Data Management for High Speed, Distributed Data Acquisition

Jeff Maggio
Principal Investigator

SBIR Exchange Aug 2025

Outline

- Our Company and Capabilities
- Our Current Product Line
- Innovations from this grant
- Acknowledgements

SkuTek Instrumentation Makes… **Instrumentation!**

- Located in Rochester NY
- The Team
  - Full time: 5 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

Houghton College

Los Alamos National Laboratory

Facility for Rare Isotope Beams at Michigan State University

University of Rochester

Fermilab

Brookhaven National Laboratory

Berkeley Lab

Tennessee Tech

Augustana University

The University of Tennessee Knoxville

National Superconducting Cyclotron Laboratory

Simon Fraser University Engaging the World

Albert Einstein Center for Fundamental Physics

Wiener Plein & Baus Corp.

Brown

Universität Bern

Université Laval

Shanghai Jiao Tong University

Montclair State University

Northrop Grumman

Raja Ramanna Centre for Advanced Technology

SkuTek Instrumentation
Where Science Meets Industry

Our Capabilities: Full End-End DAQ Expertise

Interfaces

Interfaces

Digitizers

Arrays Configurable from 2 to 1000+ Channels, used in:

- Astrophysics (*LuxZeplin Dark Matter*)
- Nuclear Sciences
- Mining
- Medical Imaging Studies
- Education & Research

Run-Control And Experiment Management

- Run Control Software
- Graphical Interfaces and Monitoring
- Open APIs for controlling our hardware

Custom Firmware

- Trigger Management and Control
- Digital Signal Processing
- Network Streaming
- Custom Offloading Solutions

Developed under this grant

Data Management

- **Streaming Digitizers** offload Data at line rates with minimal deadtime
- **Data Collector** Computers receive network streams and store your data at up to 80Gb/s per Data Collector

Automatic Data Transfer

- **Globus Scheduling Software** to automatically move data your Instrument to High Performance Computing (HPC) Facilities.
- Architected to operate within Science DMZ Networks

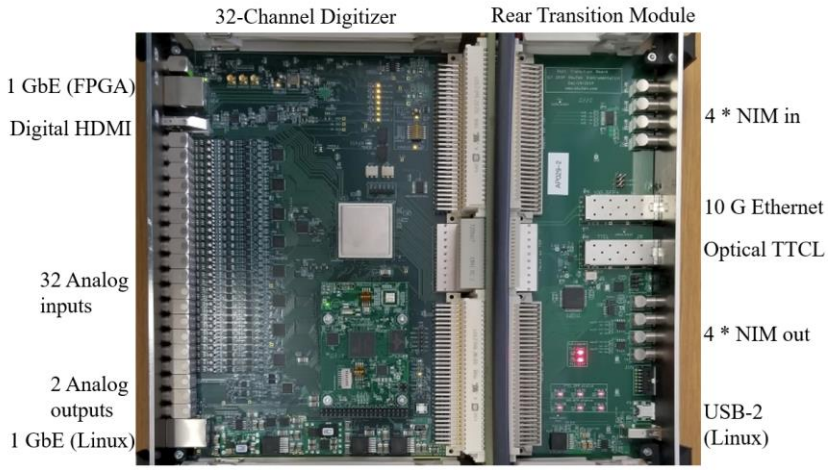
Historically We Specialized in High Performance Digitizers

Benchtop Digitizers

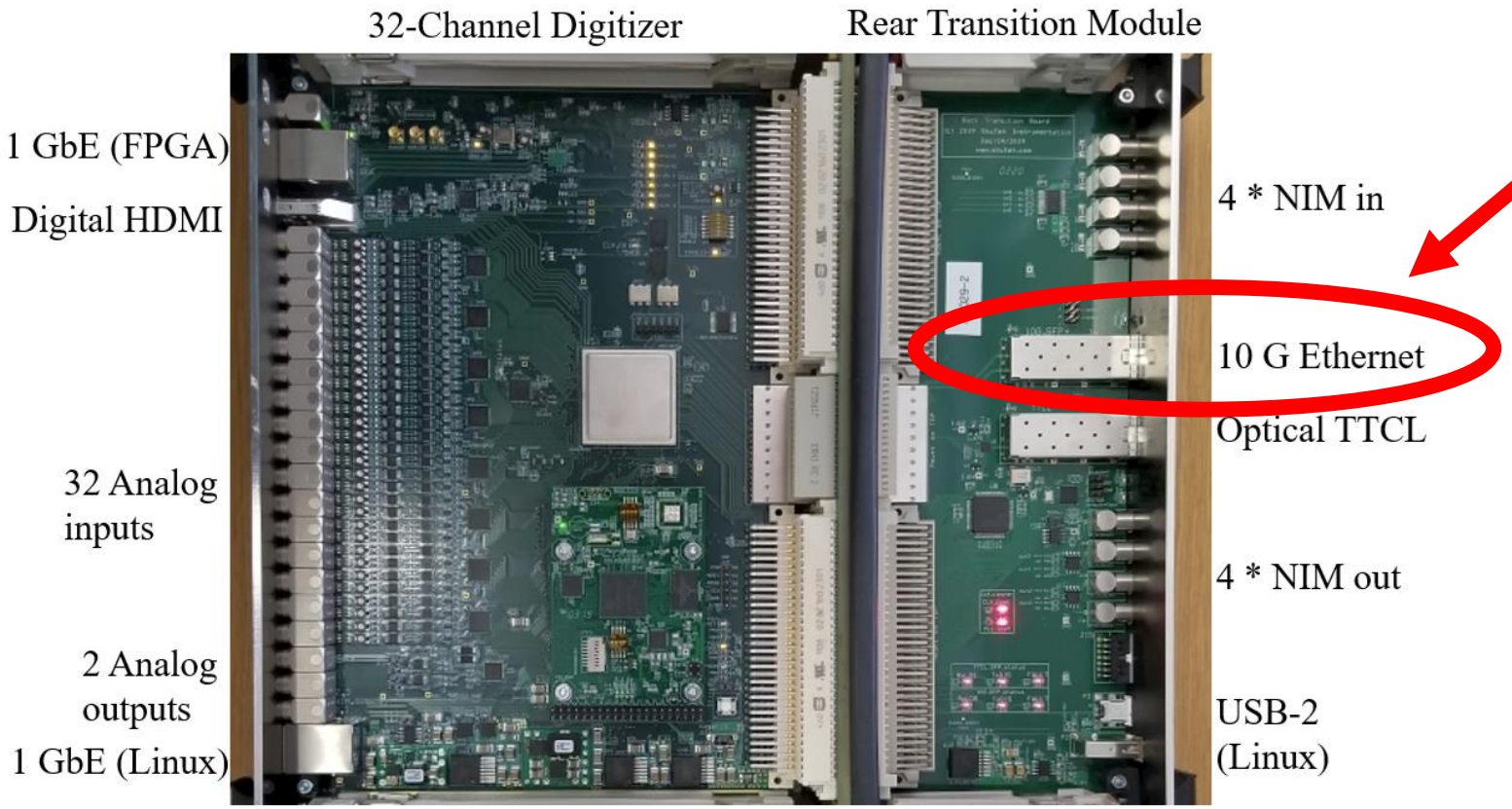
FemtoDAQ Vireo

FemtoDAQ Kingfisher

Chickadee-32 Rackmount Digitizer



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), ESnet Jefferson-lab FPGA Accelerated Transport (EJFAT)

Takeaway: DAQ systems must account for this new paradigm.

SkuTek Instrumentation
Where Science Meets Industry

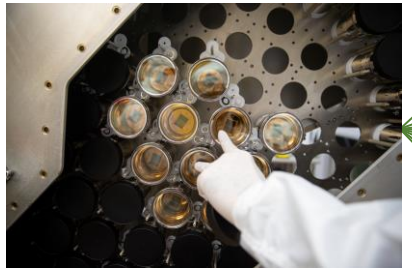
There is a New Paradigm in Data Acquisition

Instrument Facility

Offsite Computing Center

Hardware

Data Acquisition | **Data Collection** | **Data Transfer** | **Analysis and Long-Term Storage**



Detector
(pictured: LuxZeplin)

Analog Signal

Buffering and/or Recording

Facility Data Transfer Node

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

Software

user

Hardware Optimization and Tuning

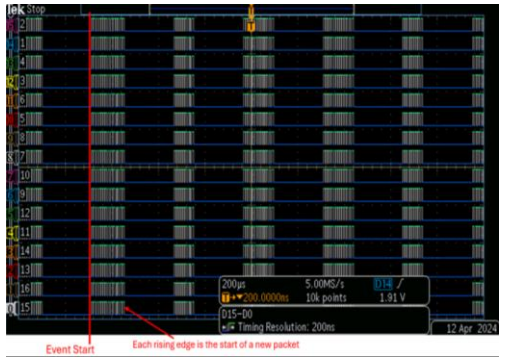
Control Interfaces and Diagnostics

All at 100 Gbps or Faster!

SkuTek Instrumentation
Where Science Meets Industry

Our Data Management Solutions are both Hardware and Software

Network Testing & Simulation Tools



Data Collection Hardware

Experimental Control and Optimization Software

Data Management Utilities
dmutils

SPEW
Skutek PErformance Writing Library

SkuTek Utilities

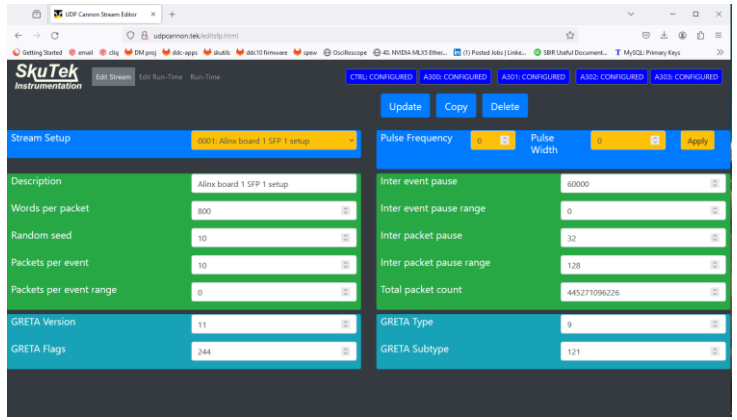
SkuTek Globus
Auto File Transfer System

How Can We be Sure Our System Works at ~100 Gbps?

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
- Controllable via a web Interface and REST API

Solidago is currently for sale!



Solidago Emulates Better than Software Equivalents

1) Hardware Solution means that no software tuning is needed

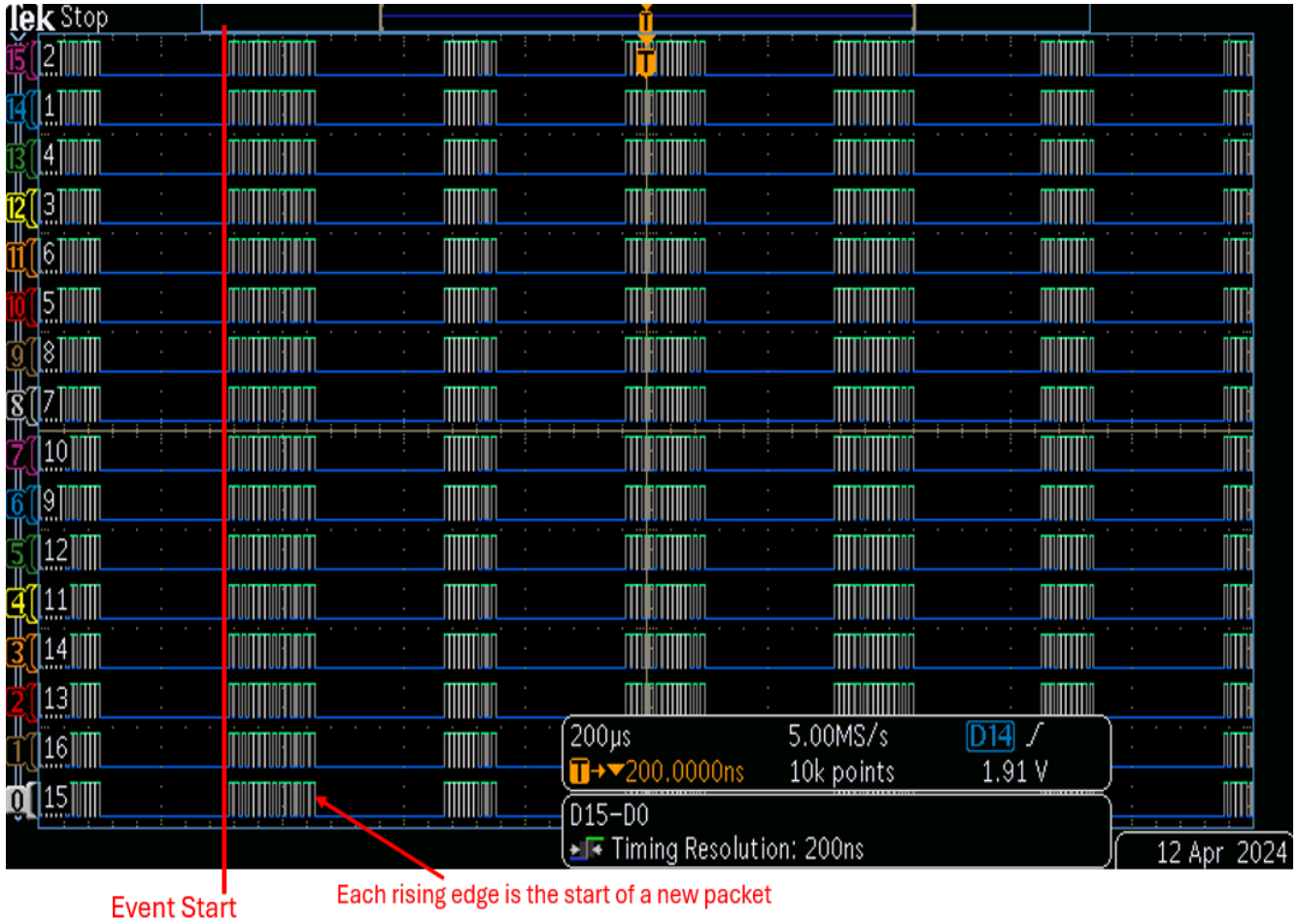
2) Streams can be synchronized with each other *(or run independently)*

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"



Possible Expansion for Solidago

We're exploring commercialization of it's firmware

ESnet Jefferson-lab FPGA Accelerated Transport (EJFAT) *expressed interest in the technology as licensed firmware.*

We're exploring a Phase IIB for developing this commercial opportunity

The Modern Data Acquisition Chain

Instrument Facility

Offsite Computing Center

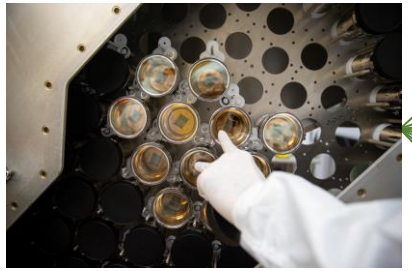
Hardware

Data Acquisition

Data Collection

Data Transfer

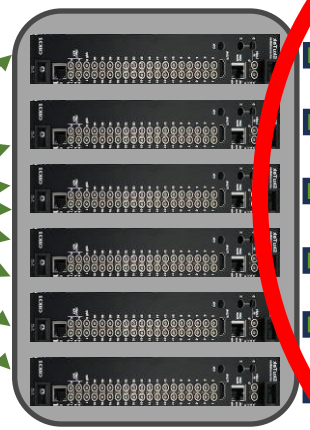
Analysis and Long-Term Storage



Detector

(pictured: LuxZeplin)

Analog Signal



Buffering and/or Recording

Facility Data Transfer Node

High Performance Computing (HPC) Facility

(pictured: NERSC)

Software

user

Hardware Optimization and Tuning

Control Interfaces and Diagnostics

Data Collector Computers

Liatris

80 Gbps Data Collector for high density digitizer arrays

Monarda

10 Gbps Data Collector for smaller setups

Handling Data at these Rates is Tricky

1) Hardware needs to "tuned" before it can meets spec

- 100 Gbps Hardware != 100 Gbps Computer
- We control 80+ parameters across kernel settings, hardware drivers, network configuration, and software methods that influence data handling speeds
- You'll see this discussed throughout DOE literature

2) Developing High Performance Software is a Time-Intensive Task

- Off the shelf solutions often don't scale
- You have to be very careful not to over-tune for a particular application

It's a challenge to maintain performance across the diversity of data acquisition conditions

Making Data Management Flexible and Portable

Data Management Utilities
dmutils

Data Management Utilities (**dmutils**) is a python-language library for tuning hardware.

- dmutils performs complex or esoteric tuning operations behind ease to use functions.
- All 80+ tuning parameters we control for are configured through dmutils
- The human effort for tuning hardware is drastically reduced and automatable with scripts.

SPEW
Skutek PErformance Writing Library

Skutek **PE**rformance **W**riting library (**SPEW**) is a C-language library for writing data to disk

- SPEW provides a common interface to multiple different writing strategies using built-in tools.
- Single line changes in source code influence buffering, memory alignment, metadata syncing, block-usage, etc.
- The human effort to write optimal code for different storage mediums is drastically reduced.

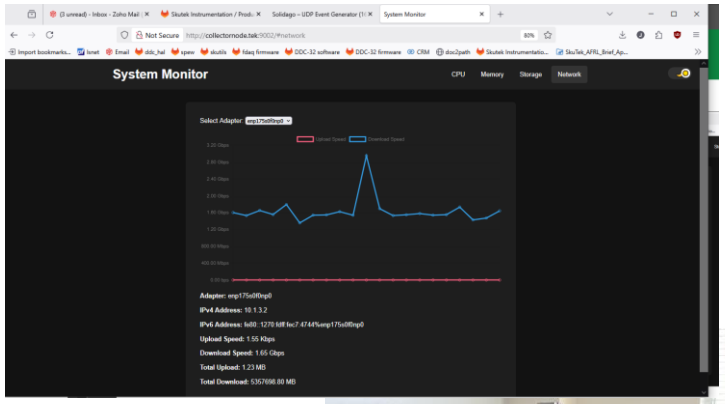
Takeaway:

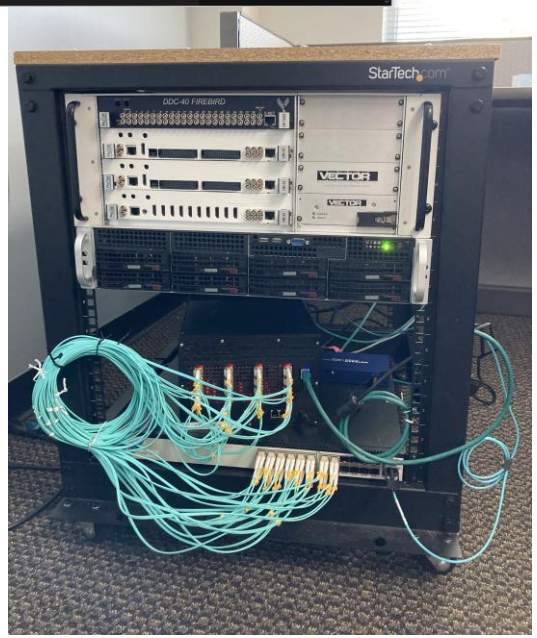
We can rapidly adapt our software to different data acquisition environments

Data Collector Computers In Summary

You can:

- Receive and record data streams from hundreds of channels simultaneously.
    - *Total sustainable rate is: 80 Gbps * number of data collectors*
- Monitor event rates and performance through a web interface
- Mount it in the same rack as your VME digitizers

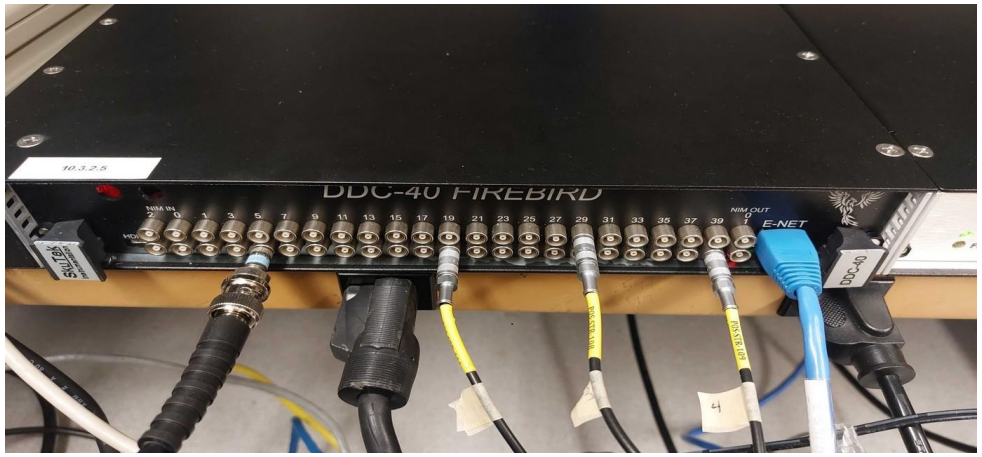




Our software makes us well-poised to adapt this technology for future applications and new hardware

We will be ready to give demos of this system by end of NCE

We Just Sold our First Data Collector!



Digitizer
(Data Acquisition)

Data Collector Computer
(Data Collection)

Digitizer and Data Management Solution deployed for
Aaron Manalaysay and Taurean Zhang at Lawrence Berkley National Lab

The Modern Data Acquisition Chain

Instrument Facility

Offsite Computing Center

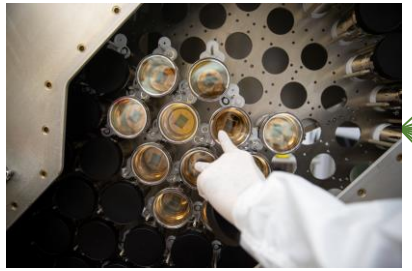
Hardware

Data Acquisition

Data Collection

Data Transfer

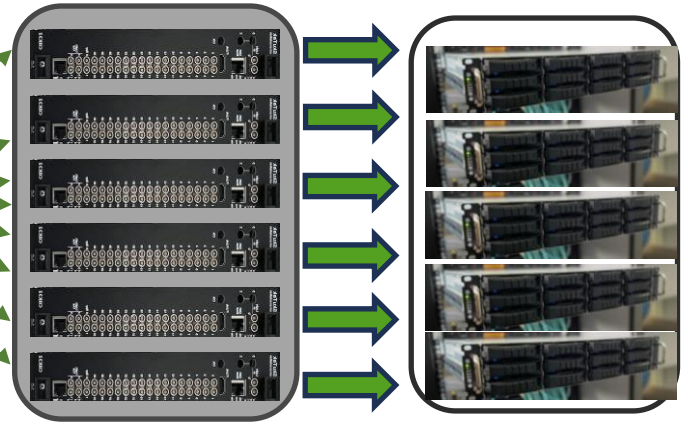
Analysis and Long-Term Storage



Detector

(pictured: LuxZeplin)

Analog Signal



Data Collector Servers

Facility Data Transfer Node

High Performance Computing (HPC) Facility

(pictured: NERSC)

Software

Data Management Utilities
dmutils

SPEW
Skutek PErformance Writing Library

Hardware Optimization and Tuning

Control Interfaces and Diagnostics

We Had to Change our Data Transfer Plans

Last year we were planning on developing another server known as a **Data Transfer Node (DTN)**

- *DTNs were pioneered by DOE's Energy Sciences Network (ESnet)*
- *DTNs generally use commercial software known as Globus for file transfer*

This would have been a Hardware Product building off of ESnet's innovations and documentation

However, Community Interest In a Hardware Product was Lackluster…

We switched approaches…

globus

Logo for Globus
(NOT a SkuTek product)

Instead, We Developed Software to Automatically Schedule Transfers on Existing DTN Infrastructure

Our Globus Scheduling Software will interface initiate transfers **using existing Data Transfer Nodes** (DTNs)

- *It can operate while data collection is ongoing.*
- *This control software can run on any computer with network access to Data Collectors*

We do this by interfacing with the Globus API

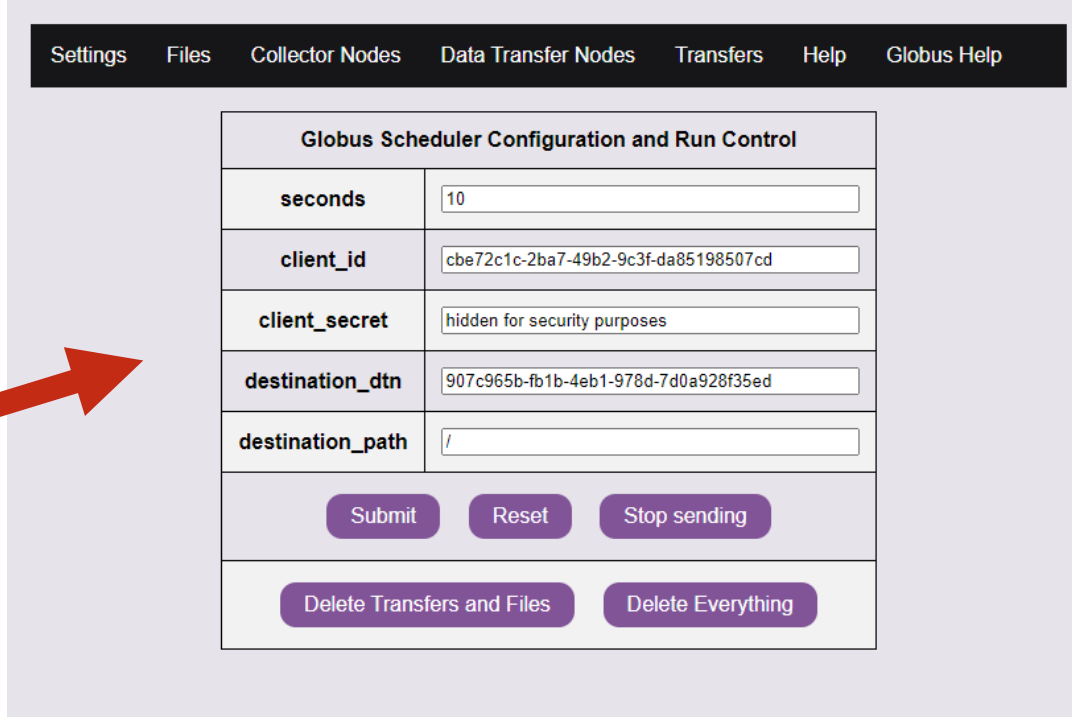
- *Globus Software is utilized extensively across DOE DTN Infrastructure*

Minimal Setup is Required

SkuTek Globus Auto File Transfer System

globus

Logo for Globus (NOT a SkuTek product)



The Modern Data Acquisition Chain

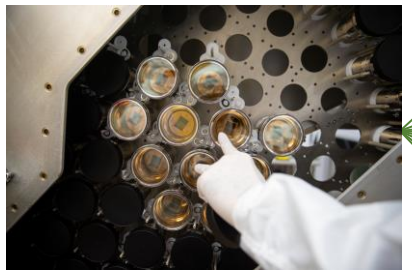
SkuTek Instrumentation
Where Science Meets Industry

Instrument Facility

Offsite Computing Center

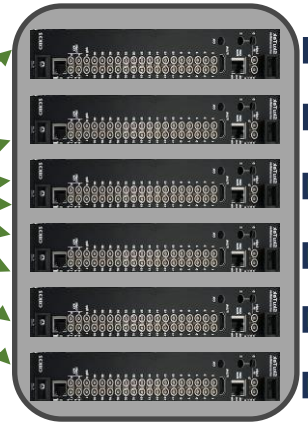
Hardware

Data Acquisition | **Data Collection** | **Data Transfer** | **Analysis and Long-Term Storage**



Detector
(pictured: LuxZeplin)

Analog Signal



Data Collector Servers

Facility Data Transfer Node

SkuTek Globus Auto File Transfer System

globus

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

Software

Data Management Utilities
dmutils

SPEW
Skutek PErformance Writing Library

Hardware Optimization and Tuning

user

Control Interfaces and Diagnostics

Introducing the Skutek Utilities Library (skutils)!

Modern distributed data acquisition systems can be composed of thousands of channels and dozens of digitizers.

Data Collection Systems have to be configured simultaneously.

We need robust remote-control software.

SkuTek Utilities (aka skutils) is a Python library which runs on your desktop to control SkuTek hardware remotely.

The need for a remote-control toolkit was brought to our attention by FRIB

A quick example:

First install with python-pip:

```
jeff@demo:~$ pip install skutils
```

```python
import skutils

# Connect to your digitizer with the IP address or hostname
digitizer = skutils.FemtoDAQController('vireo-000019.tek')

# Configure data acquisition settings
channel = 0
digitizer.setTriggerSensitivity(channel, 128)
digitizer.setEnableTrigger(channel, True)

# Trigger Data Collection and Recording
digitizer.configureRecording(channels_to_record=[channel],
                             file_recording_format = "igorph"
                             )

# Start collecting data!
digitizer.start()
```

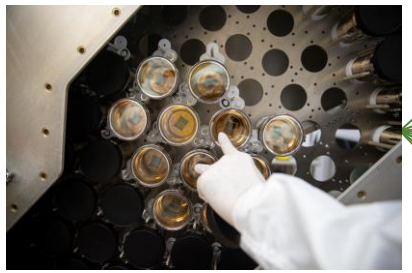
The Modern Data Acquisition Chain

Instrument Facility

Offsite Computing Center

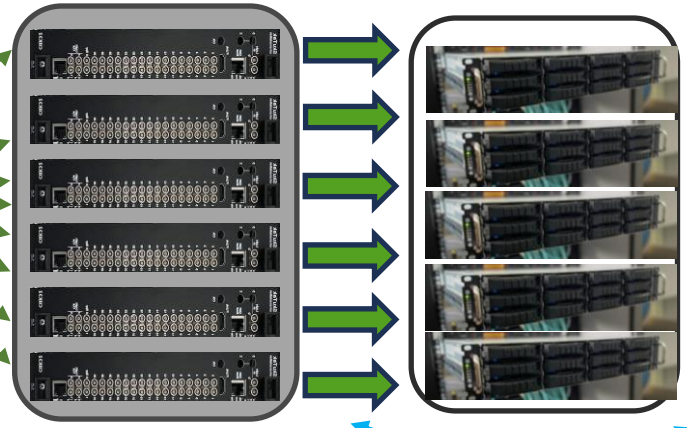
Hardware

Data Acquisition | **Data Collection** | **Data Transfer** | **Analysis and Long-Term Storage**



Detector
(pictured: LuxZeplin)

Analog Signal



Facility Data Transfer Node

SkuTek Globus Auto File Transfer System

globus

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

Data Collector Servers

Software

Data Management Utilities

SPEW
Skutek PErformance Writing Library

Hardware Optimization and Tuning

user

Control Interfaces and Diagnostics

SkuTek Utilities

SkuTek will be able to provide for every stage in the data acquisition chain :)

Thank you!

Special Thanks to Michelle Shinn and Manouchehr Farkhondeh

This work was supported by the

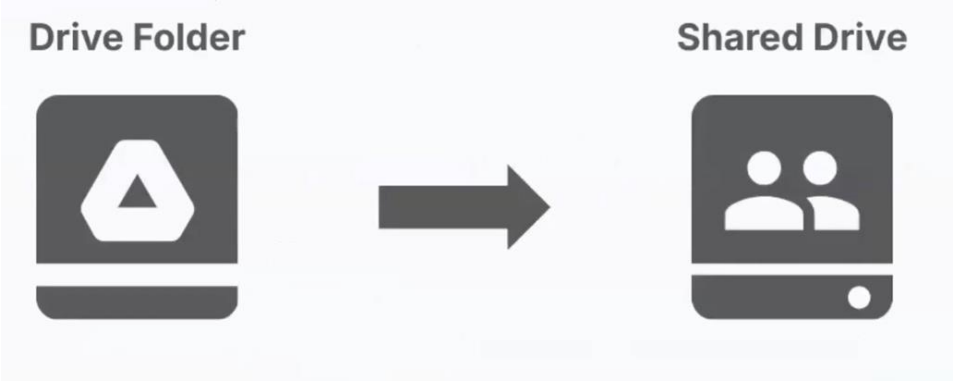
US Department of Energy, Office of Nuclear Physics

Award Number: DE-SC00021502

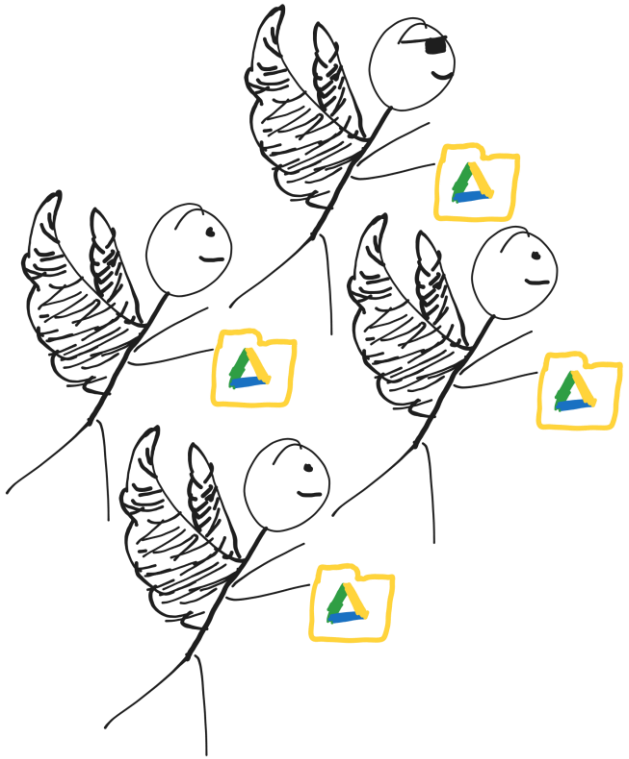
Backup Slides

Google Drive Migration Software

SkuTek, in collaboration with LZ, developed an application to migrate the Lux-Zeplin Collaboration's shared personal "**Drive Folder**" to a "**Shared Drive**"



Cloud data was migrated for 400+ users and custom apps. Over >100,000 files with no data lost. All old links to shared files were preserved.



A flock of physicists migrating their files

*Graphic of our product's use provided by Dr. Eli Mizrachi of the LZ collaboration

Skutek PErformance Writing library (SPEW)

Skutek PErformance Writing library (SPEW) is a C-language library for writing data to disk

But why…? Doesn't every language have a way to write to disk?

SPEW provides a common interface to multiple different writing strategies using built-in tools.

There are parameter flags that affect buffering, dumping, memory alignment, metadata syncing, etc0

This makes it quick for humans to optimize disk writing without rewriting software

The best strategy varies wildly depending on hardware and software conditions

With SPEW, we can rapidly modify our software stack to change deadtime, speed, disk degradation, memory usage, and determinism of disk I/O depending on customer needs

Cut down on Jargon

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 Direct	72.71 Gbps	11.53 Gbps
THEORETICAL MAX (manufacturer spec)	128 Gbps	12.8 Gbps

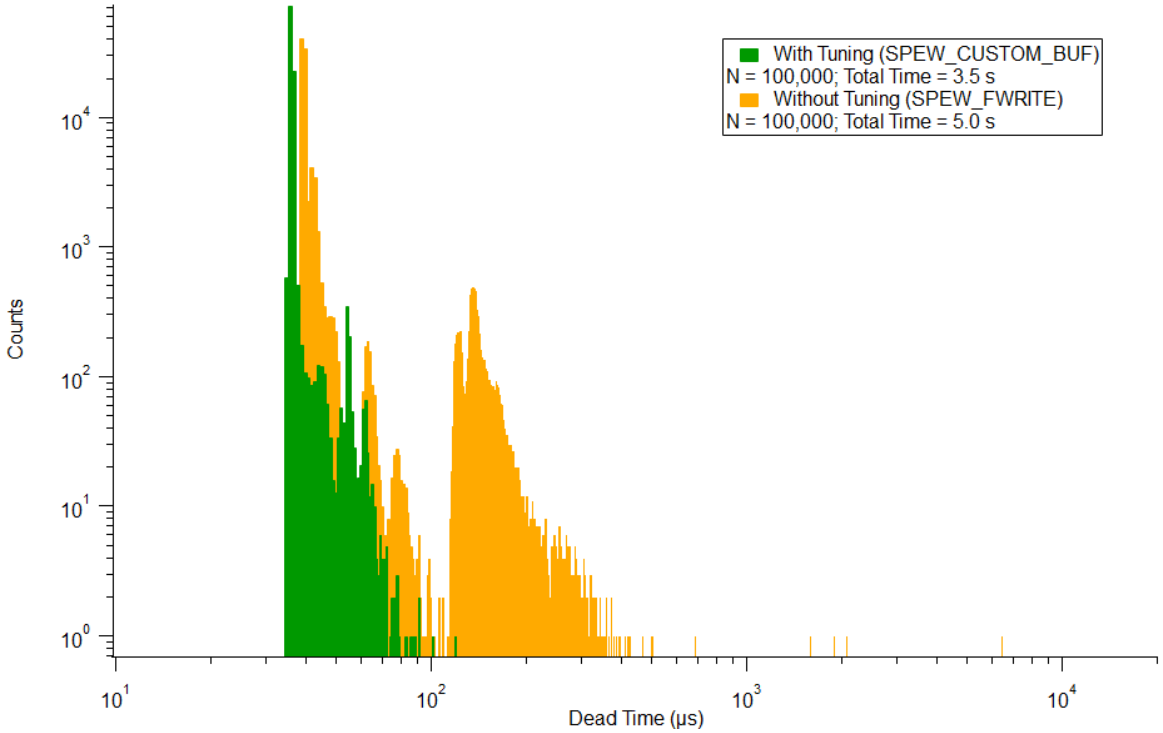
Example: 70% writing speed improvement caused by tweaking SPEW flags in a single line of code

We improved DAQ deadtime using SPEW

On our digitizers, we noticed large deadtimes in data acquisition caused by recording data to disk

Disk I/O tuning using SPEW helped us improve mean data acquisition deadtime by ~30% and cuts deadtime variance from 39us to 2.5us

This optimization was performed for researchers using our FemtoDAQ Vireo at FRIB



Data acquisition deadtime on FemtoDAQ Vireo with different SPEW modes

Less deadtime (left side of figure) is better

This work was performed by Mr. Jackson Hebel

Our dmutils library lets us Tune Computers Quickly

Examples:

- Automatically determining which CPU cores would be fastest to run on
  used extensively on data collectors

- Automatically optimizing cache performance by managing CPU interrupts
  critical in low-latency networking

- Allows us to store all 80+ startup parameters in a single place config file instead of individually in different subsystems
  makes setting up new hardware much faster