



下一代AI算力模组标准化 加速产业化落地

汪新新

新华三服务器系统部 资深系统工程师



国内算力模组标准化的迫切需求

基于计算节点尺寸，最大化算力模组尺寸；对算力模组的接口、规格做定义，给XPU厂家最大化灵活设计空间和多规格选择；补齐国内算力模组标准化的空白



国内算力模组标准化空白

国内长期依赖国外OCP-OAM模组规范开发产品，但后续国外OCP-OAM规范不再更新，亟需国内算力模组规范带动AI产业标准化。



XPU芯片厂家多

国内XPU厂家达到20+家，且各家芯片规格差异大，一家一设计给AI超节点产业落地带来巨大挑战。



高密硬件设计复杂度高

超节点整机柜属于高密产品，寸土寸金，极致的空间内做最大化的兼容会给硬件设计标准化带来挑战。

国内算力模组标准化的迫切需求

尺寸

功耗

互联

连接器

接口定义

OCP- 《OAM2.0规范》, 迭代?

面板布局

内部布局

UBB尺寸

功耗

接口定义

GCC-?

拓扑

机柜

节点

供电

散热

互联

固件

部署交付

GCC-
《液冷整机柜系统架构设计规范-A 型》

机柜接口

供电接口

接地接口

液冷接口

风冷接口

GCC- 《液冷整机柜接口规范》

建筑要求

供配电系统

制冷系统

布线

交付和运维

GCC- 《AIDC基础设施规范》

8卡-OAM服务器模组标准化历程

通过标准解决实际问题，极具前瞻性，灵活调整规范方案

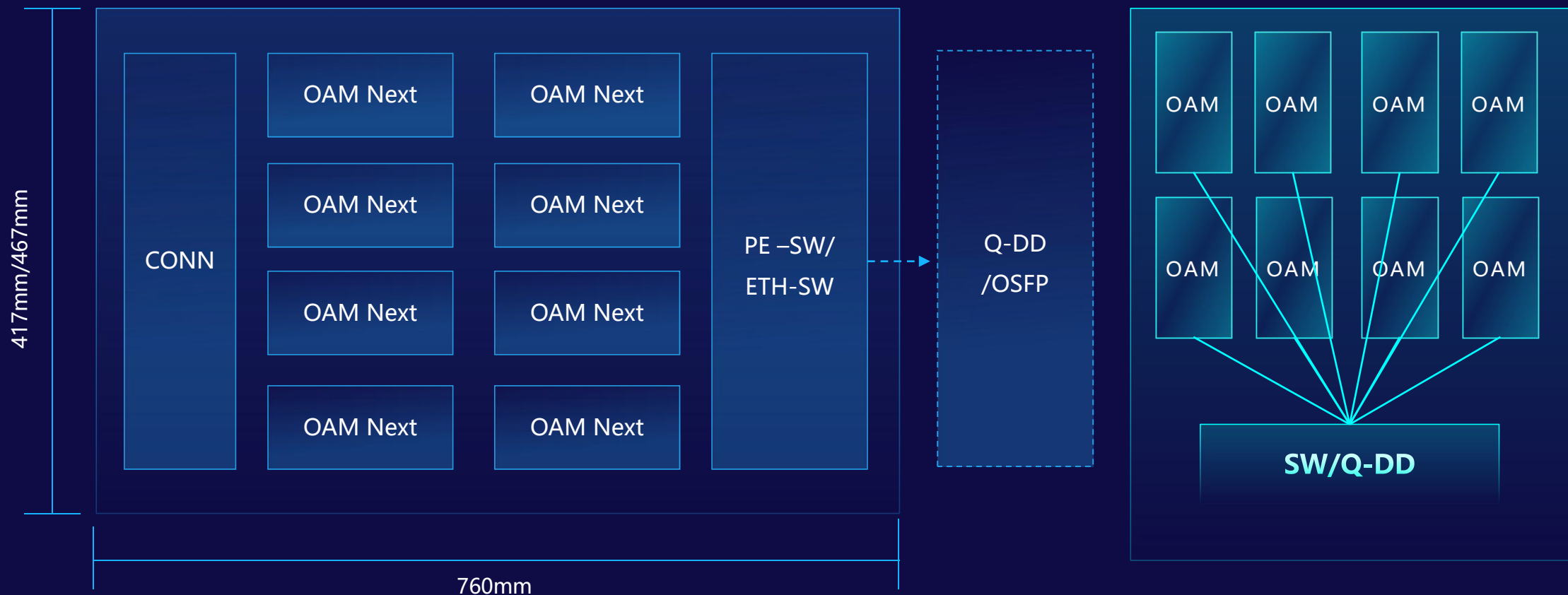


下一代算力模组标准化分析



服务器算力模组标准化分析—UBB Next

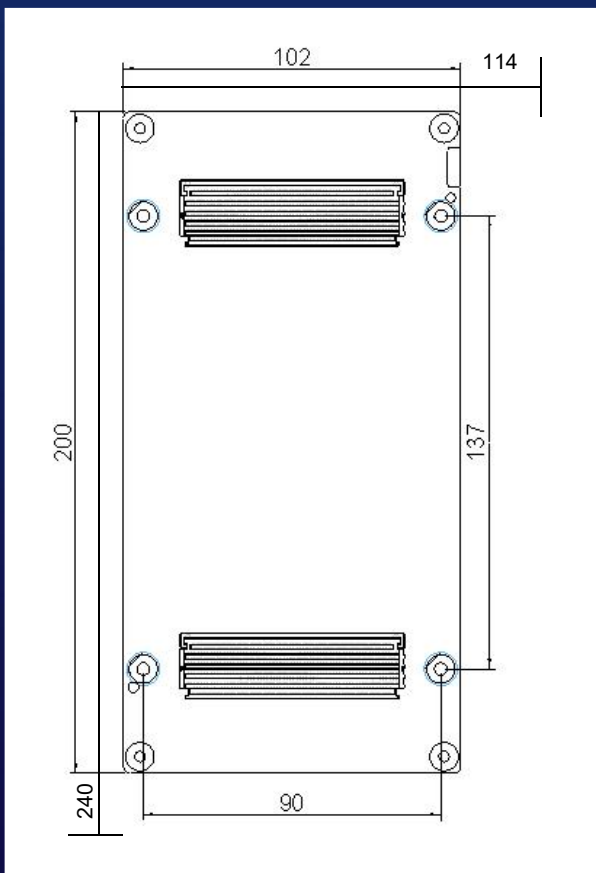
在UBB内实现SW互联，同时支持全scale-out，重新定义UBB Next的尺寸和接口



注：所有规格均为初步拟定，非定版，欢迎一起探讨。

服务器算力模组标准化分析—OAM Next

重新定义OAM Next的尺寸和接口

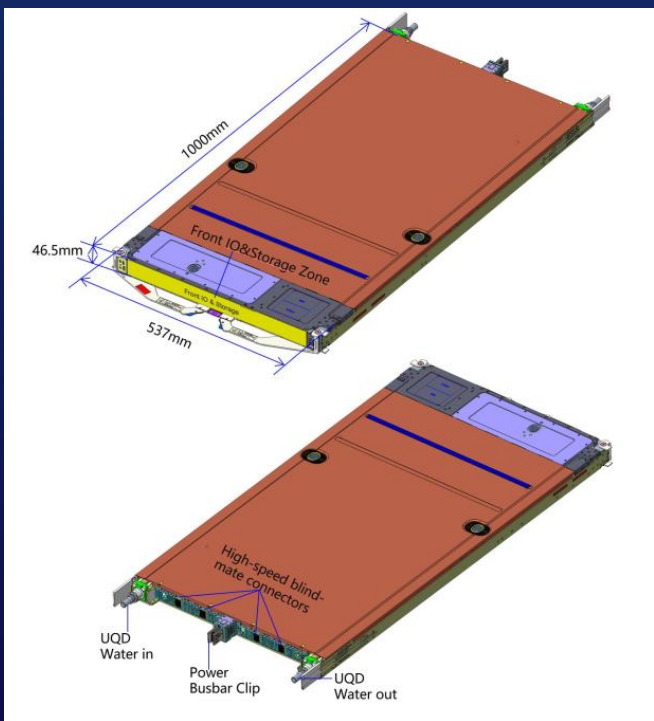


规格	参数
尺寸 (不含NPC/NPO)	200mm*102mm (规格1) --19inch 200mm*114mm (规格2) --19inch旋转90度/21inch 240mm*114mm (规格3) --21inch
尺寸 (含NPC/NPO)	300mm*102mm --19inch 300mm*114mm --21inch
供电	48V/54V and 3.3V
功耗	3000W, 仅液冷
互联接口	224G, 最大可支持7*X16
上行接口	1 or 2*PCIe6.0*X16/1 or 2*PCIe7.0*X16

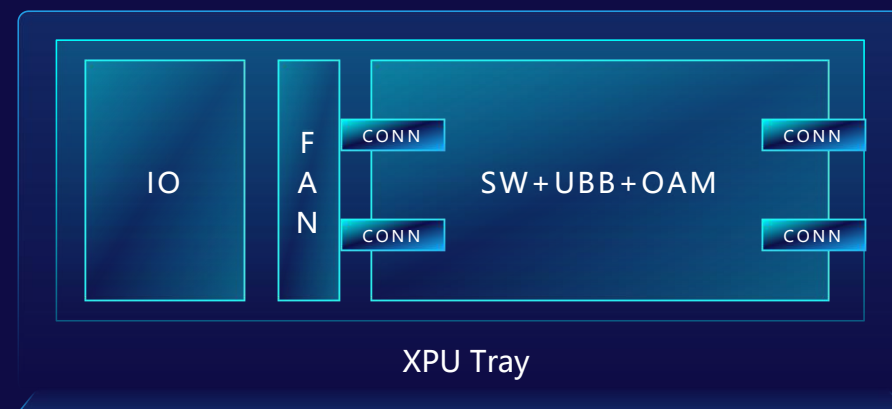
注：所有规格均为初步拟定，非定版，欢迎一起探讨。

当前超节点 (scale up) 算力模组标准化分析

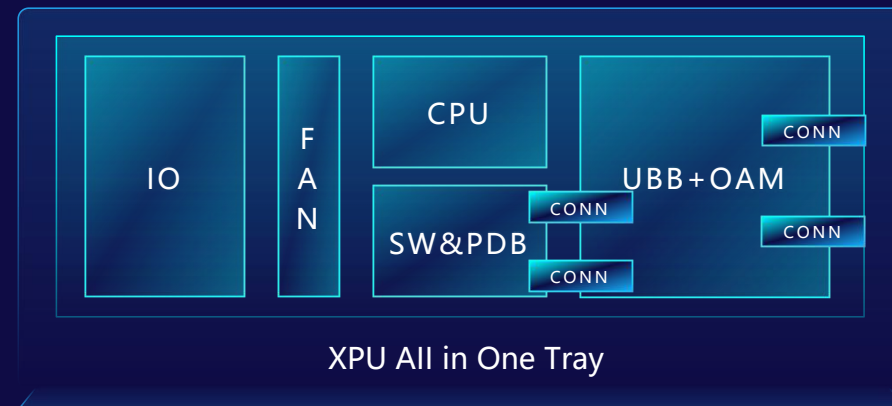
- 大禹架构已对XPU Tray和XPU All in One Tray定义了尺寸, 前后面板初步布局, 扳手设计, 浮动模块等
- 可以进一步定义节点内部的空间布局, 尺寸定义, 功耗定义, 液冷管路设计, 接口定义



XPU Tray & XPU All in One Tray



XPU Tray

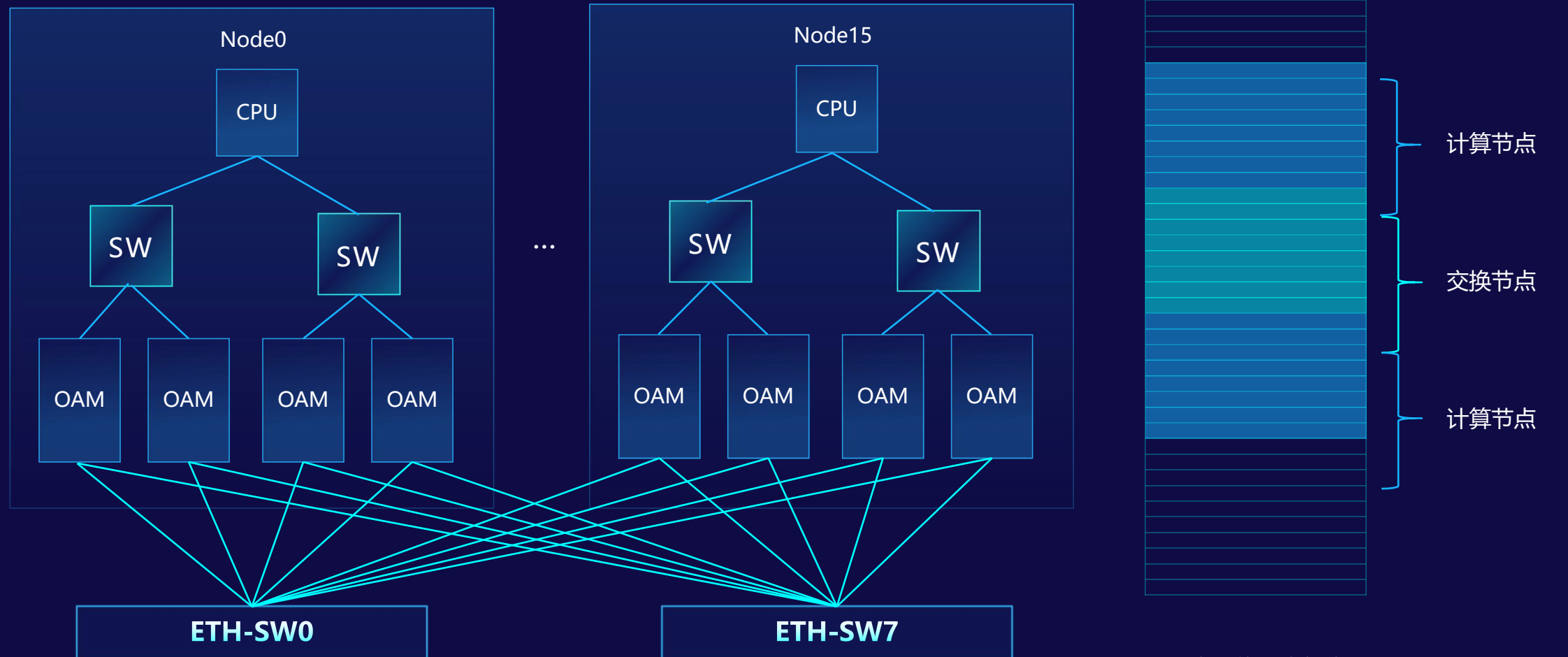


XPU All in One Tray

注: 所有规格均为初步拟定, 非定版, 欢迎一起探讨。

当前超节点 (scale up) 算力模组标准化分析

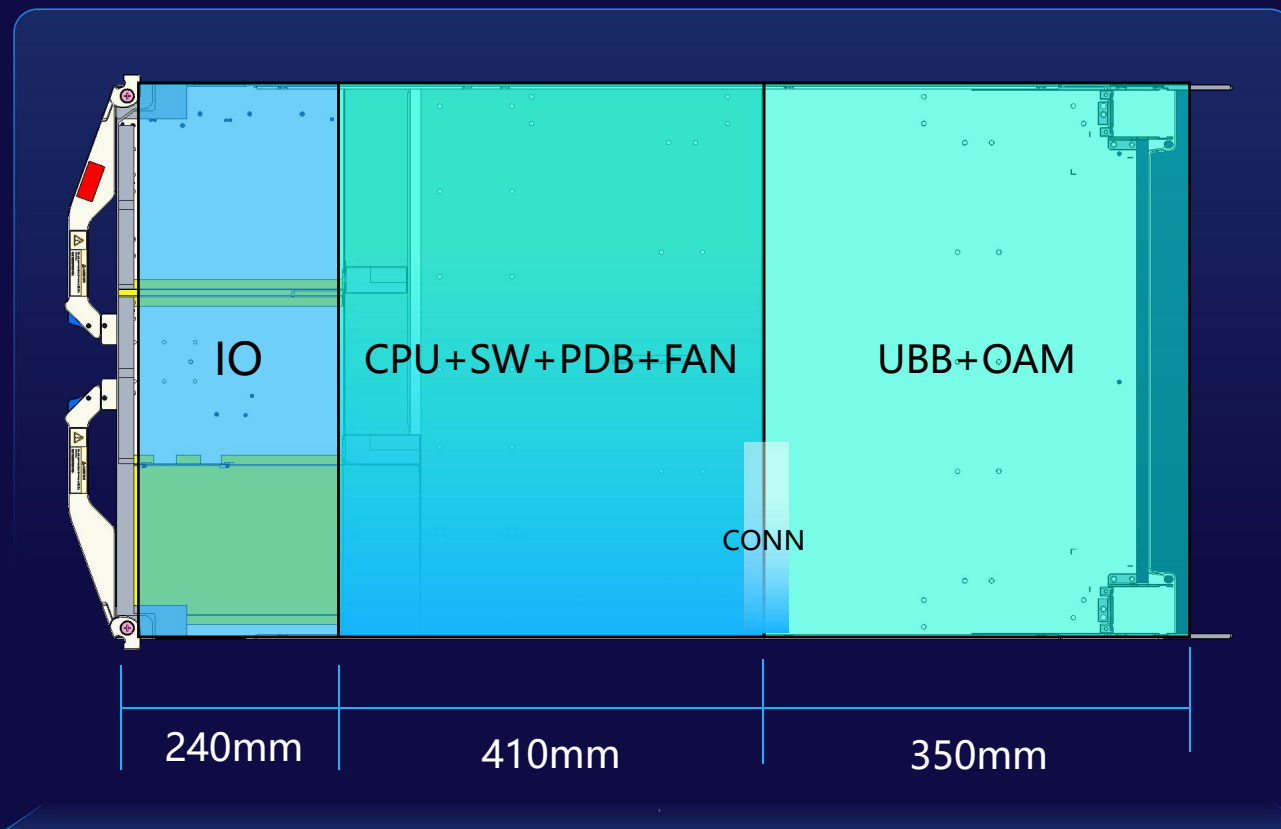
一体机方案：单柜最大可部署64*GPU



注：所有规格均为初步拟定，非定版，欢迎一起探讨。

当前超节点 (scale up) 算力模组标准化分析

一体机方案：算力模组与主机连接器对插

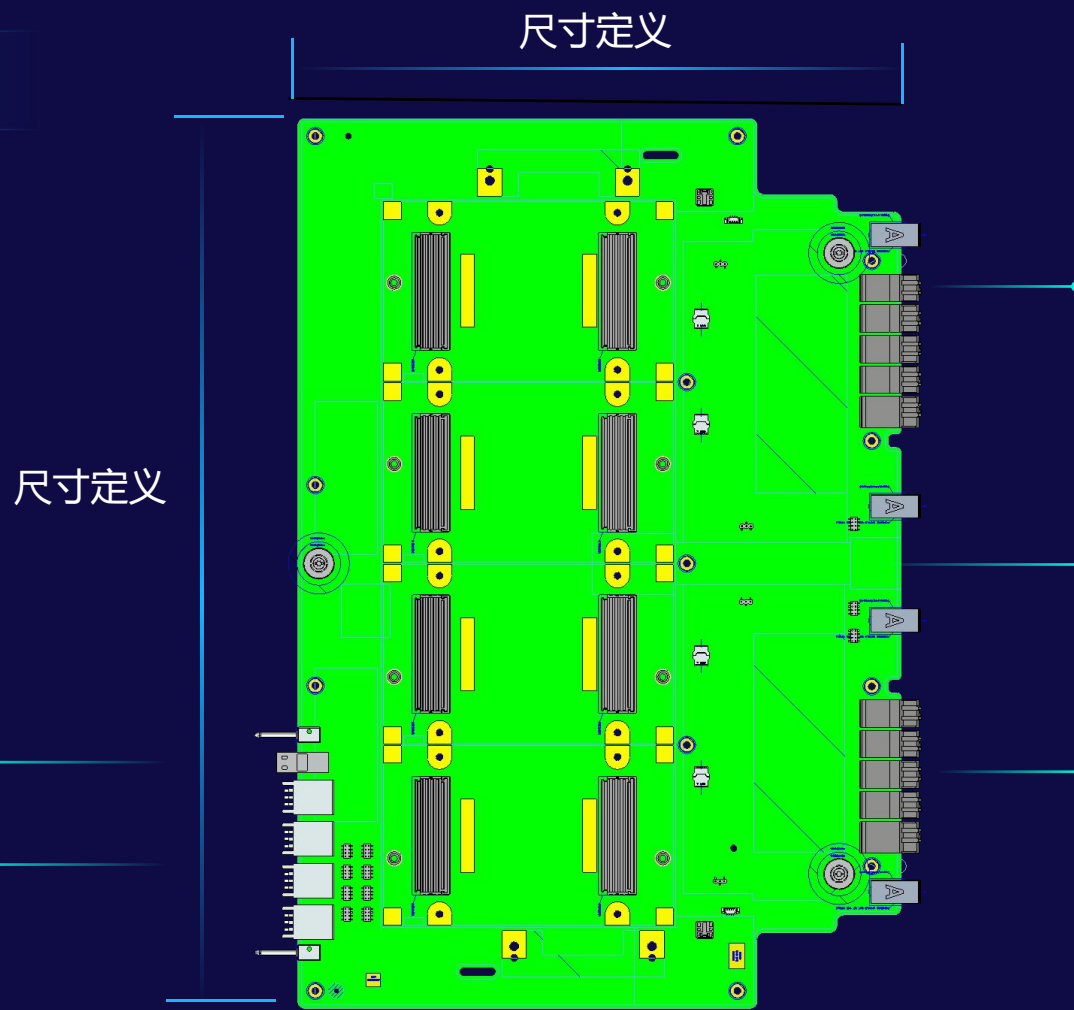


区域	规格	参数
左	尺寸	240mm*510mm
	IO配置	NIC*4+SSD*6+DPU*1+IO*1+raid卡*1
	IO布局	NIC SSD DPU
中	尺寸	410mm*510mm
	连接器选型	厂家、型号
	接口Pinmap	与UBB的接口明确定义
右	尺寸	350mm*510mm
	连接器选型	厂家、型号
	接口Pinmap	与SW的接口, 与Cable tray的接口

注：所有规格均为初步拟定，非定版，欢迎一起探讨。

当前超节点 (scale up) 算力标准化分析

一体机方案:



- 4*PCIe5.0 / PCIe6.0*X16
- Pinmap: Pinmap定义
- SI: 损耗分配
- 连接器选型

- 协议: 112G, 224G, PCIe等
- Pinmap: 典型pinmap定义
- SI: 损耗分配
- 连接器选型

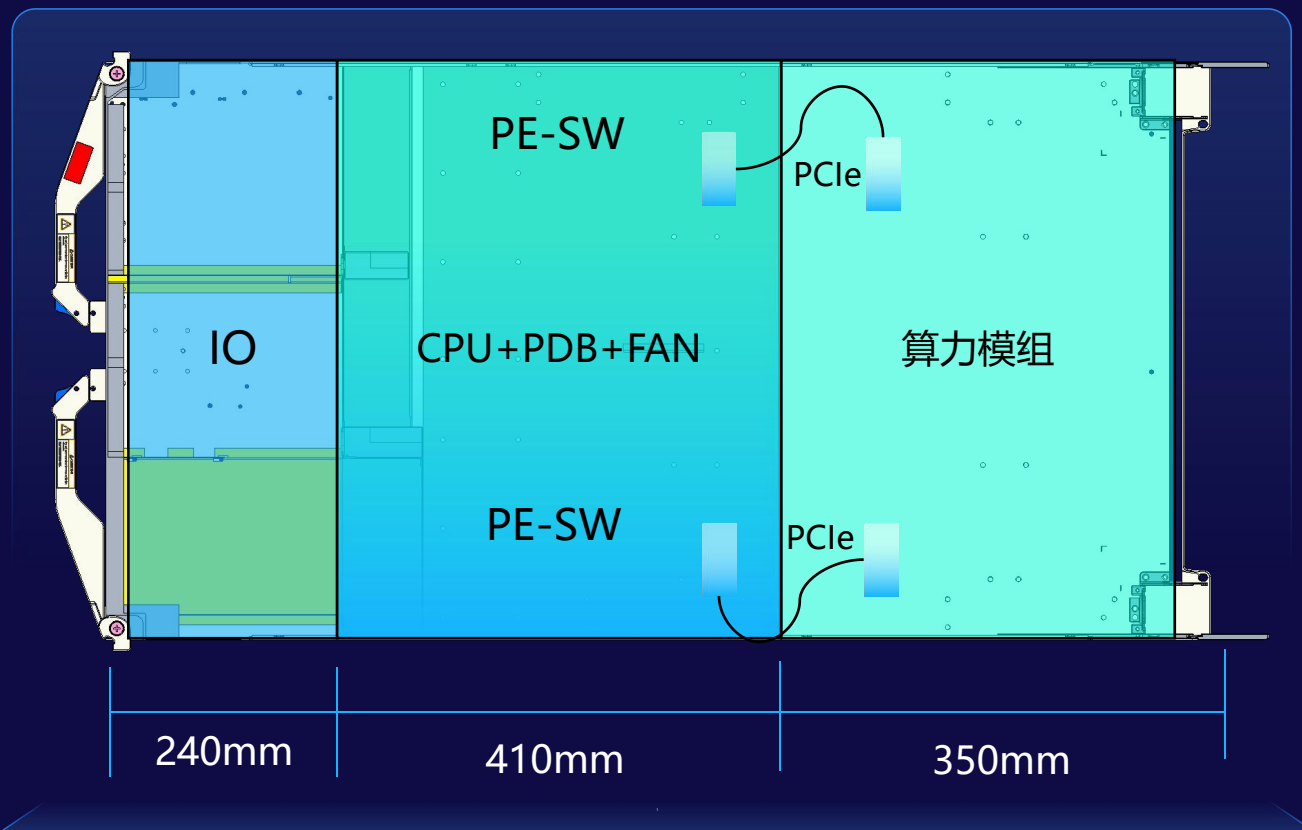
- Busbar (OCP ORV3)
功耗: 支持的最大电流

- 协议: 112G, 224G, PCIe等
- Pinmap: 典型pinmap定义
- SI: 损耗分配
- 连接器选型

注: 所有规格均为初步拟定, 非定版, 欢迎一起探讨。

当前超节点 (scale up) 算力模型标准化分析

一体机方案：算力模组与主机线缆连接

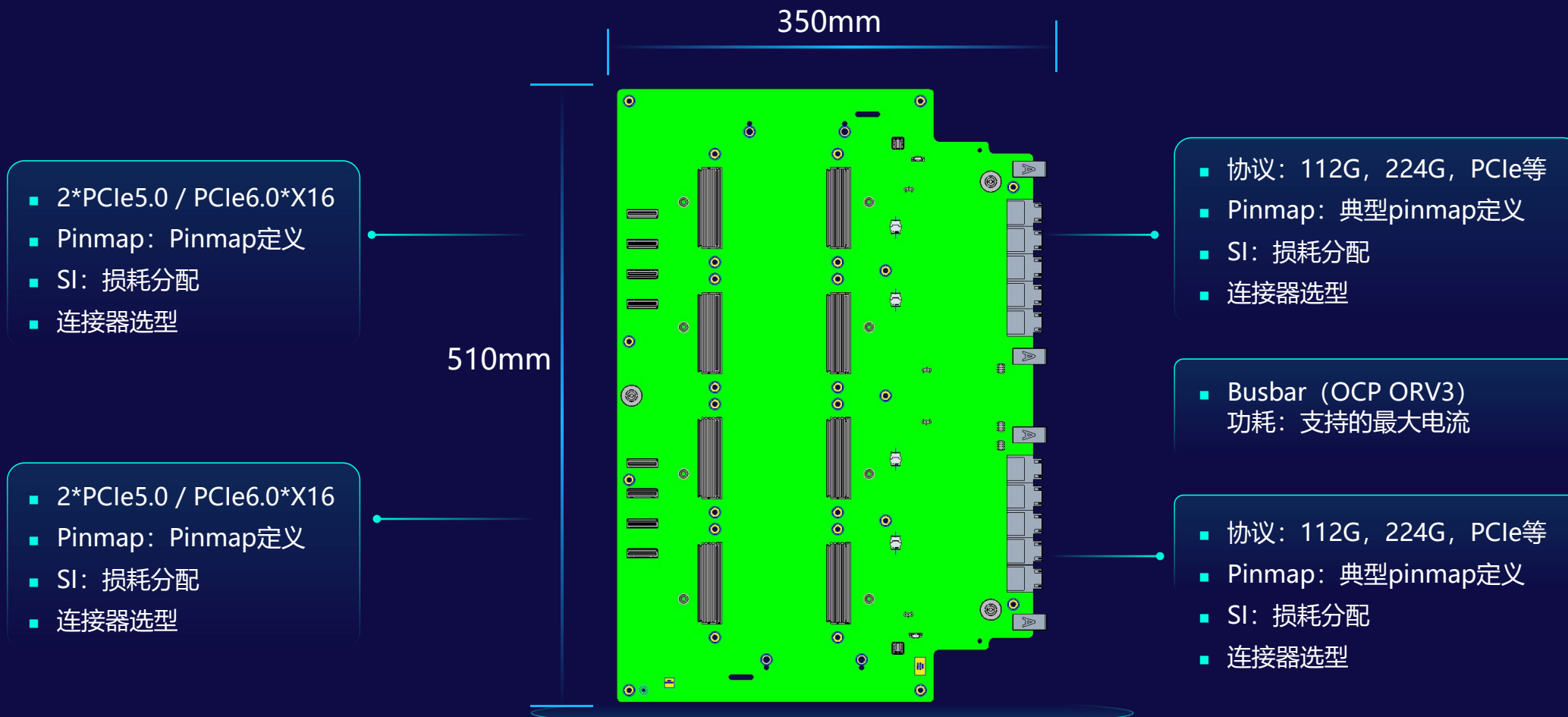


区域	规格	参数
左	尺寸	240mm*510mm
	IO配置	NIC*4+SSD*6+DPU*1+IO*1+raid卡*1
	IO布局	NIC SSD DPU
中	尺寸	410mm*510mm
	连接器选型	厂家、型号
	接口Pinmap	与UBB的接口明确定义
右	尺寸	350mm*510mm
	连接器选型	厂家、型号
	接口Pinmap	与SW的接口, 与Cable tray的接口

注：所有规格均为初步拟定，非定版，欢迎一起探讨。

当前超节点 (scale up) 算力模组标准化分析

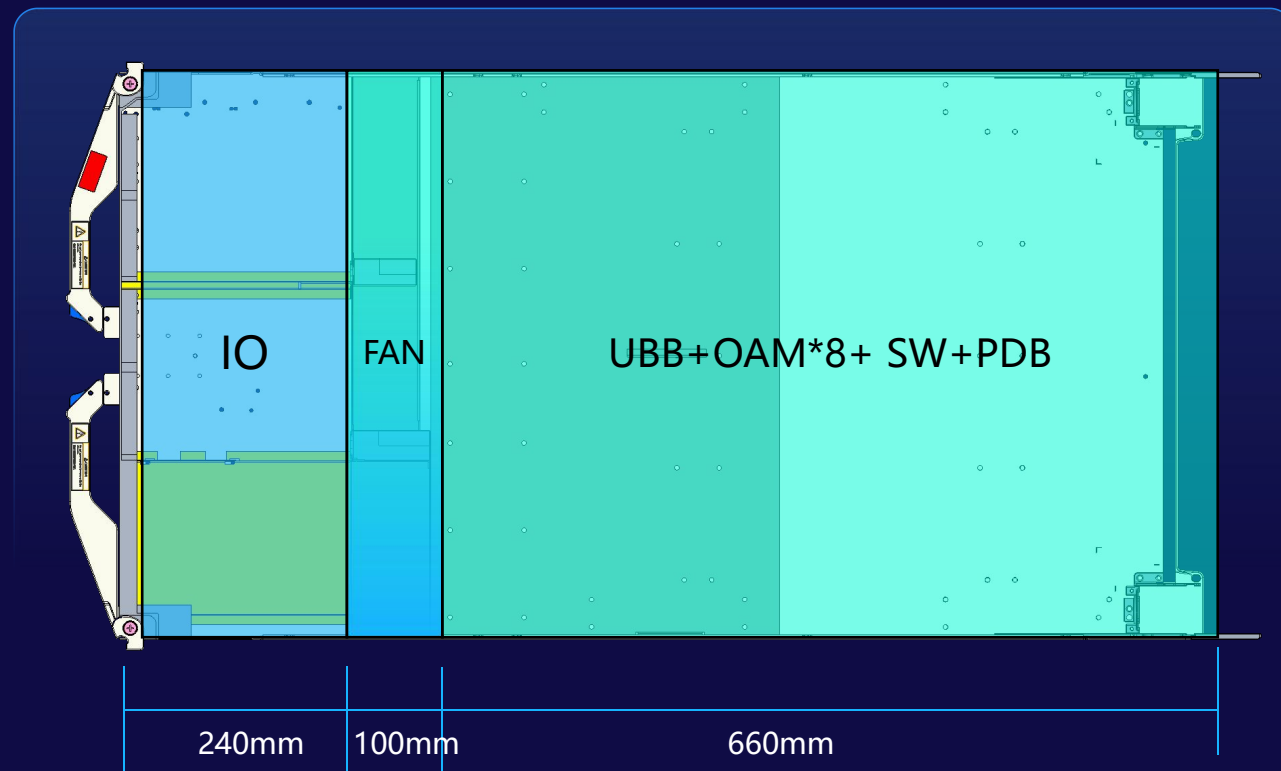
一体机方案：算力模组与主机线缆连接



注：所有规格均为初步拟定，非定版，欢迎一起探讨。

当前超节点 (scale up) 算力模组标准化分析

BOX方案:

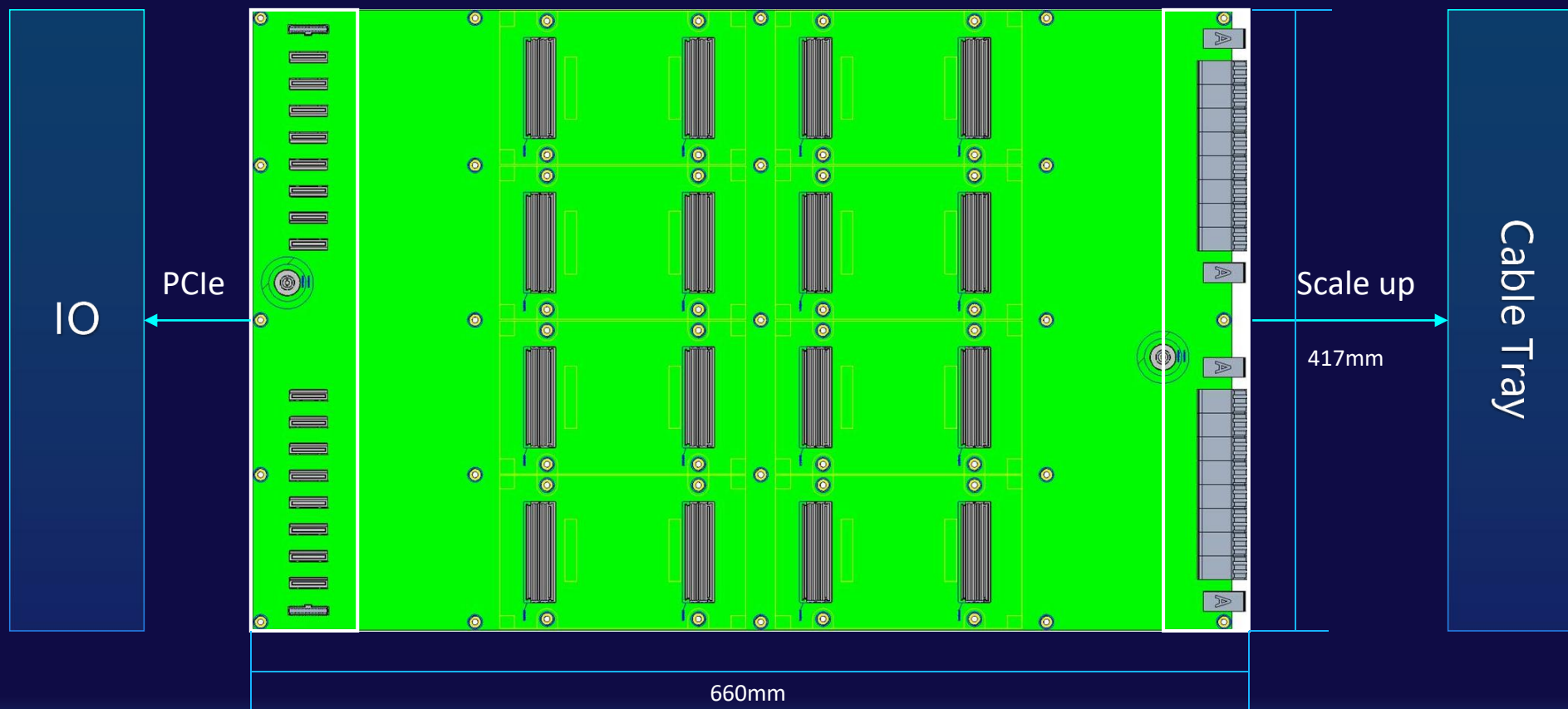


区域	规格	参数
左	尺寸	240mm*510mm
	IO配置	NIC*4+SSD*4+IO*1+CDFP*4
	IO布局	NIC SSD CDFP NIC
右	尺寸	660mm*510mm
	连接器选型	厂家、型号
	接口Pinmap	与IO侧的PCIe接口, 网卡接口, 硬盘接口等

注: 所有规格均为初步拟定, 非定版, 欢迎一起探讨。

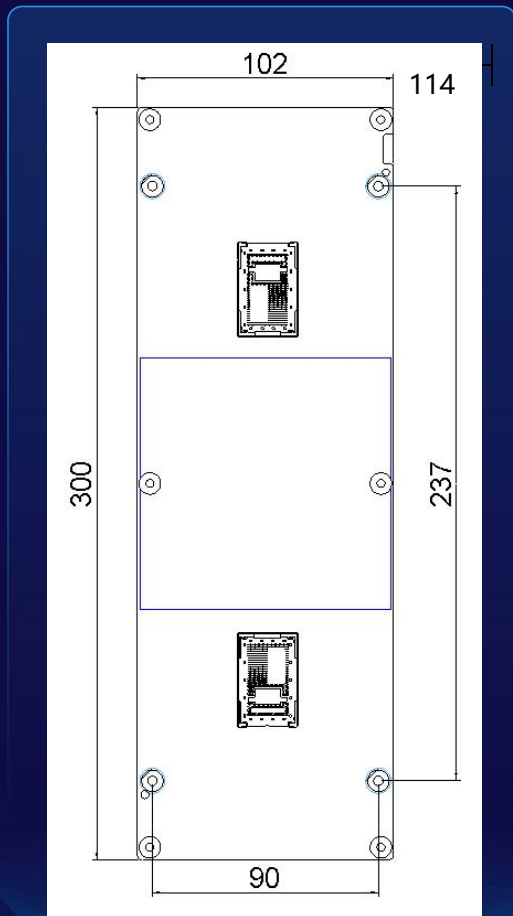
当前超节点 (scale up) 算力模组标准化分析

BOX方案：算力模组与主机线缆连接



注：所有规格均为初步拟定，非定版，欢迎一起探讨。

下一代超节点 (scale up) 算力模组标准化分析

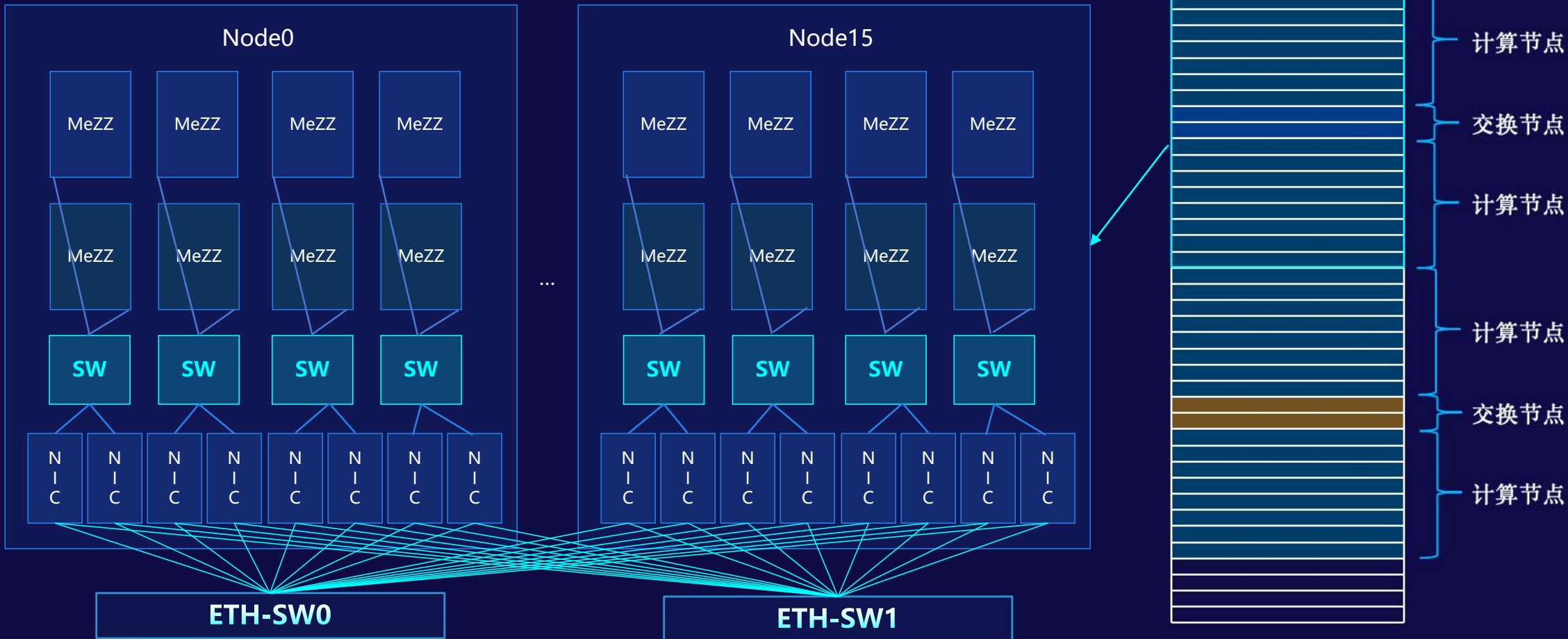


规格	参数
尺寸 (不含NPC/NPO)	200mm*102mm (规格1) --19inch 200mm*114mm (规格2) --19inch旋转90度/21inch 240mm*114mm (规格3) --21inch
尺寸 (含NPC/NPO)	300mm*102mm --19inch 300mm*114mm --21inch
供电	48V/54V and 3.3V
功耗	3000W
互联接口	224G, 最大可支持7*X16
上行接口	1 or 2*PCIe6.0*X16/1 or 2*PCIe7.0*X16

注：所有规格均为初步拟定，非定版，欢迎一起探讨。

超节点 (scale out) 算力模组标准化分析

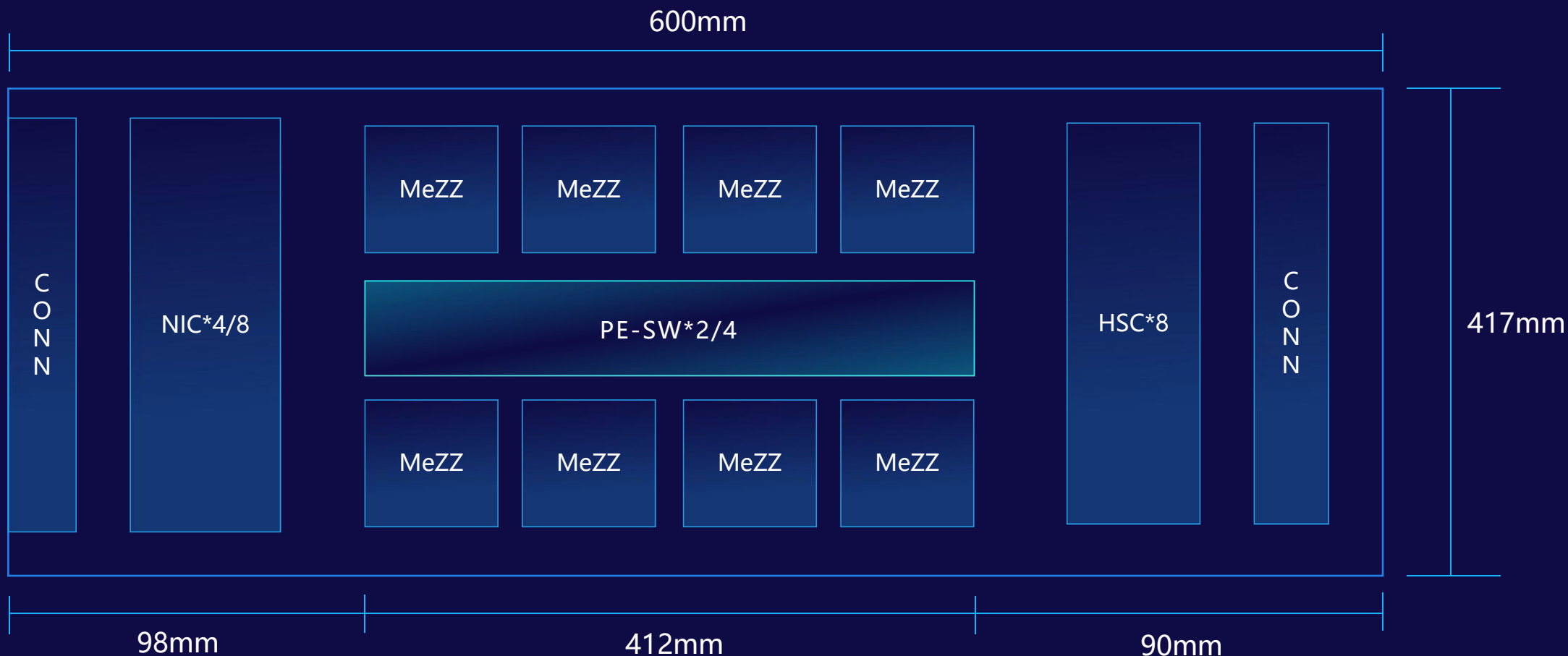
单柜最大可部署256*GPU



注：所有规格均为初步拟定，非定版，欢迎一起探讨。

超节点 (scale out) 算力模组标准化分析

对scale-out超节点的算力模组模块进行尺寸和接口定义



注：所有规格均为初步拟定，非定版，欢迎一起探讨。

算力模组标准化规格构想汇总

架构	类别	规格	参数	
服务器	OAM Next	尺寸 (不含NPC/NPO)	200mm*102mm (规格1) --19inch 200mm*114mm (规格2) --19inch旋转90度/21inch 240mm*114mm (规格3) --21inch	
		尺寸 (含NPC/NPO)	300mm*102mm --19inch 300mm*114mm --21inch	
		供电	54V and 3.3V	
		功耗	3000W, 仅液冷	
		互联接口	224G, 最大可支持7*X16	
	UBB Next	上行接口	1 or 2*PCIe6.0*X16/1 or 2*PCIe7.0*X16	
		尺寸	417mm/467mm*760mm	
		供电	54V	
		接口	8 or 16*PCIe6.0*X16/8 or 16*PCIe7.0*X16	
		超节点 (scale up)	一体机	尺寸
上行接口	4*PCIe5.0/PCIe6.0 *X16			
互联接口	4*32*112G/224G			
BOX	尺寸		660mm*417mm	
	上行接口		4*PCIe5.0/PCIe6.0 *X16	
	互联接口		8*32*112G/224G	
下一代超节点	尺寸等		参考含NPC/NPO的OAM Next	
超节点 (scale out)	1U/2U		尺寸	417mm*600mm,服务器1U/2U
			上行接口	4 / 8*PCIe6.0*X16
		互联接口	8*32*112G/224G	

注：所有规格均为初步拟定，非定版，欢迎一起探讨。

THANK YOU



THANK YOU

THANK YOU