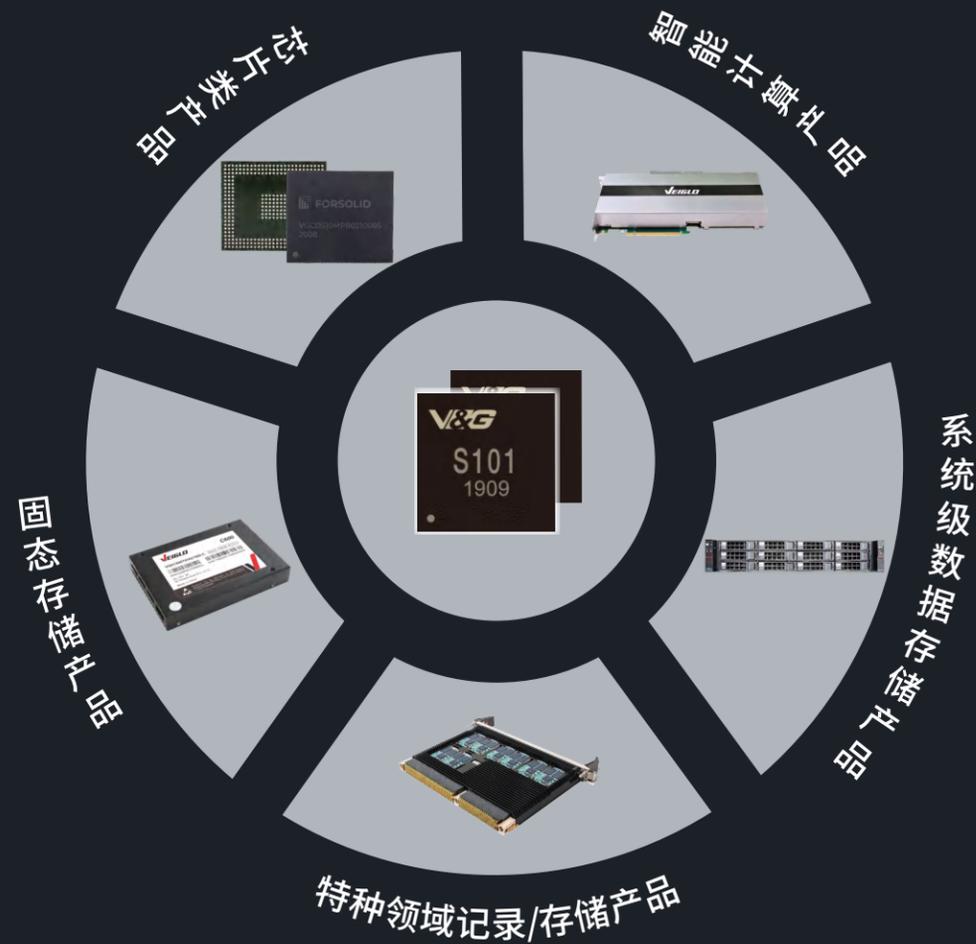


智能计算产品目录

INTELLIGENT COMPUTING
PRODUCT CATALOG



上海威固信息技术股份有限公司
V&G INFORMATION SYSTEM CO., LTD

服务热线 **6210 5185** 传真 86-021-6266 5667
86-021- 地址 上海市青浦区高泾路599号B座2层

版本号:VGZS202501



www.veiglo.cn
上海威固信息技术股份有限公司



“让数据更有价值”

——致力于成为全球领先的智能
存储与数据应用企业”

Contents 目录

公司介绍	01
技术沉淀	02
核心技术	02
荣誉与资质	04
客户与股东	05
产品介绍	
-AI训练类产品	07
-AI推理类产品	11
-AI服务器	21
-全国产嵌入式加固智算平台	23



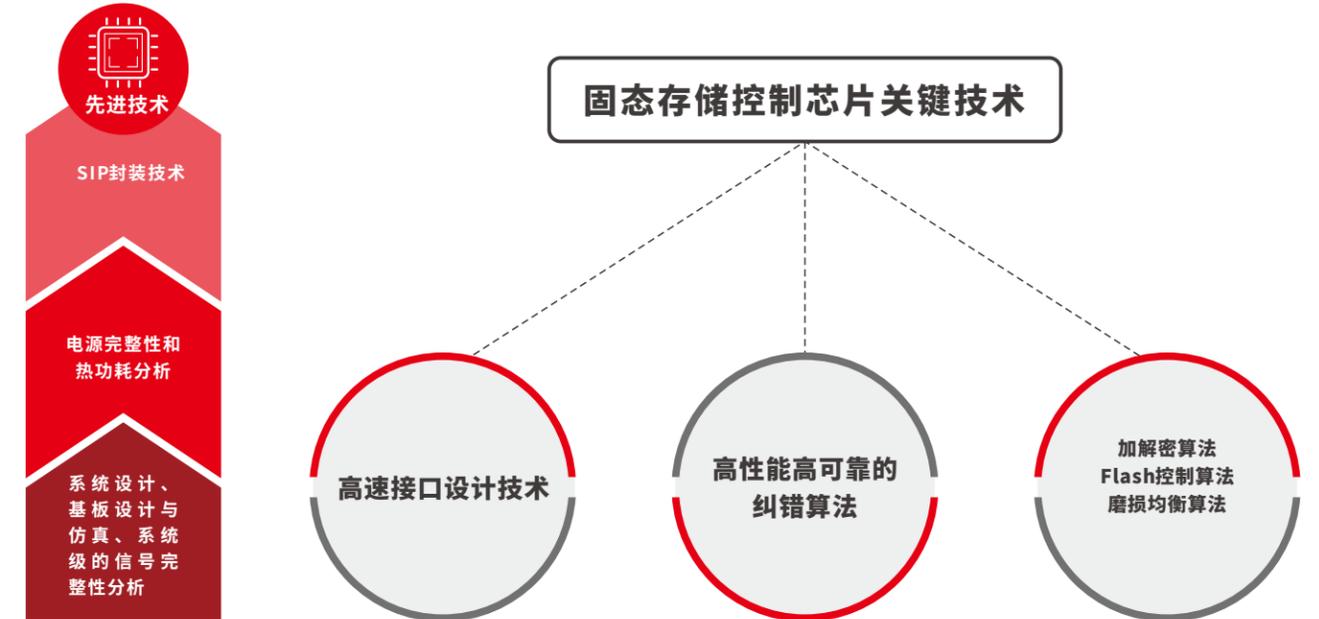
01 / COMPANY INTRODUCTION 公司介绍

上海威固信息技术股份有限公司 (简称“威固信息”) 成立于2013年。自成立以来威固信息始终立足于对固态存储技术的深入研究，努力为用户提供安全可靠、稳定耐用、智能可扩展的固态存储产品。公司产品以闪存控制技术为核心，打造了工业类存储、企业级存储、特种存储、消费类存储的全系列产品。

以闪存控制技术为核心 打造威固全系列产品线



02 / TECHNOLOGY OF PRECIPITATION 技术沉淀



经过多年的积累，威固信息已取得固态存储控制芯片关键技术的突破，完全掌握了高速接口设计技术、高性能高可靠的纠错算法以及加解密算法、高可靠的Flash控制算法、磨损均衡算法等核心技术。在单芯片固态存储产品方面，威固信息已掌握其核心的SIP封装技术，拥有系统设计、基板设计与仿真、系统级的信号完整性分析、电源完整性和热功耗分析等专业能力，在国内处于先进水平。

威固信息是上海市高新技术企业、上海市科技小巨人企业、国家级“专精特新”小巨人企业，荣获了多项行业内专业奖项和荣誉，公司投资建设的德清·威固科技园，构建了一个涵盖了集成电路、智能装备、大数据、人工智能的全新高科技产业生态集群。公司利用自身掌握的计算存储融合解决方案，以存储为基础，面向大数据、云计算、人工智能应用领域拓展，致力于大数据时代为用户提供智能存储和计算融合的软硬件综合解决方案，主要面向车载、船舶、航空、航天等高端装备信息化应用领域及大数据采集、挖掘、分析、处理、可视化等应用领域，并为各类企业数据中心提供更加高速与智能的数据应用解决方案。

03 / CORE TECHNOLOGY 核心技术

1. 闪存控制器技术

威固信息长期专注于固态存储控制器及固态存储设备的研制，经过多年的积累，已取得固态存储控制芯片关键技术的突破，完全掌握了高速接口设计技术、高性能高可靠的纠错算法以及加解密算法、高可靠的Flash控制算法、磨损均衡算法等核心技术。

闪存控制

- 数据纠错：检测并纠正读取过程中发生的错误，确保数据被正确读取、防止数据损坏；
- 磨损均衡：先进的动态/静态损耗均衡算法，有效地将闪存的使用范围分散到整个区域，大大提高了NAND闪存预期寿命；
- 坏块管理：有效的坏块管理算法，可实现对坏块检测并管理使用中出现的坏块，提高数据的可靠性；
- 垃圾回收：控制系统通知SSD永久删除不再使用的数据块，以防止未使用的数据始终占据存储块。

硬盘加密：加密数字签章算法，通过复杂的算法对使用者进行身份识别认证，保护数据安全；亦可通过数据加密算法变换数据存取编码规则，实现数据加密。

数据擦除：安全擦除功能通过清空存储介质（NAND闪存）中所有存储单元的用户数据，同时可以将硬盘完全还原为未初始化状态，硬盘的整体性能在一定程度上还原。

SMART信息管理：固态存储控制器自动检测SSD的运行状况、健康管理信息，可预判硬盘潜在故障。

掉电保护：在意外掉电的情况下，通过硬件电路监测通知控制器，在外部电路的支持下，控制器将HOST端已写入数据和FTL管理数据快速写入NAND，以实现写入数据的完整性。

2.系统级封装 (SIP) 技术



威固通过SIP系统级封装方式，以多种功能芯片/裸die进行并排或叠加的封装方式，将处理器、存储颗粒等功能芯片集成在一个封装内，从而实现一个完整的存储器，可替代传统的固态硬盘产品，具有体积小，容量大，性能高的特点。威固已掌握其核心的SIP封装技术，拥有系统设计、基板设计与仿真、系统级的信号完整性分析、电源完整性和热功耗分析等专业能力。

- 多芯片封装基板版图设计
- 芯片特种封装设计
- 芯片封装高速信号仿真
- 芯片力学仿真
- 芯片热仿真
- 封装焊接仿真
- 封装测试技术
- 芯片应用解决方案

3.高速电路工艺



存储产品对硬件电路的设计要求日益严格，产品中所涉及的外部SATA、PCIe等高速差分接口，内部DDR、NAND等高速并行接口，传输速度快、排布密度大，同时需要兼顾散热、小型化、抗振性等指标。为此威固信息为高速电路设计配备了专门的设计人才、仿真软件、测试工具，已保证每一件产品都能在用户的环境下稳定工作。

- 10~20层高密度电路板设计
- 10Gbps以上高速电路设计
- 高密度盲埋孔电路板设计
- 高速电路板信号仿真
- 电路板热仿真

4.NAND筛选



NAND FLASH作为当前固态存储最重要的存储介质，关系着产品的性能、品质、寿命等关键因素。威固信息为保证用户手中每一片产品的质量，在原材料的筛选分级上就投入了大量的技术和设备，用一套先进的算法加上完整的软硬件环境，保证每一颗NAND FLASH在安装到印刷电路板上之前都是无瑕疵的原材料。

- **宽温删选**
快速、有效的筛选闪存颗粒，并划分其可靠性等级
- **Read-retry**
通过read-retry操作测量闪存存储单元阈值电压分布
- **定制化**
提供定制测试pattern选项，测试流程更具灵活性
- **原始数据**
采集测量芯片未纠错错误率、编程操作时间、读操作时间、擦除操作时间、iddq电流
- **深度测试**
通过设置系统参数实现闪存颗粒耐久、数据保留预测等测试流程
- **寿命预测**
通过人工智能神经网络算法，实现闪存颗粒寿命的精准预测

5.数据加密及数据销毁



威固信息的SSD控制器具有数据加密和数据擦除功能，并针对客户极高机密数据存储要求设计制作了带物理销毁功能的固态硬盘，以保护用户的高度机密数据。

- ◎**数据加密**
SSD固件是安全的关键，恶意程序要想窃取数据就需要先过固件加密认证这一关，在固件中，带有认证固件的数字签章算法，在固件执行前须经由复杂的解密以及认证算法所认证，凡是经过篡改的固件程序将会被此算法所拒绝，并且无法执行于固态硬盘中。
加密认证算法在固件程序认证时，会经由国际安全加密认证的硬件安全加密模块加密，并且将加密认证所需的私人密钥储存于硬件安全模块中，且其私人密钥将不能被任何人所访问，被加密算法处理的数据因此无法被普通的设备识别。

- ◎**数据销毁**
特殊行业应用中需要对固态硬盘内的数据进行快速删除，在保密行业应用中称为数据销毁。威固信息可提供两种途径出发数据销毁功能，即HOST命令方式，通过数据传输总线输入数据销毁指令，和主控接收I/O口电平信号触发数据销毁。数据销毁后的固态硬盘清空为未初始化状态。

- ◎**硬盘物理销毁**
在保密性要求极其强烈的应用中，数据销毁无法被确认为安全的机密保护方式。威固信息为用户提供了物理销毁功能，即通过电压电流将NAND FLASH的存储单元击毁，保证数据不被泄露。

04 / Honor and Qualification 荣誉与资质



中国雷达行业协会会员单位



软件定义卫星技术联盟会员单位



上海国产高性能处理器平台产业技术创新战略联盟会员单位



中国计算机学会抗恶劣环境计算机专委会会员单位



专精特新“小巨人”企业



高新技术企业证书



ISO9001



知识产权管理体系认证证书



上海市“专精特新”中小企业

05 / Customer 合作客户



06 / Shareholder 威固主要机构股东



	标准PCIe卡	嵌入式子卡	6U VPX加速卡	整机设备
AI训练类产品	威讯103 	威讯153 	威讯163 	—
	天威100 			
AI推理类产品	威速103 	威速153 	威速163 	智盒D2000 
	威速204 	威速173 (MXM) 		
	天智100 		威速273 (MXM) 	智盒D3000 
	天智200 	XMC(传导/风冷) 即将推出		

天威100 PCIe

产品简介:

天威100是威固信息自主研发的高性能2/3长,双宽PCIe GPGPU卡,采用国产GPGPU芯片,单精度浮点算力不低于36TFLOPS@FP32,热设计功耗250W。

本卡主要用于AI训练和高性能计算环境,兼容CUDA,用户可以很轻易的将过往代码移植到天威100。

天威100卡接口为PCIe x16插槽,用户另外还须提供一路专用8PIN供电插头为本卡供电。



硬件参数	
芯片算力	36 TFLOPS@fp32 147 TFLOPS@fp16/bf16 支持int32,int16 计算多精度数据类型支持标准/混合训练
RAM	32GB@1.2GHz HBM2
SENSOR	电流传感器
电源接口	PCIe插槽标准供电 GPU专用8PIN供电电压:12V 供电电流:≥25A
信号接口	PCIe标准插槽 传输协议:PCIe GEN4.0 x16
功耗	最大功耗:≤250Watt 空闲功耗:50Watt 注意:功耗大小和具体应用相关,根据应用的不同,运行平均功耗为60Watt~200Watt
尺寸	264.70×111.35×38.92mm
工作温度	环境温度:0°C~40°C 注:最高工作环境温度和用户的散热相关,40°C为参考值 GPGPU芯片的内部结温和外壳温度差值为10~15°C。 GPGPU芯片固件对温度处理的方式为: 1.结温85°C:推荐的最高工作温度 2.结温90°C:告警温度 3.结温95°C:降频温度 4.结温100°C:关机温度

订购信息:

产品名称	计算主频	RAM	温度级别	备注
天威100	1500MHz	32GByte	商业级	TDP: 250W

威讯103(全功耗)

产品简介:

威讯103是威固信息自主研发的高性能GPGPU卡,2/3长、双宽PCIe卡形式。搭载国产GPGPU芯片,单精度浮点性能高达36TFLOPS,热设计功耗250W。

该产品实现100%全国产化,符合WKB1.0标准。适用于AI训练和高性能计算,支持CUDA,便于代码移植。

产品需安装于标准PCIe x16插槽,并连接8 PIN供电插头。



硬件参数	
芯片算力	36 TFLOPS@fp32 147 TFLOPS@fp16/bf16
RAM	32GB@1.2GHz HBM2
电源接口	PCIe插槽标准供电 GPU专用8PIN供电电压:12V 供电电流:≥25A
信号接口	PCIe标准插槽 传输协议:PCIe GEN4.0 x16
功耗	最大功耗:≤250Watt
尺寸	264.70×111.35×38.92mm(不含固定架)
工作温度	环境温度:-40°C~55°C 注:最高工作环境温度和用户的散热相关,55°C为参考值 GPGPU芯片的内部结温和外壳温度差值为10~15°C。 GPGPU芯片固件对温度处理的方式为: 1.结温85°C:推荐的最高工作温度 2.结温90°C:告警温度 3.结温100°C:关机温度
散热方式	被动风冷

订购信息:

产品名称	型号	计算主频	RAM	备注
威讯103	RevHCU-6CE-32G-103	1500MHz	32GB	TDP:250W 100%全国产

威讯153(嵌入式低功耗)

产品简介:

威讯153是威固信息自主研发的GPGPU嵌入式子卡,尺寸为125mm*95mm,搭载国产GPGPU芯片。提供6TFLOPS@FP32 75W和12TFLOPS@FP32 100W两种算力版本。

该产品实现100%全国产化,符合WKB1.0标准。适用于AI训练和高性能计算,支持CUDA,便于代码移植。

威讯153接口简单,仅需PCIe信号和12V供电,即可集成至各类平台,助力用户快速部署AI训练和高性能计算单元,加速产品开发,降低开发风险。



硬件参数	
芯片算力	6 TFLOPS@fp32 24 TFLOPS@fp16 TDP:75W
	12 TFLOPS@fp32 48 TFLOPS@fp16 TDP:100W
支持int32, int16 计算, 支持多精度数据类型支持标准/混合训练	
ROM	1MB SPI FLASH
RAM	16GB@800MHz HBM2E
SENSOR	电流传感器
电源接口	连接器: 昱丰YFCSLH-030-1.5-G-D-A-K 供电电压: 12V 供电电流: ≥10A
信号接口	连接器: 昱丰YFCSLH-030-1.5-G-D-A-K 传输协议: PCIe GEN4.0 x16
功耗	最大功耗: 根据不同算力分为75W、100W两个版本。 空闲功耗: 45 Watt 注意: 功耗大小和具体应用相关。
尺寸	125mm × 95mm
工作温度	环境温度: -40°C~55°C 注: 最高工作环境温度和用户的散热相关, 55°C为参考值 GPGPU芯片的内部结温和外壳温度差值为10~15°C。 GPGPU芯片固件对温度处理的方式为: 1. 结温85°C: 推荐的最高工作温度 2. 结温90°C: 告警温度 3. 结温100°C: 关机温度

订购信息:

产品名称	型号	计算主频	RAM	备注
威讯153	RevHCU-1CE-16G-153	1500MHz	16GB	TDP:75W 100%全国产
威讯153	RevHCU-2CE-16G-153	1500MHz	16GB	TDP:100W 100%全国产

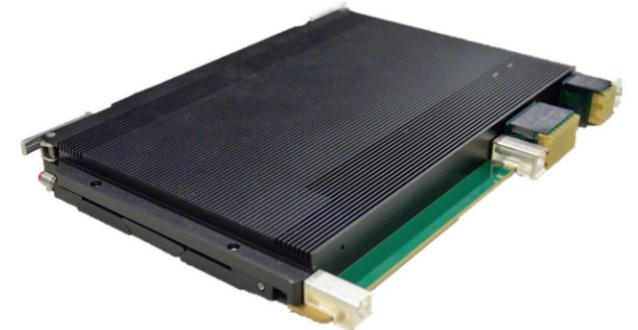
威讯163(VPX)

产品简介:

威讯163是威固信息按照欧式6U标准自主研发的GPGPU VPX模块,搭载国产GPGPU 芯片,提供两种算力版本,12TFLOPS@FP32 100W至36T@FP32 220w。低算力版风冷散热,高算力版需用户自行设计LFT(穿透式液冷)结构件。

该产品实现100%全国产化,符合WKB1.0标准。适用于AI训练和高性能计算,支持CUDA,便于代码移植。

威讯163接口简便,只需PCIe信号、12V及3.3V辅助供电,便于集成至各类处理平台,助力用户快速集成智算算力至嵌入式系统,加速开发,降低风险。



硬件参数	
芯片算力	12TFLOPS@fp32 TDP:100W
	36TFLOPS@fp32 TDP:220W
支持int32, int16 计算, 支持多精度数据类型支持标准/混合训练	
ROM	1MB SPI FLASH
RAM	16GB@800MHz HBM2E
电源接口	供电电压: 12V、3.3V (可选) 供电电流: ≥25A
信号接口	传输协议: PCIe GEN4.0 x16(可根据需要调整为x8), 自适应PCIe Gen3.0、2.0。
功耗	最大功耗: 根据不同算力和内存大小, 分为100W、220W两个版本。 空闲功耗: 50Watt
尺寸	233mm × 160mm × 25.4mm
工作温度	环境温度: -40°C~55°C 注: 最高工作环境温度和用户的散热相关, 55°C为参考值 GPGPU芯片的内部结温和外壳温度差值为10~15°C。 GPGPU芯片固件对温度处理的方式为: 1. 结温85°C: 推荐的最高工作温度 2. 结温90°C: 告警温度 3. 结温105°C: 关机温度

订购信息:

产品名称	型号	计算主频	RAM	备注
威讯163	RevHCU-2CE-16GB-163	1500MHz	16GB	TDP:100W 全国产
威讯163	RevHCU-6CE-32GB-163	1500MHz	32GB	TDP:220W 全国产

威速103 PCIe

产品简介:

威速103是威固信息自主研发的AI推理加速卡，半高半长单宽PCIe卡形式。搭载国产AI推理芯片，支持FP32、FP16、INT8等多精度推理混合计算。单精度浮点性能高达16TFLOPS，半精度浮点性能高达64TFLOPS，显存16GByte。

该产品实现100%全国产化，符合WKB1.0标准。适用于AI推理和高性能计算，兼容CUDA，便于代码迁移。

该产品配套软件栈集成多种深度学习框架及调试调优工具，为人工智能应用的开发和部署提供了便利。



硬件参数	
芯片算力	16 TFLOPS@fp32 64 TFLOPS@fp16 256 TOPS@int8
视频解码	最大支持128路并发 (1080P @30fps) 支持H.264/H.265/VP9/AVS2
图像解码/编码	2000/500 fps (JPEG)
处理单元主频	1.0GHz
RAM	16GB@1600MHz HBM2E
SENSOR	电流传感器
电源接口	PCIe插槽标准供电 供电电流: ≥7A
信号接口	PCIe标准插槽 传输协议: PCIe GEN4.0 x16
功耗	最大功耗: ≤75Watt 空闲功耗: 19Watt
尺寸	167.65 × 68.90 × 19.46 mm
工作温度	环境温度: -40°C~65°C 注: 最高工作环境温度和用户的散热相关, 65°C为参考值 GPGPU芯片的内部结温和外壳温度差值为5~10°C。 GPGPU芯片固件对温度处理的方式为: 1. 结温95°C: 推荐的最高工作温度 2. 结温100°C: 关机温度

订购信息:

产品名称	型号	计算主频	RAM	备注
威速103	RevMRU-3CE-16G-103	1000MHz	16GB	TDP: 75W 100%全国产

威速153(子卡)

产品简介:

威速153是威固信息自主研发的AI推理嵌入式子卡，尺寸为125mm*95mm，搭载国产AI推理芯片。单精度浮点性能高达16TFLOPS，半精度浮点性能高达64TFLOPS，显存16GByte。该产品实现100%全国产化，符合WKB1.0标准。适用于AI推理和高性能计算，兼容CUDA，便于代码迁移。

威速153接口简便，仅需PCIe信号和12V供电，即可集成至多种平台，助力用户快速部署AI推理单元，加速产品开发，降低开发风险。



硬件参数	
芯片算力	16 TFLOPS@fp32 64 TFLOPS@fp16 256 TOPS@int8
视频解码	最大支持128路并发 (1080P @30fps) 支持H.264/H.265/VP9/AVS2
图像解码/编码	2000/500 fps (JPEG)
处理单元主频	1.0GHz
ROM	1MB SPI FLASH
RAM	16GB@1600MHz HBM2E
SENSOR	电流传感器
电源接口	连接器: 昱丰YFCSLH-030-1.5-G-D-A-K (国产版本) 供电电压: 12V 供电电流: ≥7A
信号接口	连接器: 昱丰YFCSLH-030-1.5-G-D-A-K (国产版本) 传输协议: PCIe GEN4.0 x16
功耗	最大功耗: 75 Watt 空闲功耗: 19 Watt 注意: 功耗大小和具体应用相关。
尺寸	125mm × 95 mm
工作温度	环境温度: -40°C~65°C 注: 最高工作环境温度和用户的散热相关, 65°C为参考值 GPGPU芯片的内部结温和外壳温度差值为5~10°C。 GPGPU芯片固件对温度处理的方式为: 1. 结温95°C: 推荐的最高工作温度 2. 结温100°C: 关机温度

订购信息:

产品名称	型号	计算主频	RAM	备注
威速153	RevMRU-3CE-16G-153	1000MHz	16GB	TDP: 75W 100%全国产

威速163(VPX)

产品简介:

威速163是威固信息按照欧式6U标准自主研发的AI推理加速VPX模块,搭载国产AI推理芯片。单精度浮点性能高达16TFLOPS,半精度浮点性能高达64TFLOPS,显存16GByte。

该产品实现100%全国产化,符合WKB1.0标准。适用于AI推理和高性能计算,兼容CUDA,便于代码迁移。

威速163接口简便,只需PCIe信号、12V及3.3V辅助供电,便于集成至各类处理平台,助力用户快速集成AI推理算力至嵌入式系统,加速开发,降低风险。



硬件参数	
芯片算力	16 TFLOPS@fp32 64 TFLOPS@fp16 256 TOPS@int8
视频解码	最大支持128路并发 (1080P @30fps) 支持H.264/H.265/VP9/AVS2
图像解码/编码	2000/500 fps (JPEG)
处理单元主频	1.0GHz
ROM	1MB SPI FLASH
RAM	16GB@1600MHz HBM2E
SENSOR	电流传感器
电源接口	供电电压:12V、3.3V(可选) 供电电流:≥7A
信号接口	传输协议:PCIe GEN4.0 x16 (可根据需要调整为x8), 自适应PCIe Gen3.0、2.0。
功耗	最大功耗:75 Watt 空闲功耗:19 Watt
尺寸	233mm×160mm×25.4mm
工作温度	环境温度:-40°C~65°C 注:最高工作环境温度和用户的散热相关,65°C为参考值 GPGPU芯片的内部结温和外壳温度差值为5~10°C。 GPGPU芯片固件对温度处理的方式为: 1.结温95°C:推荐的最高工作温度 2.结温100°C:关机温度

订购信息:

产品名称	型号	计算主频	RAM	备注
威速163	RevMRU-3CE-16G-163	1000MHz	16GB	TDP:75W 全国产

威速173(MXM)

产品简介:

威速173是威固信息自主研发的MXM TYPE-A高性能AI推理核心卡。威速173核心卡将高性能AI推理芯片、电源模块以及其他元器件集成于一块MXM卡,实现超强计算性能、大容量内存、方便快捷的二次设计等特点的异构并行计算解决方案。

威速173核心卡采用标准MXM V3.1规范进行设计,只需提供PCIe信号和12V供电即可实现高性能AI推理单元到各类平台的应用。本产品帮助用户快速将高性能AI推理单元应用到自己的嵌入式模块中,加速产品的开发进程,降低产品的开发风险。



硬件参数	
芯片算力	16 TFLOPS@fp32 64 TFLOPS@fp16 256 TOPS@int8
视频解码	最大支持128路并发 (1080P @30fps) 支持H.264/H.265/VP9/AVS2
图像解码/编码	2000/500 fps (JPEG)
处理单元主频	1.0GHz
RAM	16GB高性能HBM2E内存
SENSOR	电流传感器
电源接口	接口:MXM V3.1 供电电压:12V
信号接口	接口:MXM V3.1 传输协议:PCIe GEN4.0 x16
功耗	最大功耗:75Watt 空闲功耗:19Watt
尺寸	82mm×70 mm
工作温度	环境温度:-40°C~65°C 注:最高工作环境温度和用户的散热相关,65°C为参考值 高性能AI推理芯片的内部结温和外壳温度差值为5~10°C。 高性能AI推理芯片固件对温度处理的方式为: 1.结温95°C:推荐的最高工作温度 2.结温105°C:关机温度

订购信息:

产品名称	型号	计算主频	RAM	备注
威速173	RevMRU-3CE-16G-173	1.0GHz	16GB	TDP:75W 100%全国产

威速204

产品简介:

威速204是威固信息自主研发的PCIe全高、2/3长、单宽人工智能推理加速卡，采用国产AI推理芯片。加速卡支持FP32, FP16, INT8等多种精度推理。单精度浮点性能高达24 TFLOPS, 半精度浮点性能高达96 TFLOPS, 显存32 GByte, 热设计功耗150W。

威速204配套软件栈集成了多种主流的深度学习编程框架以及调试和调优工具，为人工智能应用的开发和部署提供了便利。



硬件参数	
芯片算力	24 TFLOPS@fp32 96 TFLOPS@fp16 384 TOPS@int8
视频解码	最大支持128路并发 (1080P @30fps) 支持H.264/H.265/VP9/AVS2
图像解码/编码	2000/500 fps (JPEG)
处理单元主频	1.5GHz
ROM	1MB SPI FLASH
RAM	32GB@1600MHz HBM2E
SENSOR	电流传感器
电源接口	PCIe插槽标准供电, GPU专用8PIN 供电电压12V 供电电流: ≥14A
信号接口	PCIe标准插槽 传输协议: PCIe GEN4.0 x16
功耗	最大功耗: ≤150 Watt 空闲功耗: 36 Watt
尺寸	264.70 × 111.35 × 19.46 mm (不含固定架)
工作温度	环境温度: -40°C~65°C
散热方式	被动风冷

订购信息:

产品名称	型号	计算主频	RAM	备注
威速204	RevMRU-4CE-32G-204	1500MHz	32GB	TDP:150W 全国产

威速273 (MXM)

产品简介:

威速273是威固信息自主研发的MXM TYPE-A高性能AI推理核心卡。威速273核心卡将高性能AI推理芯片、电源模块以及其他元器件集成于一块MXM卡，实现超强计算性能、大容量内存、方便快捷的二次设计等特点的异构并行计算解决方案。

威速273核心卡采用标准MXM V3.1规范进行设计，只需提供PCIe信号和12V供电即可实现高性能AI推理单元到各类平台的应用。本产品帮助用户快速将高性能AI推理单元应用到自己的嵌入式模块中，加速产品的开发进程，降低产品的开发风险。



硬件参数	
芯片算力	24TFLOPS@fp32 96 TFLOPS@fp16 384TOPS@int8
视频解码	最大支持128路并发 (1080P @30fps) 支持H.264/H.265/VP9/AVS2
图像解码/编码	2000/500 fps (JPEG)
处理单元主频	1.5GHz
RAM	32GB高性能HBM2E内存
SENSOR	电流传感器
电源接口	接口: MXM V3.1 供电电压: 12V
信号接口	接口: MXM V3.1 传输协议: PCIe GEN4.0 x16
功耗	最大功耗: 150Watt 空闲功耗: 36Watt
尺寸	82mm × 70 mm
工作温度	环境温度: -40°C~65°C 注: 最高工作环境和用户的散热相关, 65°C为参考值 高性能AI推理芯片的内部结温和外壳温度差值为5~10°C。 高性能AI推理芯片器件对温度处理的方式为: 1. 结温95°C: 推荐的最高工作温度 2. 结温105°C: 关机温度

订购信息:

产品名称	型号	计算主频	RAM	备注
威速273	RevMRU-4CE-32G-273	1.5GHz	32GB	TDP:150W 100%全国产

天智100

产品简介:

天智100是威固信息自主研发的PCIe半高半长单宽人工智能推理加速卡,采用国产AI推理芯片。加速卡支持FP32, FP16, INT8等多种精度推理。单精度浮点性能高达16TFLOPS,半精度浮点性能高达64TFLOPS,显存16GByte。

天智100配套软件栈集成了多种主流的深度学习编程框架以及调试和调优工具,为人工智能应用的开发和部署提供了便利。



硬件参数	
芯片算力	16 TFLOPS@fp32 64 TFLOPS@fp16 256 TOPS@int8
视频解码	最大支持128路并发(1080P @30fps) 支持H.264/H.265/VP9/AVS2
图像解码/编码	2000/500 fps (JPEG)
处理单元主频	1.0GHz
RAM	16GB@1600MHz HBM2E
SENSOR	电流传感器
电源接口	PCIe插槽标准供电 供电电流:≥7A
信号接口	PCIe标准插槽 传输协议:PCIe GEN4.0 x16
功耗	最大功耗:≤75Watt 空闲功耗:19Watt
尺寸	167.65×68.90×19.46 mm
工作温度	环境温度:0°C~40°C 注:最高工作环境温度和用户的散热相关,65°C为参考值 GPGPU芯片的内部结温和外壳温度差值为5~10°C。 GPGPU芯片固件对温度处理的方式为: 1.结温95°C:推荐的最高工作温度 2.结温100°C:关机温度

订购信息:

产品名称	协处理单元主频	RAM	温度级别	备注
天智100	1.0GHz	16GByte	商业级	TDP:75W

天智200 PCIe

产品简介:

天智200是威固信息自主研发的PCIe全高、2/3长、单宽人工智能推理加速卡,采用国产AI推理芯片。加速卡支持FP32, FP16, INT8等多种精度推理。单精度浮点性能高达24 TFLOPS,半精度浮点性能高达96 TFLOPS,显存32 GByte,热设计功耗150W。

天智200配套软件栈集成了多种主流的深度学习编程框架以及调试和调优工具,为人工智能应用的开发和部署提供了便利。



硬件参数	
芯片算力	24 TFLOPS@fp32 96 TFLOPS@fp16 384 TOPS@int8
视频解码	最大支持128路并发(1080P @30fps) 支持H.264/H.265/VP9/AVS2
图像解码/编码	2000/500 fps (JPEG)
处理单元主频	1.5GHz
RAM	32GB@1600MHz HBM2E
SENSOR	电流传感器
电源接口	PCIe插槽标准供电, GPU专用8PIN 供电电压12V 供电电流:≥14A
信号接口	PCIe标准插槽 传输协议:PCIe GEN4.0 x16
功耗	最大功耗:≤150 Watt 空闲功耗:36 Watt
尺寸	264.70×111.35×19.46 mm (不含固定架)
工作温度	环境温度:0°C~40°C
散热方式	被动风冷

订购信息:

产品名称	计算主频	RAM	温度级别	备注
天智200	1500MHz	32GByte	商业级	TDP:150W

智盒二代(D2000)

产品简介:

智盒是一款全国产边缘智能计算产品。不同于一般的边缘计算盒,本产品具有极大的算力资源,面向低吞吐、高算力需求市场。另外,本产品也可以作为用户后期使用、开发高性能智算产品的演示平台。

本品有如下特点:

- 对尖端国产元器件的应用开发;
- 搭载算力高达16TFLOPS@FP32、256TOPS@int8的国产GPGPU;
- 高性能通用并行计算技术;

该产品由GPGPU模组、飞腾 D2000处理器组成。



智盒二代(D3000)

产品简介:

智盒二代是一款全国产边缘智能计算产品。不同于一般的边缘计算盒,本产品具有极大的算力资源,面向低吞吐、高算力需求市场。另外,本产品也可以作为用户后期使用、开发高性能智算产品的演示平台。

本品有如下特点:

- 对尖端国产元器件的应用开发;
- 搭载算力高达24TFLOPS@FP32、384 TOPS@int8的国产GPGPU;
- 高性能通用并行计算技术;

该产品由GPGPU模组、飞腾 D3000处理器组成。
内部所有元器件均采用国产元、进口兼容,可实现国产化率100%。



硬件参数	
CPU	飞腾D2000 8核(8 * FT663) FTC663 2.2GHZ 主频, 4M L2 CACHE; 64BIT 双通道DDR4; 最大容量 32GB 兼容 ARM V8 指令系统, 支持 64 位和 32 位指令 兼容 ARM V8 虚拟化体系结构, 支持业界主流的 KVM、Xen 虚拟机 支持单精度、双精度浮点运算指令 支持 ASIMD 处理指令
GPGPU	16 TFLOPS@fp32 64 TFLOPS@fp16/bf16 256 TOPS@int8 16GB 显存
视频解码	最大支持128路并发(1080P @30fps) 支持H.264/H.265/VP9/AVS2
图像编码/解码	2000/500 fps (JPEG)
数据接口	USB: 4路USB3.0, 兼容USB 2.0 显示: 一路HDMI接口, 一路DVI接口 千兆网: 1路RJ45
供电接口	供电要求: 12V外接220V转12V适配器供电
功耗	最大功耗: ≤150Watt 空闲功耗: ≤60Watt
散热方式	主动风冷
工作温度	环境温度: 0°C~45°C

订购信息:

型号	GPGPU配置	显存	备注
智盒二代(D2000)	天智100	16GB	整机功耗小于220W

硬件参数	
CPU	飞腾D3000 8核(8 * FT862) • FTC664 2.5GHZ 主频, 4M L2 CACHE; • 64BIT 双通道DDR4; 最大容量 32GB • 兼容 ARM V8 指令系统, 支持 64 位和 32 位指令 • 兼容 ARM V8 虚拟化体系结构, 支持业界主流的 KVM、Xen 虚拟机 • 支持单精度、双精度浮点运算指令 • 支持 ASIMD 处理指令
GPGPU	24 TFLOPS@fp32 96 TFLOPS@fp16/bf16 384 TOPS@int8 32GB 显存
扩展能力	可扩展一路半高半长PCIe模块
视频解码	最大支持128路并发(1080P @30fps) 支持H.264/H.265/VP9/AVS2
图像编码/解码	2000/500 fps (JPEG)
数据接口	USB: 4路USB3.0, 兼容USB 2.0 显示: 1路HDMI显示 千兆网: 2路RJ45
供电接口	供电要求: 12V外接220V转12V适配器供电
功耗	最大功耗: ≤220Watt 空闲功耗: ≤60Watt
散热方式	主动风冷
工作温度	环境温度: 0°C~45°C

订购信息:

型号	GPGPU配置	显存	备注
智盒二代(D3000)	天智200	32GB	整机功耗小于250W

AI服务器 (VG-TW100C)

产品简介:

全新一代AI 服务器VG-TW100C是威固信息新推出的一款4U双路机架式加速计算服务器。

- 该服务器拥有24个DDR4 DIMM 插槽;
- 支持8个全高全长双宽GPU加速卡;
- 支持8块2.5英寸热插拔SAS/SATA/NVME 硬盘,板载支持2个M.2 SSD硬盘;
- 支持多种不同网络选择:以太网、InfiniBand, OPA;
- 可实现高性能和I/O 灵活性,满足不同应用程序的互联需求;
- 板载集成BMC;
- 支持IPMI 2.0 远程管理;
- 支持4个1600W&2000W 冗余电源。



产品特性:

· 应用场景: 持多模态大模型、视频生成大模型、具身智能、基础AI研究范式等的训练及推理,可为低空经济、自动驾驶汽车、智能城市、医疗健康、大数据、高性能计算和虚拟现实等重点行业领域提供算力支持。

- 超强性能:**
 - 最大支持2-4颗ARM/X86处理器,最大功率350W;
 - 每CPU支持12个内存通道,总共支持24 DIMM 插槽,标配内存512GB;
 - 标配8张自主可控 TW100 GPU加速卡。

- 关键特性:**
 - 标配GPU卡单卡算力最高可达147 TFLOPS@fp16, 32GB高性能HBM2内存,并支持GPU卡升级;
 - 兼容CUDA,无需修改CUDA代码即能完成迁移;
 - 支持主流AI框架:TensorFlow、Pytorch、PaddlePaddle等;
 - 支持Ubuntu、CentOS、麒麟等主流操作系统。

- 稳定品质:**
 - 4个1600W&2000W冗余电源,保证服务器高品质运行;
 - 支持通过RAID 卡超级电容,提供掉电数据保护功能;
 - 功能全面的集中管理软件,实现智能远程管理。



硬件参数	
形态	4U 机架式服务器
CPU	支持2-4颗ARM/X86处理器, TDP 最大350W
内存	标配512G, 24个DDR4 DIMM插槽; 支持单根128GB/64GB/32GB/16GB 等规格内存条,低电压1.2V; 支持3200/2933/2666MHz ECC DDR4 RDIMM/LRDIMM
加速卡	标配8张自主可控 TW100 GPU加速卡; 单卡算力为36TFLOPS@FP32, 147 TFLOPS@fp16, 295 TOPS@int8;意单卡显存32GB
硬盘	前置8个2.5英寸SATA/SAS热插拔硬盘位,板载支持2个M.2 SSD盘;标配960GB电子盘*2
网卡	主板集成4个千兆网口和1个维护网口; 标配4个万兆网口
IO接口	显示:支持最高分辨率1920×1080, VGA接口(后置); USB:2个USB3.0接口(后置),2个USB2.0(前置); 串口:1个RS232系统通用串口(后置); 管理:1个BMC管理接口
扩展插槽	前置4个双宽PCIe x16设备; 后置8个双宽GPU设备
硬盘控制器	板载SAS和SATA控制器, 支持SAS3.0和SATA3.0
标准接口	USB3.0*2、USB2.0*2、千兆网口*4、万兆网口*2,管理网口*1、VGA*1、电源*2、DB9串口*1
维护	支持IPMI2.0 支持BMC管理,可实现远程开机、温度传感器状态监控,支持在线BIOS及BMC更新,支持BMC、KVM-over-IP、Serial-over-Lan (SOL)和IPMI2.0等服务器带外管理功能;支持在线故障诊断;支持远程加关电和整机复位;支持远程更新CPU 固件、BIOS 和安装操作系统
适配软件	支持centos、ubuntu、麒麟等主流操作系统;部署智算综合应用台,兼容CUDA,无需修改CUDA代码即能完成迁移;支持TensorFlow、Pytorch、PaddlePaddle等主流AI框架
电源	支持4 个1600W&2000W 冗余电源, 支持热插拔
工作温度	10°C~+35°C
存储温度	-40°C~+55°C
湿度	工作时相对湿度8%~90% (不凝结) 非工作时湿度5%~95% (不凝结)

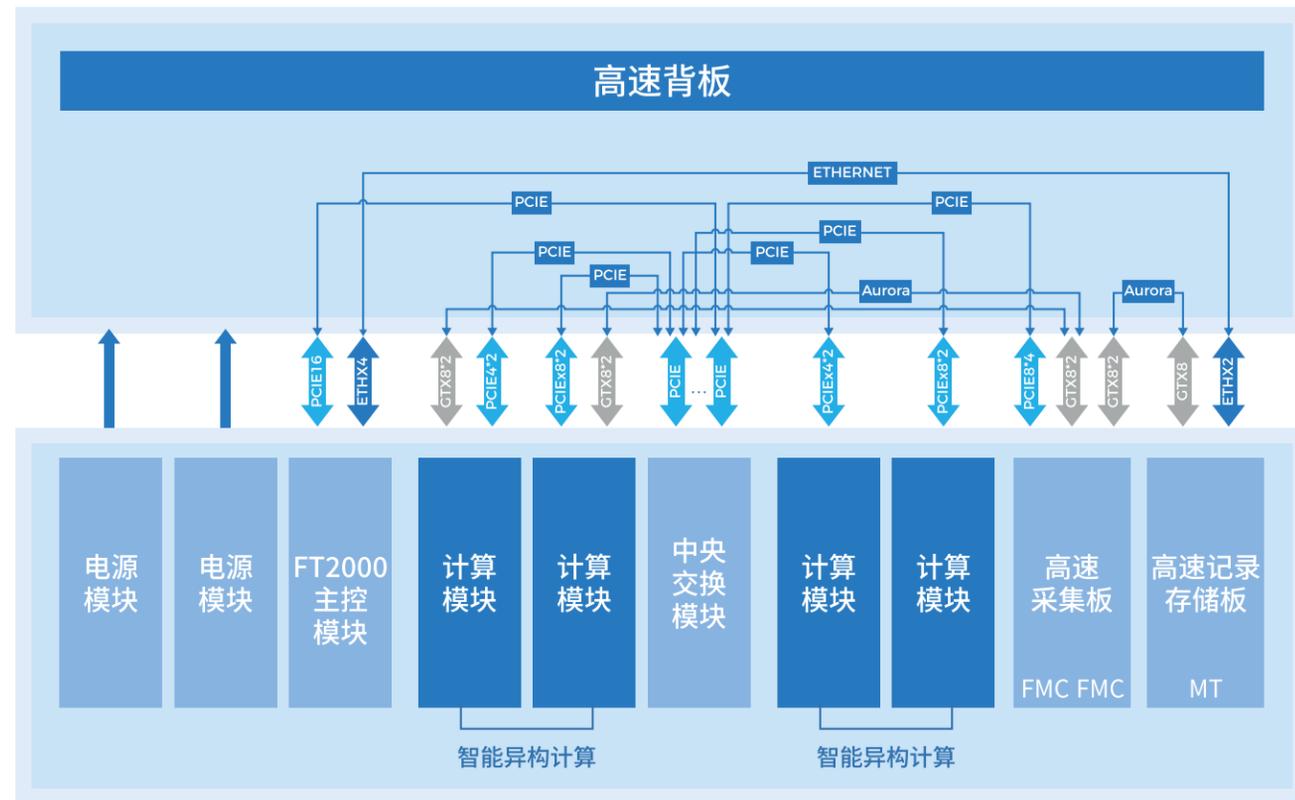
全国产嵌入式加固智算平台

概述:

嵌入式加固智能计算平台, 基于VPX加固架构, 包含一块飞腾CPU板、一块高速采集板、一块16T存储记录板、四块异构计算板(单板最高24TFLOPS@fp32)和两块电源板。
平台将计算算力进行模块化封装, 可灵活混插(FPGA+DSP异构计算模块与GPGPU计算模块可互换), 具备较高的可扩展性和灵活性, 整机采用国产化和加固设计, 架构和技术上具备更好的适应性, 满足严苛使用环境。



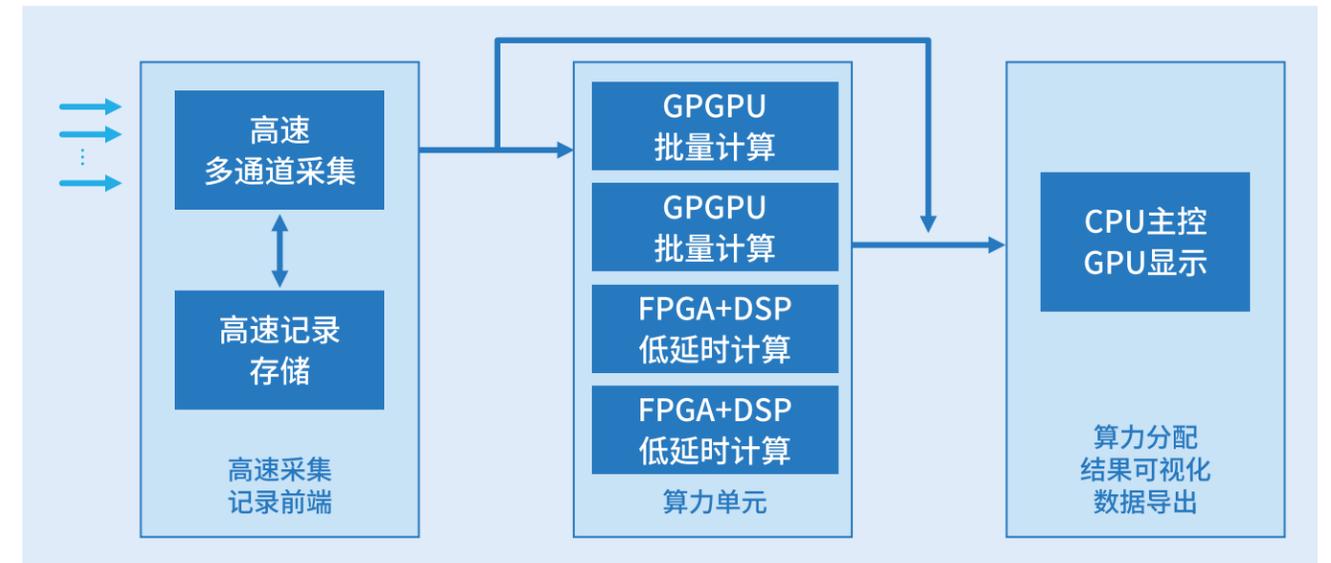
产品框图:



产品特点

嵌入式加固智能计算平台可采集多路高速数据流(图像信号、射频信号等), 通过高速PCIe接口进行点对点直接传输至多个计算模块进行实时低延迟的计算, 原始信号也可送至高速记录存储模块进行保存, 同时第三路通过PCIe输出给CPU主控模块进行可视化显示。

智能计算平台的计算模块可动态重构, 根据任务的计算量和实时性要求, 在FPGA+DSP异构计算模块和GPGPU计算模块之间动态分配任务, 智能化适应不同场景需求。



产品基本特性:

- 加固9U高度VPX10槽位, 遵循VITA48.1/VITA48.2规范
- 计算容量: 4槽位, 支持安装配置4路计算模块, FPGA异构计算模块GPGPU模块可混合配置
- 整机采用IPMI2.0对各计算模块进行智能化状态管理
- 计算模块接入带宽:
 1. GPGPU计算模块接入有效带宽12GB/s或6GB/s(取决于插槽位置)
 2. FPGA异构计算模块接入有效带宽6GB/s
- 最大计算能力:
 - 满配GPGPU计算模块算力96T@fp32, 384T@fp16, 算力可灵活调度, 按通道分割, 按功能分割, 延时优先等
 - 满配FPGA异构计算模块包含V7 690T×8+DSP 6678×16, 支持在线重构
- 高速信号采集输入接口: FMC×2, 遵循FMC HPC规范, 接口模块可灵活更换
 1. 可支持JESD 204B 高速ADC/DAC接入, 可选AD9208、AD9172等射频前端接入子卡
 2. 支持多路cameralink红外图像与偏振光图像接入, 可选多路cameralink红外或可见光前端接入子卡
 3. 支持多通道高速光口数据输入输出, 用于高速数据源记录和回放
- 计算数据延迟: 支持高带宽低延时PCIe P2P传输, 采集至计算最低延迟小于5ms
- 数据记录带宽: 5GB/s
- 数据回放与记录:
 1. 万兆以太网数据回放: 500MB/s (SDK方式)
 2. 提供16路记录逻辑通道
 3. 提供16路回放逻辑通道
- 功耗: 1300W (GPGPU计算模块满配)
- 运行系统: 银河麒麟V10版本
- 主控CPU型号: FT2000+64
- 工作温度: -40°C~+65°C