Data Management for High Speed, Distributed Data Acquisition

Jeff Maggio

Principal Investigator

SBIR Exchange Aug 2024

Outline

- Our Company and Capabilities
  - Team
- Our Current Product Line
  - Digitizers & Logic Modules
- Data Management Research
  - Performance networking
  - Data Storage
  - User Software
- Acknowledgements and Future Plans

SkuTek
Instrumentation
Where Science Meets Industry

SkuTek Instrumentation Makes… **Instrumentation!**

- The Team
  - Full time: 4 Research Engineers + 1 Engineering Associate
  - Part time: 2 Other Senior Engineers, 1 Manager, 1 EE consultant
  - Interns rotating in and out constantly

- Our Focus
  - Electronics & Data Acquisition (DAQ) for High Energy Physics, Astrophysics, and Nuclear Physics.

We Serve National and International Customers

University of Rochester

Los Alamos National Laboratory

Brookhaven National Laboratory

Houghton College

Fermilab

Augustana University

The University of Tennessee Knoxville

Albert Einstein Center for Fundamental Physics

NSCL National Superconducting Cyclotron Laboratory

SFU Simon Fraser University — Engaging the World

Brown

Universität Bern

上海交通大學 Shanghai Jiao Tong University

Wiener Plein & Baus, Corp.

Northrop Grumman

Raja Ramanna Centre for Advanced Technology

Université Laval

SkuTek Instrumentation
Where Science Meets Industry

Our Capabilities: full end-end DAQ expertise

- Electronic Design
- Firmware Development
- Digital Signal Processing
- Detector Assemblies
- **Data Management Systems**

SkuTek Instrumentation
Where Science Meets Industry

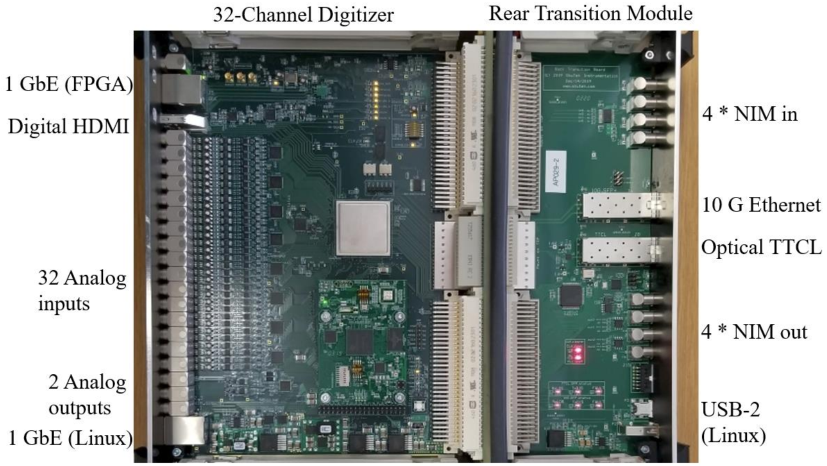
Historically We Specialize in High Performance Digitizers

Benchtop Digitizers

FemtoDAQ Kingfisher

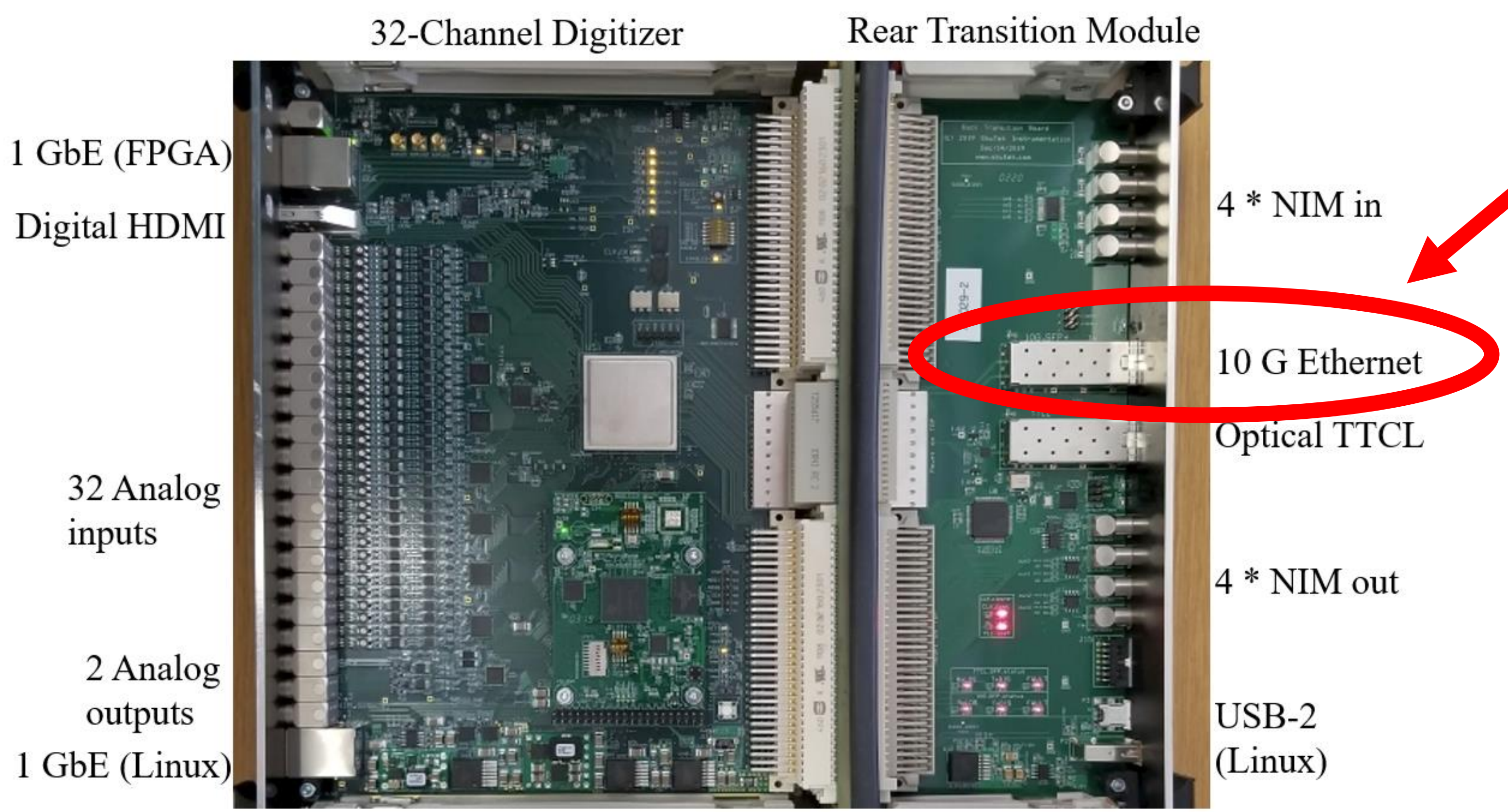
FemtoDAQ Vireo

Chickadee-32 Rackmount Digitizer



SkuTek Instrumentation
Where Science Meets Industry

Modern Digitizers Produce a Lot of Data



Chickadee-32 Digitizer
Top View

- 10 Gbps readout from our Chickadee-32 digitizer
  - 1.2 Gigabytes every second
- Imagine thousands of channels…
  - Data rates can be measured in hundreds of TB/hour

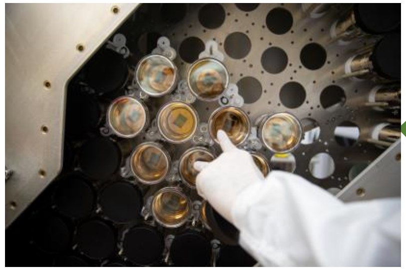
Scientific Data Demands are getting Bigger

1) The DOE's Energy Science Network (ESNet) estimates data rates and volumes will increase by several orders of magnitude this decade. [6]
   - Driven by new instrumentation, larger channel counts, AI & machine learning, etc

2) There is growing adoption of the "Distributed Computing Infrastructure" (DCI) data model.
   - Data processing occurs at High Performance Computing (HPC) centers, often geographically separated from the experimental facility.
   - Examples: Square Kilometer Array, Cherenkov Telescope Array, Linac Coherent Light Source (LCLS), Gamma Ray Energy Tracking Array (GRETA), LuxZeplin, etc. [2,3,4,6]

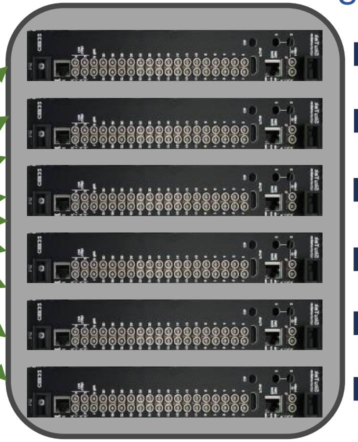
Takeaway: DAQ systems must account for this new paradigm.

There is a New Paradigm in Data Acquisition



Detector
(pictured: LuxZeplin)

Analog Signal

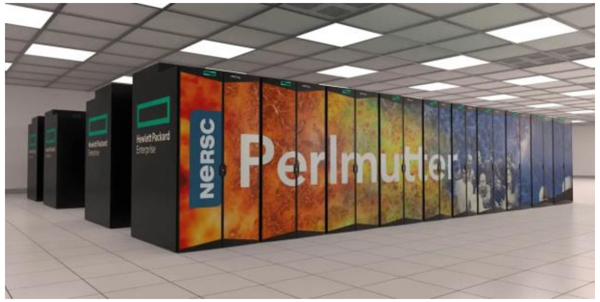


Data Acquisition

Digitized UDP Streams

Data Management

File Transfer



High Performance Computing (HPC) Facility
(pictured: NERSC)

Data Streaming → Stream Capture → Long Distance Transfer

SkuTek Instrumentation
Where Science Meets Industry

We are Developing 4 Data Management Products to Fit the New Paradigm

Solidago

DAQ Emulator / UDP Event Generator

Stream Concentration Solutions

Lossless Concentration of 100+ Gbps UDP streams

Not discussed in this talk
(work completed last year)

Liatris

80 Gbps Collector Node (Data Buffer / Recorder)

Salix

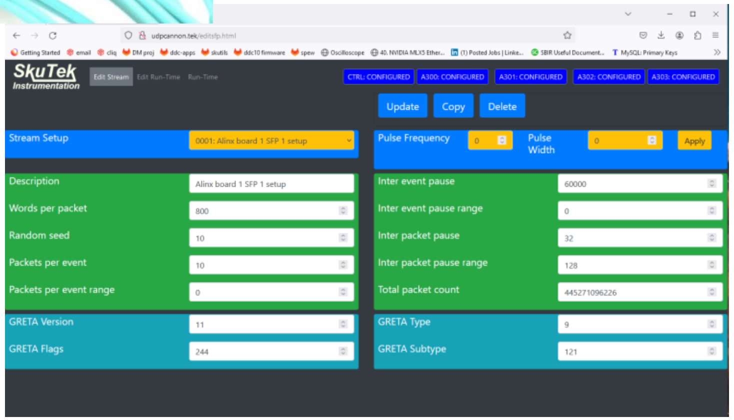
40 Gbps Data Transfer Node

SkuTek Instrumentation
Where Science Meets Industry

How do you Simulate the Data Environment of a Particle Accelerator?

Introducing the Solidago UDP Cannon!

- Emulates data streams from 16 digitizers (~512 channels)
- 0-160 Gbps programmable streaming rate (up to 20GB/s)
- Utilizes GRETA packet formats *(a standard in SkuTek and DOE DAQ systems)*
- Controllable via a web Interface and REST API



Solidago is currently for sale!

Solidago (Data Streaming) → **Liatris** (Stream Capture) → **Salix** (File Transfer)

Solidago Emulates Better than Software Equivalents

1) Hardware Solution means that no software tuning is needed

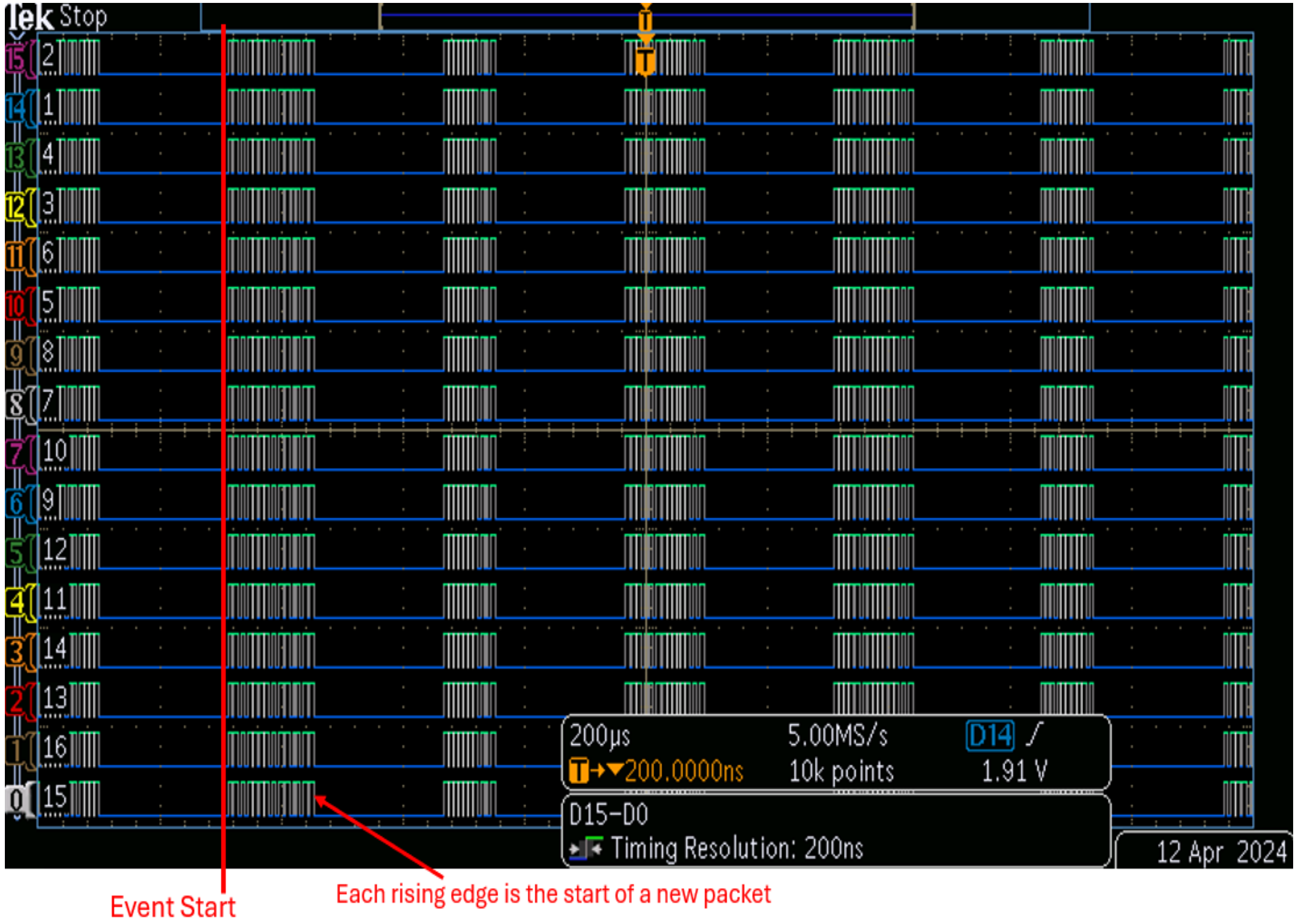
2) Streams can be synchronized with each other *(or run independently)*

3) The pattern of each stream is programmable and randomizable

The result:

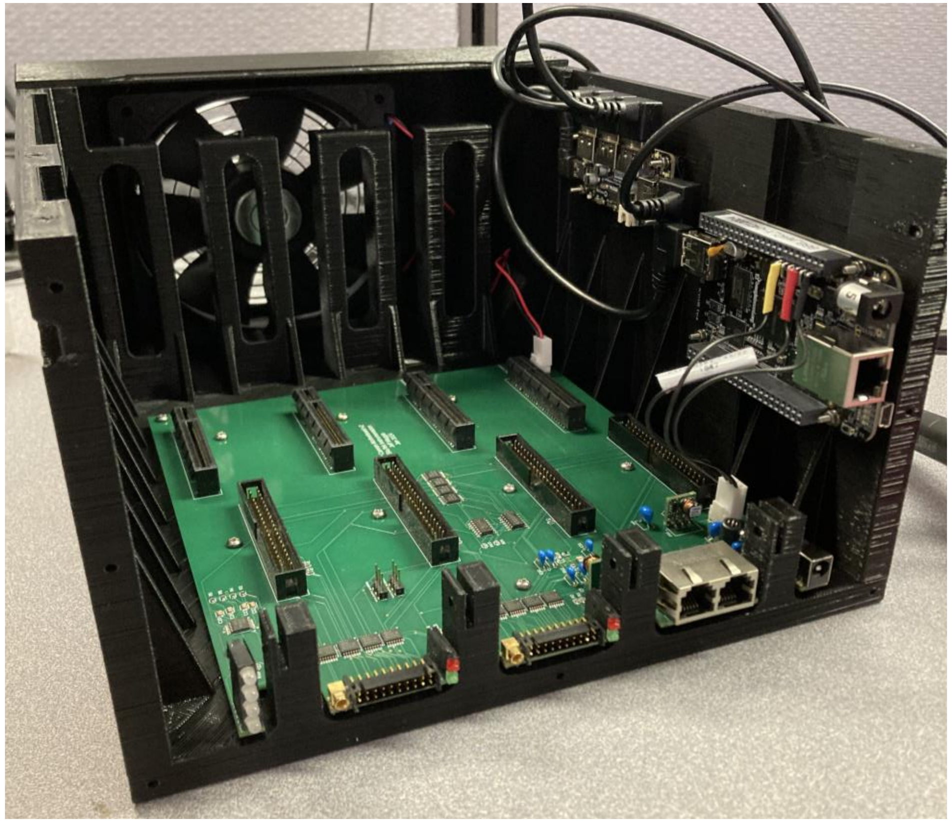
Solidago can mimic the event pattern you'd see in a pulsed particle accelerator

UDP Traffic can be precisely timed in realistic traffic "bursts"



SkuTek Instrumentation
Where Science Meets Industry

Solidago Internals are Compact



Solidago (Data Streaming) → **Liatris** (Stream Capture) → **Salix** (File Transfer)

Capturing High Speed Streams with Liatris

Liatris Collector Node

This is a Linux server which captures and records data streams locally

- Up to 80 Gbps sustained lossless UDP reception and recording
- Captures 256+ SkuTek Digitizer Channels

Liatris buffers data in files before movement to HPC facilities for processing

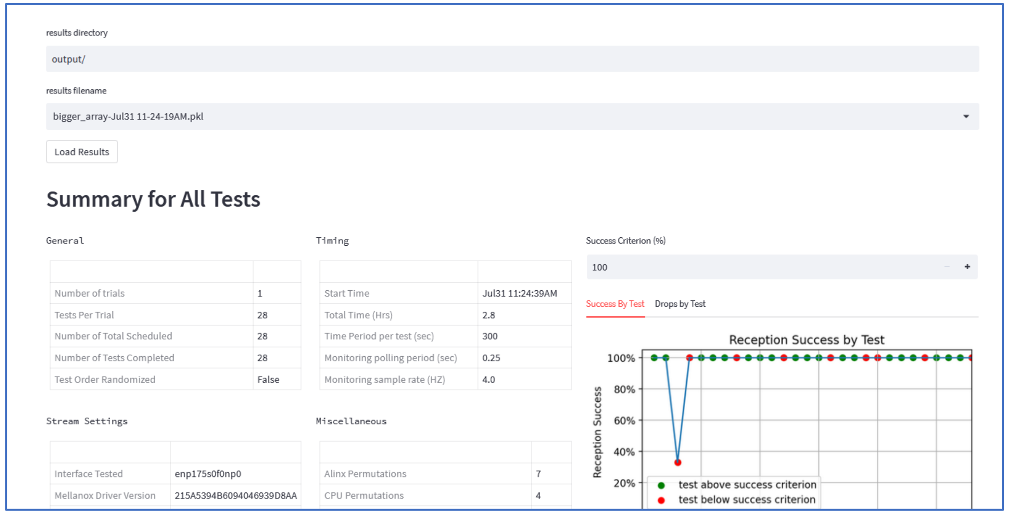
Programmable via a web GUI and REST api

Solidago (Data Streaming) → **Liatris** (Stream Capture) → **Salix** (File Transfer)

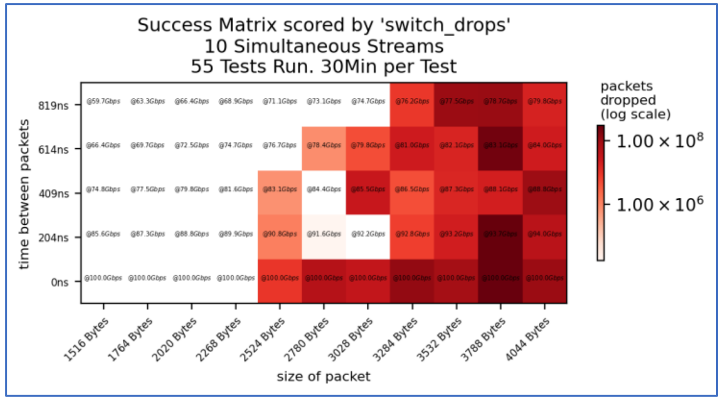
The Devil is in ~~the Details~~ *Linux Network Tuning*

100G capable hardware ≠ 100G capable computer.

- Performant networking requires “tuning”.
- We found 60+ parameters across kernel settings, hardware drivers, and network configuration that influence networking speeds.



Using Solidago, We developed software to automate testing different network configurations.



We achieved high speed performant networking in Linux.

0-80 Gbps : 100% success

80-90 Gbps : >99% success

100 Gbps : >90% success

Making Reliable And Scalable Stream Capture Software

Managing Software Development is a Difficult Balance

- Simple tasks can spiral into months of work
- **It's critical to reduce points of failure!**

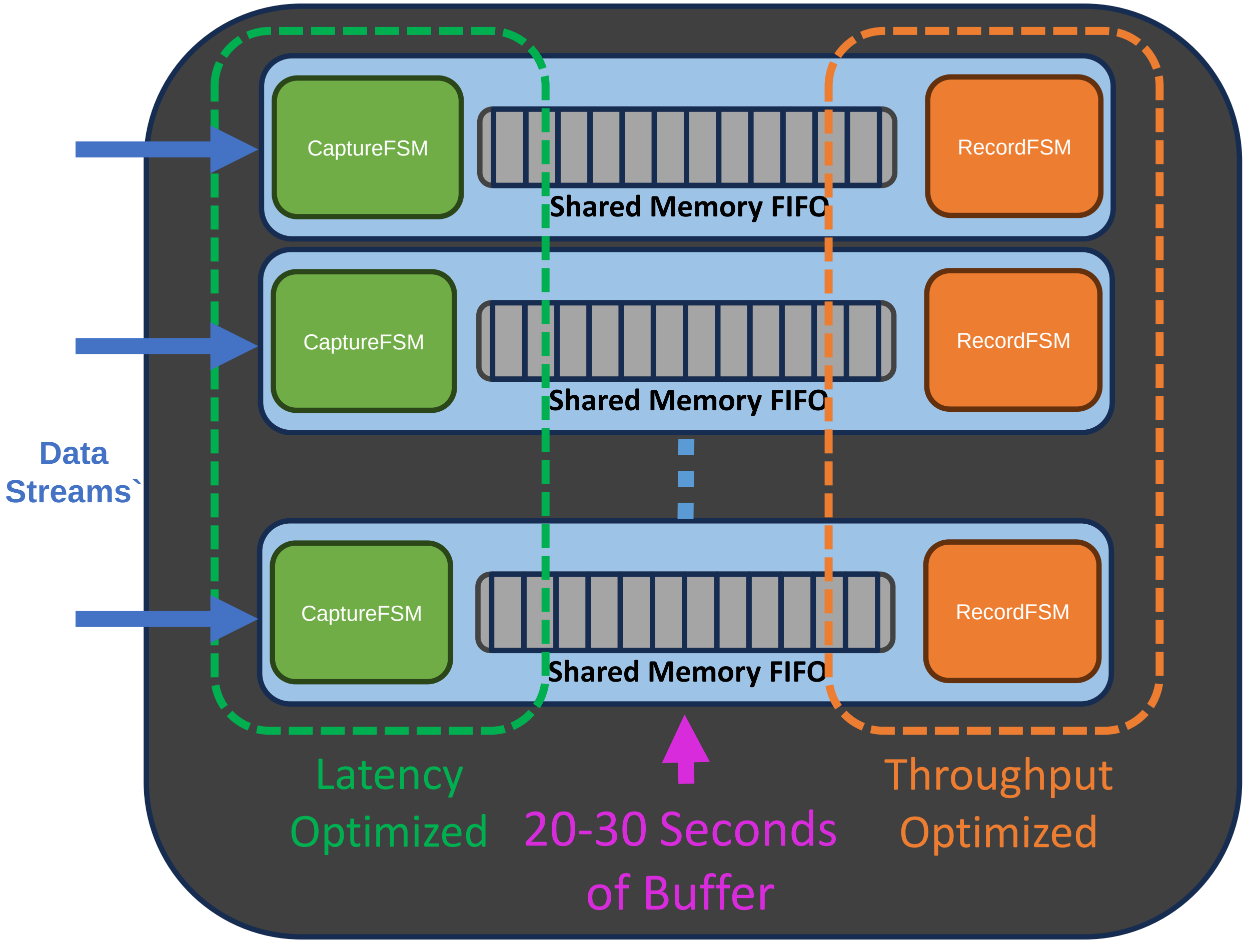
Our Software is as firmware-like as possible

- Independent Software Systems continuously run as *state machines*
- Behavior is defined by global variables modifiable by any privileged process

We Developed a Novel FIFO to connect Software Components

- This is a zero-copy Fifo. Data is never copied
- Allows simultaneous writing and reading

→ The result is a low overhead and simple communication mechanism between software subsystems



SkuTek Instrumentation
Where Science Meets Industry

Once Captured, We Have to Write Data to Disk

Writing to disk efficiently presents 2 major challenges:

1. Write performance can vary widely depending the code and storage environment *(OS, device contention, interface)*
2. Different experiments will have different storage requirements.
   - High-Capacity HDD RAID array
   - High Speed local SSDs
   - Networked Remote Object Store

```
FILE* fp = fopen("example.file","w");

fwrite(data_ptr, size, nmemb, fp);

fflush(fp);

fclose(fp);
```

```
SPEW_file* spew_fp = SPEW_open("example.file","w");
SPEW_initialize(spew_fp, SPEW_BUFFERED_STREAM, buffer_size);

SPEW_write(data_ptr, size, nmemb, spew_fp);

SPEW_flush(spew_fp);

SPEW_close(spew_fp);
SPEW_free(spew_fp);
```

The Solution:

Skutek PErformance library for Writing (SPEW)

SPEW provides a common interface to a variety of different disk writing techniques

- SPEW is a C library designed to mimic standard POSIX writing functions

Solidago (Data Streaming) → **Liatris** (Stream Capture) → **Salix** (File Transfer)

SkuTek Instrumentation
Where Science Meets Industry

SPEW Simplifies I/O Optimization

Using SPEW is straightforward.

- Often only requiring 2 additional lines of code over POSIX standards

SPEW itself is relatively simple code

- Its power is in its unified interface for a variety of writing strategies
- Best I/O practices are handled behind the scenes for the user

Currently used across our product line!!!

- In some cases on Liatris, it has improved our write speed by 70%

We are investigating commercializing SPEW by itself

```c
FILE* fp = fopen("example.file","w");

fwrite(data_ptr, size, nmemb, fp);

fflush(fp);

fclose(fp);
```

```c
SPEW_file* spew_fp = SPEW_open("example.file","w");
SPEW_initialize(spew_fp, SPEW_BUFFERED_STREAM, buffer_size);

SPEW_write(data_ptr, size, nmemb, spew_fp);

SPEW_flush(spew_fp);

SPEW_close(spew_fp);
SPEW_free(spew_fp);
```

SPEW writing strategy	SSD Array (4 SSDs)	HDD Array (8 HDDs)
Linux Fwrite Default (control)	68.61 Gbps	11.52 Gbps
POSIX Custom Buffer	112.68 Gbps	11.73 Gbps
Direct Bufferless	109.71 Gbps	11.52 Gbps
Direct Buffered	116.11 Gbps	11.63 Gbps
Threaded Buffers	72.71 Gbps	11.53 Gbps
THEORETICAL MAX (manufacturer spec)	128 Gbps	12.8 Gbps

Solidago (Data Streaming) → **Liatris** (Stream Capture) → **Salix** (File Transfer)

Moving Data Between Computing Centers

Salix Data Transfer Node

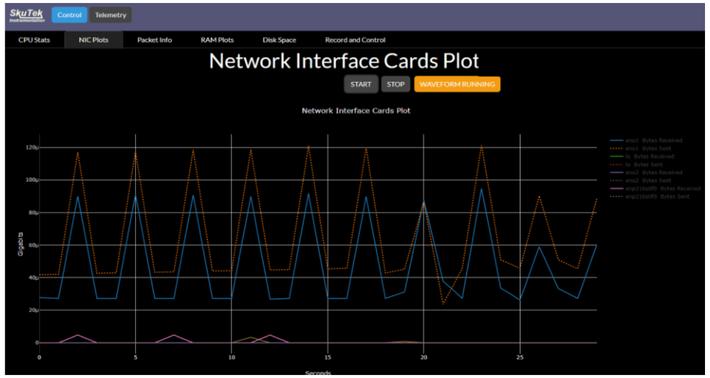
ESNet has published detailed designs for “Data Transfer Nodes” (DTNs) -Computers they pioneered for mass data movement. ***(We need to commend them for all this work!)***

We made a few innovations on ESNet reference designs:

1. **Python tuning library: “dmutils”** *(data management utils)*
   - Abstracts out the complicated process of applying tuning to the Linux kernel and networking hardware.
   - Can grid-search network configurations to find optimal profiles
   - Tuning profiles can be saved and applied in a single line
2. **Integration with Liatris**
   - We’re working on a mechanism to automatically transfer data files as soon as Liatris completes writing them
3. **REST API & Web Interface**
   - Easily query performance, stage transfers, or apply tuning profiles

```
from dmutils.nic import EthtoolManager

# Apply interrupt configuration settings
ethtool_manager = EthtoolManager("enp175s0f0np0")
ethtool_manager.isolate_queue_interrupt(queue_num, socket_cpu)
ethtool_manager.set_queue_target_port(queue_num, listening_port, pipeline_id)
```



Solidago (Data Streaming) → **Liatris** (Stream Capture) → **Salix** (File Transfer)

Testing *Salix* With Commercial Partnership Resources

Problem: As a small company, we don't have direct access to long-distance high-performance networks

Solution: Commercial Partnership!

1. *A potential partner in Maryland who we can't disclose publicly yet*
2. *National Energy Research Scientific Computing Center (NERSC)*

If all goes according to plan, our prototype ***Salix DTN*** will be traveling to Maryland later this year for cross-country data movement testing!

We're very grateful to our potential partner and NERSC for their collaboration! This involves no small amount of work on their part!

We will move data to NERSC and demonstrate usage on the Perlmutter supercomputer

Our ***Salix DTN*** is planned to be installed at a 40G Internet2 Uplink

Solidago (Data Streaming) → **Liatris** (Stream Capture) → **Salix** (File Transfer)

SkuTek Instrumentation
Where Science Meets Industry

There's Still a Lot to Do

1) Liatris and Salix codebases are in late-stage development

2) Users Interfaces and APIs are in early-stage development

- *Usability is critical! And GUIs are slow to develop.*
- *DOE is aiming to standardize computing APIs*
    - *We want to collaborate and help define best practices*

3) We want more in-situ proof of success

- *We're seeking partners to test/utilize these systems*

We looking into a Phase IIA/B Application to support continued development

Thank you!

Our PMs: Michelle Shinn and Manouchehr Farkhondeh

Friendly Faces in DOE: Mario Cromaz, Giordano Cerriza, and John Anderson

EXTRA thanks to our Interns: Iris Bassin and Charlie Vitkus


Iris Bassin (left) Charlie Vitkus (right)

Backup Slides

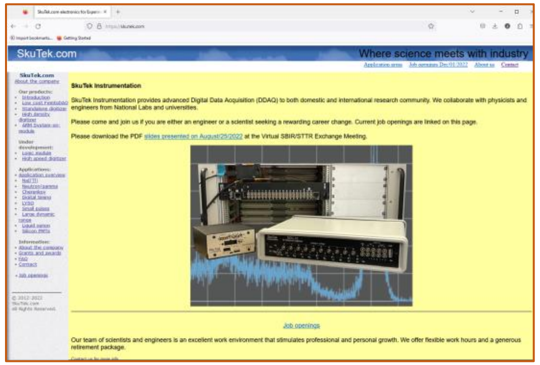
SkuTek Instrumentation
Where Science Meets Industry

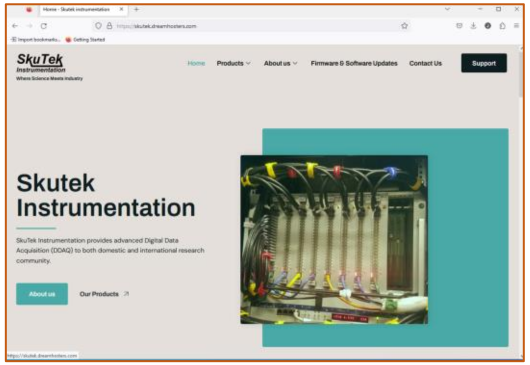
TABA funding has been fantastic

- TABA funding has helped our business has develop considerably!

- We're building our brand by sponsoring conferences

- We have a new website focused on showcasing our products and providing support





Visit us at www.skutek.com!

References

1. Sands, A. E. (2017). Managing Astronomy Research Data: Data Practices in the Sloan Digital Sky Survey and Large Synoptic Survey Telescope Projects. UCLA. Retrieved from https://escholarship.org/uc/item/80p1w0pm
2. Wang, Ruonan, Tobar, Rodrigo, Dolensky,… Processing Full-Scale Square Kilometre Array Data on the Summit Supercomputer. United States. https://doi.org/10.1109/SC41405.2020.00006
3. Lamanna, G., Antonelli, L. A., Contreras, J. L., Knödlseder, J., Kosack, K., Neyroud, N., … Zoli, A. (2015). Cherenkov Telescope Array Data Management. doi:10.48550/ARXIV.1509.01012
4. Cromaz, M., Dart, E., Pouyoul, E., & Jansen, G. (2021). Simple and Scalable Streaming: TheGRETA Data Pipeline*. EPJ Web Conf., 251, 04018
5. “FasterData.” Fasterdata.es.net, United States Department of Energy, https://fasterdata.es.net/.
6. “ESNet.” www.es.net, United States Department of Energy, https://www.es.net/science-engagement/science-requirements-reviews/esnet-network-requirements-reviews/bes-requirements-review-2022/