NEPP Processor Enclave: Testing Artificial Intelligence & Machine Learning

Edward J. Wyrwas
edward.j.wyrwas@nasa.gov
301-286-5213
SSAI, Inc / NASA GSFC NEPP

This work was sponsored by:
NASA Electronic Parts and Packaging (NEPP) Program

NEPP Team Members

Ed Wyrwas edward.j.wyrwas@nasa.gov
Carl Szabo carl.m.szabo@nasa.gov
Scott Stansberry scott.d.stansberry@nasa.gov
Alyson Topper alyson.d.topper@nasa.gov

SSAI, Inc / NASA GSFC NEPP

Acronyms

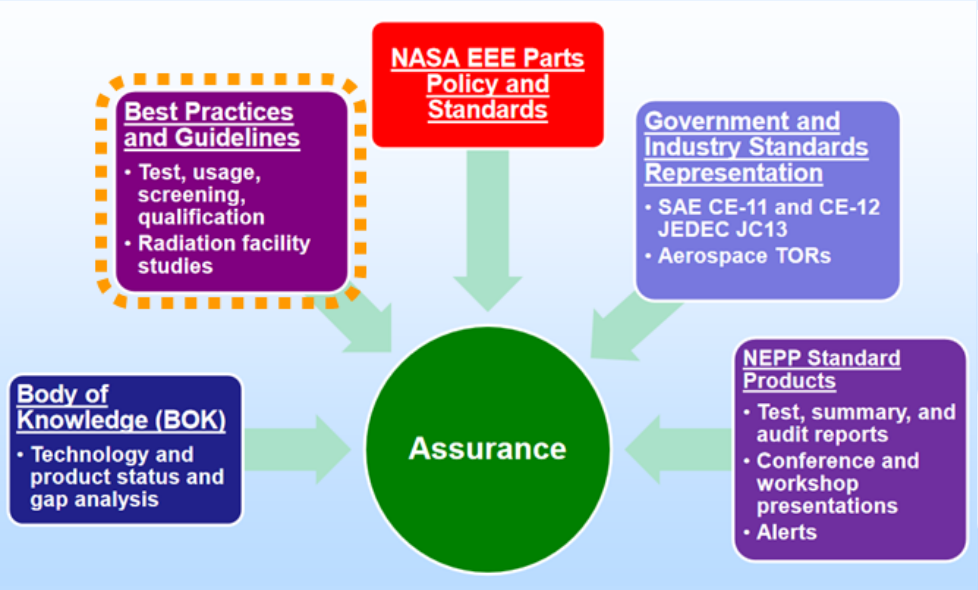
- Advanced Micro Devices (AMD)
- Application Specific Integrated Circuit (ASIC)
- Automotive, Railway and Avionic Multicore System (ARAMiS)
- Artificial Intelligence (AI)
- Advanced RISC Machine (ARM)
- Automated Test Equipment (ATE)
- Board Level Adapter Plate (BLAP)
- Body of Knowledge (BOK)
- Complementary metal oxide semiconductor (CMOS)
- Convection, Rotation and planetary Transits (COROT)
- Commercial off-the-shelf (COTS)
- Compute Unified Device Architecture (CUDA)
- Dual Data Rate (DDR)
- Device under test (DUT)
- Electrical, Electronic and Electromechanical (EEE)
- Field programmable gate array (FPGA)
- Fin Field-effect transistor (FinFET)
- General Purpose Input / Output (GPIO)
- Graphics Processing Unit (GPU)
- High Performance Computing (HPC)
- Input / Output (IO)
- Intellectual Property (IP)
- Linear energy transfer (LET)
- Key Process Indicators (KPI)
- LINear equations software PACKage (LinPack)
- Machine Learning (ML)
- Mean time to failure (MTTF)
- Multi-Bit Upset (MBU)
- National Aeronautic and Space Administration (NASA)
- NASA Electronic Parts and Packaging (NEPP)
- Numerical Python (NumPy)
- Open Compute Language (OpenCL)
- Open Graphics Language (OpenGL)
- Operating System (OS)
- Printed Circuit Board Assembly (PCBA)
- Procurement Request (PR)
- RApid Machine-lEarned Triage (RAMjET)
- Single-Bit Upset (SBU)
- Scientific Python (SciPy)
- Single-Event Effect (SEE)
- Single-Event Functional Interrupt (SEFI)
- Single-Event Upset (SEU)
- Single-Event Upset Cross-Section (σ_{SEU})
- Simultaneous Localization And Mapping (SLAM)
- System on Chip (SOC)
- System on Module (SOM)
- Transiting Exoplanet Survey Satellite (TESS)
- Technical Operation Report (TOR)

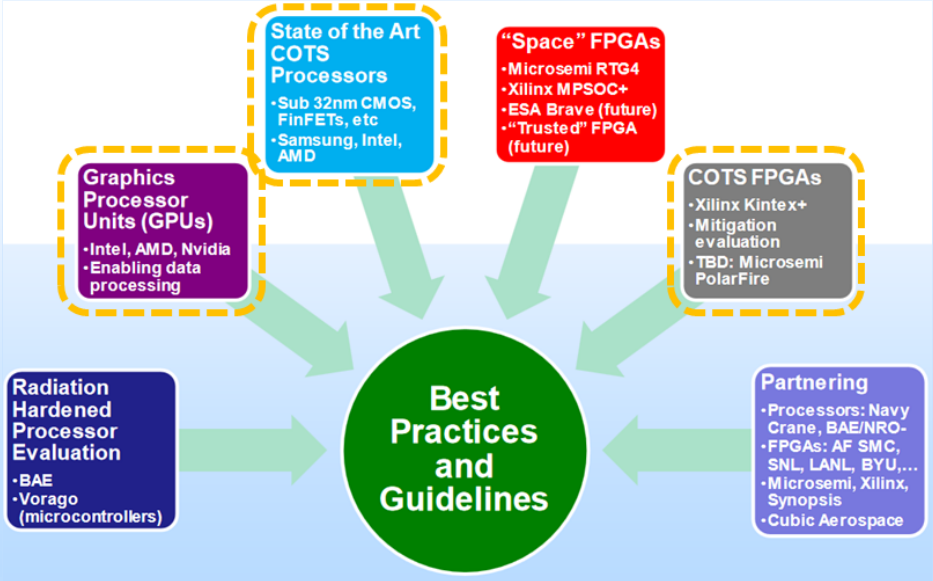
Agenda

- Processor Enclave
- COTS Approach
  - System View
  - Evaluation and Report
  - Test Suite
  - Thermal Control
- Roadmap

NEPP Mission Statement

Provide NASA's leadership for developing and maintaining guidance for the screening, qualification, test, and reliable use of Electrical, Electronic, and Electromechanical (EEE) parts by NASA, in collaboration with other government agencies and industry.



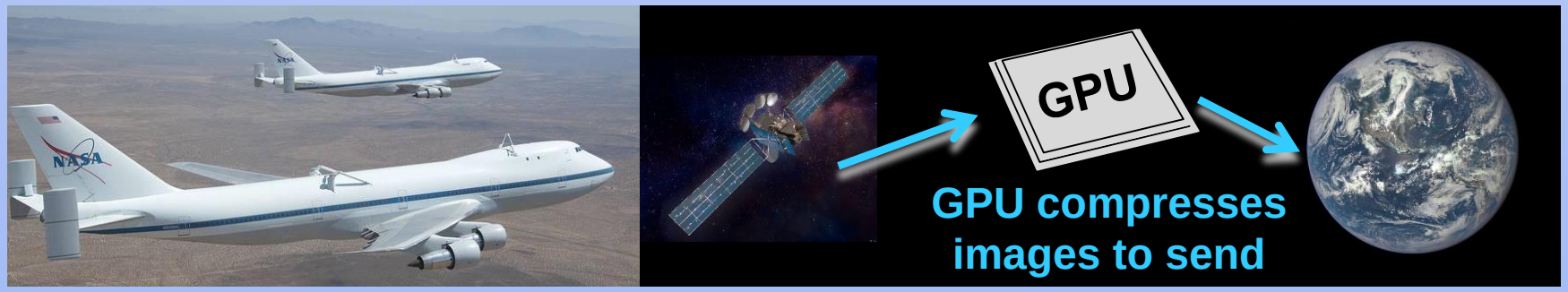


Processor Enclave

- Computational device families are converging and multiple EEE components are required for a complete subsystem
  - Need other components to test the GPU and its memory
- Using high-level languages, applications are accelerated by running the sequential part of their workload on the CPU and accelerating parallel processing on embedded "soft" IP and coprocessor devices (e.g. GPU, Neural Processor)
- We may still use some legacy parts with well-known reliability and radiation tolerances, but we also test leading edge computational components
  - Microprocessors (e.g., x86, x64, ARM, Power Arch.)
  - GPUs (e.g., NVIDIA, AMD, Qualcomm)
  - Memories (e.g., 3D Xpoint, PCM, DDR3, DDR4, Flash)

Aerospace Applications

- In the ARAMiS project (2013), leading industry companies and research institutions demonstrated the usability of multicore technology in safety-critical applications based on prototypical implementations.
  - https://www.offis.de/en/offis/project/aramis.html
- COROT2: first GPU-powered satellite
  - https://www.lume.ufrgs.br/bitstream/handle/10183/159210/001023249.pdf
- We are currently using this technology for:
  - Simultaneously Localization and Mapping (SLAM) for autonomous maneuvers
  - Data compression and encryption
  - Remote Sensing and Classification (AI & Machine Learning)



What Should We Do About COTS?

- For high performance computing (HPC) devices and their cohort (the Processor Enclave), we tend to base our testing on the two extremes:

Test Structures

If we have control over the design and fabrication, then those tools enable creation of automated test equipment (ATE) to perform a high speed test. Testing is therefore *easy*.

Black Box ASIC

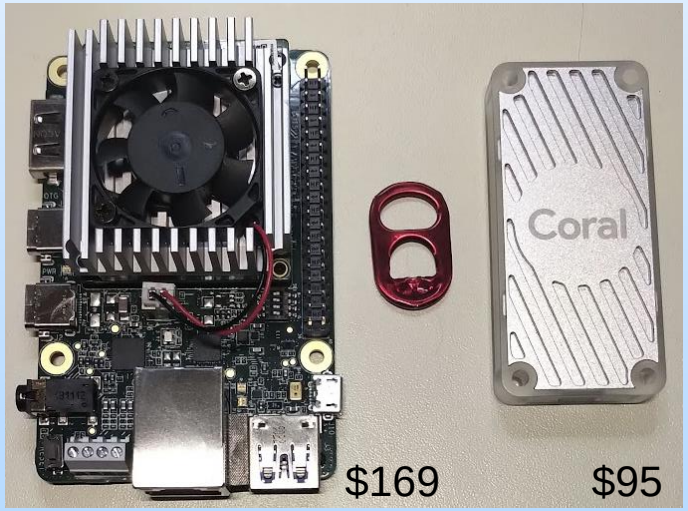
If we only have a COTS device, we are limited in how we can exercise/stress, monitor and detect. The whole PCBA is tested as a single unit, but some things are generalized.

Evaluating COTS

- What can we do with COTS in the gray area in between?

Test Structures		Black Box ASIC

- Typical COTS Development kit
  - Power Adapter (maybe)
  - Carrier Board with peripherals & IO
  - System on Module
  - Vendor (A) software package
  - Vendor (B) operating system
  - Limited schematic, manuals, etc.

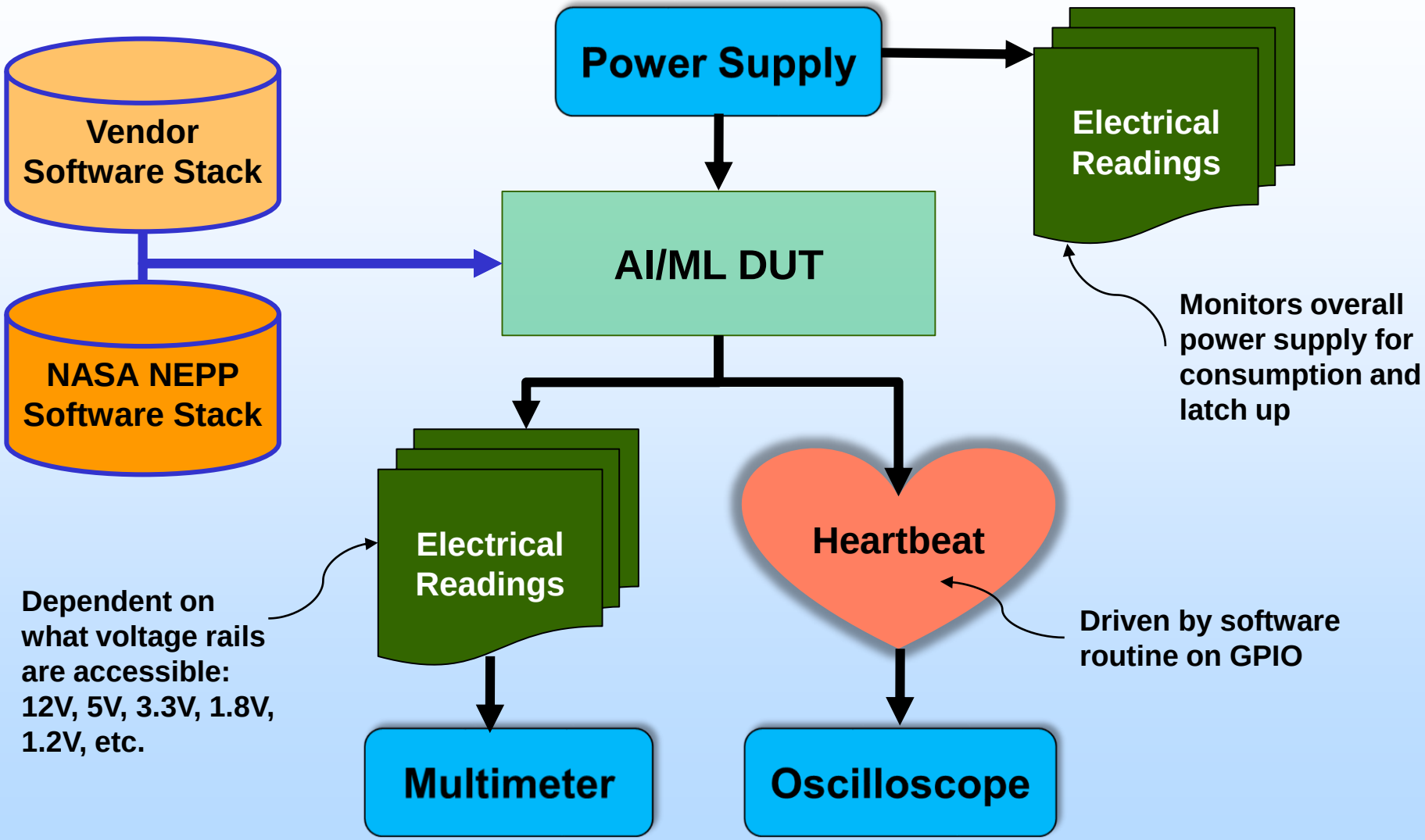


What's the Approach?

- Use the COTS kit, < $1000 each
- Use legacy capabilities for
  - Power injection at supply, rail, load
  - Power monitoring at supply, rail, load
  - Temperature monitoring
- Use new capabilities for
  - Temperature control
  - "Computer hardware" performance monitors
- Use the same build heritage of Linux (Ubuntu 18.04) on all kits (where possible)
- Build-up and compile software applications to run on <u>all</u> kits. Use the same code base and community supported libraries. (i.e. Tensor Flow 2.0)

And Then Let's Standardize This Process

DUT Hardware Monitoring



Timeline to Identify What's Feasible

	Task Item	Level of effort	Timeline
1	Identify code to drive a GPIO heartbeat & Document this step	4 hours	1 business day
2	Prove out heartbeat monitoring with instrumentation & Document this step	4 – 8 hours	1 business day
3	Identify the main power supply rail & Document this step	1 hour	
4	Identify additional power supply rails & Document this step	Up to 8 hours	1 business day
5	Identify Point of Load power rails & Document this step	Up to 24 hours	3 business days
6	Prove out electrical monitoring with instrumentation & Document this step	4 – 12 hours	2 business days
7	Finalize sections within **Work Instruction** document	4 – 12 hours	2 business days

A difficult COTS kit should take at most 2 weeks to **evaluate testability** and identify the level of rework associated with "wiring it up"

Interested Party Survey

- Informal questions to 100s of technical folk and managers in both industry and government to evaluate this landscape
  - Majority takeaway was a concerning lack of reference documentation about how to conduct one's own test

- Other noteworthy responses
  - There is not a consistent source for troubleshooting information
  - There is not a repository of common codes to evaluate functionality
  - There are plenty of performance benchmarking tools for individual hardware

Two-fold Solution

- **Maintain Reference Material** (Teams Channel open to NASA users)
  - Wiki to assure repeatability by a variety of electrical engineering skill sets, other NASA centers and by folks less familiar with the topic
  - Work instructions written to put bounds on the level of anticipated work commitment for a flight project to use similar hardware, and to ascertain a metric to compare the varying levels of engagement required for different levels of circuit card complexity
  - Guide to reference when at a beam line or when conducting troubleshooting
- **Conduct Biweekly Telecons to Stay Relevant** (open to NASA partners)
  - Discuss current technology landscape, upcoming devices and tech
  - Discuss any pain points anyone is trying to resolve
  - Inquire about information relevant to the processor enclave
  - Coordinate efforts to evaluate specific devices or tech

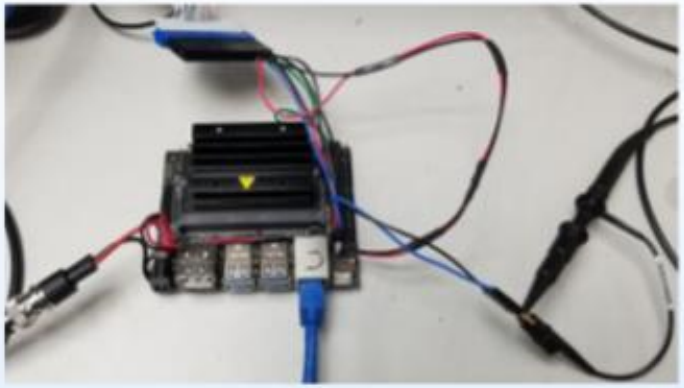
Work Instructions - Guides

- How-to and example code to generate heartbeat from board
  - What pins are being connected to?
  - What ancillary EEE components are necessary to send/capture the heartbeat?
  - What instrumentation is required to monitor the heartbeat?
  - What secondary equipment may be available to do the same thing, but as a fail-safe? (when applicable)

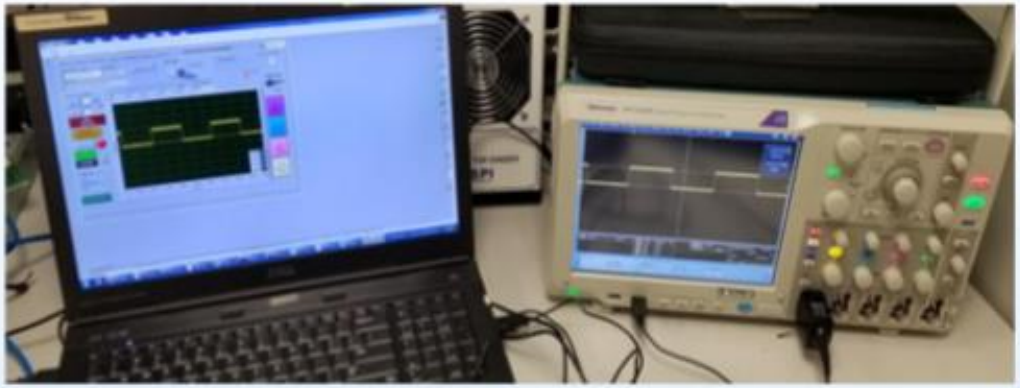
- Hierarchy of permissible resolution of voltage rails
  - Can we monitor the rails in addition to the supply?
  - Can we monitor using a current sense resistor?
  - What physical limitations (if any) exist when trying to connect to these points of interest?
  - What teardown steps (if any) were necessary to investigate the points of interest?

Work Instructions - Images

- Overview of PCBA & Inset showing areas of interest
  - Places for electrical monitoring & GPIO pins for heartbeat
- Screenshot of schematic for areas of interest
- Annotations showing an uninformed end-user the significance of x, y, z
- Photo or screenshot of each piece of the instrumentation
- Photo or screenshot of the LabVIEW interface (for each piece of instrumentation) showing how we intend to use the instrumentation
- Photo of any ancillary wiring, breakout boards, etc.



J41 GPIO breakout header to the right with scope probe



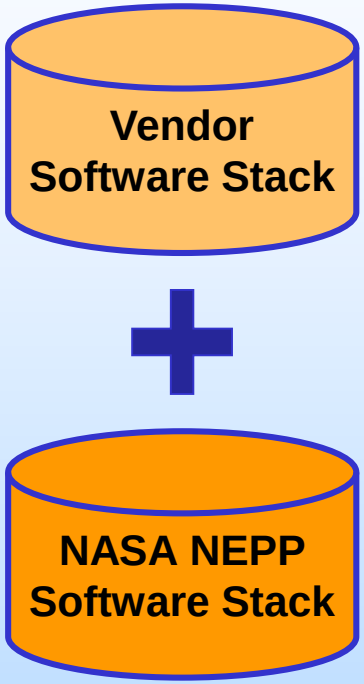
Digital Oscilloscope remotely monitored

Application Focused Payloads

- Tangential to "Test As You Fly"
  - Bit streams: sensors, camera, offline video feed
  - Math: linear algebra, matrices, encryption
  - AI/ML: signals, images, control
  - Display: colors, texture, ray tracing
  - Electrical and thermal monitoring
  - Thermal Control

- Interdependencies of the Processor Enclave
  - Part isolation: Memory, Processor, Coprocessor (i.e. GPU)
  - Software applications need a complete system, operating system and software stack

Artificial Intelligence & Machine Learning Test Vectors



- Software stack from vendor is comprised of Linux, PCBA carrier board drivers and hardware initializers, and some health prognostic tools
- Software stack from NASA NEPP is comprised of
  - 2x Mathematics tests (CUDA, OpenCL)
  - 1x Graphics buffer test (OpenGL)
  - 1x Linear Algebra test (Linpack)
  - 4x Artificial Intelligence tests (Python including SciPy and NumPy, Keras and TensorFlow 2.0)
- **Android OS applications for Math and Graphics are in development too. The TensorFlow Lite one is complete.**

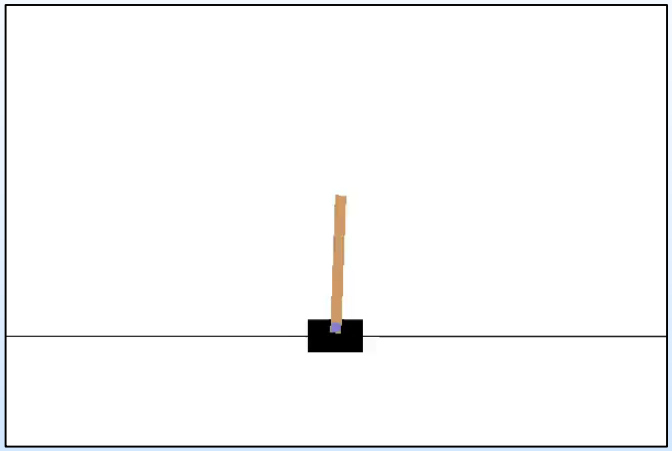
Artificial Intelligence & Machine Learning Test Vectors

- Mechanical Control Simulation
  - OpenAI Gym CartPole (PID test)
- Diverse Categorical Classification
  - Landsat object identification (22 categories)
- Binary Classification
  - Sky images (moon vs sun)
- Signals Processing and Classification
  - TESS Exoplanet Survey data

PID & Control Systems

OpenAI Gym CartPole

- A pole is attached by an un-actuated joint to a cart, which moves along a frictionless track. The system is controlled by applying a force of +1 or -1 to the cart. The pendulum starts upright, and the goal is to prevent it from falling over. A reward of +1 is provided for every timestep that the pole remains upright.
- Similar control system to reaction wheels (fly wheels), propulsion and electromagnetic adjusters on a spacecraft



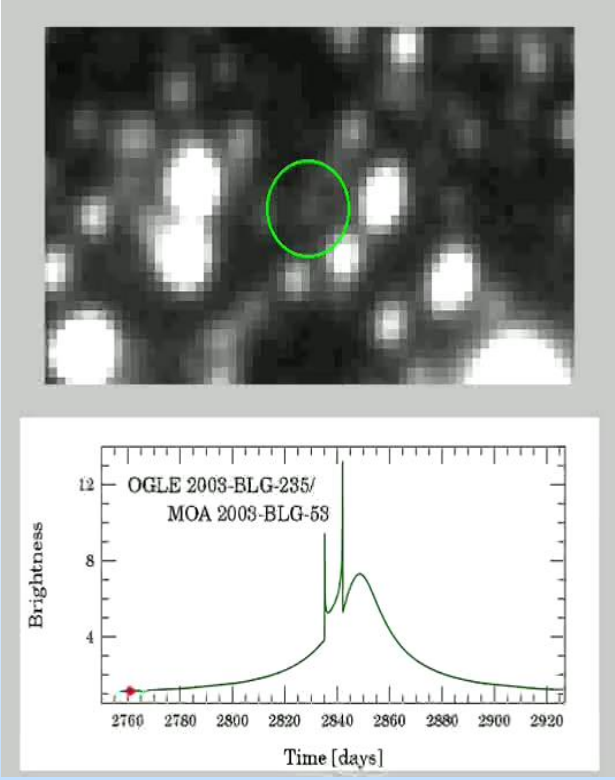
Using AI for Classification

- Training and deployment of two Keras-based neural networks which classify images
- Binary classification using Flickr.com images
  - categories are: sun or moon (and edge cases)
- Diverse classification using Landsat images
  - categories include: agricultural, airplane, baseball diamond, beach, buildings, chaparral (vegetative desert), dense residential, forest, freeway, golf course, harbor, intersection, medium residential, mobile homes, overpass, parking lot, river, runway, sparse residential, storage tanks and tennis court

Baseball Diamond	Intersection	Dense Residential

TESS Exoplanet Survey data

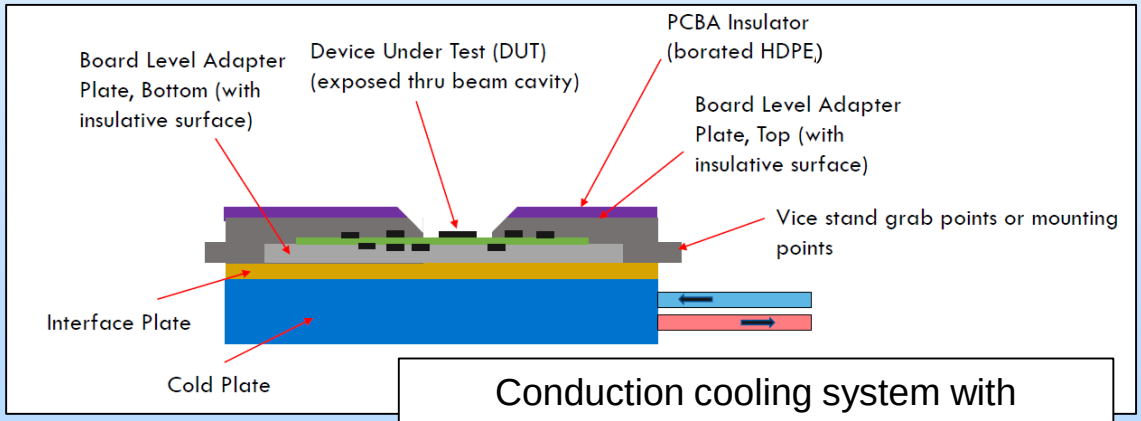
- Dr. Richard Barry is evaluating whether low power GPUs can be used to perform an automated assessment on gravitational microlensing by extrasolar planets
  - RApid Machine-lEarned Triage (RAMjET) – An artificial intelligence running on NASA's 7000 Pflop climate supercomputer that detects new planets and black holes.
  - https://richardkbarry.com
- RAMjET code has been deployed on low power AMD and NVIDIA hardware
  - A separate experimental GPU branch is supported on GitHub with the installation scripts for the various dependencies and to bring up TensorFlow 2.0 on ARMv8 hardware.
  - https://github.com/golmschenk/ramjet



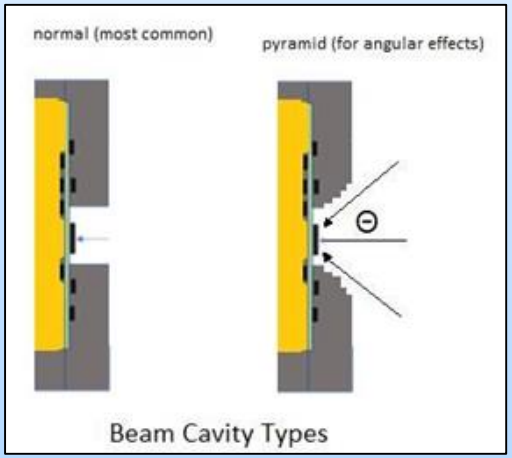
https://richardkbarry.com/?page_id=93

System Level Constraints

- To assure each DUT within the test campaign experiences the same stress profile, NASA NEPP has designed a pseudo-COTS cooling plate to achieve this
  - Guarantees we can record system-level temperatures at the DUT and at various points of interest on the PCBA
  - Permits the test operator to select the idle or loading temperature of the DUT
- A Board Level Adapter Plate (BLAP) was fabricated for the AMD e9173 PCIE GPU and NVIDIA Jetson TX2 SOC in FY20.
- Proposed work for FY21 includes a second cooling rig, one BLAP for a collection of Neural Network devices and at least two BLAP for noteworthy GPU & SOC



Conduction cooling system with swappable DUT-specific adapter plates



Cold Plate Adapter

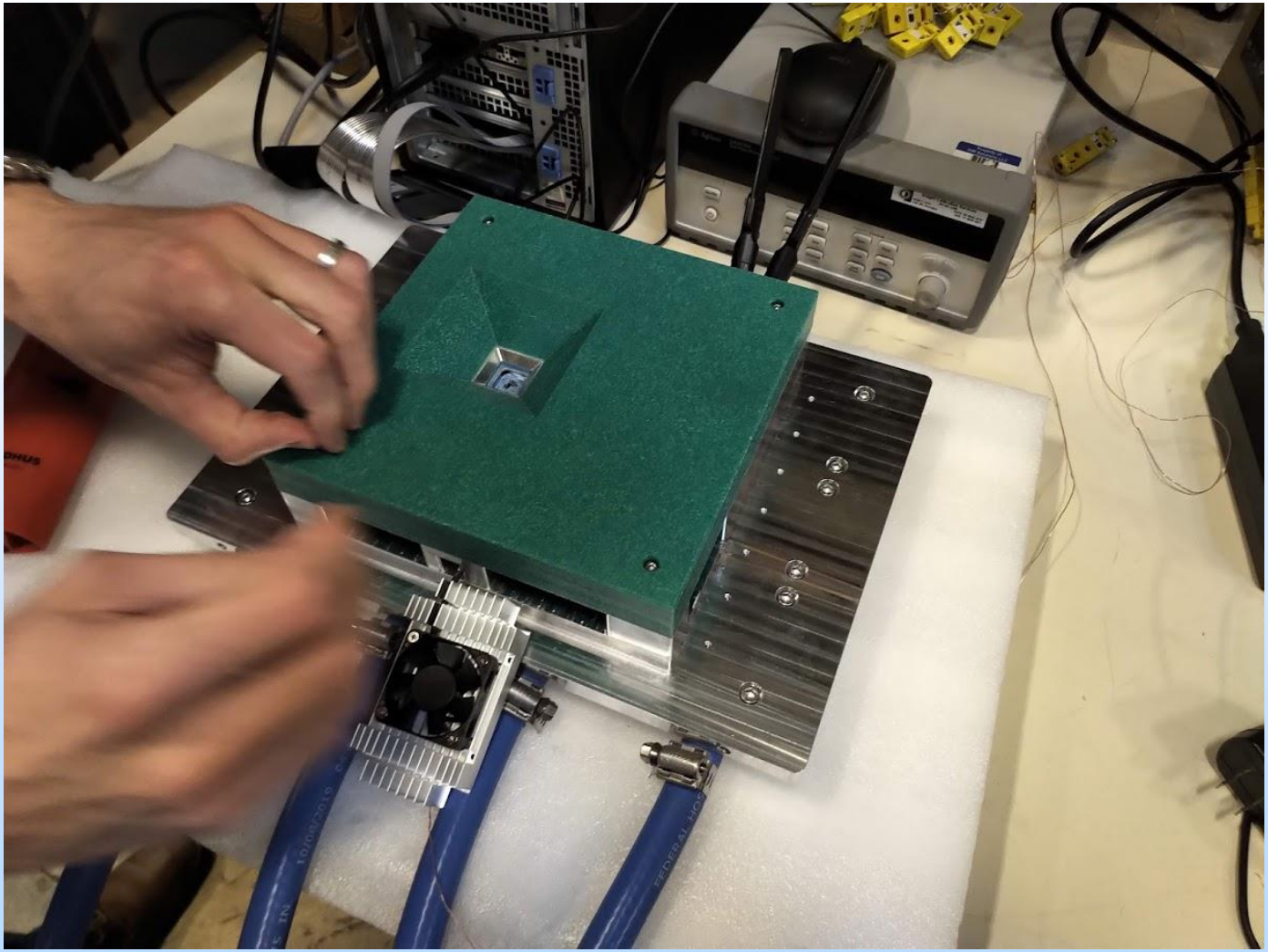


Image of NVIDIA Jetson TX2 embedded in adapter plates and set up on cold plate

Evaluation Timeline

*Radiation Testing (RT)
*Cold Block (CB) Adapter Development
*Test Feasibility Evaluation (FE)
*Continuation or
*Termination or

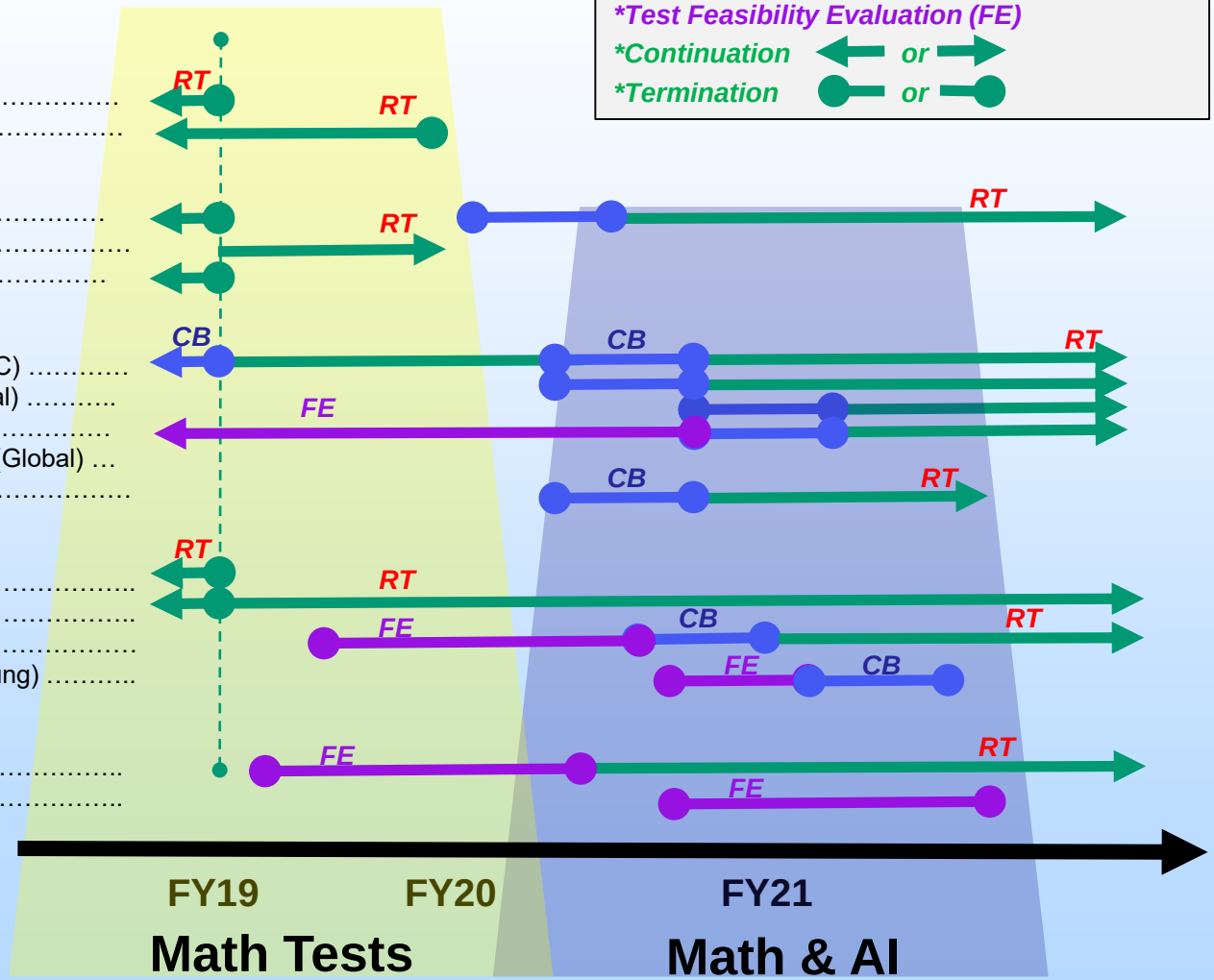
Microprocessors
- ≤14nm++ Intel (Intel) ……
- ≤10nm AMD (Global) ……

Microprocessors with embedded GPUs
- 14nm++ Intel (Intel) ……
- ≤ 10nm AMD (Global) ……
- 10nm Qualcomm 835 (Samsung) ……

GPUs
- 14/16nm NVIDIA GTX 1050/1080 (TSMC) ……
- 14nm AMD Radeon RX580/Vega (Global) ……..
- 14nm Intel Discrete GPU (Intel) ……
- 14nm Low Power, AMD Radeon e917x (Global) …
- 12nm NVIDIA RTX 2080 (TSMC) ……

System on Chip
- 20nm NVIDIA Tegra X1 (TSMC) ……..
- 16nm NVIDIA Tegra X2 (TSMC) ……..
- 12nm NVIDIA Tegra Xavier (TSMC) ……
- ≤ 10nm Qualcomm Snapdragon (Samsung) ……..

Neural Network Devices
- Google TPU……..……..
- Many, many others……..……..



FY20-21 Hardware Selection

- FPGA Co-Processor Devices
  - Xilinx Zynq SoC
  - Xilinx MPSOC
  - Xilinx Versal
- Microprocessors
  - Intel Core i7 6500U
  - AMD Ryzen 3 v1202B
  - AMD Ryzen Low Power R1102G, R1305G
- GPU SOCs
  - NVIDIA Jetson TX2, TX2i, Nano
  - NVIDIA Xavier AGX, NX
- SOCs
  - NXP i.MX 8QuadMax
  - Edgeless EaI Series
  - Allwinner F1C200s
  - Samsung Exynos
  - Qualcomm Snapdragon Snapdragon 765G, 865+
- Neural Network Devices
  - Google TPU
  - Intel (Movidius/Nervana/Loihi)
  - RockChip Rk1808, Rk3399Pro
  - Nepes neuralStic_2304
  - LiteMicro Ethos-U55
  - Sipeed Maix
  - Amlogic S905X3
  - Gyrfalcon lightspeeur SPR2801 AI
  - Xilinx Alveo
  - Texas Instruments Sitara AM57x

- Spreadsheet of viable DUTs exists
  - Several vendors have been contacted for more information on specifications and availability
  - Some NDAs are being signed

FY20-21: Processor Enclave Testing

Description:

- This is a task over all device topologies and processes
- The intent is to determine inherent radiation tolerance and sensitivities
- Identify challenges for future radiation hardening efforts
- Investigate new failure modes and effects
- Testing includes total dose, single event and reliability. Test vehicles will include CPU, SOC and GPU devices from NVIDIA and other vendors as available
  - Compare to previous generations
  - Investigate failure modes/compensation for increased power consumption

FY20-21 Plans:

- Finish development of universal test suite which includes math (FFT, LinPack, Pi), output buffer (colors), memory hierarchy and neural networks
- Probable test structures for SEE and TID are COTS from:
  - NVIDIA (<16nm)
  - AMD (<14nm)
  - Intel (~10nm)
  - Samsung (<14nm)
  - Others (<45nm)
- Tests: characterization pre, during and post-rad

Schedule:

Microelectronics T&E	FY 21											
	S	O	N	D	J	F	M	A	M	J	J	A
On-going discussions for test samples												
Design & Fabrication of Cold Plate Parts												
SEE and TID testing												
Analysis and Comparison			◇			◇			◇			◇

Lead Center/PI: GSFC/SSAI/Wyrwas
Co-Is: Steve Guertin, Andrew Daniel, Carl Szabo, Scott Stansberry

Deliverables:

- Test reports and quarterly reports
- Expected submissions for publications
- Cold plate and board level adapters

NASA and Non-NASA Organizations/Procurements:

- Source procurements:
  - Proton (MGH, Provision)
  - Heavy Ions (LBNL, TAMU)
  - TID (GSFC)
  - Laser (NRL)

FY20-21: AI & Machine Learning Devices

Description:

- This is a task over all device topologies and processes
- The intent is to determine inherent radiation tolerance and sensitivities and provide a baseline of current market offerings
- Identify challenges for future radiation hardening efforts
- Investigate new failure modes and effects
- Testing includes total dose, single event and reliability. Test vehicles will include 'simple interface' devices with USB or similar connectivity.
  - Identifies viable suppliers and provide insight into product obsolescence roadmaps
  - Investigate failure modes and mitigation strategies for increased power consumption and latch up events

FY20-21 Plans:

- Deploy AI and Machine Learning test suite developed in FY20, making refinements where necessary
- Probable test structures for SEE:
  - 45nm down to 7nm
- Tests:
  - Electrical characterization pre, during and post-rad
  - Functional testing with AI/ML accuracy and other KPI

Schedule:

Microelectronics T&E	FY 21											
	S	O	N	D	J	F	M	A	M	J	J	A
On-going discussions for test samples												
Science community collaborations												
SEE and TID testing												
Analysis and Comparison			◇			◇			◇			◇

Lead Center/PI: GSFC/SSAI/Wyrwas
Co-Is: Alyson Topper

Deliverables:

- Test reports and quarterly reports
- Expected submissions for publications
- Additional cold plate and board level adapters

NASA and Non-NASA Organizations/Procurements:

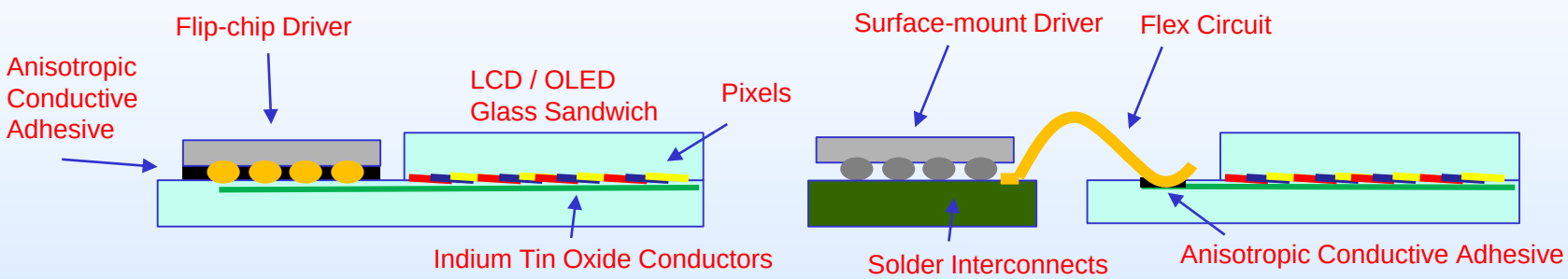
- Source procurements:
  - Proton (MGH, Provision)
  - Heavy Ions (LBNL, TAMU)
  - TID (GSFC)
  - Laser (NRL)

FY20-21: Chip-On-Glass Display Controllers

- The new suit that will be worn on Artemis missions is called the Exploration Extravehicular Mobility Unit, or xEMU for short.
- Not only will there be display drivers within the helmet, but touchscreen displays throughout the vehicles.

https://www.nasa.gov/feature/a-next-generation-spacesuit-for-the-artemis-generation-of-astronauts

FY20-21: Chip-On-Glass Display Controllers



– While a GPU or other display driver is necessary to render an image or text on a display, there is still sensitive circuitry upstream before the display's pixels are reached.

FY20-21: Chip-On-Glass Display Controllers

Description:

- This is a task over all device topologies and processes
- The intent is to determine inherent radiation tolerance and sensitivities and to Identify challenges for future radiation hardening efforts
- Investigate new failure modes and effects
- Testing includes total dose, single event and reliability.
  Test vehicles will include LCD and OLED display screen families smaller than 16” diagonal that meet JSC requirements

FY20-21 Plans:

- Test pixel sensitivity of transistor drivers and on-glass display controller
- Probable test structures for SEE:
  - < 350nm
- Tests:
  - Irradiation of small quadrant of pixels and COG
  - Characterization pre, during and post-rad
  - Functional testing with colors and refresh rates

Schedule:

Microelectronics T&E	FY 21											
	S	O	N	D	J	F	M	A	M	J	J	A
On-going discussions for test samples												
Ongoing disc. w/ xEMU/Artemis/Gateway												
SEE and TID testing												
Analysis and Comparison			◇			◇			◇			◇

Lead Center/PI: GSFC/SSAI/Wyrwas

Deliverables:

- Test reports and quarterly reports
- Expected submissions for publications
- Additional cold plate and board level adapters

NASA and Non-NASA Organizations/Procurements:

- Source procurements:
  - Proton (MGH, Provision)
  - Heavy Ions (LBNL, TAMU)
  - TID (GSFC)
  - Laser (NRL)

Summary of Deliverables

- Estimation of effort
  - Resource requirements to go from PR to running apps
- Work instructions for each DUT
  - Instrumentation reference
  - Challenges and noteworthy remarks
- Test Plan(s)
  - Software package description
  - Electrical monitoring requirements
  - Thermal constraints and solutions
- Test Reports
  - Discussion and Findings
  - Beam line results
  - Guidance for Usage

Acknowledgements

- *This work has been sponsored by*

  NASA Electronic Parts and Packaging (NEPP) Program

  NASA Office of Safety & Mission Assurance

- *Thanks is given to the NASA Goddard Space Flight Center's Radiation Effects and Analysis Group (REAG) for their technical assistance and support.*

Teams Channel

- **“NEPP – Processor Enclave”**
- **NASA users can join with the team code: 1a2iu5y**
- **26 members and growing**
- **Contact Steve Guertin (JPL) for our biweekly call**