Standardizing GPU Radiation Test Approaches

Edward J Wyrwas

Edward J Wyrwas[1], Carl Szabo[3], Kenneth A LaBel[2], Michael Campola[2] and Martha O'Bryan[3]

1. Lentech, Inc. Greenbelt, MD; 2. NASA Goddard Space Flight Center, Greenbelt, MD; 3. AS&D, Inc., Beltsville, MD

National Aeronautics and Space Administration

Abstract: *A standardized test method has been created to characterize and stress graphics processing units (GPU) during radiation effects testing.*

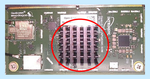
Introduction

While some Graphical Processing Units (GPUs) are discrete components (i.e. GTX 1050), others take the form of an IP block or embedded engine within a System on Chip (SoC) device such as the Qualcomm Snapdragon™ 820 which contains a Qualcomm Adreno™ 530 GPU. Within this device are various functional blocks which can be exercised with software payloads. NVidia's Jetson™ TX1 SoC is provided on a System on Module (SOM). Within it are Central Processing Unit (CPU) cores and an nVidia GPU which can be accessed similarly to a discrete GPU coprocessor. While the packaging is different, each one of these GPUs needs to be tested using the same standardized code.

Table 1: Comparison of GPU Types

Part Model	GTX 1050	APQ8096	Jetson TX1
Manufacturer	nVidia	Qualcomm	nVidia
Technology	16nm FinFET	14nm FinFET	20nm CMOS
REAG ID	GSFC 17-039	JPL	GSFC 16-038
Board Model	EVGA 02G-P4-6152-KR	Intrinsyc Open-Q 820	699-82180-1000-100 U
Packaging	Flip Chip, BGA, PCB	Flip Chip, BGA, SOC	Flip Chip, BGA, SOM
Memory Capacity	2GB GDDR5, >8GB DDR4	3GB LPDDR4	4GB LPDDR4
Performance	1.86 TFLOPs	0.50 TFLOPs	1.00 TFLOP
Test Bench OS	Windows 2016	Android 6 Marshmallow	Linux for Tegra

GTX 1050 Card



APQ8096 SOC (under heatsink)

TX1 SOM

Figure 1: Comparison of GPU Types

A universal test bench is under development to provide a standardized approach to test GPUs with minimal variation between device types. The test bench must perform comparably under proton, heavy-ion, laser and total ionizing dose tests.

- Proton testing evaluates SEE-induced parametric variations such as transients, SEFIs, and accessible device power-states.
- Heavy-ion testing determines effects of different levels of Linear Energy Transfer (LET) on the device. Because the process technology is mixed architecture and is smaller than 180 nm CMOS it may be susceptible to destructive SEE in its embedded sensors.
- Laser testing exposes a specific area of the chip to laser pulses and the focused light (about 1 micron in diameter) moves across the surface in a controlled pattern.
- Total ionizing dose (TID) testing is performed in an accelerated environment and characterizes the long-term radiation effects on the device and determines whether dose-rate sensitivity exists.

Acronyms

AMD	Advanced Micro Devices, Inc.	GSFC	Goddard Space Flight Center
BGA	Ball Grid Array	IP	Internet Protocol
CMOS	Complementary Metal–Oxide–Semiconductor	JPL	Jet Propulsion Lab
COTS	Consumer Off The Shelf	KVM	Keyboard, Video & Mouse
CPU	Central Processing Unit	LPDDR#	Low Power Double Data Rate (memory)
DDR#	Double Data Rate (memory)	OOM	Out of Memory
DUT	Device Under Test	PCB	Printed Circuit Board
FinFET	Fin Field Effect Transistor	SEE	Single Event Effects
FTP	File Transfer Protocol	SEFI	Single Event Functional Interrupt
GB	Gigabyte	SOM	System on Module
GDDR#	Graphics Double Data Rate (memory)	TFLOPs	Tera-Floating Point Operations
GPU	Graphics Processing Unit	TSMC	Taiwan Semiconductor Manuf. Company
		YOLO	You Only Look Once

Device Preparation

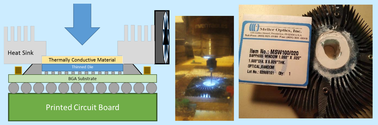


Figure 2: Obverse Side Modifications
Sapphire window design (left); Milling COTs heatsink (center); Final product (right)

Figure 3: Obverse Side Modifications
Die thinning (left); New heatsinks for other components around central BGA (right)

Figure 4: Reverse Side Modifications
400W Cooling on Bare NVIDIA GTX 1050

Figure 5: Reverse Side Modifications
180W Cooling on Lidded AMD Ryzen CPU

To conduct heavy-ion and laser testing, a custom tooled cooling solution was created to permit access to the thinned die from the obverse side while absorbing the heat through the reverse side of the printed circuit board. This orientation permitted nominal operation from both the DUT GPU and a control GPU (with stock cooling solution) within the test bench. The cooling solution created for GPU testing is also a verified solution to test COTS CPU devices such as an AMD Ryzen™ microprocessor which contains a GPU. An alternative cooling method, from the obverse side of the PCB, can also be employed using a thin (20 mil) thermally-conductive sapphire window and heat sink combination.

The DUT preparation described allows an ideal situation to be developed for both soldered and socketed components. Additionally, it is radiation tolerant by design so that the system can be used in open air, in a vacuum or radiation chamber. A direct path for radiation is created through thinning and polishing of the die. The cooling solution allows the device to operate under load while maintaining a temperature appropriate for the test (i.e. 20°C). The die can be thermally imaged and superimposed onto an optical image of the active regions (mirrored in the case of a flip-chip device, of course) to provide a feature map. A laser test can correlate radiation response from a proton or heavy ion test to a very specific area on the die and be marked on the feature map.

Test Bench Configuration

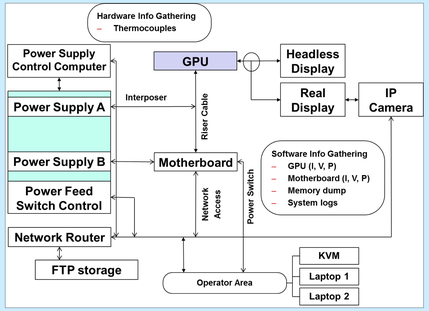


Figure 6: System-Level Organization

Radiation testing mirrors the logistics and electrical monitoring associated with reliability and qualification testing. Ideally, a component or component on a minimalist daughterboard should be used with a pin and socket interconnect to a carrier board. This is often the case with discrete components (e.g. diodes) when undergoing radiation tests. Practical repeatability is often overlooked in test creation due to resource constraints and haste. The monitoring should take place from the carrier board for consistency and mitigation against handling damage. Power supplies should be controlled and monitored with software so that timings or intervals between operational steps are consistent between each DUT and each test run. Electrical control using network-based software controlled COTS relays permits rapid creation of test benches without intensive development. These are a few broad examples of system-level control that facilitates autonomy.

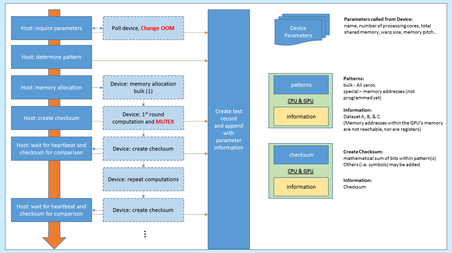


Figure 7: Test Execution and Logical Flow

A simple mutex is being used within the kernel to allocate a single integer worth of memory space, do nothing with it and then release the memory space. This keeps the process resident in the GPU pipeline for no less than 30 seconds per interval. The mutex can be adjusted to increase this by one to two orders of magnitude. The mutex memory allocation can only be performed in a serial fashion. Therefore, while all the computation calls can take place in parallel, each and every thread must wait in line to allocate the single integer space.

The operating system has watchdogs (e.g. out of memory (OOM) daemon) to ensure no task can take down the system.

- In Windows, the driver subsystem will crash and Windows will reset it thus ending the processing job.
- In Linux, it will identify the delayed process and terminate it.

Because we are using a mutex, the system thinks the GPU has halted. There are ways around this by adjusting the memory priority of applications and hardware drivers in the system.

Three types of payloads have been created for the GPU test bench: Neural Network, Math-Logic and Colors. The neural network is a convolutional neural network (CNN) which can avoid processor optimizations that recursive neural networks (RNN) primarily benefit from. Math-Logic uses mathematics and conditional logic statements to exercise memory hierarchy. The Colors payload assesses corruption in the output image presented to a display.

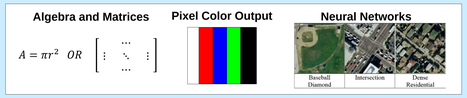


Figure 8: Software Payload Types

- Convolutional Neural Network (CNN) to identify land usage objects using a dataset modified from [4] for use with a "You Only Look Once" (YOLO) algorithm for object identification in still images and live stream video. The CNN was configured to be trainable on both GPU and CPU microprocessor types. Twenty one image categories were identified across the dataset. The figure above shows three such categories. The YOLO algorithm provides an accuracy rating and the most likely image classification.
- Mathematical and logical payloads such as pi, polynomial arithmetic, Markov permutations (folding algorithms) and algebra are leveraged to fill the computational and memory components of the device while preventing hardware optimizations to manipulate the software bit stream.
- Graphics, texture and color rendering tests have been developed. Graphics memory tends to be directional in that it behaves as read-only. The simplest test allows monitoring of this memory by triggering a pixel color change with automatic screen compare for pixel-change identification. The most complex of these tests performs a burn-in to the rendering logic of the device. Pixel corruption or display artifacts are monitored and recorded.

Discussion & Conclusions

Test portability plays a major role in standardizing a test. It isn't beneficial to have an expansive lab setup that cannot be affordably and easily transported. Radiation testing often requires trips to other facilities. The hardware selected for the test benches are COTS computers that can run Windows and Linux. This permits a test bench computer to be procured near the test facility in case of a failure during freight shipping. The software is compiled and packaged with all its dependencies and licenses so there is nothing to install. The test bench software also includes the software necessary to produce and retain run logs with unique identifiers and template-based formatting of data (i.e. voltages, memory maps and bit streams).

The GPU test bench and its software payloads have been written with attention to open-source or public domain-sourced code snippets and hardware components such that these tests could be recreated by other engineers. This standardized approach to testing mitigates the hardware optimizations found in newer generation microprocessors whereas an apples to apples comparison would otherwise not be possible. This approach involves rapid development, quicker procurement using modular system and network components, using COTS, in house development using public domain material, and software that can be easily updated to accommodate new DUTs while maintaining the ability to test older DUTs. The goal of the test is not to confirm that a newer GPU is better than an older GPU (which optimization will most certainly do), but rather whether the fabrication technology itself is more susceptible to radiation effects.

Acknowledgment

The authors acknowledge the sponsor of this effort: NASA Electronic Parts and Packaging Program (NEPP). The authors thank members of NASA GSFC's Radiation Effects and Analysis Group (REAG) who contributed to the creation of the test bench: Stephen R. Cox, Noah Burton, Alyson D. Topper, Ray Ladbury and Martin Carts.

References

1. Edward Wyrwas, "Proton Testing of nVidia GTX 1050 GPU," https://nepp.nasa.gov/files/28629/NEPP-TR-2017-Wyrwas-17-039-GTX1050-2017Apr-TN45745.pdf
2. NASA/GSFC Radiation Effects and Analysis home page, http://radhome.gsfc.nasa.gov
3. NASA Electronic Parts and Packaging Program home page, http://nepp.nasa.gov
4. Yi Yang and Shawn Newsam., "Bag-Of-Visual-Words and Spatial Extensions for Land-Use Classification," ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems (ACM GIS), 2010. The original satellite images were from the USGS National Map Urban Area Imagery collection for various urban areas in the USA. This material was based on work supported by the National Science Foundation under Grant No. 0917069
5. Joseph Redmon., "YOLO9000: Better, Faster, Stronger," arXiv preprint arXiv:1612.08242, 2016.
6. Interactions of Ions with Matter website, http://www.srim.org/

Presented by Edward Wyrwas at the 2018 Institute of Electrical and Electronics Engineers (IEEE) Nuclear and Space Radiation Effects Conference (NSREC), Kona, Hawaii, July 16-20, 2018