



## Workgroup Computing Solution for HPC

### The Latest in GPU Computing

The NextIO vCORE™ Express Lite GPU Computing Solution is designed for entry-level GPU computing yet delivers all the “must have” features for the technical and enterprise workgroup computing space. Compared to typical multi-core CPUs, Tesla 20-series computing systems deliver equivalent performance at 1/10th the cost and 1/20th the power consumption.

### Powerful and Scalable Solution

Designed with two Fermi-based NVIDIA Tesla 20-series computing processors with 12GB of total GPU RAM in a standard 1U chassis. The vCORE Express Lite scales to solve the world’s most important computing challenges – more quickly and accurately.

### vCORE Express Lite Delivers:

- Full GPU communication bandwidth – Dedicating full bandwidth to a single GPU gives new options to the HPC community. Choices are no longer limited to a select set of servers that have the cooling and power necessary to drive GPU computing.
- Accurate provisioning – For projects that cannot effectively utilize multiple GPUs per server, the 2 GPU option gives the HPC community the ability to provide only the needed capacity to a project.
- Cost effective entry – Traditionally, entry cost has been a barrier of entry into the GPU market. By delivering only 2 GPUs, the cost of building a test or starter cluster is dramatically reduced while maintaining the flexibility to test different configurations on a server.



#### Lower TCO

1/10th the cost and 1/20th the power consumption of traditional CPUs

#### Cost Effective

Removes cost barriers to entry-level GPU computing

#### Standards Based

PCIe host connections and GPUs from industry leading vendor NVIDIA

#### High Performance

Full PCIe x16 bandwidth to each GPU

#### Server Independent

Manage server and GPU lifecycles independently

#### Serviceable

Easy to remove, replace, or upgrade without opening servers

Form Factor	1U 19" Chassis
Number of TESLA GPUs	2 M2075 or 2 M2090
GPU Memory Speed	1.55 GHz
GPU Memory Interface	384-bit
GPU Memory Bandwidth	148 GB/sec
Double Precision Floating Point Performance	1.03 TFLOPS (Peak) M2075 1.33 TFLOPS (Peak) M2090
Single Precision Floating Point Performance	2.06 TFLOPS (Peak) M2075 2.66 TFLOPS (Peak) M2090
Total Dedicated Memory*	12GB GDDR5
Power Consumption	500W TDP (Typical)
System Interface	PCIe x16 Gen2

\*Note: With ECC on, a portion of the dedicated memory is used for ECC bits, so the available user memory is reduced by 12.5%. (e.g. 6GB total memory yields 5.25GB of user available memory.)

## Product Ordering Information

SKU	Components	Description
C210-MN2075-2-B C210-MN2090-2-B	vCORE Express Lite vCORE Express Lite	2075 2 NVIDIA M2075 GPUs Barebones w/ 3yr Warranty 2090 2 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](mailto:info@nextio.com) or contact your authorized NextIO reseller.

## NextIO vCORE™ Features and Benefits

<p>996 CUDA Cores (448 per GPU for M2075) 1024 CUDA Cores (512 per GPU for M2090)</p>	<p>Delivers up to 665 of double precision peak performance in each GPU, enabling 1.33 TeraFLOPs of double precision performance in 1 U of space. Single precision peak performance is over 2.5 TeraFLOPs.</p>
<p>ECC Memory</p>	<p>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.</p>
<p>System Monitoring Features</p>	<p>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.</p>
<p>vCORE Express 2070 Includes 24GB of GDDR5 Memory (6GB per GPU)</p>	<p>Maximizes performance and reduces data transfers by keeping larger data sets in local memory that is attached directly to the GPU. Delivers up to 665 of double precision peak performance in each GPU, enabling 1.33 TeraFLOPs of double precision performance in 1 U of space. Single precision peak performance is over 2.5 TeraFLOPs.</p>
<p>NVIDIA Parallel DataCache™</p>	<p>Accelerates algorithms such as physics solvers, ray-tracing, and sparse matrix multiplication where data addresses are not known beforehand.</p>
<p>NVIDIA GigaThread™ Engine</p>	<p>Maximizes the throughput by faster context switching that is 10X faster than previous architecture, concurrent kernel execution, and improved thread block scheduling.</p>
<p>Asynchronous Transfer</p>	<p>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.</p>
<p>CUDA Programming Environment with Broad Support of Programming Languages and APIs</p>	<p>Choose C, C++, OpenCL, DirectCompute, or Fortran to express application parallelism and take advantage of the “Fermi” GPUs innovative architecture.</p>