SUPERMICRO GPU-OPTIMIZED GRID, HPC & WORKSTATION SOLUTIONS

GTC March 18, 2013
Presented by: Don Clegg
VP, Marketing & Business Development
GTC 2013 Press Release – San Jose, California, March 18, 2013

Supermicro®

Powering the Cloud with NVIDIA GRID™
Server Platforms Provide Virtualized GPU Clusters with NVIDIA GRID K1 and K2

4U FatTwin™

12x GPUs
3x GPUs per Node
SYS-F627G542T+FTJ+ST3P1+

5x GPUs
4U Tower 2x 150W CPU
NVIDIA® Maximus™ Certified
SYS-7047GR-TRF

www.supermicro.com/GPU
Presentation Agenda

Supermicro Overview

Why Supermicro

GPU-Optimized Roadmap:
- FatTwin™
- 1U / 2U / 3U Rackmount
- 4U SuperWorkstation™
- SuperBlade™

Benefits Summary
As GPU-enabled computing matures, selecting the best hardware platform is more essential than ever. Successful enterprises understand the importance of optimizing compute power and density, I/O bandwidth and latency, plus electrical power-efficiency and cooling to ideally match the intended application within the specified budget.

Supermicro, with its industry-leading building-block solutions, delivers the most comprehensive range of GPU-optimized platforms on the market. This presentation, featuring Supermicro's FatTwin™, Twin™, SuperBlade™ and rack/tower building blocks will highlight some of the most important architectural innovations to consider when selecting the best GPU platforms for HPC, Cloud/GRID and Workstation applications.
Translation

- Supermicro delivers the most versatile GPU-enabled solutions with Industry Leading:

  *Performance per Watt*
  *Performance per *
  *Performance per \(ft^3\)*

- This empowers Supermicro Partners to deliver the best solutions for

  GRID, HPC and Workstations
Supermicro Overview

- Founded in 1993, headquartered in Silicon Valley, USA (NASDAQ 2007: SMCI)
- Supermicro is the market leader in server technology innovation and green computing
- Supermicro has the broadest server portfolio in the industry; designs are in-house / “under the same virtual roof”
- Supermicro’s application-optimized high-efficiency GPU, Server, Blade, Storage, embedded and workstation solutions are deployed globally for data centers and mission critical IT infrastructures
Supermicro Global Support

Main Campus (San Jose)
Fremont Facility

Corporate locations (USA HQ, Taiwan, Netherlands)
Local support in more than 70 countries
Distribution and VAR alliances
Commitment to help customers achieve the most optimized GPU, HPC, Cloud and Datacenter Solutions
Asia Science & Technology Park – In Full Operation

Production Capacity

- 30K Server Systems per month integration capacity
- 2x current manufacturing capacity
- Local Asia Operations and Logistics Support – Speedy Response
- 1.2M sq. ft. (2M sq. ft. when completed) dramatically increases global reach
The Supermicro Time-to-Market Advantage

Server Building Block Solutions®

Resource Optimized

Customer Driven Design Expertise

Application Optimized

>550 Motherboards

>1300 Chassis

> 350 Cooling Modules

> 140 Power Supplies

Open CPU/Memory

Operating Systems/Applications

Industry Standard

Customer Specific
The Strongest Application Optimized Products

- Broadest IT, HPC & datacenter solutions available in the industry
New Technologies and Innovation

- GPU Optimized Products
  - First with NVIDIA Kepler K10/K20/K20X and GRID K1/K2
- FatTwin™ Architecture
  - New FatTwin™ Architecture: 8/4/2 nodes in 4U, 17% Performance Increase
  - 8HS 3.5" HDDs in 1U, Front I/O, 135W CPU, PUE Optimized,
  - GPU GRID and Hadoop, models, First to Market, Density/Cost Optimized
- Workstation Products
  - Maximus 2.0 (NVIDIA Tesla + Quadro)
  - Optimized for High Frequency Trading (HFT): hyper-speed
- IPC and Embedded Applications
  - Increased product breadth, dedicated FAE and Sales teams
- UP Solutions
  - Volume products. X10 in plan
- Switch Products:
  - Volume products, Blade FCoE announced
- MicroCloud™ and MicroBlade (TBA)
  - 12-node system in production (Very strong demand)
- Power Subsystems
  - 200W, 1000W BBP based Server/Storage solutions available
- Software
  - Windows 8 Readiness
Why Choose Supermicro GPU Solutions

- Highest density
- Support top CPU (150W) & GPU
- Optimized cooling & system design

Best product specs

Variety of Choice

Supplier / Customer Relationships

Extensive Validation & expertise

- 1U, 2U, 3U, 4U, 7U
- UP, DP
- 8, 16, 24 DIMM
- Rack mount, Tower, Blade, FatTwin

- 25%+ market share
- Best quality
- Best performance
- Best reliability

BEST Performance per Watt
Performance per $
Performance per ft$^3$
Technology Progression – GPU Application Optimization

- **Telsa S1070**
  - 1U 4-GPU Standalone box
  - PCI-E x16

- **6016TT-TF 1U Twin™**
  - The fastest 1U server in the world

- **SW7046GT-TRF**
  - The first optimized CPU/GPU integrated Hybrid System

- **SC827HD-R1400B**
  - Next generation CPU & GPU support

- **2U Twin**
  - 2U CPU / with QDR IB onboard

- **2U, 6-GPU Integrated***
- **1U, 4-GPU Integrated***

- **4U, 4 Nodes, 12-GPU Integrated**

- **2008 2009 2010/11 2012/13**
  - The most powerful PSC

- **7U, 10 Nodes, 20-GPU**

- **4U, 4 Nodes, 12-GPU Integrated**

- **2U, 6-GPU Integrated***

- **1U, 4-GPU Integrated***

- **7U, 10 Nodes, 20-GPU**

- **4U, 4 Nodes, 12-GPU Integrated**

- **2U, 6-GPU Integrated***

- **1U, 4-GPU Integrated***

- **7U, 10 Nodes, 20-GPU**
Supermicro GPU Solutions

- **GPU Server (Passive GPU)**
  - 1U UP – Value
    - 1017GR-TF
    - 5017GR-TF
  - 1U/2U DP – Scalable, High Density
    - 1027GR-TQF
    - 1027GR-TQFT
    - 1027GR-TRF
    - 1027GR-TRFHT
    - 2027GR-TRFH
    - 2027GR-TRFHT
  - 3U & Above – Powerful
    - 7047GR-TRF
    - 7047GR-TPRF
    - 6037R-72RFT+
    - GPU FatTwin
    - GPU Blade

- **GPU Workstation (Active GPU)**
  - 1U
    - 5037A-iL
    - 5037A-i
  - 2U
    - 7037A-iL
    - 7037A-i
    - 7047A-T
  - 4U
    - 7047GR-TRF
Fat Twin – Front or Rear I/O

Fat Twin – Front I/O

GPU Optimized Configurations

4U with 4-node/8-node
PUE optimized – 130W/135W
More flexible, dense storage
- 2.5”/3.5” SAS/SATA; 8 HDD/1U
More add-on cards

Fat Twin – 1U/2U/4U Nodes

FatTwin™ Overview

17%+ Lower Power Consumption*
30%+ More Storage Capacity*
Up to 5 Add-on Cards per Node

3 Double-Width GPUs per node
Full Support for Dual Xeon 135W
Power & Cooling Optimized
Many more features...

Standard 19” Racks, Front Access Server Nodes, Industry Standard IPMI 2.0
Highest Efficiency, Redundant Platinum-Level, Power Supply (95%+)

* See Supermicro for detailed information solutions
Leading Architecture - Fat Twin GPU Solutions

FatTwin = Best Performance / Watt / $ / ft³

12 Double Width GPU in 4U enclosure

- 4 nodes in 4U, Front I/O
- 3 GPUs per node, total 12 GPUs supported
- 3 PCI-E Gen3 x 16 slots
- 2 additional PCI-E Gen3 x8 slots
- Dual Intel Xeon CPU, 135W CPU supported
- Up to 16 DIMM, 512GB memory
- 6 x 2.5” or 2 x 3.5” hot-swap SAS/SATA/SSD
- 10GbE onboard option
- High-efficiency platinum level 1620W redundant power supply
- Hot-swappable cooling fans

SYS-F627G3-F73+/FT+
1U DP Rackmount GPU Server

Up to 3 Double Width GPU cards + 1 LP add on card
3 GPUs: SYS-1027GR-TRF
SYS-1027GR-TSF
SYS-1027GR-TRFT

10Gbe onboard
SYS-1027GR-TRFT

1800W redundant power supply (Single for 1027GR-TSF)

Intel® Xeon® Sandy Bridge
E5-2600 Series Socket R
Platinum level high-efficiency

Mellanox ConnectX QDR/FDR InfiniBand

4 Hotswap 2.5” SATA/SSD Drive

Optimized Cooling System Design

Up to 3 GPU Cards
K1, K2, K10, K20, K20X

X9DRG-HF/HTF
1U UP Rackmount GPU Server

Up to 2 Double Width GPU cards + 1 LP add on card
SYS-1017GR-TF
SYS-5017GR-TF

Intel® Xeon® Sandy Bridge E5-2600 Series Socket R

Platinum level high-efficiency 1400W power supply

Optimized Cooling System Design

Mellanox ConnectX QDR/FDR InfiniBand

Up to 2 GPU Cards K1, K2, K10, K20, K20X

X9SRG-F
K1 / K2 2U Reference Designs

Enterprise Reference Architecture for Designers

4 User (Dedicated GPU) Virtual Workstation
2 GRID K2 boards (4 NVIDIA High-end GPUs)
Dual CPU socket 2U server
64GB System memory (16GB per user)
1100W per 2U

Reference Architecture for Power Users

8 User (Dedicated GPU) VDI server
2 GRID K1 boards (8 NVIDIA GPUs)
Dual CPU socket 2U server (32 HT cores)
64GB System memory 800W per 2U

Enterprise VDI Reference Architecture for Knowledge Workers

8-32+ User (Dedicated or Shared GPU) VDI server
2 GRID K1 boards (8 NVIDIA GPUs)
Dual CPU socket 2U server
Minimum 128GB System memory
800W per 2U
2U Rackmount GPU Server

Up to 6 Double Width GPU cards + 1 LP add on card
- 6 GPUs: SYS-2027GR-TRFH
- 4 GPUs: SYS-2027GR-TRF
- 1 LP add on card: SYS-2027GR-TRFT

10 Hotswap 2.5” SATA/SSD Drive

10GbE onboard
- SYS-2027GR-TRFT
- SYS-2027GR-TRFHT

Optimized Cooling System Design

Intel® Xeon® Sandy Bridge
- E5-2600 Series Socket R

Platinum level high-efficiency
- 1800W redundant power supply

Mellanox ConnectX QDR/FDR InfiniBand

Up to 6 GPU Cards K1, K2, K10, K20, K20X

X9DRG-HF/HTF
3U Rackmount GPU Server
SYS-6037R-72RFT+

- Model Name: SYS-6037R-72RFT+
- CPU: 2 Xeon E5-2600 series
- Memory: Up to 24 DDR3 1600 Reg. ECC, 768GB
- HDD: 8 H/S 3.5" (Optional 2.5"x2 at rear)
- Controller: LSI 2208 SAS2 8ch
- I/O slots: 2 FHFL e3.0x16 (Opt: Dual M-series GPU Support), 3 FHFL e3.0x8, 1 FHHL e3.0x8
- NIC: 2 10GbE (X540)
- IPMI: IPMI 2.0 & KVM w/ LAN
- PSU: 1280W Redundant Platinum, Optional Battery Unit for 2+1
- Size: 3U x 25.5" (648mm)

Optional rear fan

Up to 2 GPU Cards K1, K2, K10, K20, K20X

SAS LSI 2208 Hardware RAID controller onboard for 8 SAS HDDs

1000W Battery Backup Power

Optional 2.5" HDD bay

Up to 24 DDR3 768GB

X9DR7-TF+

2 10GbE ports

Up to 2 GPU Cards K1, K2, K10, K20, K20X
Tower / 4U Rackmount GPU Server
SYS-7047GR-TPRF

❖ Best Cooling → 150W CPU

❖ More Slots...
Up to 4 Double Width GPU cards
+ 2 PCIe3.0x8, 1 PCIe2.0x4

4 GPU Cards

❖ Passive GPU for Server / Workstation
Only from Supermicro!
Tower / 4U Rackmount GPU Workstation
SYS-7047GR-TRF

- Up to 4 Double Width GPU cards
  + 2 PCIe3.0x8, 1 PCIe2.0x4
- 8 Hotswap 3.5" SATA/SSD Drive
- 4 C-series GPU Cards
- Intel® Xeon® Sandy Bridge E5-2600 Series Socket R
- Platinum level High-efficiency 1620W redundant power supply
- Optimized Cooling System Design
- Mellanox ConnectX QDR/FDR InfiniBand
- 8 Hotswap 3.5" SATA/SSD Drive

X9DRG-QF
High Performance UP SuperWorkstation
SYS-5037A-i

- **Gold level** high-efficiency 900W power supply
- 2 C-series GPU cards
- Intel® Xeon® Sandy Bridge E5-1600/2600 Series Socket R (Single socket)
- Support up to 256GB DDR3 RDIMM or 64GB DDR3 UDIMM
- 2 Front USB 3.0
- 90° Rotatable HDD Cage Design

- 2 PCIe3.0 x16 slots,
- 1 PCIe3.0 x 4 in x8 slot,
- 1 PCIe2.0 x 4 in x8 slot,
- 1 PCI-32 slot

**Intel inside® Xeon®**
Recap: Supermicro = Best/Most GPU Solutions

- **GPU Server (Passive GPU)**
  - 1U UP – Value
    - 1017GR-TF
    - 5017GR-TF
  - 1U/2U DP – Scalable, High Density
    - 1027GR-TQF
    - 1027GR-TQFT
  - 3U & Above – Powerful
    - 7047GR-TPRF

- **GPU Workstation (Active GPU)**
  - 1U UP – Value
    - 5037A-iL
    - 5037A-i
  - 2U DP – Scalable, High Density
    - 7037A-iL
    - 7037A-i
  - 3U & Above – Powerful
    - 7047A-T
    - 7047GR-TRF
See our products in Booths 207 & 306
THANKS

Q & A