Open AI Infra
A Chapter of GCC

兆瓦级 800V AI 机柜供电研究报告-柜内DC部分解读

Global Computing Consortium
全球计算联盟

1.芯片的功耗趋势

2.HVDC进柜 三种路线分析

3.HVDC进柜的核心挑战

4.HVDC进柜挑战的应对策略

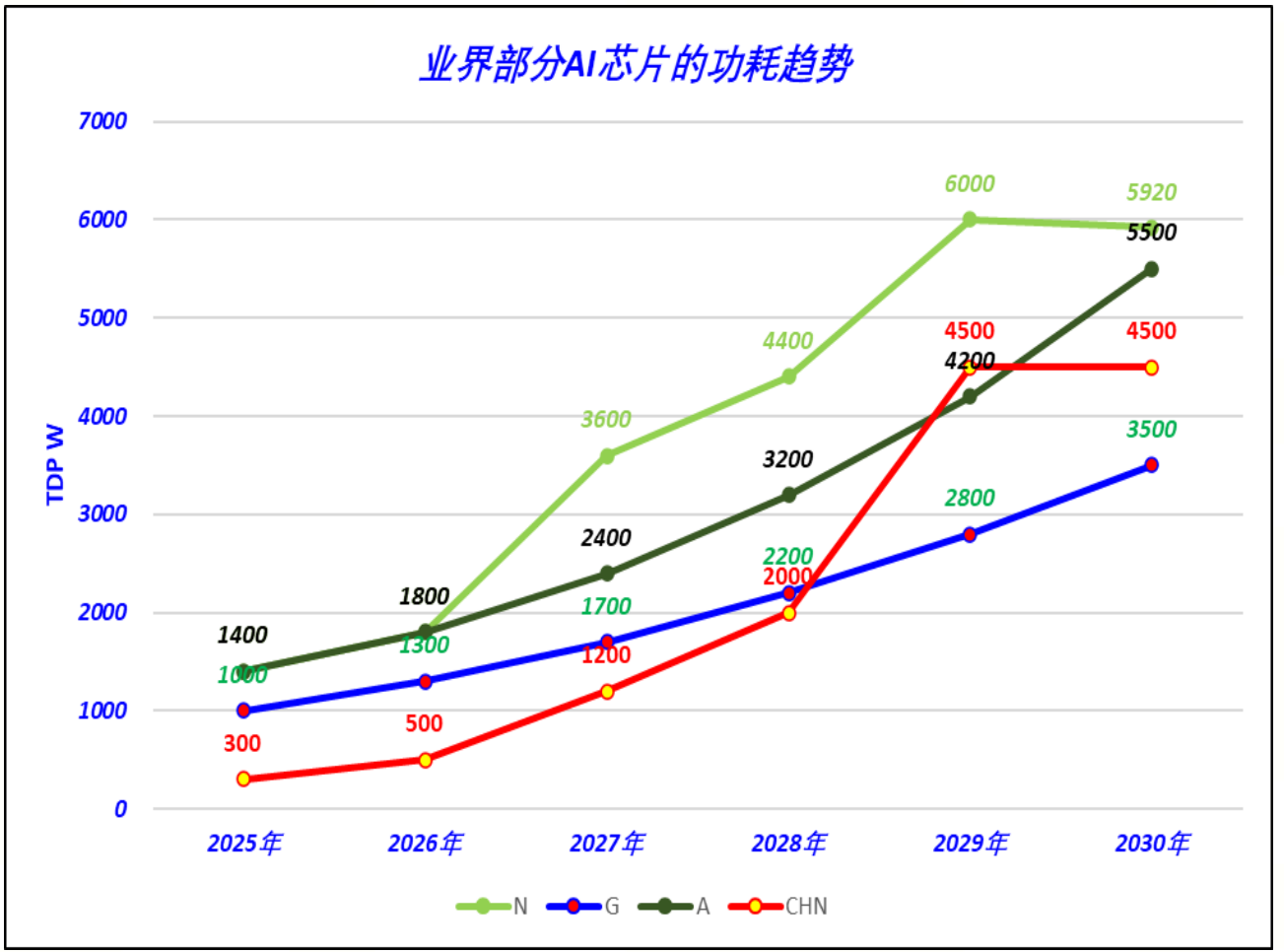
芯片功耗趋势

Vendor	Chip Model	Year	TDP / Power
Nvidia	Blackwell Ultra (GB300)	2025	1,400W
Nvidia	Rubin	2026	1,800W
Nvidia	Rubin Ultra	2027	3,600W
Nvidia	Feynman	2028	4,400W
Nvidia	Feynman Ultra	2029	6,000W
Nvidia	Post-Feynman	2030	5,920W

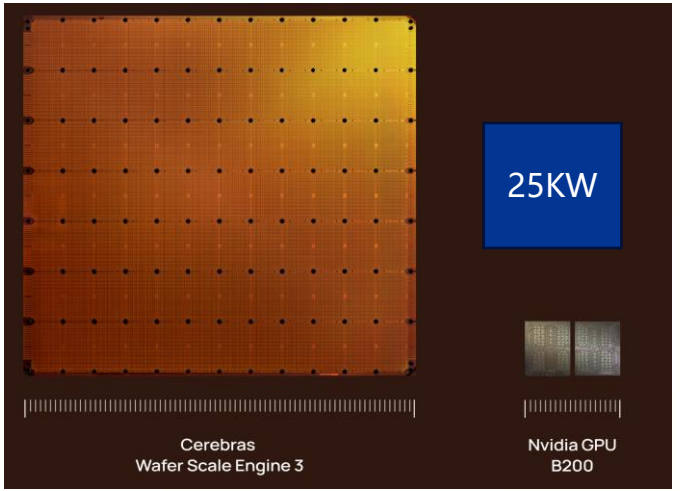
Vendor	Chip Model	Year	TDP / Power
Google	TPU v7 (Ironwood)	2025	1,000W
Google	TPU v8t	2026	1,300W*
Google	TPU v9	2027	1,700W*
Google	TPU v10	2028	2,200W*
Google	TPU v11	2029	2,800W*
Google	TPU v12	2030	3,500W*

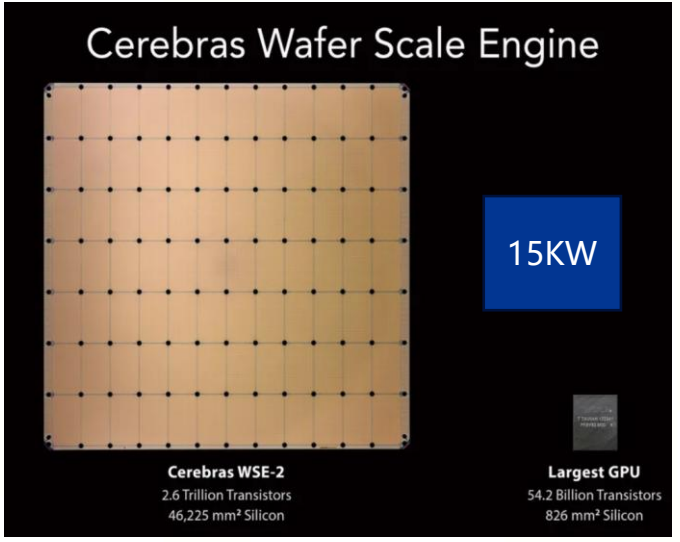
Vendor	Chip Model	Year	TDP / Power
AMD	Instinct MI355X	2025	1,400W
AMD	Instinct MI400	2026	1,800W*
AMD	Instinct MI500	2027	2,400W*
AMD	Instinct MI600	2028	3,200W*
AMD	Instinct MI700	2029	4,200W*
AMD	Instinct MI800	2030	5,500W*

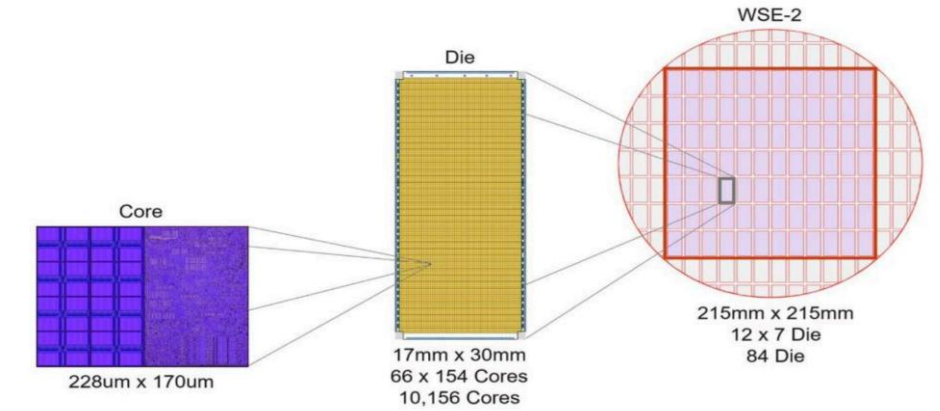
Vendor	Chip Model	Year	TDP / Power
CHN	XXX	2025	300W*
	XXX	2026	500W*
	XXX	2027	1200W*
	XXX	2028	2000W*
	XXX	2029	4500W*
	XXX	2030	4500W*



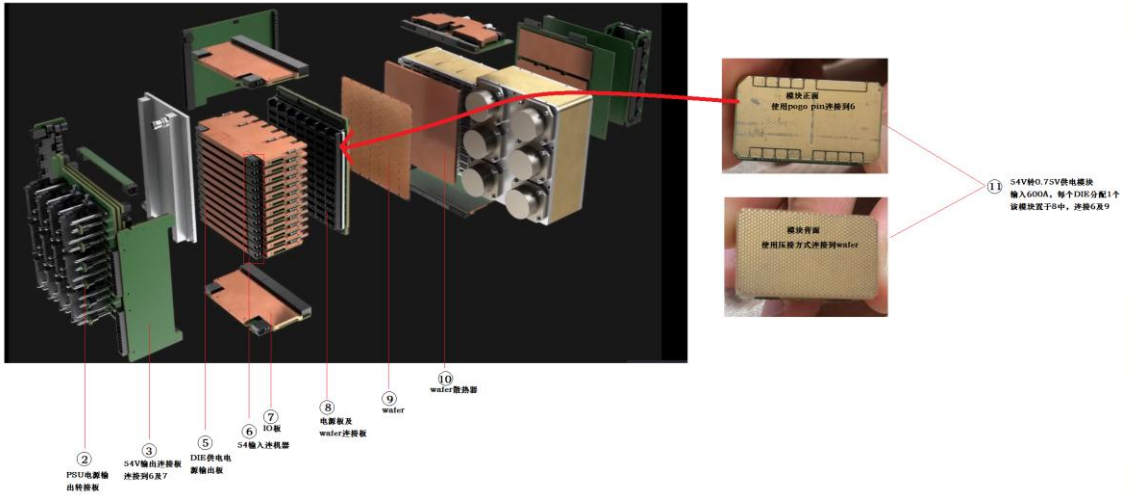
芯片功耗趋势



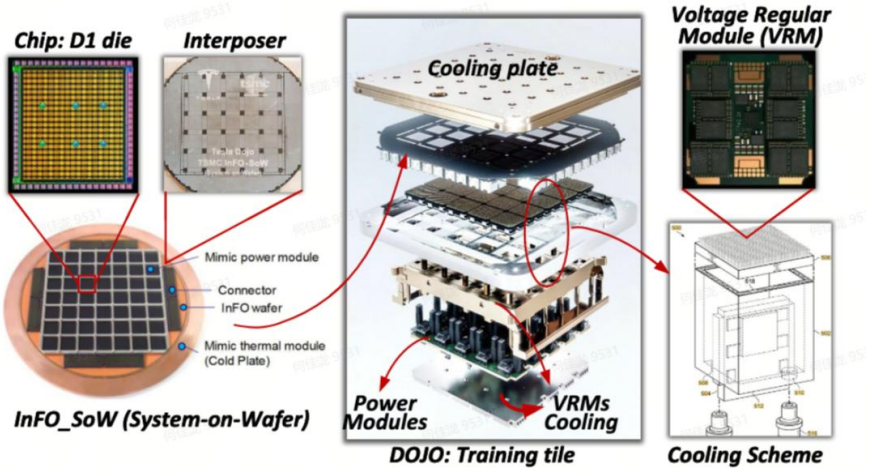




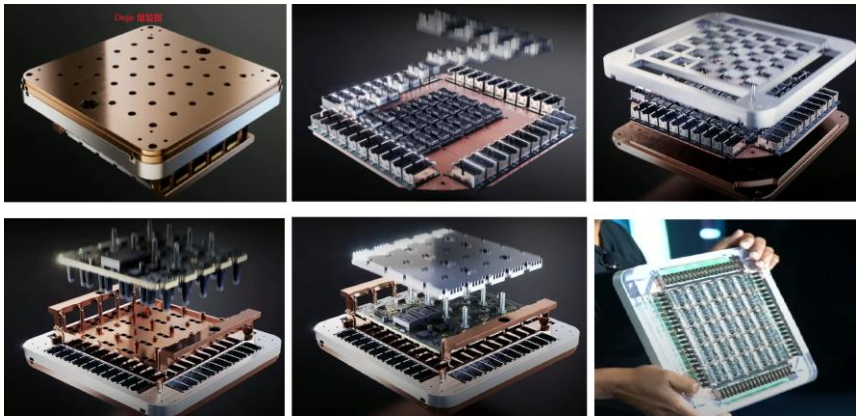
Scaling from small core to entire wafer. WSE: wafer-scale engine.

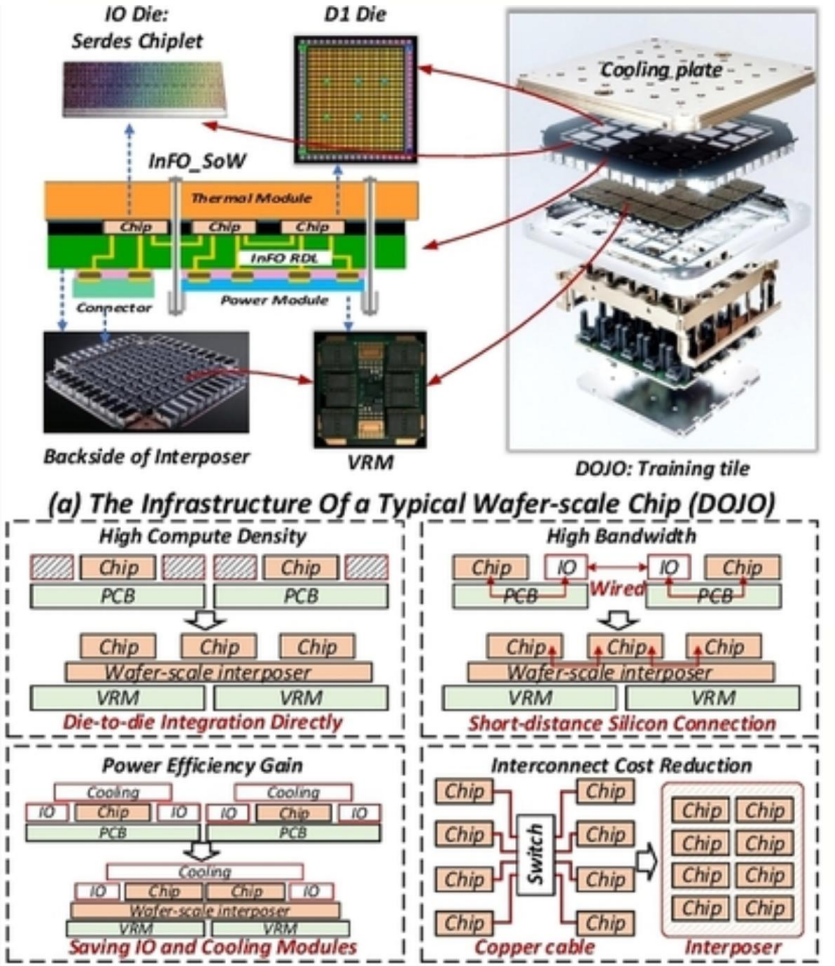


芯片功耗趋势



(a) The Infrastructure Of a Typical Wafer-scale Chip (DOJO)

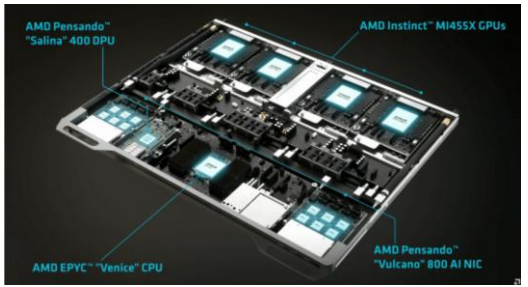


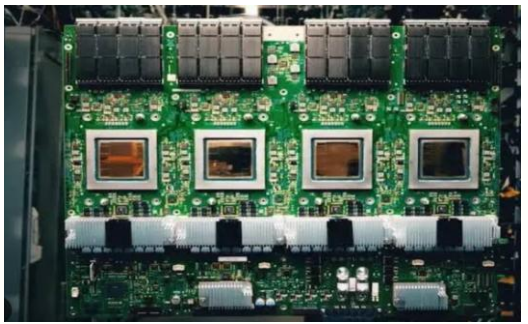


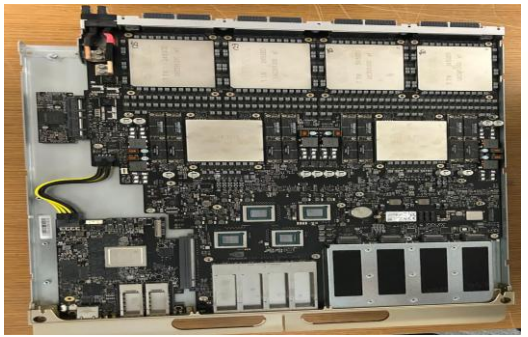
DOJO1
15KW

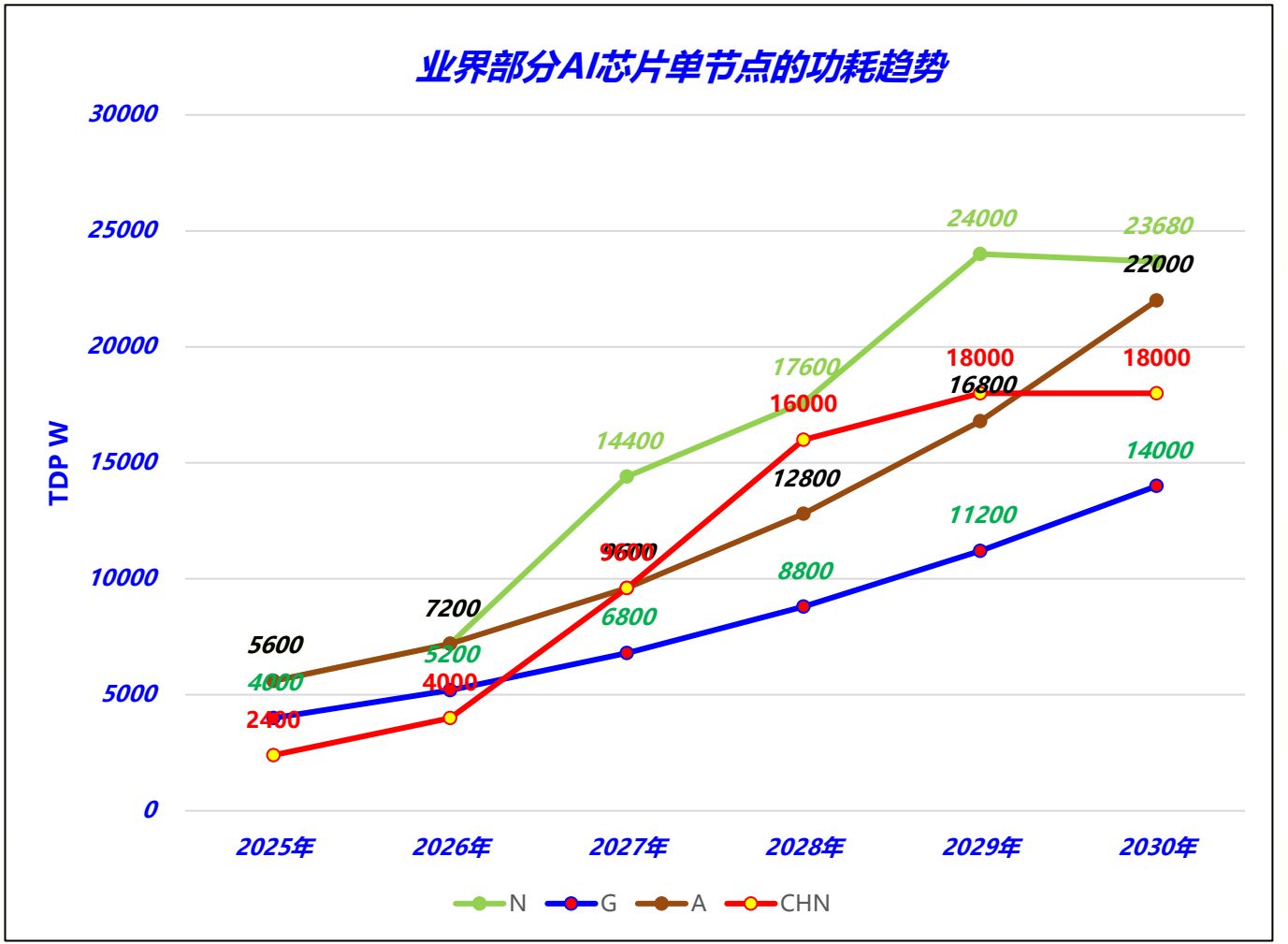
DOJO2
12KW

芯片功耗趋势

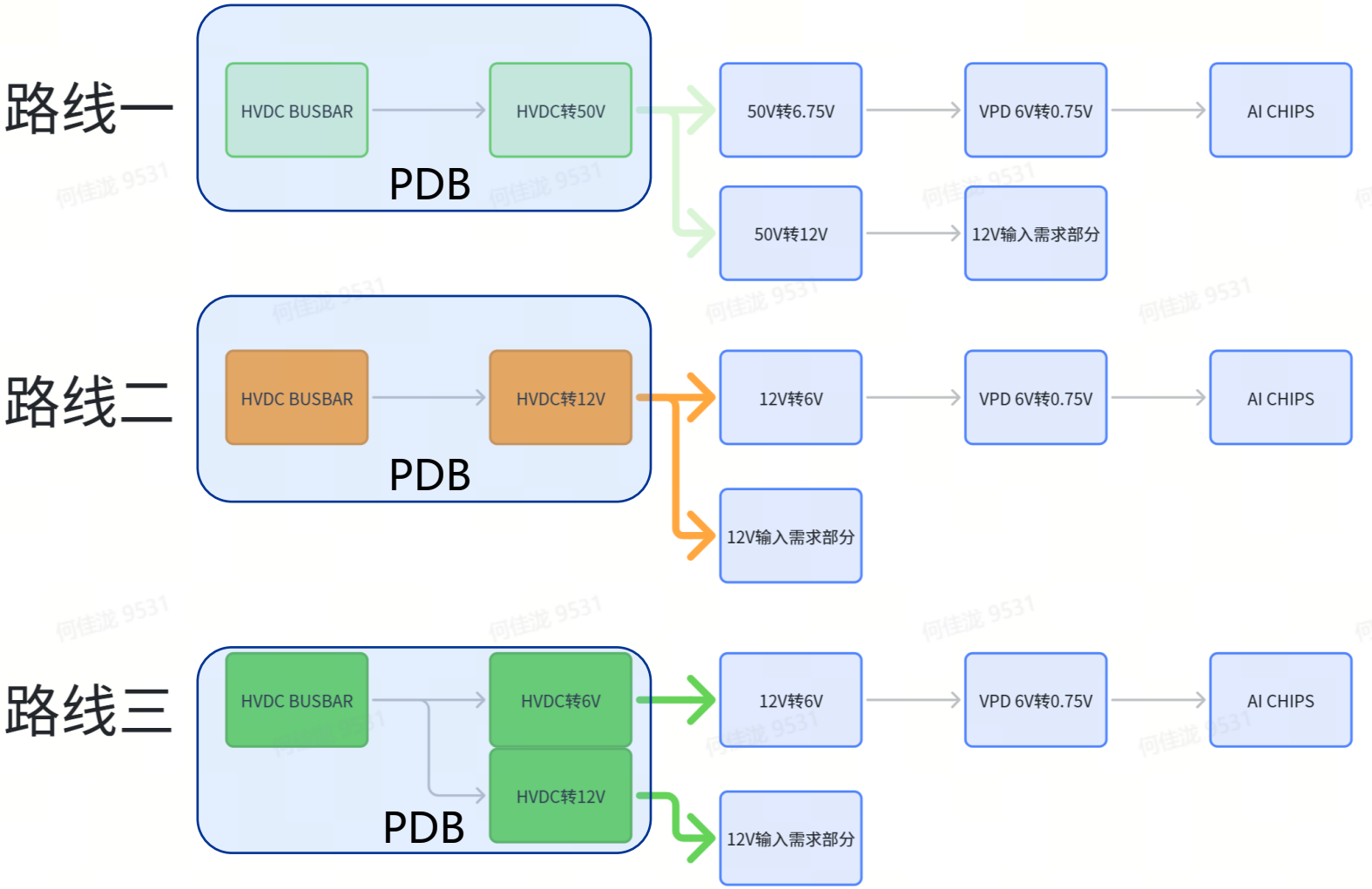

AMD Pensando™
"Salina" 400 DPU
AMD Instinct™ MI455X GPUs
AMD EPYC™ "Venice" CPU
AMD Pensando™
"Vulcano" 800 AI NIC






业界部分AI芯片单节点的功耗趋势
TDP W
30000
25000
20000
15000
10000
5000
0
2025年
2026年
2027年
2028年
2029年
2030年
N
G
A
CHN
5600
4000
2400
7200
5200
4000
14400
9600
6800
17600
12800
8800
24000
16800
11200
23680
22000
14000
18000
16000
18000

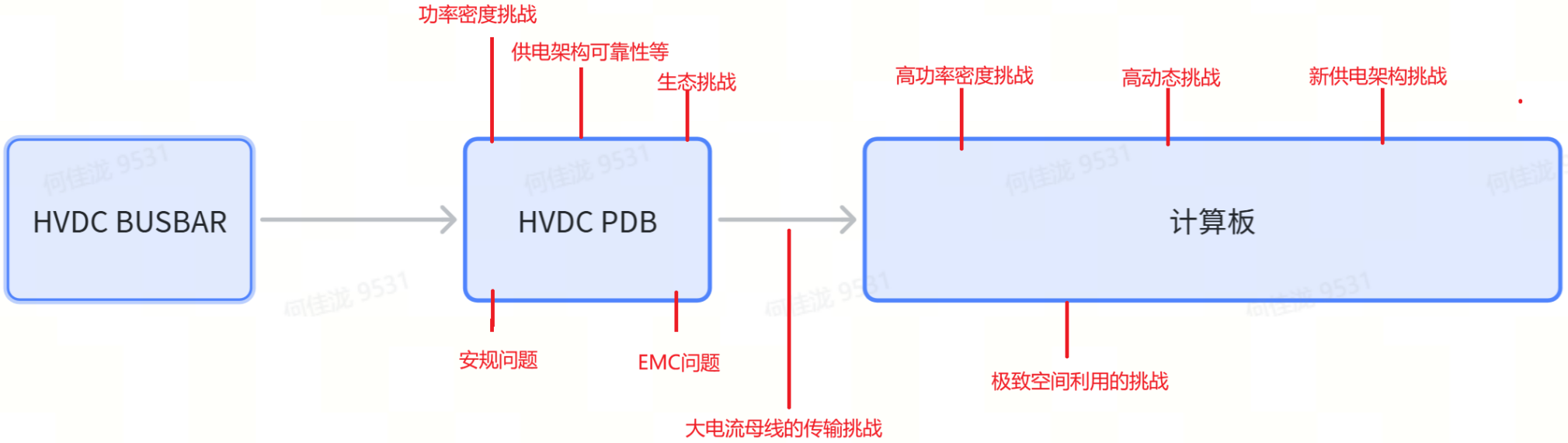
HVDC进柜 三种路线分析



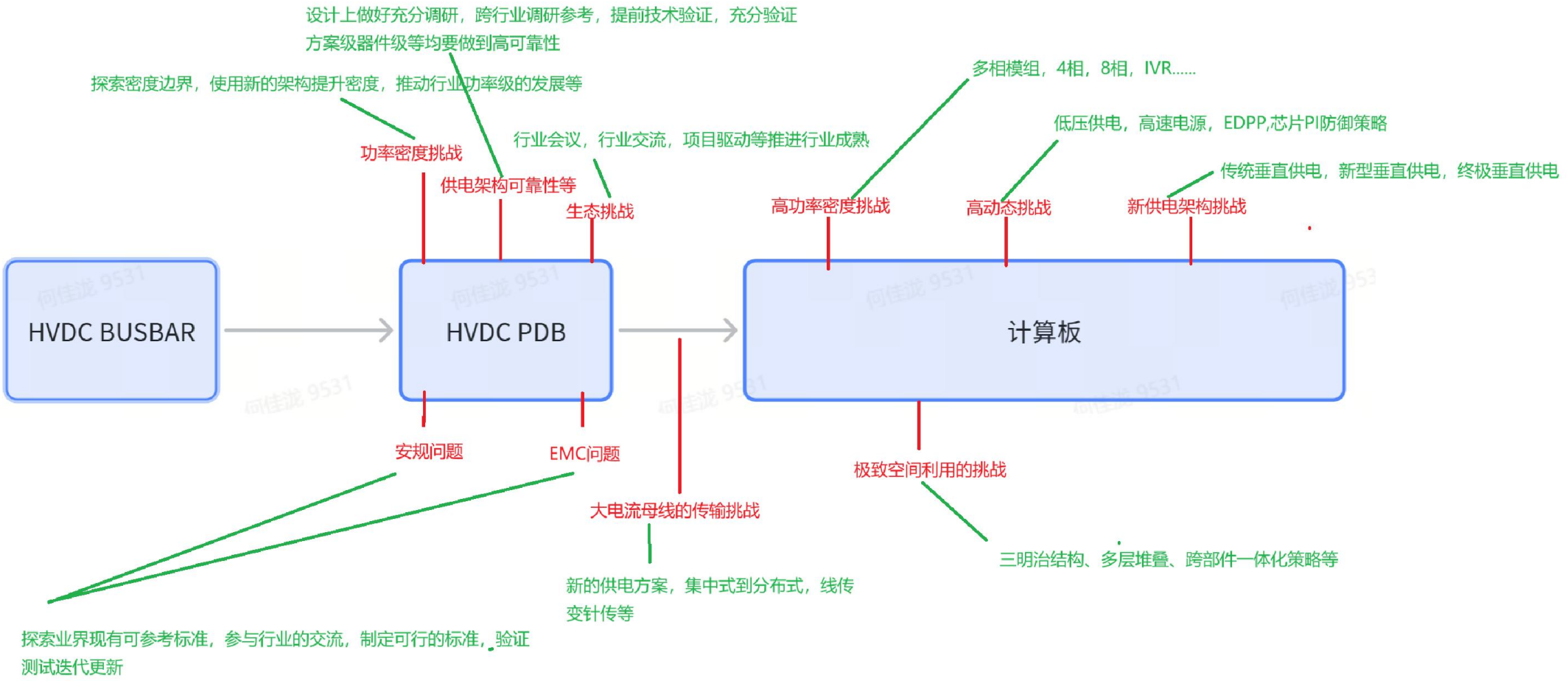
HVDC进柜 三种路线分析

三种转换路线	兼容性	优势	主要劣势
800V→54V	高	1、兼容现有54V集中供电架构 2、计算节点不需要变更 3、生态较成熟，落地快可以应对短期非线性需求	转换级数最多
800V→12V	低	12V需求端无需单独转换可以直接提供	输出电压不匹配未来VR低输入电压需求
800V→6.75V	差	1、计算节点无需50V转6.75V，少一级转换 2、转换电压匹配VR侧的电压发展需求 3、转换链路更少，全链路效率更优	1、母线电流太大传统架构传输母线难度极大 2、12V需求需要单独转换后级的计算板的供电架构无法复用传统架构

HVDC进柜的核心挑战



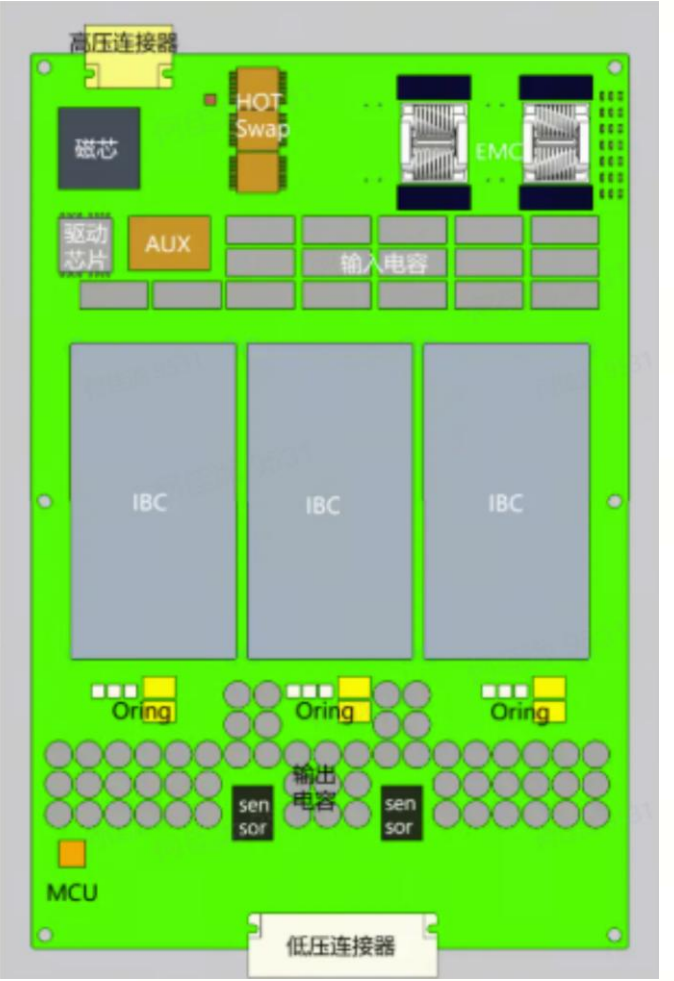
HVDC进柜挑战的应对策略



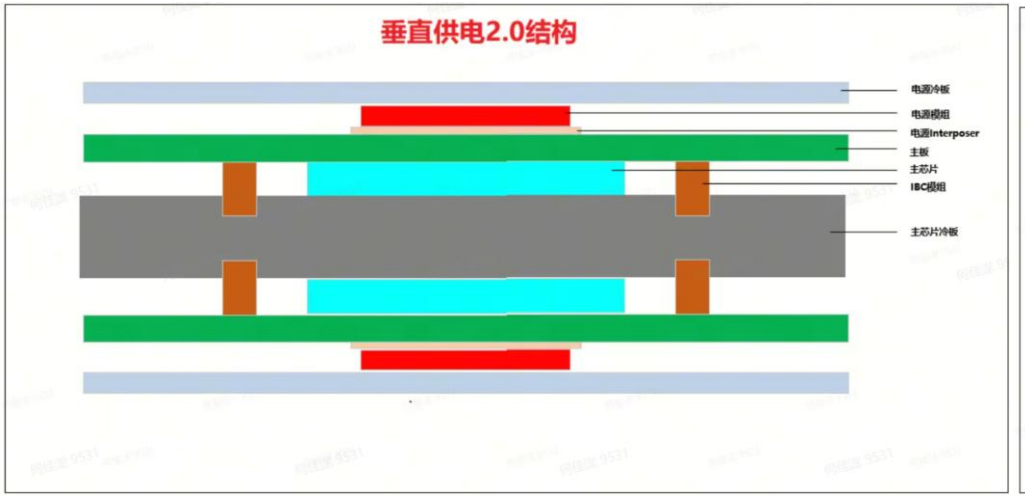
HVDC进柜挑战的应对策略-功率密度的探索

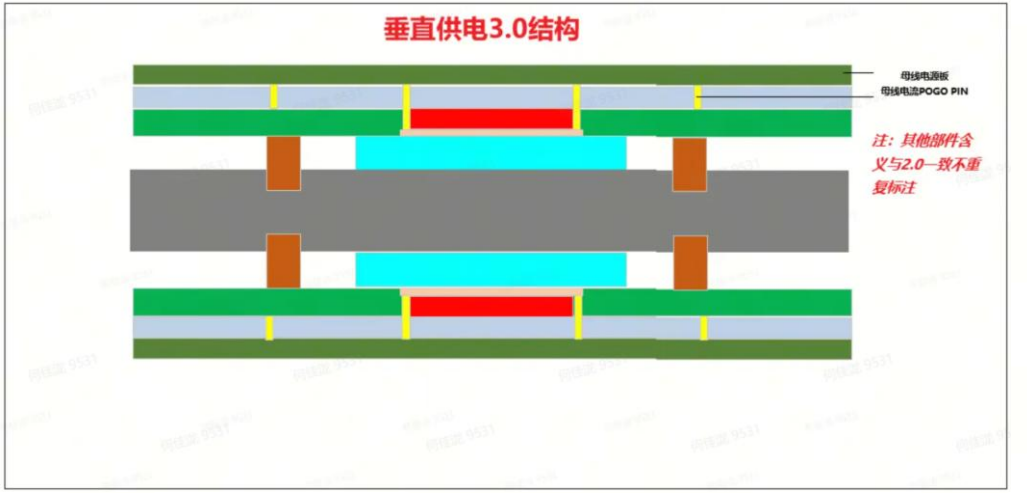
输出参数 Output Characteristics	
峰值功率 Peak power	<=150%，偏离平均值50ms
输出电压	54V ±5%
保持时间 Hold-up time	6mS
输出电容	500uF/kW，共计15mF
最大输出纹波&噪声 Max output ripple&Noise	500mVrms
输出连接器 Output connector	型号型号:TBD 主板连接器: GPU连接器：规格型号TBD Tray_present：开路拉高至3.3V，对接由GPU托盘拉低 PWR_RST_L：接收到GPU托盘的TRAY POWER RESET拉高信号后，MCU输出低电平重启HV IBC
Transients	
最大输出电压下跌及过冲 Maximum output voltage undershoot and overshoot	负载阶跃、负载掉线或开机、关机：+/-3% during load step,load shed or turn on, turn off：+/-3%
直流输出瞬态响应负载电流条件 DC output transient response shall be measured by applying the following load current conditions	2.负载变化频率: 范围从 100Hz 到 1MHz 2. Load frequency: Vary from 100Hz to 1MHz 3.负载斜率: 10A/us，下降和爬升 (最大斜率可根据负载频率和占空比限制) 3. Load slew rate: 10A/us for falling and rising (maximum slew rate could be limited based on load frequency and duty) 4.负载占空比: 范围从 10% 到 90% 4. Load duty: Vary from 10% to 90%
负载电容 Load capacitance capability	最大500uF/kW Up to 500uF/kW

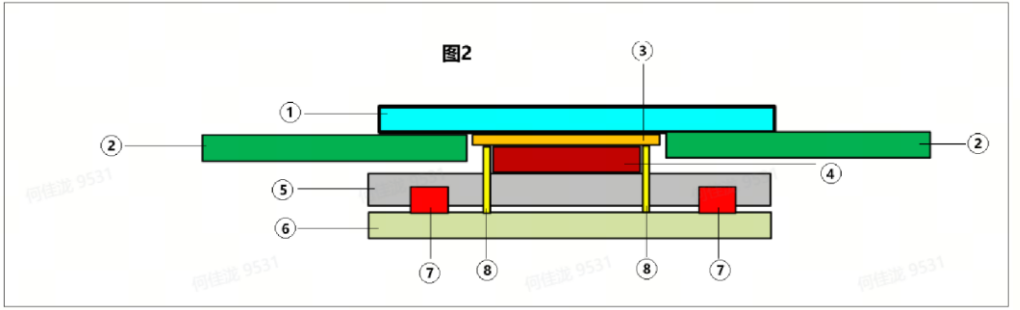
电气 Electrical	
PDB模块 Power Distribution Board	a. TDP - 30kW. b. EDP - 45kW@50mS c. Size - 330(L) x220(W) x 34(H) mm（带冷板）
输入电压 Input voltage	800V ± 5%
电压转换率 Voltage conversion ratio	800V:54V

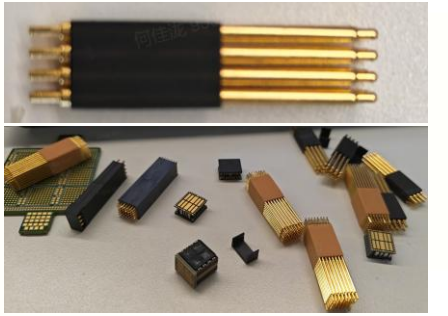


HVDC进柜挑战的应对策略-高密探索









Open AI Infra

Thank You !

Global Computing Consortium
全球计算联盟