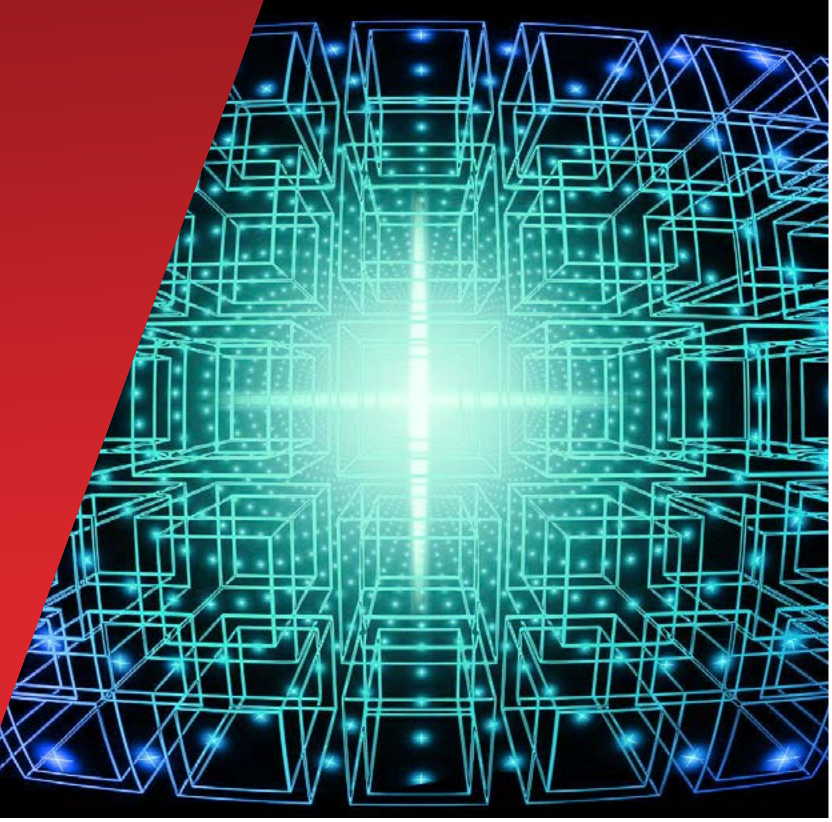


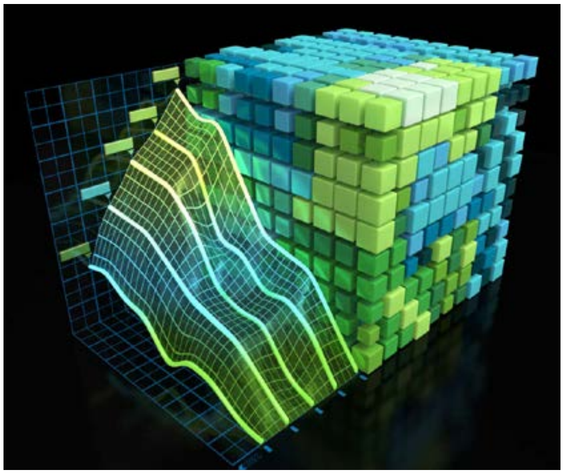
Aitech

Process High Volumes of Data in Rugged Embedded Systems, No Matter the Requirements

April 2020

With embedded processing requirements ramping up across all applications utilizing embedded electronics, new means of handling the vast amounts of data being generated have come to fruition, providing never-before seen levels of computation, in near real time.

One of the forerunners in this evolution is GPU accelerated computing, originally developed by NVIDIA for the gaming industry. But the application of NVIDIA's Compute Unified Device Architecture (CUDA) into non-gaming related embedded computing systems has redefined the parameters of data processing. Data-intense applications that typically operate in a stable, and most times temperature-controlled, environment, like gaming, higher education, and telecommunications, easily jumped on the GPGPU (general purpose calculation on graphics processing unit) bandwagon.



NVIDIA's CUDA architecture is reinventing processing abilities in embedded systems

Graphic courtesy of NVIDIA

But where does that leave system designers who need to build computing platforms subjected to extreme shock and vibration as well as extreme/wide temperature and humidity fluctuations that range from sub-zero to triple digits…in addition to supporting crucial, lifesaving and security-focused applications?

Defining Application Challenges

Knowing how to transition this exciting processing technology into the realm of harsh environments is not insurmountable. In fact, if you apply the same principles of ruggedization employed in developing other rugged embedded systems, a pattern of thinking will surely fall into place. The key then lies in making sure you are employing GPU accelerated computing in the best way for your defense, military or mission-critical application. View the White Paper: GPU Accelerated Computing Technology Update.

Taking stock of your system requirements is a good place to start. Today's military systems are utilizing far more resources in a much smaller footprint, typically referred to as optimized SWaP—size, weight and power—while needing to keep costs low. In addition, these applications function in extremely harsh environments, and carry with them the need to operate reliably all the time, every time. This dichotomy has plagued many electronic engineers developing critical military, defense and space systems for decades, but, as history has shown, these challenges can not only be mitigated, but met, as well.

Rugged GPGPU

It's important to always keep a close eye on balancing your system needs and processing requirements with the rugged aspects of the application environment. By relying on proven ruggedization techniques as well as verified testing methodologies, GPU accelerated computing can offer unique advantages in system performance, even in the harshest of environments. Below is a review of the top system requirements generally facing a design engineer and where rugged GPGPU fits in.

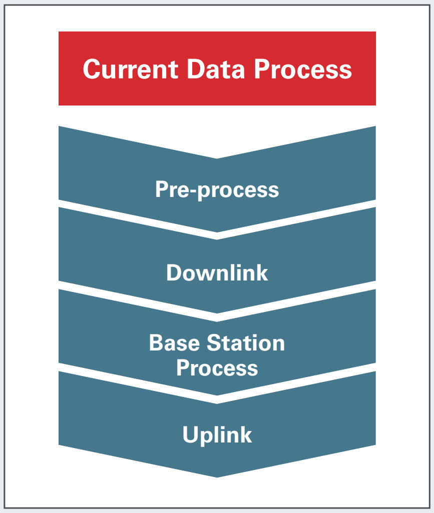
Data Processing

The myriad of inputs an embedded system is required to process continues to grow. On top of that, for mission-critical applications, accuracy, reliability and speed are essential to a real-time, decision-making process, whether it's a response required by an actual human or through artificial intelligence. Because it is based on a parallel architecture, GPU accelerated computing processes tens of thousands of data points simultaneously, versus only hundreds using serial processing. Even a typical multi-core CPU-based architecture only offers a handful of cores running in parallel. When integrated into a ruggedized system, GPGPUs can meet the growing data requirements of today's military, defense and space applications.

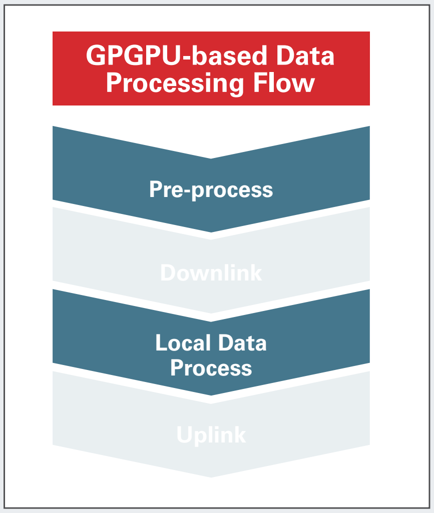
Reducing Data Latency

A traditional flow of data processing for remote applications involves a "pre-process + downlink + base station process + uplink" configuration that affects data latency, as shown in Diagram A. With a high performance, low power GPU accelerated computing system, "decision latency" can be calculated as depicted in Diagram B, providing a more real-time response than the previous methodology:

Dia. A



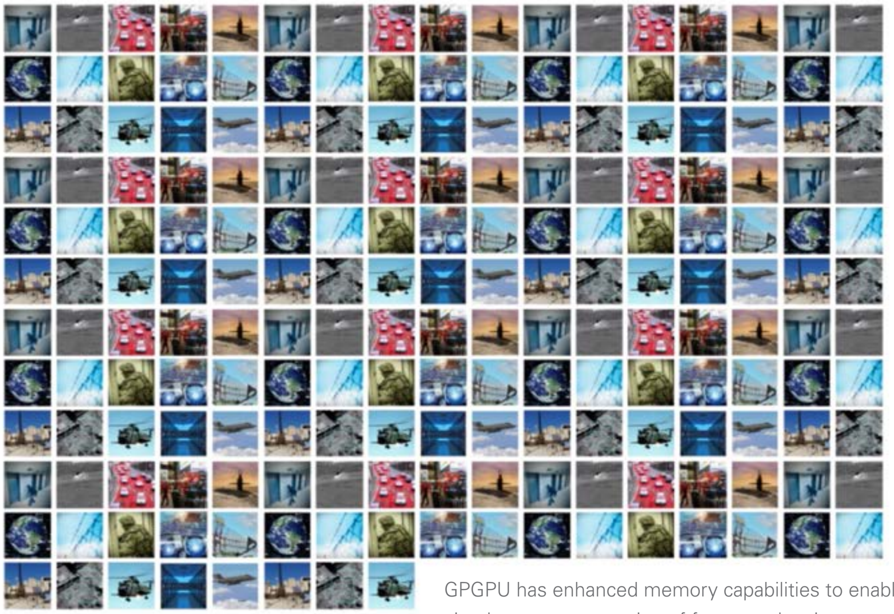
Dia. B



1. Input video data and preprocessing (preprocess can be data encoding, recording, etc.)
2. Local data process; can be accelerated by CUDA API and Deep Learning capabilities

Memory and Storage

Where exactly is all this data supposed to be stored? Unlike more stationary operating environments, rugged, mobile and mission-critical systems operate in places with volatile connections and may need to house data onboard until the network is restored. Think about a satellite in orbit or a ground vehicle in a remote area. The data can't necessarily be transmitted real-time, but when it is, it needs to go fast and it needs to be right. Onboard storage capacities have been increasing, thanks to compact, high density Flash-based modules, high-speed NVME protocols and secure PCIe-based interfaces now available. In addition, memory capabilities have been enhanced by GPGPUs, allowing processing closer to the edge, near the sensor, where it is needed most. So, when the data is transmitted, it is potentially already pre-processed and actionable.



GPGPU has enhanced memory capabilities to enable simultaneous processing of far more data input

Power Consumption

The question is not only one of managing power resources across compact, high density systems, but rather how does GPU accelerated computing increase the power-to-performance ratio of these systems. Some GPGPU boards are very power efficient (especially the ones based on NVIDIA's Jetson family, due to the fact that the CPU is ARM-based), with some boards offering the same consumption requirements as CPU and GPU boards together. However, GPGPU boards can process far more data using thousands of parallel CUDA cores. So, the power-to-performance ratio is what is affected in a very positive manner. More processing is available to the application for the same, and sometimes slightly less, power.

Proven Use Cases

GPGPU computing is quickly becoming a platform for advanced computing intelligence, due to the hundreds of high-performance cores that provide unprecedented parallel processing capabilities, using general purpose programming languages, such as NVIDIA's CUDA API. With the proper ruggedization techniques in place, military, defense and space programs can benefit from GPGPU computing systems, and the deep learning capabilities they provide.

If you are building a rugged, SWaP-optimized system that needs to capture graphics from several HD-SDI/Composite video inputs and manage data from a lot of I/O interfaces, while providing great image processing capabilities, a rugged GPGPU-based system using CUDA parallel computing is an area you should explore. Below are some documented use cases of this exceptional technology that show how reliable and effective it can be in mission-critical, harsh environments, both manned and unmanned.

Air: Autonomous Electric Vertical Takeoff and Landing (eVTOL) Aircraft

Prototypes of pilotless eVTOLS are being rapidly developed, with many utilizing platforms that already exist, such as drones or unmanned helicopters, then integrating leading edge technologies to achieve the needed functionality of these air transport vehicles. Rugged GPU accelerated computing is at this forefront. In fact, the technology is advancing so quickly that systems are moving to next gen architectures as development is taking place (Watch the Video: NVIDIA's AGX Xavier). In this instance, accommodating the increased sensor processing integrated into the unit is cause for the upgrade to replace typical CPU-based embedded computing architectures.

Land: Ground Mobile Platform

Relied upon to send mission-critical data from the battlefield to a forward battle command, or directly to the soldiers on the ground themselves, tanks and other ground vehicles incorporate several onboard cameras and data collection points to aid in the decision-making process. In one instance, a rugged GPGPU system is capturing images from six cameras—four composite and two HD-SDI—then performing simultaneous image processing applied to object recognition and classification as well as situation awareness. The system is using CUDA to save this sensitive data on internal fast NVME SSD that can be transmitted back to the command center instantly and when needed. The multiple video inputs are processed simultaneously in a low-power, small form factor (SFF), rugged system, which provides a performance-per-watt (PPW) factor that is critical in determining a go/no go for program deployment.

Sea: Onboard Observation System

Unlike a ground vehicle, when a ship is docked at port, maneuverability is quite limited. Some could equate a large vessel tethered to a dock as a 'sitting duck.' Defense organizations are looking for ways to make sure that onboard personnel are as well-equipment to protect against unauthorized incursions or to ensure approaching vessels are quickly recognized and/or identified as friend or foe to ensure the ship is safe and protected from attacks while at rest. A GPGPU-based, multiple camera installation that monitors movement all around a ship's perimeter would help detect potential intrusions and alert the bridge crew to threats or security breaches.

Because of the girth these vessels, multiple camera locations as well as reliable data transmission is paramount. In addition, because the ship will then return to sea, the camera system needs to be sea-worthy, ready to withstand salt spray, ocean waves and high degrees of shock and vibration. Rugged GPU accelerated computing fits the bill perfectly by providing this level of high video processing performance and subsystem ruggedization features to survive these types of applications both at dock and out at sea.

Space: Atmospheric Reentry Program

Space is known for being the birthplace of many a disruptive technology, so it seems fitting that GPU accelerated computing would find a home here as well. One of the more recent applications includes an inflatable heat shield targeted to retrieve reusable engines or for the delivery of heavier cargo to the surface of Mars. A GPGPU system collects video data and telemeters it to ground- or satellite-relay stations for near real-time monitoring of the entire reentry from deployment to touchdown. ■

19756 Prairie Street
Chatsworth, CA 91311
Toll Free: (888) Aitech-8 (248-3248)
Ph: (818) 700-2000
Fax: (818) 407-1502
sales@aitechsystems.com

No. 91 Prestige South End Road
560004 Basanvagudi
Bengaluru, Karnataka, India
Ph: +91-80-4866-8105
Fax: +91-80-4866-8106
Indiasales@aitechsystems.com

4 Maskit Street
PO Box 4128
Herzliya, 4673304 Israel
Ph: +972 (9) 960-0600
Fax: +972 (9) 954-4315
sales@aitechsystems.com

AitechSystems.com