NextIO vCORE™ Express
WORKGROUP COMPUTING SOLUTION FOR HPC

The Latest in GPU Computing
The NextIO vCORE™ Express GPU computing solution delivers "must have" features for the technical and enterprise workgroup computing space including ECC memory for uncompromised accuracy and scalability, and 7X the double precision performance compared to NVIDIA Tesla 10-series GPU computing products. Compared to typical multi-core CPUs, Tesla 20-series computing systems deliver equivalent performance at 1/10th the cost and1/20th the power consumption.

Powerful and Dense Solution
Designed with four Fermi-based NVIDIA Tesla 20-series computing processors with up to 24GB of total GPU RAM in a standard 1U chassis. The vCORE™ Express scales to solve the world's most important computing challenges – more quickly and accurately.

The vCORE Express achieves:
- Denser Server Clusters - Separating GPUs from servers over high speed PCI Express, more dense server GPU clusters can be realized in the data center.
- Easier Lifecycle Management - GPUs and servers each have different lifecycles. vCORE Express allows the GPU and the server to be managed separately while minimizing downtime.
- Lower Total Cost of Ownership - Compared to multi-core CPU clusters, vCORE Express is about a 1/10th of the cost-factoring in the number of servers and infrastructure for equivalent performance to GPU computing systems. When compared to servers with embedded GPUs, the savings associated with lifecycle management will often justify the price of a vCORE system.

Form Factor	1U 19" Chassis
Number of TESLA GPUs	4 M2070 or 4 M2090
GPU Memory Speed	Express 2070: 1.55 GHz Express 2090: 1.85 GHz
GPU Memory Interface	384-bit
GPU Memory Bandwidth	Express 2070: 148 GB/sec Express 2090: 177 GB/sec
Double Precision Floating Point Performance	1.3 Tflops (Peak)
Single Precision Floating Point Performance	2.06 Tflops
Total Dedicated Memory*	24GB GDDR5 / 12GB for M2050
Power Consumption	900W TDP (Typical)

Product Ordering Information

SKU	Components	Description
C210-MN2070-B	vCORE Express 2070	4 NVIDIA M2070 GPUs Barebones w/ 3yr Warranty
C210-MN2090-B	vCORE Express 2090	4 NVIDIA M2090 GPUs Barebones w/ 3yr Warranty
Option SKU	**Option Components**	**Option Description**
HIC-X16A	P797 Gen2 PCIe x16 HIC	vCORE Extreme PCIe 2.0 x16 Active Host Interface card (supports x16 cables lengths up to 2 meters)
HIC-DX16A	P984 Gen2 PCIe x16 DHIC	Dual cable connect - PCIe 2.0 x16 Active Host Interface card (supports x16 cables lengths up to 2 meters)
CAB-X1605M	0.5 Meter x16 to x16 Cable	0.5 Meter x16 to x16 Cable
CAB-X1620M	2 Meter x16 to x16 Cable	vCORE Extreme 2 Meter x16 to x16 Cable

For assistance in determining the correct configuration for your organization please contact NextIO at 1-877-7-NEXTIO or via email at info@nextio.com or contact your authorized NextIO reseller.

NextIO vCORE™ Features and Benefits

Express 2070: 1536 CUDA Cores (448 per GPU) Express 2090: 2048 CUDA Cores (512 per GPU)	Delivers up to 665 Gigaflops of double-precision peak performance in each GPU, enabling 2.6 TeraFLOPS of double precision performance in a 1U of space. Single precision peak performance is over 5.3 TeraFLOPS.
ECC Memory	Meets a critical requirement for mission critical applications with uncompromised computing accuracy and reliability. Offers protection of data in memory to enhance data integrity and reliability for applications.
System Monitoring Features	Simplifies management and remote monitoring post-installation via NVSMI. Status lights on the front and rear of the unit ensures IT staff can see the status on either side of the rack.
vCORE Express 2070/2090 Includes 24GB of GDDR5 Memory (6GB per GPU)	vCORE Express 2070/2090 includes 6GB/GPU for supporting the most common GPU applications.
NVIDIA Parallel DataCache™	Accelerates algorithms such as physics solvers, ray-tracing, and sparse matrix multiplication where data addresses are not known beforehand.
NVIDIA GigaThread™ Engine	Maximizes the throughput by faster context switching that is 10X faster than previous architecture, concurrent kernel execution, and improved thread block scheduling.
Asynchronous Transfer	Turbocharges system performance by transferring data over the PCIe bus while the computing cores are crunching other data. Even applications with heavy data-transfer requirements, such as seismic processing, can maximize the computing efficiency by transferring data to local memory before it is needed.
CUDA Programming Environment with Broad Support of Programming Languages and APIs	Choose C, C++, OpenCL, DirectCompute, or Fortran to express application parallelism and take advantage of the "Fermi" GPUs innovative architecture.