Happy December 2020 🎄
Congrats to the Chang‘e5 team!
ANNOUNCING
NVIDIA A100 80GB
Supercharging The World’s Highest Performing AI Supercomputing GPU

- 80GB HBM2e
  For largest datasets and models

- 2TB/s +
  World’s highest memory bandwidth to feed the world’s fastest GPU

- 3rd Gen Tensor Core

- Multi-Instance GPU

- 3rd Gen NVLink
NVIDIA SELENE

Now Featuring NVIDIA DGX A100 640GB

- 4,480 A100 GPUs
- 560 DGX A100 system
- 850 Mellanox 200G HDR switches
- 14 PB of high-performance storage
- 2.8 EFLOPS of AI peak performance
- 63 PFLOPS HPL @ 24GF/W
TODAY’S AI DATA CENTRE

- 50 DGX-1 systems for AI training
- 600 CPU systems for AI inference
- $11M
- 25 racks
- 630 kW
DGX A100
DATA CENTRE

- 5 DGX A100 systems for AI training and inference
- $1M
- 1 rack
- 28 kW

$1M  28 kW
1/10th COST
1/20th POWER
NVIDIA CUDA-X AI ECOSYSTEM
EXPANDING NGC

NEW CONTAINERS FOR A100 & ARM
Now

HPC Simulation & Visualization
- AutoDock 4
- Chroma
- GROMACS
- LAMMPS
- NAMD
- ParaView
- RELION
- VMD

AI Frameworks (A100)**
- TensorFlow
- PyTorch
- mxnet

NEW FEATURES
Now

- NGC Private Registry
- Multi-arch support for x86, Arm and Power
- Higher HPC app performance w/ NVTAGS

NGC Container Environment Modules

NGC-READY SYSTEMS FOR A100
Starting Q3

- AWS
- Azure
- Oracle Cloud
- Alibaba Cloud
- Atos
- Dell
- Hewlett Packard Enterprise
- Lenovo
- Supermicro

Learn More - ngc.nvidia.com | NGC Private Registry | NVTAGS | NGC Container Environment Modules

* Available week of June 22
** Available starting with v20.06
RAPIDS is a set of open source software libraries which gives you the freedom to execute end-to-end data science and analytics pipelines entirely on GPUs.

www.rapids.ai
SIMNET v.0.2

AI-accelerated Physics Simulation Toolkit

- Solve larger problems faster with XLA and AMP support, and Multi-GPU, Multi-Node implementation
- Models Multiple Physics in Forward, Inverse and Data Assimilation simulations with accuracy & convergence
- Parameterized system representation to solve multiple scenarios simultaneously
- APIs for implementing new Physics, Geometry, and Domains and detailed User Guide examples
FOURIER NEURAL OPERATORS

A176 - AI Fanless Small FF Supercomputer

A176 – Used for:
• Video processing
• Recording data

The ball is a recorder that is attached to A176 via a cable. There is an explosive device that will break the link with the A176, so the ball will fall to the earth without being destroyed.
NVIDIA AGX
Family of Systems for Embedded AI HPC
Self-driving cars
Robotics
Smart Cities
Healthcare

NVDLA.org
Orbital Sidekick
# JETSON TX2i

**MODULE FOR INDUSTRIAL ENVIRONMENTS**

<table>
<thead>
<tr>
<th>FEATURES</th>
<th>JETSON TX2</th>
<th>JETSON TX2i</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>GRAPHICS</strong></td>
<td>NVIDIA Pascal™, 256 CUDA cores (Up to 1.3 GHz)</td>
<td>NVIDIA Pascal™, 256 CUDA cores (Up to 1.26 GHz)</td>
</tr>
<tr>
<td><strong>CPU</strong></td>
<td>HMP Dual Denver 2/2MB L2 + Quad ARM® A57/2MB L2 (Up to 2.0 GHz)</td>
<td>HMP Dual Denver 2/2MB L2 + Quad ARM® A57/2MB L2 (Up to 1.95 GHz)</td>
</tr>
</tbody>
</table>
| **VIDEO**      | 4K x 2K 60Hz Encode (HEVC)  
                 | 4K x 2K 60Hz Decode (12 bit support) | 4K x 2K 60Hz Encode (HEVC)  
                 | 4K x 2K 60Hz Decode (12 bit support) |
| **MEMORY**     | 8 GB 128 bit LPDDR4  
                 | 3733 MT/s  
                 | 8 GB 128 bit LPDDR4  
                 | 3200 MT/s with ECC |
| **DISPLAY**    | 2x DSI, 2x DP 1.2 / HDMI 2.0 / eDP 1.4 | Up to 6 cameras (2 lane)  
                 | CSI2 D-PHY 1.2 (2.5 Gbps/lane) |
| **CSI**        | Gen 2 | 1x4 + 1x1 OR 2x1 + 1x2 |
| **PCIE**       | Data Storage | 32GB eMMC, SDIO, SATA |
| **OTHER**      | CAN, UART, SPI, I2C, I2S, GPIOs | CAN, UART, SPI, I2C, I2S, GPIOs |
| **CONNECTIVITY** | 1 Gigabit Ethernet, WLAN, BT | 1 Gigabit Ethernet |
| **MECHANICAL** | 50mm x 87mm x 10.4mm (400-pin)  
                 | Compatible Board to Board | 50mm x 87mm x 11.8mm (400-pin)  
                 | Compatible Board to Board Connector |
| **INPUT VOLTAGE** | 4.5V - 19V | 9V - 19V |

Introducing EGX A100 Converged Accelerator
Combining Mellanox and NVIDIA Ampere GPU Architecture

Unprecedented AI Inference Performance
- Ampere based Architecture
- 3rd Generation Tensor Core

Enhanced Security
- Secure GPU enclave protects AI model
- Line-speed TLS & IPSec Crypto Engines
- Service Mesh Offloads (SDN)

In-Line Network Acceleration
- Dual 100Gb/s Ethernet or InfiniBand
- Accelerated Switch & Packet Processing
- Time Triggered transmission tech for Telco (5T for 5G)
BIG DATA PIPELINE

Ingredients:
- Lots of data
- Lots of compute
- Software tools
- Time and patience

Method:
1. Collect raw, massive sets of data.
2. Put the data in a Data Lake.
3. Grab the data that you need and sort through.
4. Find patterns in the data.
5. Solve the problem.

INGESTION ➢ STORAGE ➢ PROCESSING ➢ SERVING

1. Obtaining and importing data
2. Organizing & storing data for future use
3. Manipulating and analyzing the data
4. Operationalizing the solution
Adding Physics to USD
TRAIN AN ALGORITHM TO EXAMINE EVERY PIXEL

GOES-16: 4k x 4k x 11 channels
Monitor Environmental Change

- drought
- flooding
- deforestation
- urbanification
- melting glaciers
- sea-level change
Use Inpainting to Repair Damaged GOES-17 Observations
Real-Time Photorealistic Digital Humans with NVIDIA RTX A6000

“Digital Domain leads the industry in pioneering real-time photo-realistic digital humans. Working with Epic Games and NVIDIA, we have continuously pushed the limits of technology. The new RTX A6000 lets us completely redefine what’s possible with real-time ray tracing and machine learning.”

Darren Hendler | Director, Digital Human Group
GALACTIC WINDS
NVIDIA IndeX and Cholla
Brant Robertson
Evan Schneider
Simulated samples of a dolly (with actual CAD model) used to train object detection and pose estimation neural networks.

Procedurally generated simulated images used for segmentation network training.

Multiple Carter robots operating simultaneously in virtual warehouse; Each operated by an independent Jetson Xavier.

ISAAC
nvidia.com/en-gb/deep-learning-ai/industries/robotics
MUCH MORE WITH ISAAC SOFTWARE
GPU Accelerated Algorithms/DNNs (GEMs)

Free Space Segmentation
3D Object Pose Estimation
Motion Planning
Stereo Depth
Stereo Visual Inertial Odometry
Super Pixels
April Tags
2D Skeleton Pose Estimation
DeepStream Integration
Planner with Costmaps
Multi Lidar Support
Navigation (LQR Path Planner)
Sensors
Robot Platforms
Audio
And more...
RICH CONTENT PORTFOLIO

Fundamentals and advanced hands-on training in key technologies and application domains

- Deep Learning Fundamentals
- Accelerated Computing Fundamentals
- Accelerated Data Science Fundamentals
- Intro to AI in the Data Center
- AI for Anomaly Detection
- AI for Autonomous Vehicles
- AI for Digital Content Creation
- AI for Healthcare
- AI for Industrial Inspection
- AI for Intelligent Video Analytics
- AI for Predictive Maintenance
- AI for Robotics
SHARE YOUR LIFE’S WORK AT GTC 2021

JOIN A GLOBAL COMMUNITY OF BRILLIANT MINDS ONLINE NEXT MARCH

NVIDIA’s GTC brings together a global community of developers, researchers, engineers, and innovators with the common goal of sharing achievements while discovering new technologies and tools that drive change around the globe.

If you work with any of our GPUs, DPUs, or software offerings is making a difference, submit a talk or poster to join us online in March.

March 15 – 25, 2021
Submit your ideas at www.nvidia.com/gtc