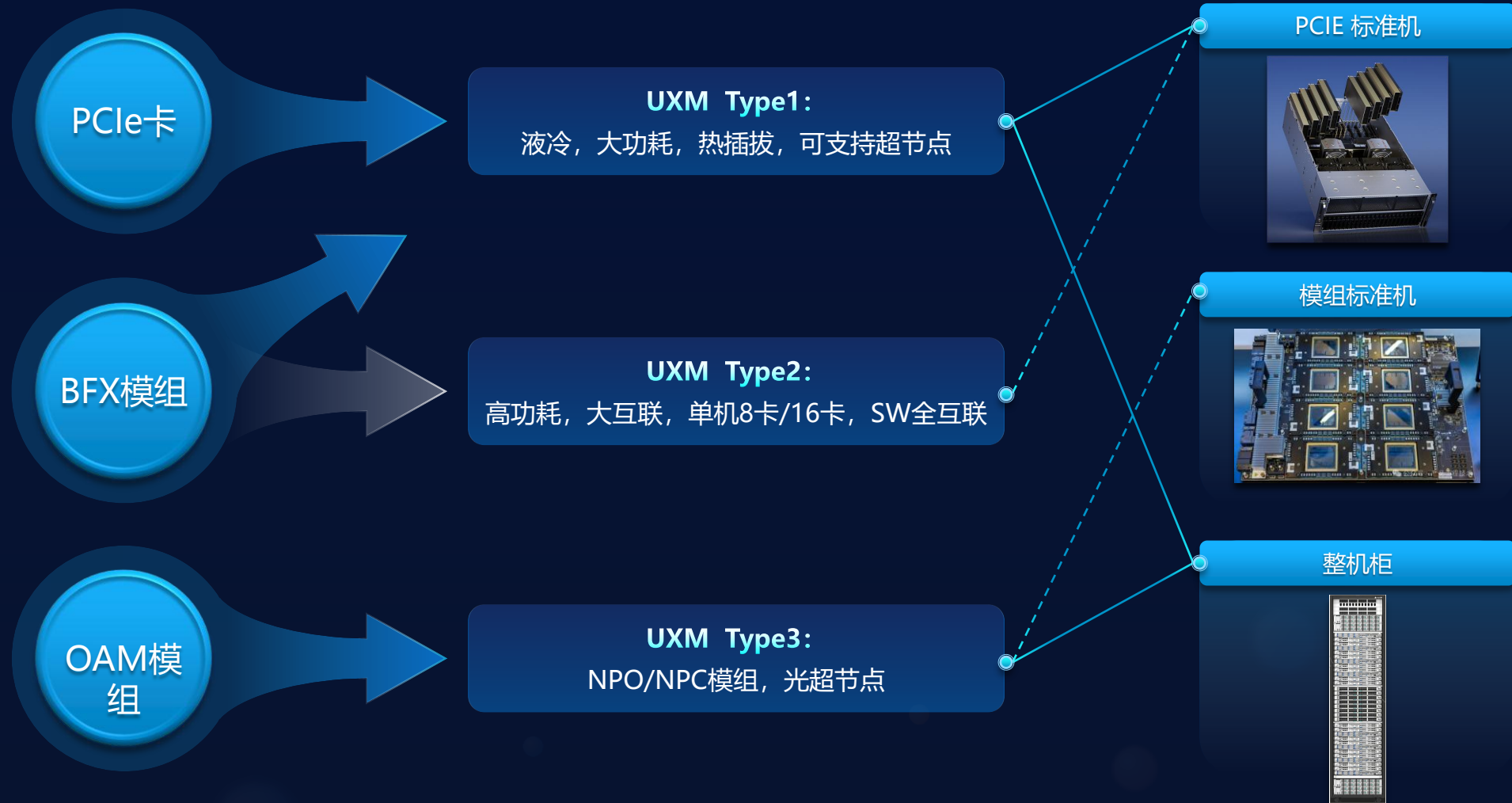
16卡服务器成为主流，停机维护故障面较大。



推理型整机柜使用PCIE标卡，需要侧插，要实现单Tray 8卡，整机装配和维护极其复杂。



新场景下，采用现有模组设计整机制造和维护难度大





# >> XPU模组规格升级需求

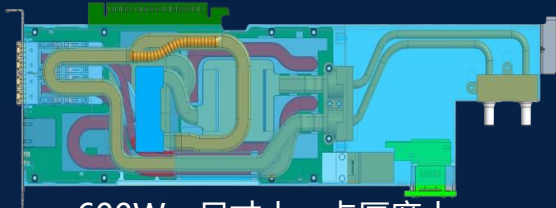
PCIe卡



600W, 风冷, 单机8卡, 弱互联



Bfx卡



600W, 尺寸小, 卡厚度小



OAM卡



1000W, 风冷&液冷, 单机8卡, 尺寸102\*170mm, 电互联



## 模组类型

## Type1

功率

2000W+

散热

原生液冷

Form Factor

支持120mm大芯片

互联

单机16卡/强互联

## 模组类型

## Type2

功率

3000W+

散热

原生液冷

Form Factor

支持100mm大芯片

互联

单机16卡

## 模组类型

## Type2

功率

3000W+

散热

原生液冷

Form Factor

支持120mm大芯片

供电

VPD

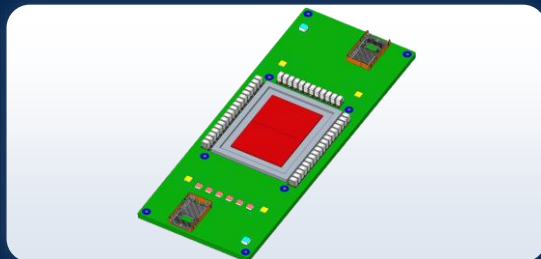
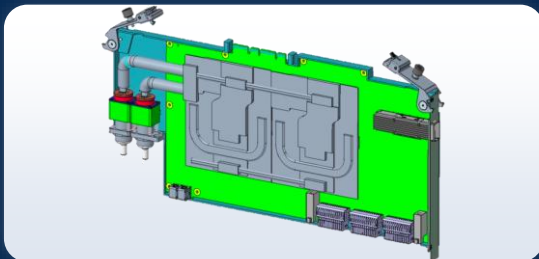
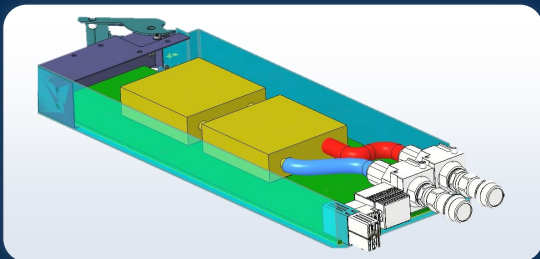
互联

光互联, 超节点64+, 强互联

# >> XPU模组标准化思路-初步规格

| 模组类型 | Type1                  | Type2                       | Type3                     |
|------|------------------------|-----------------------------|---------------------------|
| 供电   | 54V                    | 54V                         | 54V                       |
| 功耗   | 2* (600W~1200W+)       | 8卡: 单卡3000W<br>16卡: 单卡1500W | 3000W                     |
| 散热方式 | 液冷                     | 液冷                          | 液冷                        |
| 上行   | 2*PCIe6.0/PCIe7.0*X16  | 1*PCIe6.0/PCIe7.0*X16       | 1~2*PCIe6.0/PCIe7.0*X16   |
| 互联   | 60Lane~ 112G/224G      | 64Lane~ 112G/224G           | 3.2T/6.4T/12.8T NPO       |
| 场景一  | 8-16卡 互联: 液冷、双芯、易维护    | 8-16卡 互联: 三总线盲插             | 8-16卡 光互联                 |
| 场景二  | Scale out整机柜: 单节点8卡易维护 | /                           | Scale up整机柜: MW级系统64卡以上互联 |

概念设计



注: 所有规格均为初步拟定, 欢迎讨论。

## 规格建议

模组尺寸 宽约130mm 长约380mm

电源功率 ~3000W

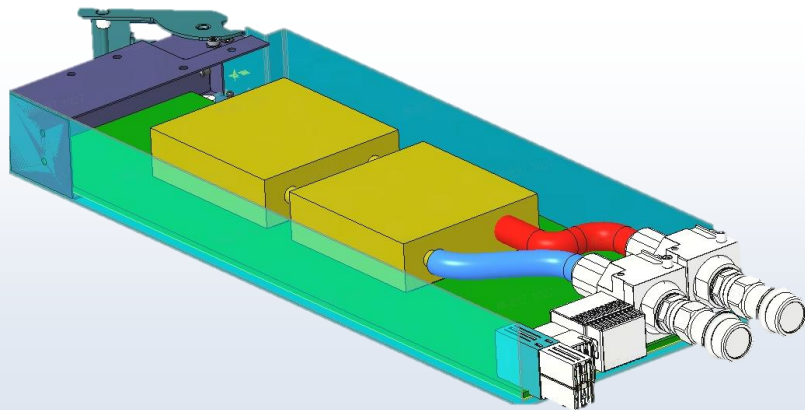
GPU最大功率 ~2400W /2\*1200W

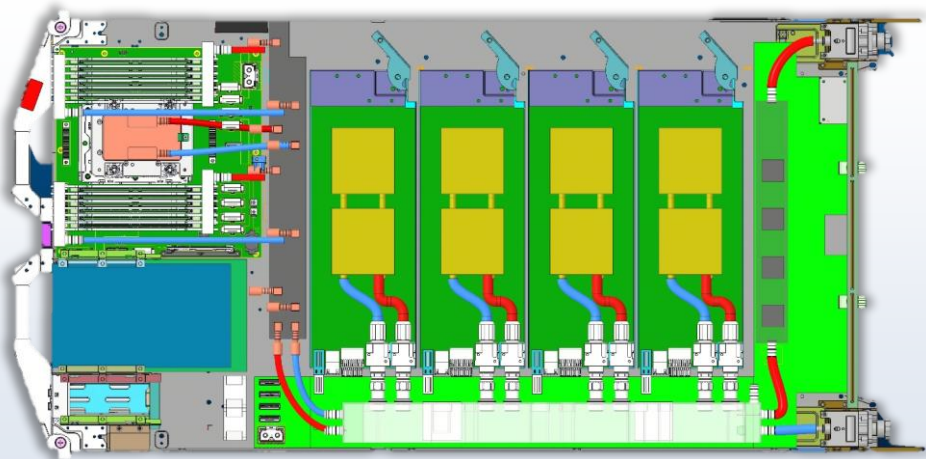
液冷最大功率 ~3000W

P C I E 接口 Pcie6.0/7.0 X16 \*2

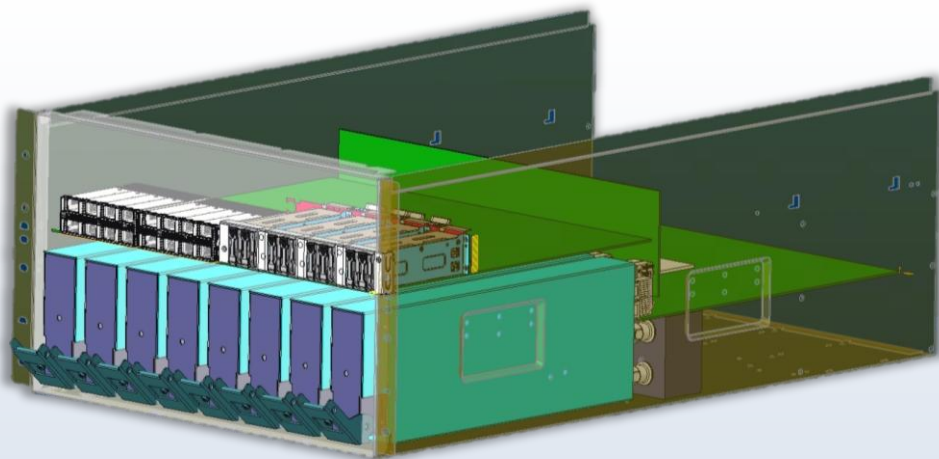
高速接口 60Lane~ 最高224G

维 护 支持盲插





整机柜Tray



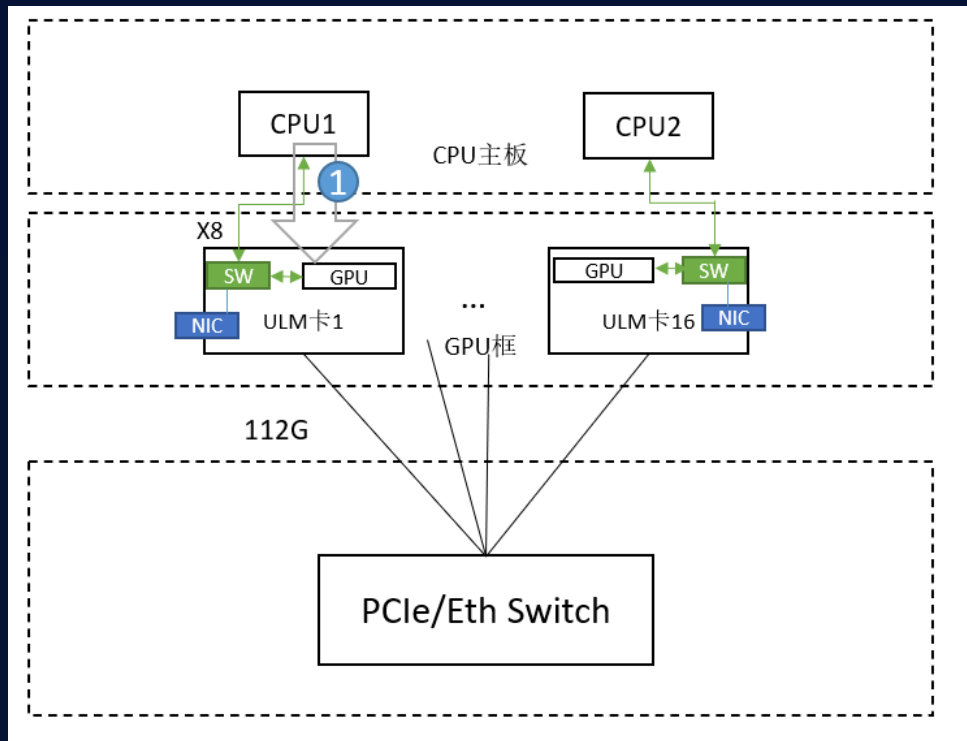
PCIE 标准机

## ■ 液冷板卡

- 类BFX模组，包括液冷板卡和底部的SW互联板；
- ### 液冷板卡
- 内部通过类似CX8的方案（PE SW+NIC）实现GPU的scale-out；
  - 液冷板卡水电网三总线盲插；

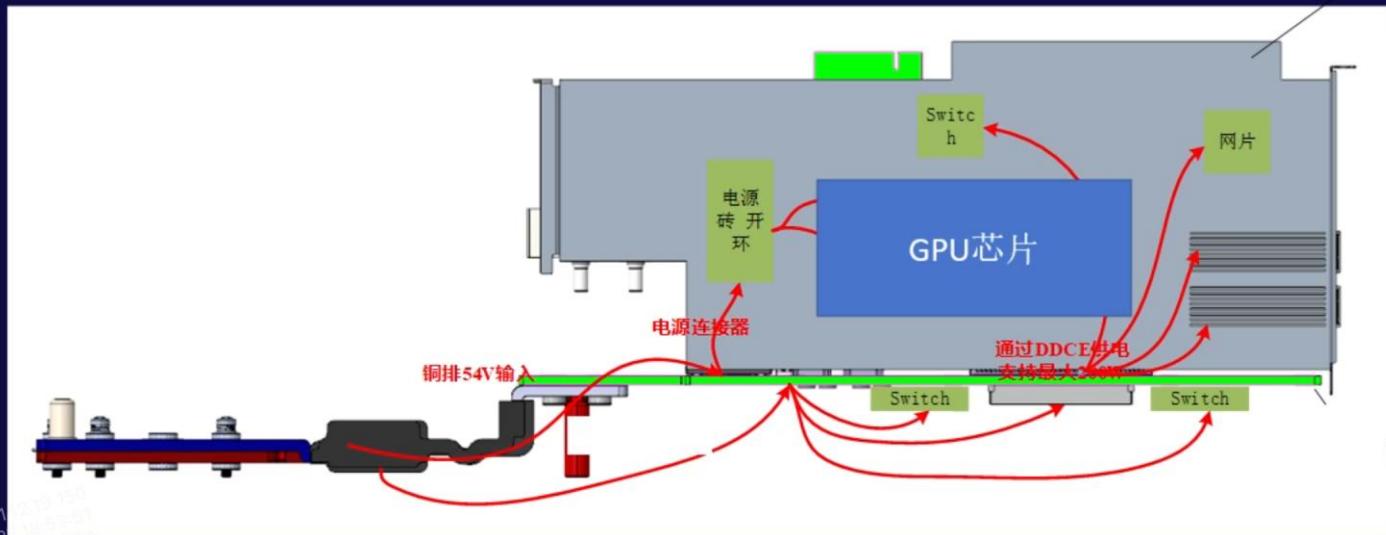
## ■ SW互联底板

- 将SW互联底板置于bottom面，通过液冷进行散热；
- SW互联板的正面为16组电源+高速插槽，盲插后与SW实现全互联。



# UXM (Unified XPU Module) Type2方案介绍

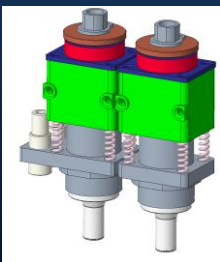
- 单卡面积提升
  - 单卡尺寸323\*180mm, 可支持国产越来越大面积的GPU芯片;
- 单卡厚度提升
  - 单卡厚度: 23.52mm, 冷板设计更优, 可支持更大功耗;
- 单卡功耗提升
  - 54V供电 (降低汇流排数量), PSU数量增加, 快接头数量增加, 可支持MQDB04, 单卡从功耗和散热最大可支持1500W+;





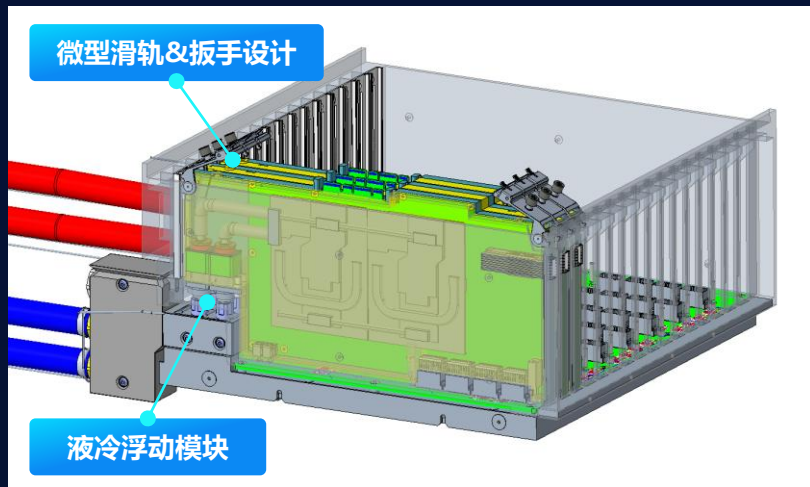
## 液冷接头浮动设计

- MQDB04自身浮动在三盲插中极值不满足
- 设计MQDB04浮动模块，增加轴向、径向浮动，满足公差分析要求



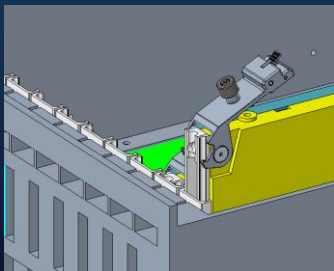
## 微型滑轨&扳手设计

## 液冷浮动模块



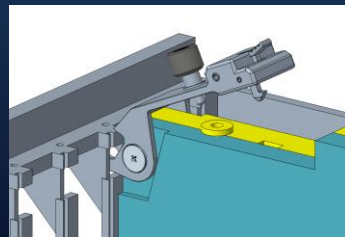
## 微型滑轨导向设计

- 压铸围框加工成本高，精度难保证
- 验证微型滑轨滑移精度，满足插入行程的导向精度要求



## 扳手设计

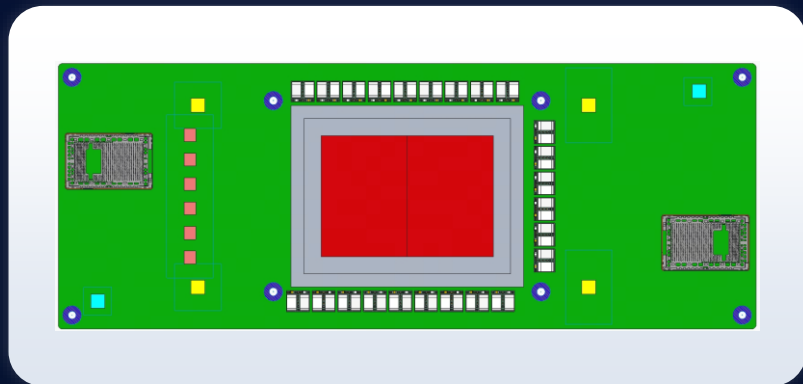
- 更换radsok及PLD连接器，增加稳定性，最大插入力增至400N
- 扳手助力比4.6，单扳手最大施力50.5N

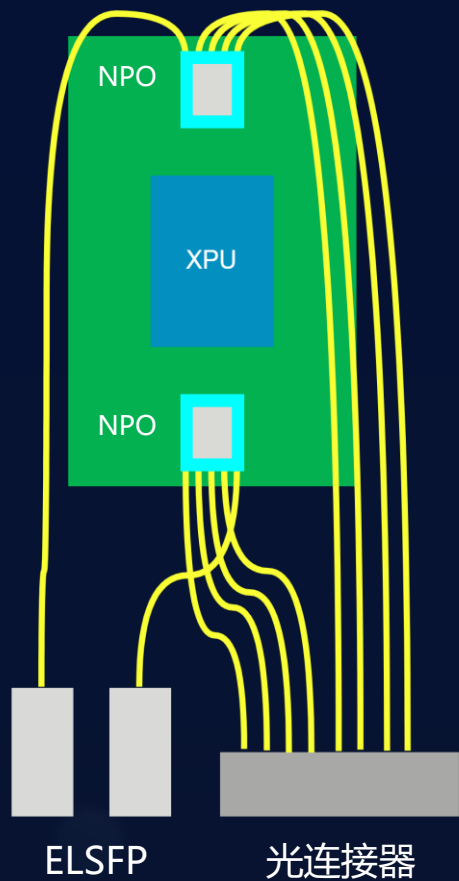


|     |    |        | 结构 | 散热  | 布局&SI | 连接器 |
|-----|----|--------|----|-----|-------|-----|
| 8卡  | 功耗 | 1500W  | 可行 | 可行  | 可行    | 可行  |
|     |    | 3000W  | 可行 | 难度中 | 可行    | 可行  |
|     | 互联 | 64lane | 可行 | /   | 可行    | 可行  |
| 16卡 | 功耗 | 1500W  | 可行 | 难度中 | 可行    | 可行  |
|     |    | 3000W  | 可行 | 难度大 | 难度中   | 可行  |
|     | 互联 | 64lane | 可行 | 可行  | 难度中   | 可行  |

# >> UXM Type3: NPO XPU模组规格定义

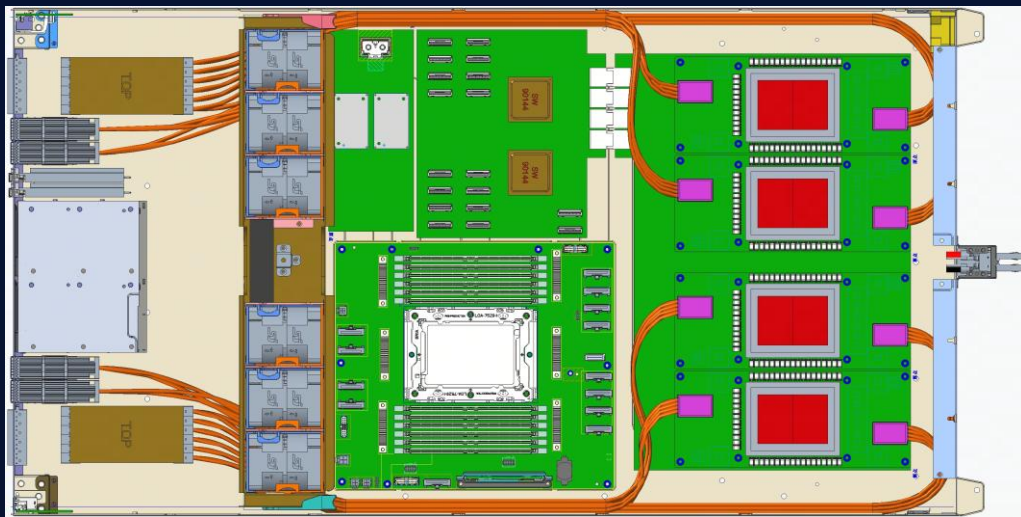
| 规格   | 参数                                    |
|------|---------------------------------------|
| 尺寸   | 长约300mm, 宽约114mm                      |
| 供电   | 54V                                   |
| 模组功率 | 3000W/支持垂直供电                          |
| 上行接口 | 1 or 2 x16 host link (PCIe Gen6/Gen7) |
| 互联接口 | 3.2T/6.4T/12.8T NPO                   |
| 散热   | 仅支持液冷                                 |





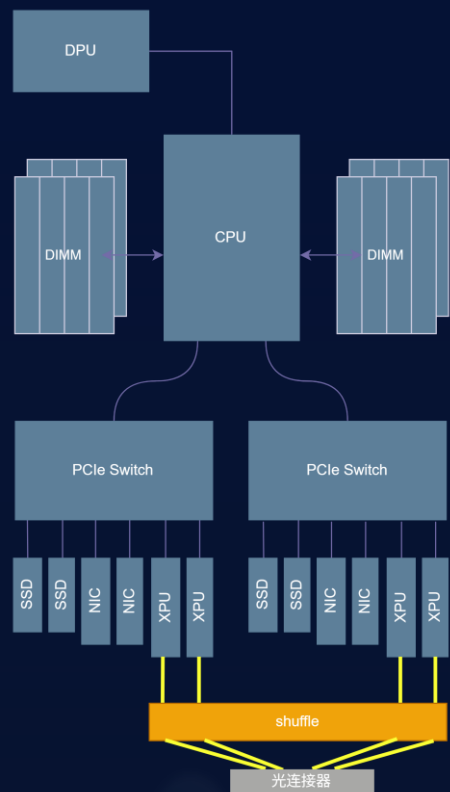
XPU模组互联示意

- NPO尽量靠近GPU，减少链路损耗；
- 突破单柜的限制，搭配NPO交换机实现GPU光互联，更大的一级scale up域；
- 节省UBB和PHY Retimer，更多的空间给到模组；
- 根据需求灵活选择XPU模组数量，支持后续扩大规模。

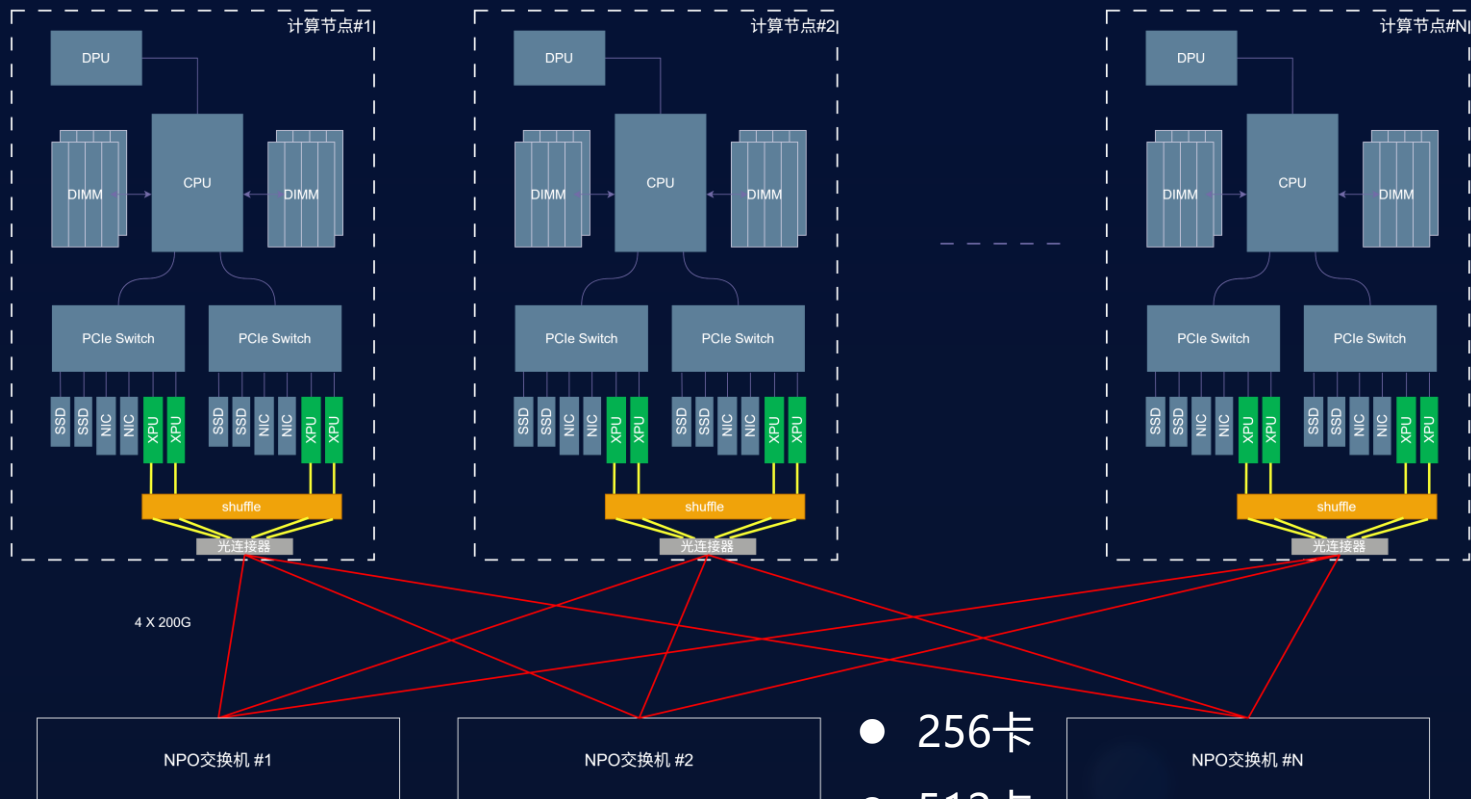


注：所有规格均为初步拟定，非定版，欢迎一起探讨。

# UXM Type3: NPO XPU模组—典型应用场景



NPO XPU模组节点框



XPU模组光互联

注：所有规格均为初步拟定，非定版，欢迎一起探讨。

## 行动呼吁

感谢您对 XPU 模组标准化的关注  
诚邀您继续参与，提供宝贵反馈，携手助力生态系统的发展

V0.5

2026年Q4

V0.8

2027年Q1



模组意见收集表（飞书）

### 反馈方式

zhangguangbin@gccorg.com

wanghaifeng@gccorg.com



模组标准化项目群（微信）

# Thank You !

