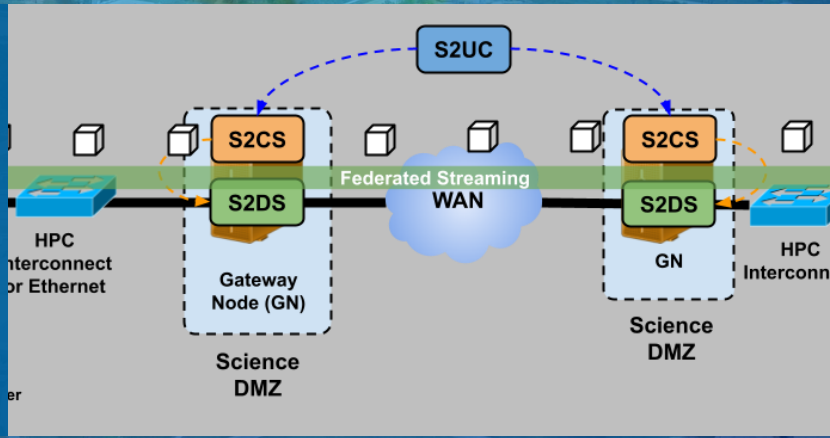
6TH GLOBAL RESEARCH PLATFORM WORKSHOP (6GRP)

SEPT. 16, 2025. CHICAGO, IL.

SCISTREAM: ENABLING APPLICATIONS TO STREAM DATA BETWEEN SCIENCE INSTRUMENTS AND HPC



JOAQUIN CHUNG
Research Scientist
Data Science and Learning Division
Argonne National Laboratory

U.S. DEPARTMENT OF ENERGY
Argonne National Laboratory is a U.S. Department of Energy laboratory managed by UChicago Argonne, LLC.

Argonne NATIONAL LABORATORY

OUTLINE

- Introduction to SciStream
  - Streaming Processing in Science Use Cases
  - Federated Scientific Instruments
  - SciStream Overview
- Using Testbeds for Prototype and Development
  - SciStream Architecture and Protocols [Chameleon Cloud, HPDC'22]
  - SciStream eBPF Data Plane [FABRIC Testbed, INDIS'22]
  - SciStream Demonstrations [SCinet, SC23 and SC24]
- Outlook

STREAMING PROCESSING IN SCIENCE USE CASES

U.S. DEPARTMENT OF ENERGY
Argonne National Laboratory is a U.S. Department of Energy laboratory managed by UChicago Argonne, LLC.

Argonne NATIONAL LABORATORY

SCIENTIFIC FACILITIES PRODUCE DATA AT UNPRECEDENTED RATES

Light sources such as the APS are expected to generate Tbps of raw data

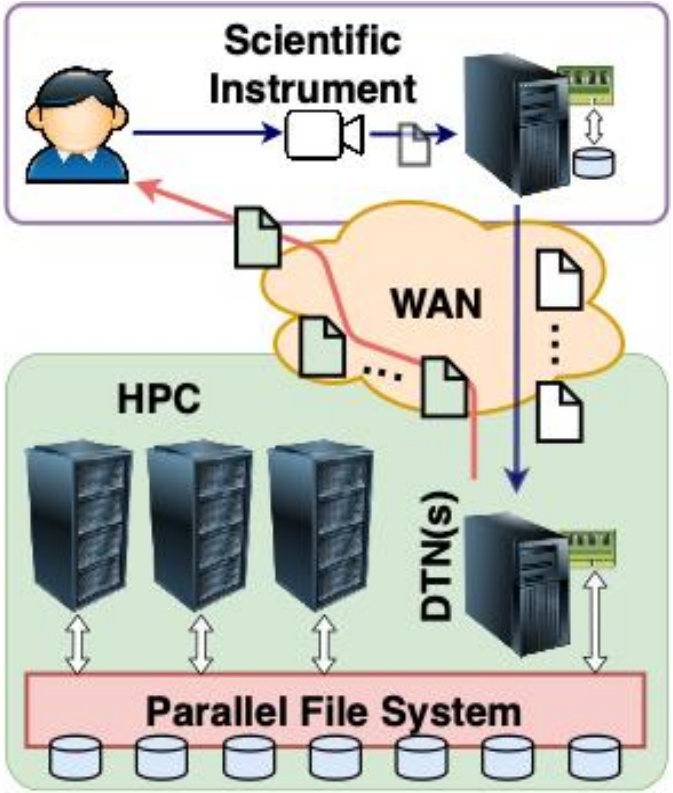
TRANSFER DATA TO HPC FACILITIES FOR PROCESSING

Experiment

Transfer data

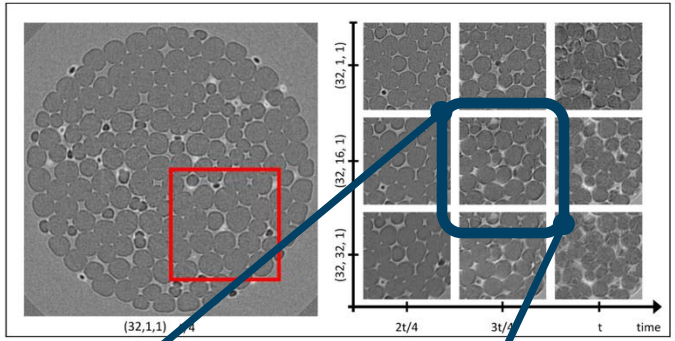
Data Analysis

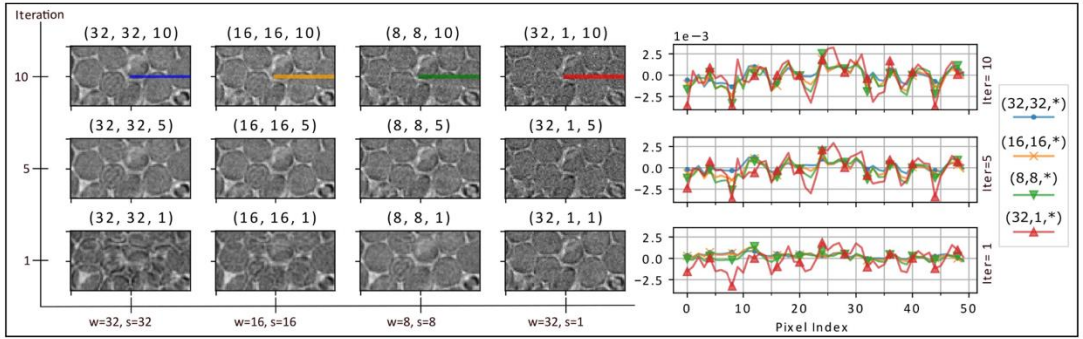
TRADITIONAL APPROACH: FILE TRANSFER & OFFLINE PROCESSING

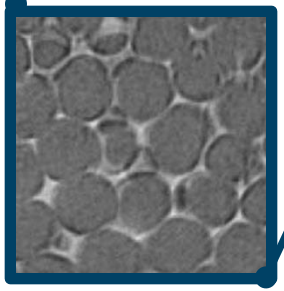


OFFLINE ANALYSIS MAY RESULT IN WASTEFUL EXPERIMENTAL TIME

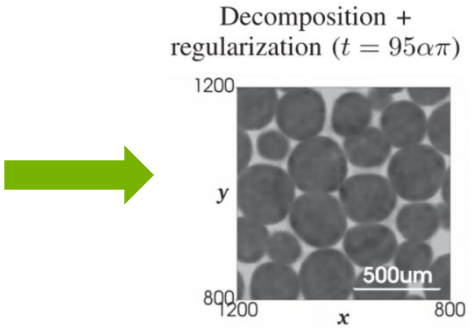
Need Complementary Data Analysis Techniques: Offline and Online



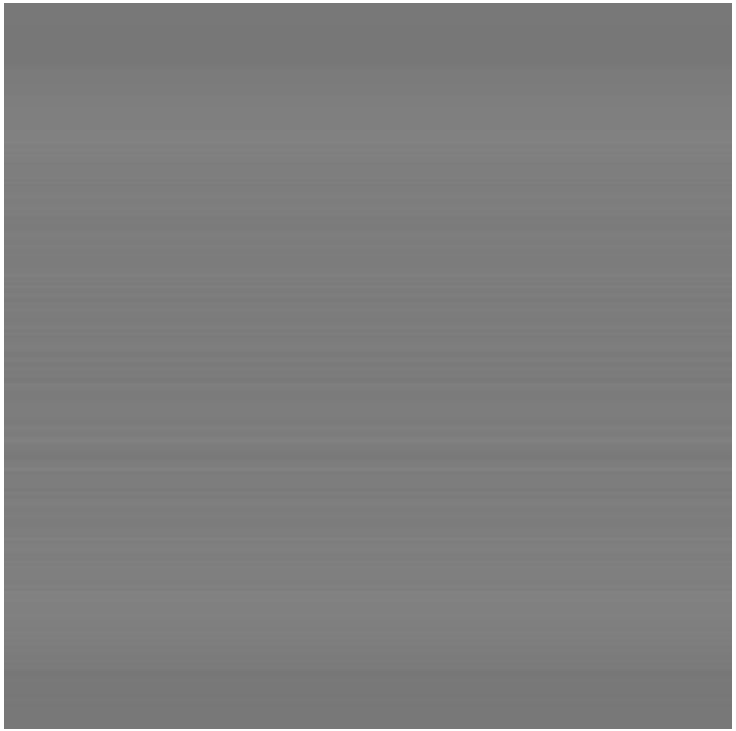




Approximate reconstruction is in real-time but not as accurate as the FBP+TV with full data.



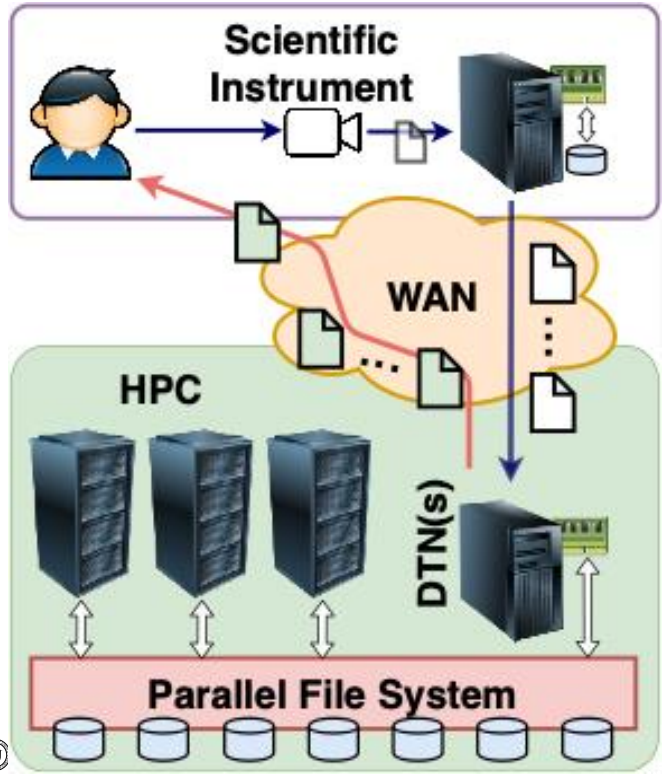
Online Analysis Can Significantly Reduces Experiment Time



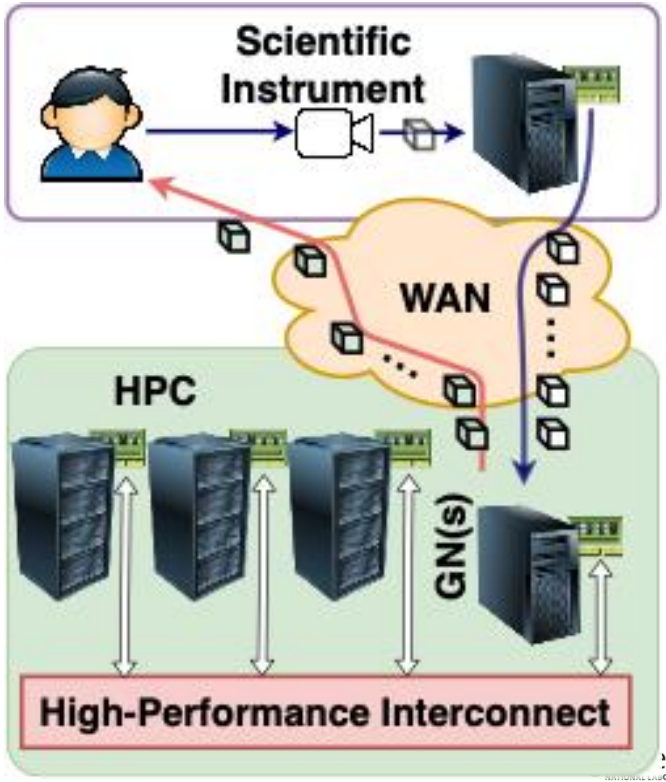
- Foam dataset with increasing temperature
- Initial stage does not require fast data acq.
  - Time (secs): [0, 17]
- The faster data acquisition is desired when features change rapidly
  - Time (secs): (17,-)
- Ideal experimentation with stream (processing)
- **Real-time processing with rapid feature detection, adjusting data acquisition**
- Other potential science use cases
- Region of Interest, on-the-fly data reduction…

Online Analysis Needs Memory-to-Memory Streaming

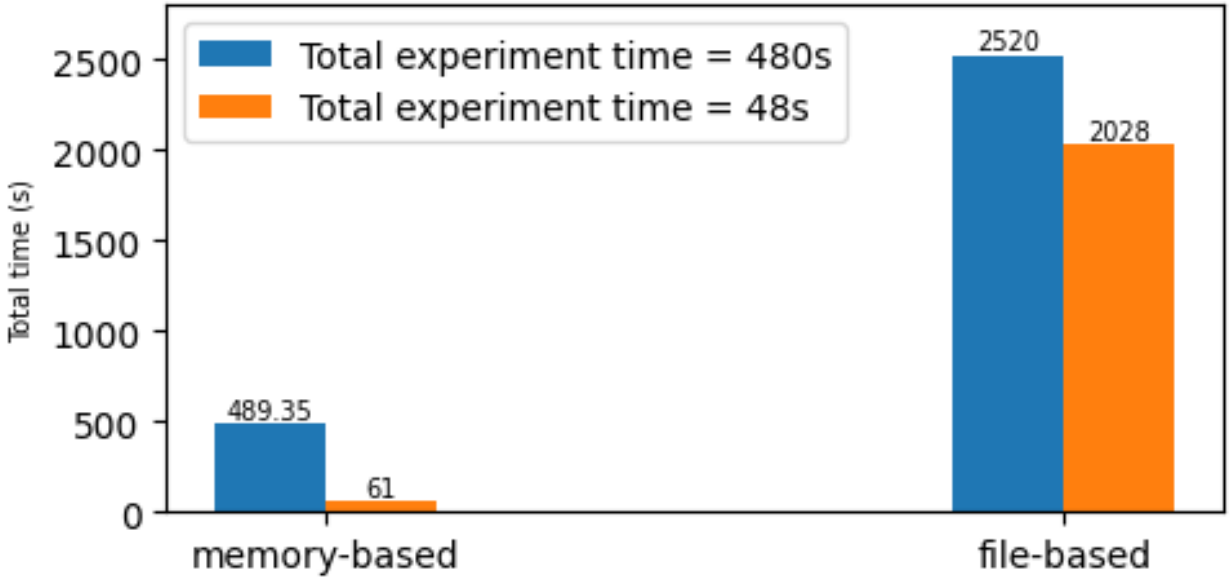
Traditional File-based Data Movement



Memory-to-memory Streaming From Instrument to HPC



ONLINE ANALYSIS: MEMORY-TO-MEMORY STREAMING VS FILE-BASED TRANSFERS



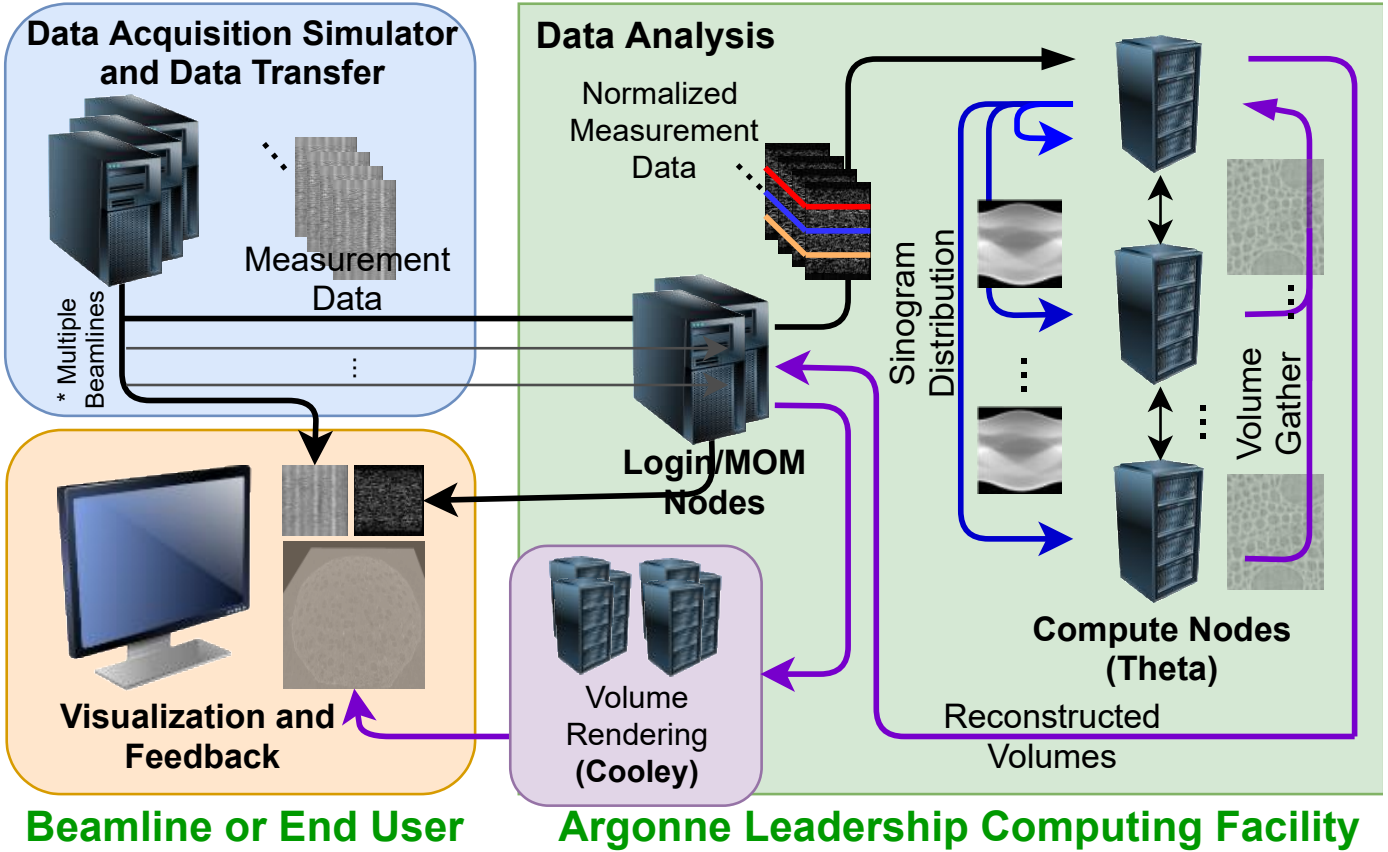
Total data generated:
1440 frames with 12 GB

FEDERATED SCIENTIFIC INSTRUMENTS

U.S. DEPARTMENT OF ENERGY
Argonne National Laboratory is a U.S. Department of Energy laboratory managed by UChicago Argonne, LLC.

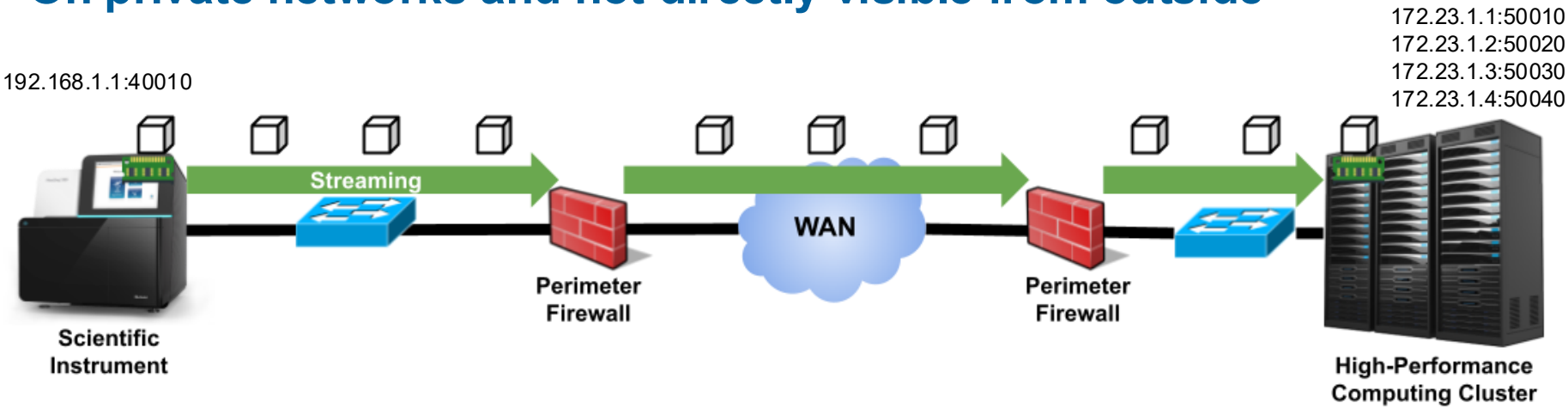
Argonne NATIONAL LABORATORY

SC19 TECH CHALLENGE: REAL-TIME STREAM ANALYSIS OVER WAN



INSTRUMENT AND HPC ARE IN DIFFERENT SECURITY DOMAINS

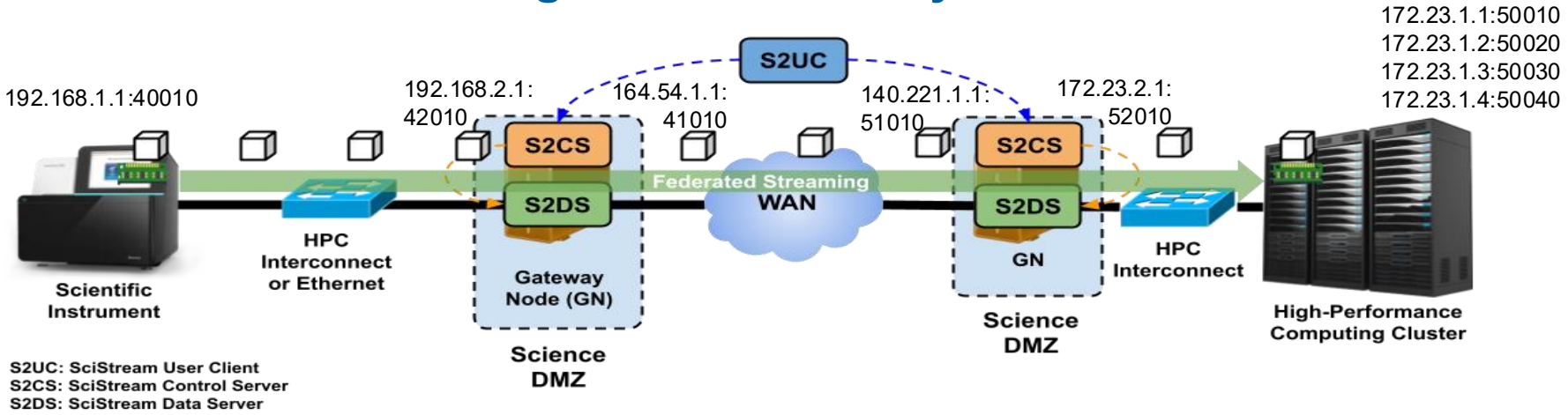
On private networks and not directly visible from outside



- Scientists need to negotiate with each institution opening TCP ports on their firewalls to enable streaming
- Scientists also need to implement the mechanisms to comply with the security policies of each institution

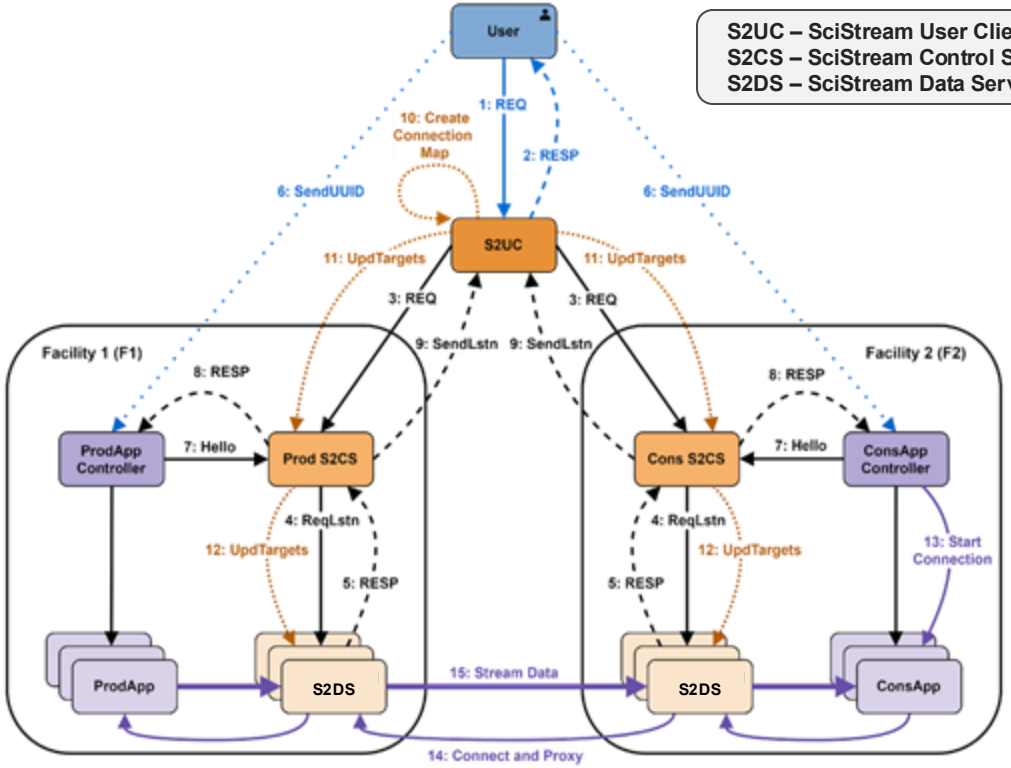
SCISTREAM

Enables secure streaming between security domains



- ✓ Automates the process of deploying secure bridges between security domain
- ✓ Keeps streaming app semantics intact
  - ✓ One end listens on one or more [IP:Port]
  - ✓ Other end connects to given [IP:Port]
- ✓ Requires only minimal changes to app to authenticate itself to SciStream
- ✓ Supports a lot of streaming applications
  - ✓ Operate at the transport protocol layer
  - ✓ One-time deployment and vetting for institutions
- ✓ Uses Globus Auth for federated id & securing data & control channel
- ✓ Supports 3rd party control for hosted orchestration and monitoring

SCISTREAM PROTOCOL AND DESIGN CONSIDERATIONS

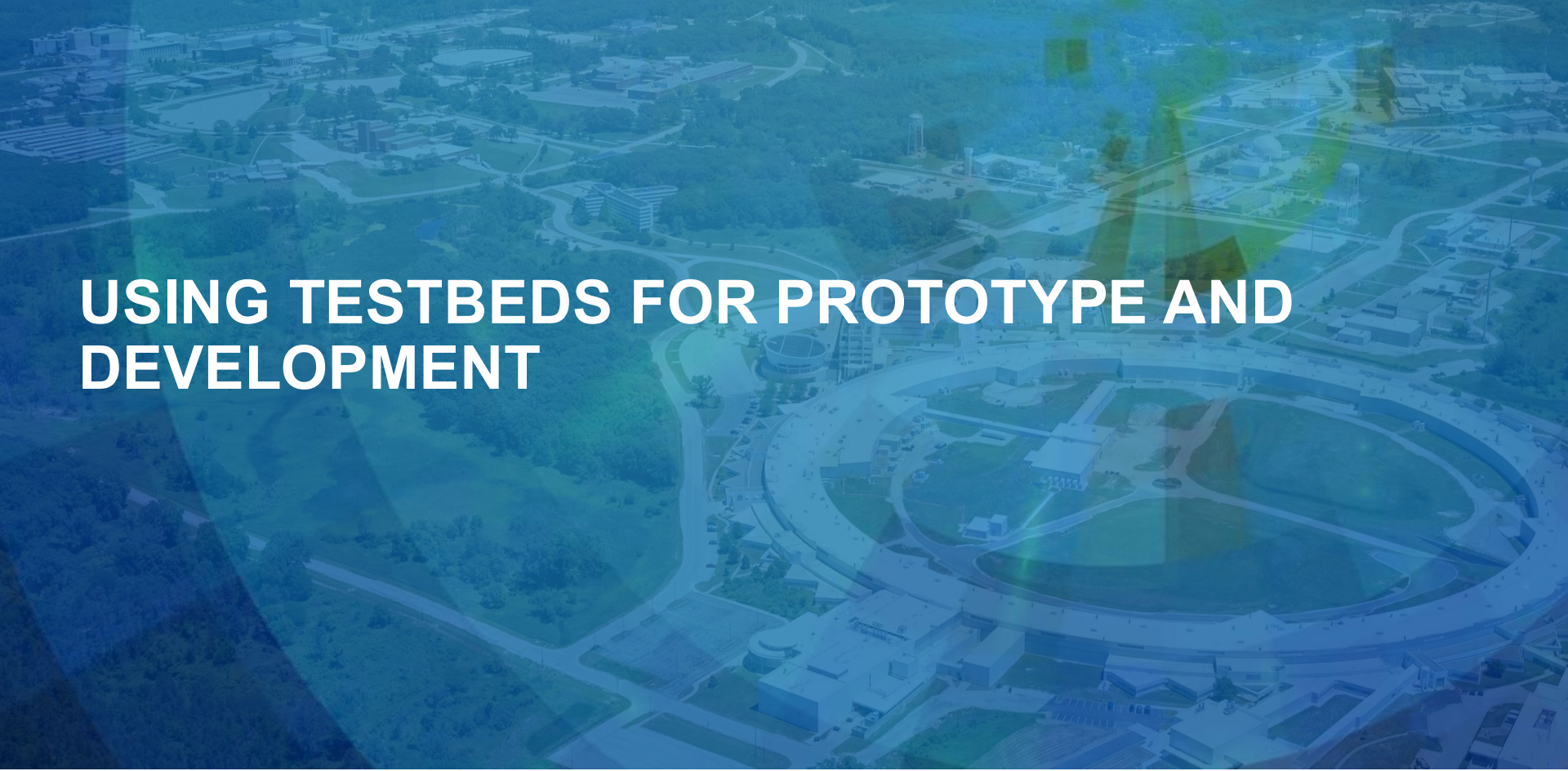


- **Third-party Streaming**
- **Secure Streaming**
- **General and Transparent Streaming**
- **Provisioned vs. Best-effort resources**

J. Chung, W. Zacherek, A. Wisniewski, Z. Liu, T. Bicer, R. Kettimuthu, and I. Foster, "SciStream: Architecture and toolkit for data streaming between federated science instruments," in Proceedings of the 31st International Symposium on High-Performance Parallel and Distributed Computing, HPDC '22, (New York, NY, USA), p. 185–198, Association for Computing Machinery, 2022.

Recent Releases

- Releases 1.0, 1.1 and 1.2
  - Integrated HAProxy, Nginx and Stunnel for production-grade forwarding quality
  - Integrated with Globus Auth
  - Data-channel authentication allowing ports to be open securely
  - Released Scistream container image
  - Implemented encryption for control and data-channel
  - TLS protocol and PSK authentication
  - Open Source Code https://github.com/scistream/scistream-proto
  - Documentation https://scistream.readthedocs.io/en/latest/scistream/



USING TESTBEDS FOR PROTOTYPE AND DEVELOPMENT

U.S. DEPARTMENT OF ENERGY
Argonne National Laboratory is a U.S. Department of Energy laboratory managed by UChicago Argonne, LLC.

Argonne NATIONAL LABORATORY

Prototype Implementation [HPDC'22]

SciStream Control Server (S2CS) and Data Server (S2DS)

S2CS:

- Implemented in Python using state machine
- Memory footprint is 10MB, data streaming request is completed in ~0.12 s while a release is completed in 0.003 s

S2DS (Implementation options):

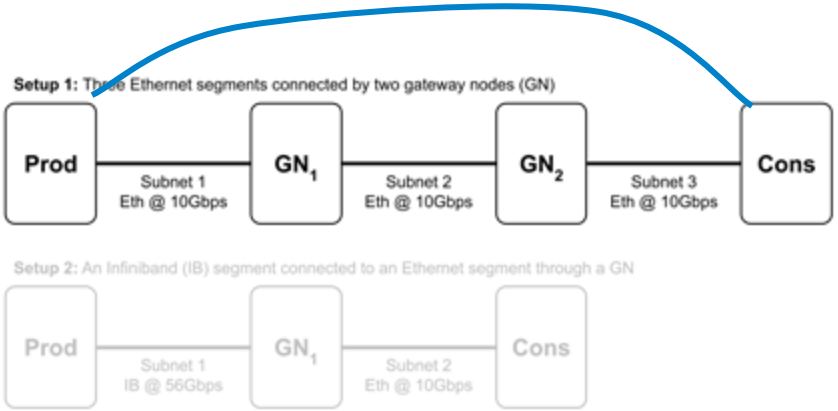
- L3 NAT or tunnels
- L4 Proxy (TCP or UDP)
- L7 (Application) Proxy

J. Chung, W. Zacherek, A. Wisniewski, Z. Liu, T. Bicer, R. Kettimuthu, and I. Foster, "SciStream: Architecture and toolkit for data streaming between federated science instruments," in Proceedings of the 31st International Symposium on High-Performance Parallel and Distributed Computing, HPDC '22, (New York, NY, USA), p. 185–198, Association for Computing Machinery, 2022.

HPDC'22 Evaluation

S2DS Implementation Options

Experimental setup on Chameleon



Methodology: compare an streaming application running on a SciStream-enabled infrastructure vs. the same application when producer and consumer have direct connectivity over the network (AKA No-SciStream)

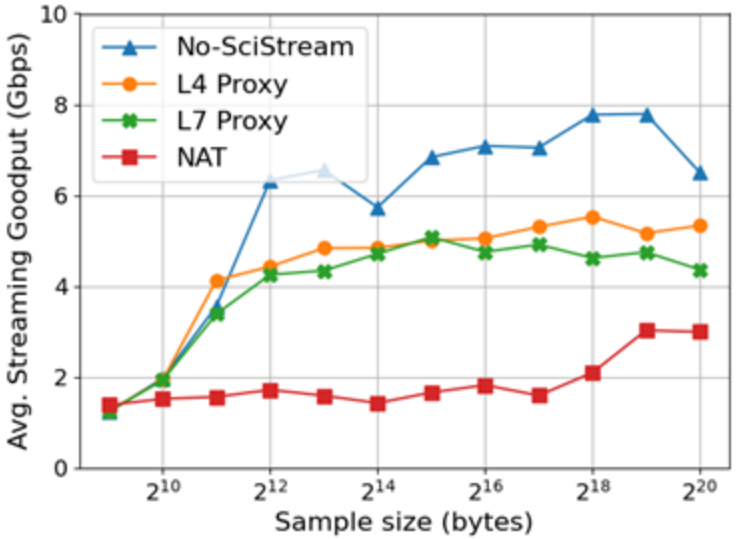
Measurements:

- Goodput evaluation
- Added latency and intermessage delay variation

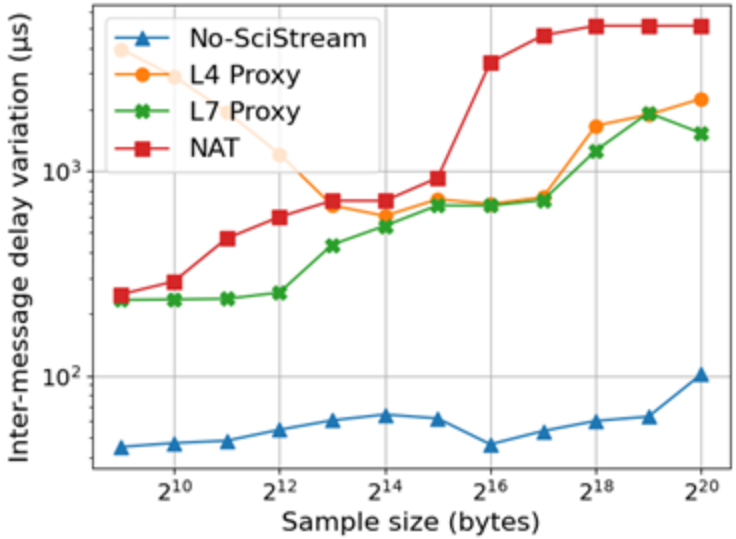
HPDC'22 Evaluation

SciStream over a real WAN (~30 ms RTT)

Streaming goodput as a function of sample size



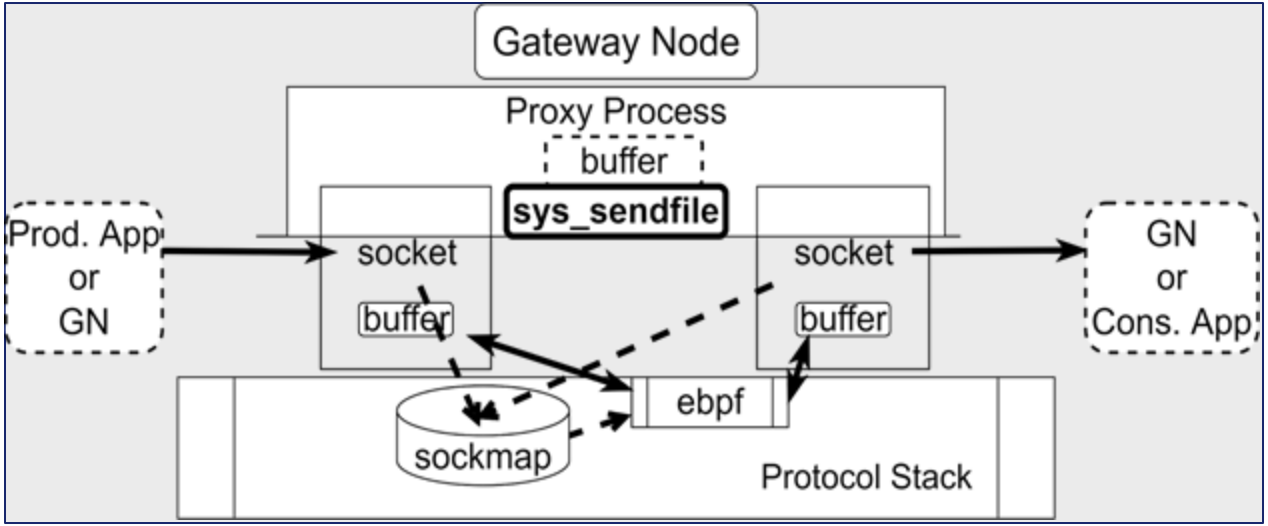
Intermessage delay variation as a function of sample size



U.S. DEPARTMENT OF ENERGY Argonne National Laboratory is a U.S. Department of Energy laboratory managed by UChicago Argonne, LLC.

Chameleon Argonne NATIONAL LABORATORY

Sockmap with eBPF on TCP Proxy - Overview [INDIS'22]



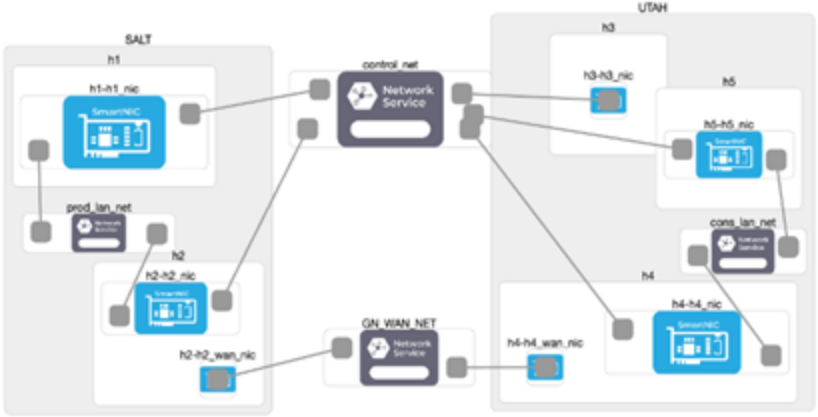
+ Sockmap will take over all the forwarding logic by **replacing the original `sk_data_ready` callback function**. The original `sk_data_ready` function, used in socket programming, is an interface between the protocol stack with proxy process, it wakes up user-space every time a packet arrives.
+ Sockmap **uses a 'Stream Parser' method to change the control logic of packets** to eBPF system. This shortens the process and achieves the redirection on kernel space only.
- Sockmap cannot change buffer size, if a large amount of data arrives, it may drop a lot of packets.
- Sockmap only supports TCP traffic, which is written by default in the kennel of Linux systems.

C. Qu, J. Chung, Z. Liu, T. Bicer, and R. Kettimuthu, "Evaluating SciStream (federated scientific data streaming architecture) on fabric," in 2022 IEEE/ACM International Workshop on Innovating the Network for Data-Intensive Science (INDIS), pp. 25–31, 2022.

OVERVIEW OF THE INDIS'22 EXPERIMENT

Evaluation of S2DS Implementation Options over real WAN

Experimental setup on FABRIC - one example on LAN (Salt to Utah)



*LAN (SALT to UTAH), Metro (MICH to STAR), Short WAN (STAR to DALL), WAN (UTAH to MAX), Long WAN(MICH to TACC)

Compared among 5 topologies with RTT of:

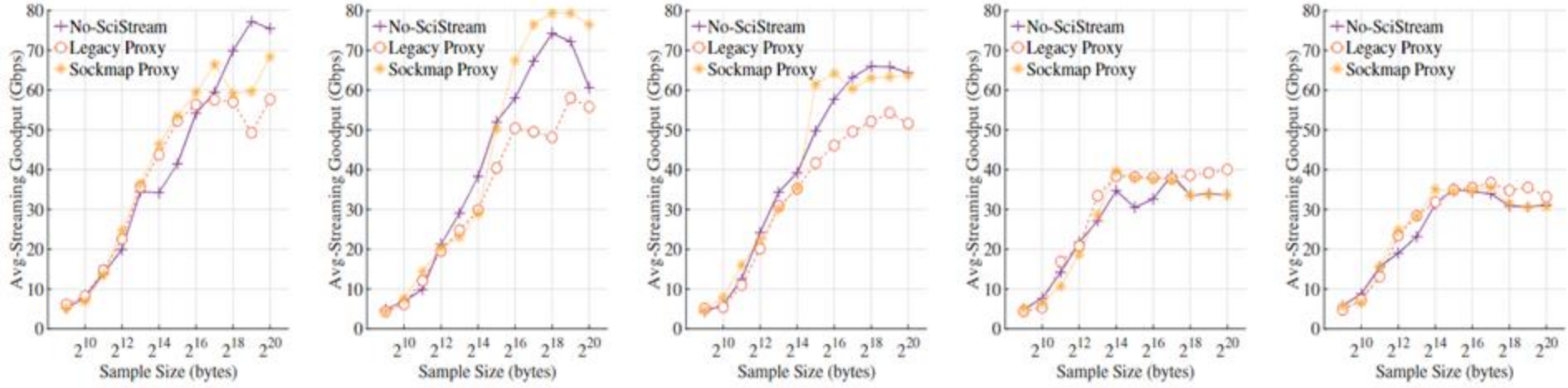
	LAN	Metro	Short WAN	WAN	Long WAN
Prod LAN	0.087	0.092	0.167	0.165	0.161
Cons LAN	0.100	0.105	0.179	0.148	0.160
GN WAN	0.253	5.293	23.998	57.848	143.370
Overall	**0.440**	**5.490**	**24.344**	**58.161**	**143.691**

Measurements:

- Goodput evaluation
- Perceived delay
- Jitter

EVALUATION [INDIS'22]

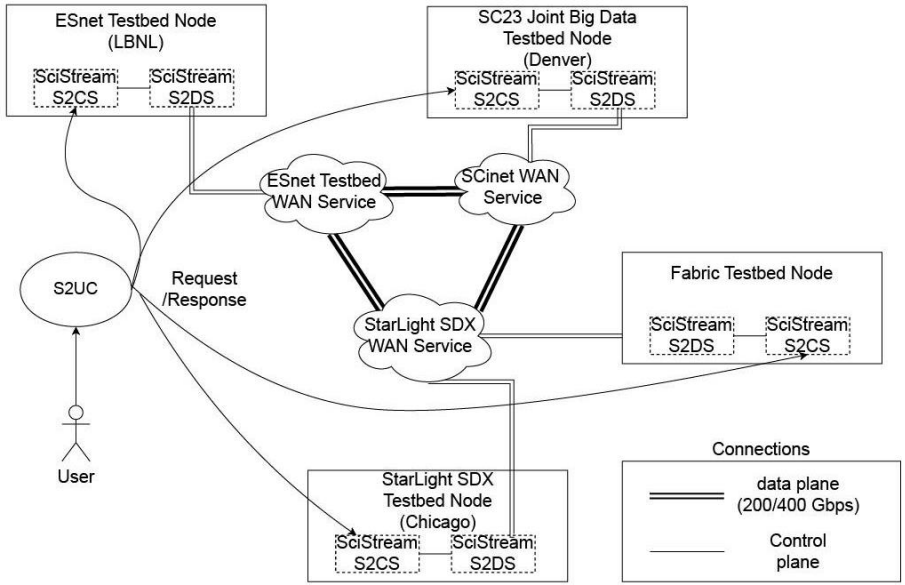
S2DS Goodput - Direct connection, Legacy SciStream and Sockmap



Goodput performance of two TCP implementations of `SciStream S2DS` (legacy and sockmap proxies) compared with the No-SciStream case over five network setups: (a) LAN, (b) metro, (c) short WAN, (d) WAN, and (e) long WAN.

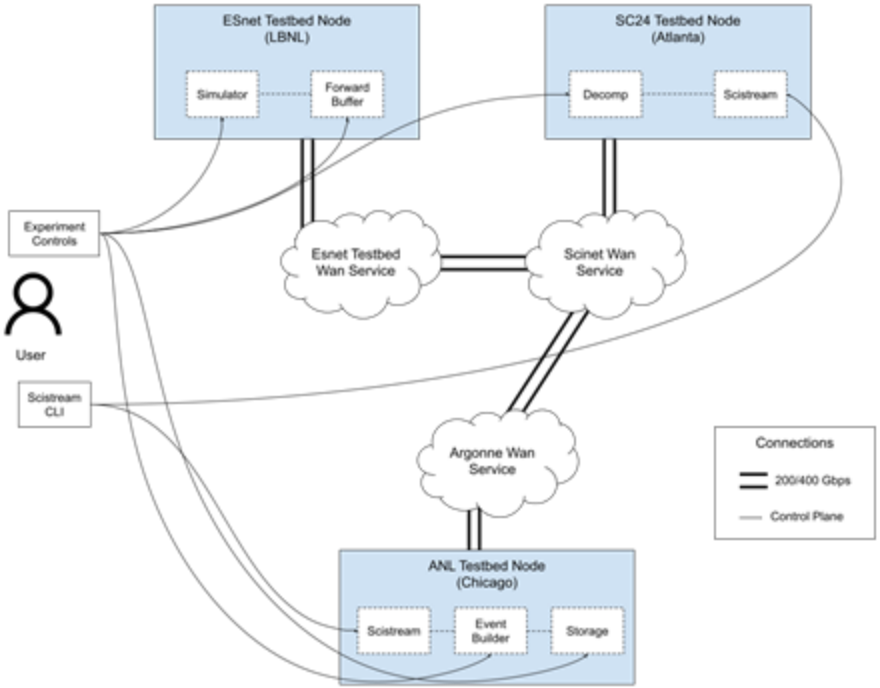
SCISTREAM DEMO AT SC23 IN COLLABORATION WITH ESNET

- **SciStream achieved up to 87 Gbps with 32 concurrent iperf streams**
  - SSH tunnels achieved 2.8 Gbps with similar concurrency
- **SciStream achieved up to 10 Gbps with 1 single iperf stream**
  - SSH tunnels achieved mere 300 Mbps



SciStream Demo at SC24 in collaboration with ESnet

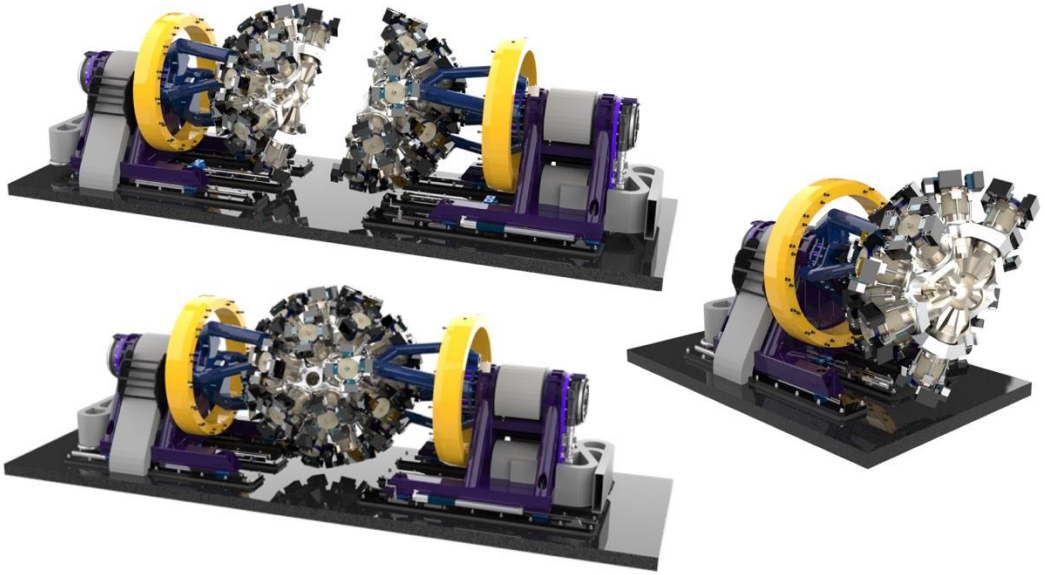
- No noticeable overhead without Data channel authentication
- Data channel authentication significantly faster than SSH tunnels
- Control integrations between DELERIA and Scistream (In Progress)



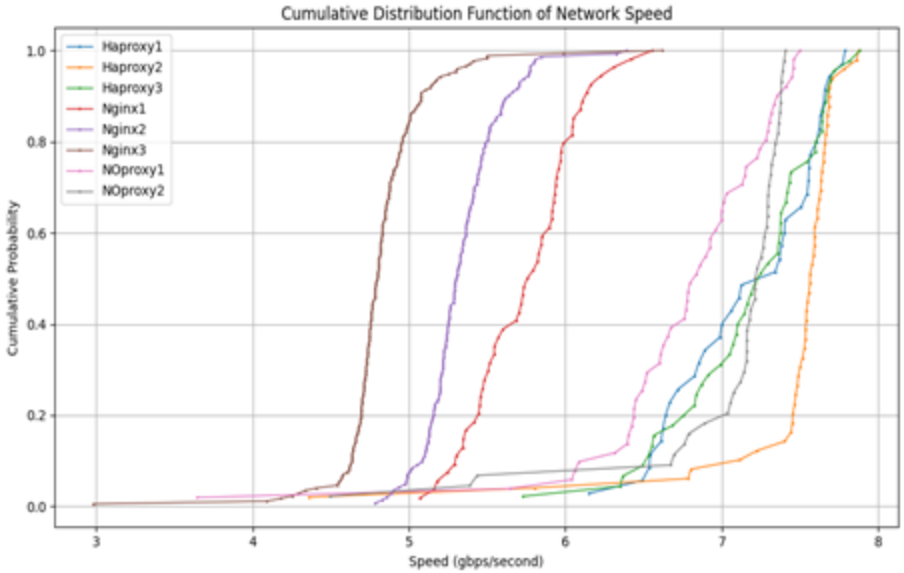
U.S. DEPARTMENT OF ENERGY Argonne National Laboratory is a U.S. Department of Energy laboratory managed by UChicago Argonne, LLC.

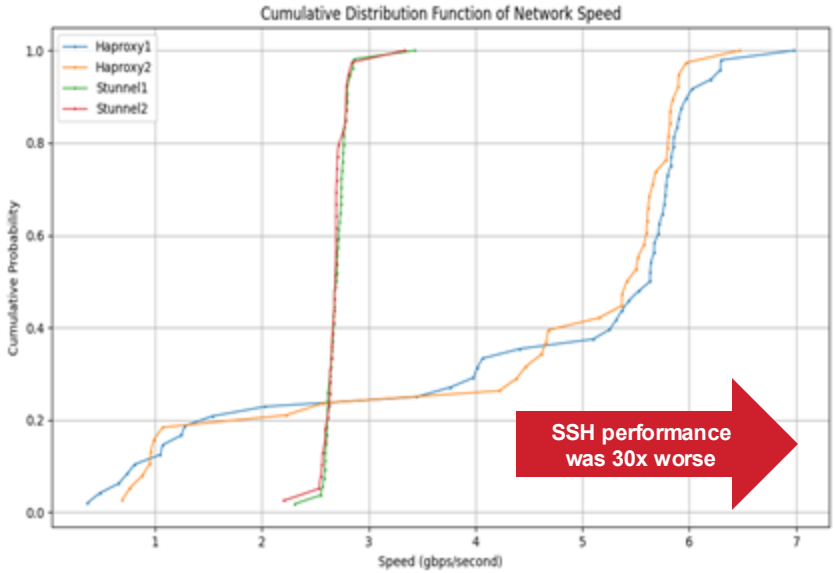
DELERIA WORKFLOW FOR GRETA

- Scientific instrument at Michigan State
- Real-time study of the energy and 3D position of gamma rays within atomic nuclei
- **DELERIA continuous streaming of live measurement data to compute nodes**
- Uses nanomsg for network communication



SciStream Demo at SC24 in collaboration with ESnet





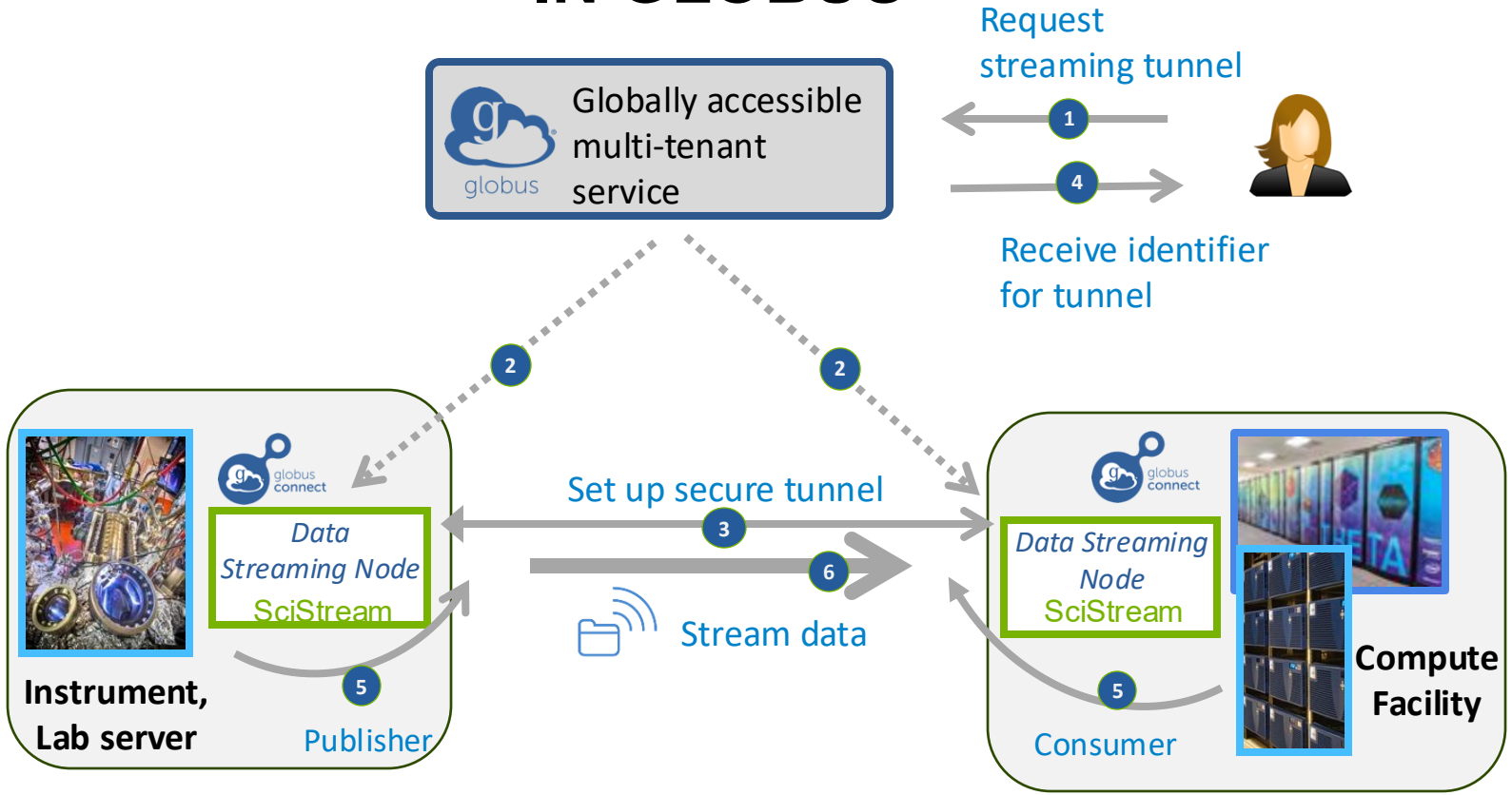
U.S. DEPARTMENT OF ENERGY Argonne National Laboratory is a U.S. Department of Energy laboratory managed by UChicago Argonne, LLC.
SCinet ESnet ENERGY SCIENCES NETWORK Argonne NATIONAL LABORATORY

OUTLOOK

U.S. DEPARTMENT OF ENERGY
Argonne National Laboratory is a U.S. Department of Energy laboratory managed by UChicago Argonne, LLC.

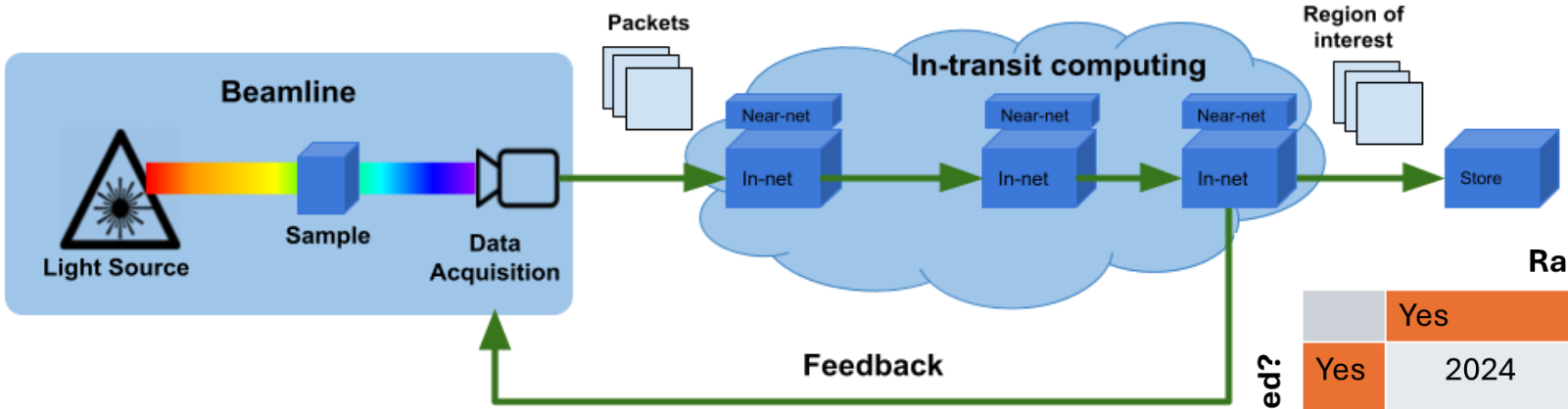
Argonne NATIONAL LABORATORY

SCISTREAM IS BEING PRODUCTIZED AS A SERVICE IN GLOBUS



COMPUTING IN TRANSIT TO IDENTIFY RARE EVENTS IN STREAMING SCIENTIFIC DATA

Collaboration with Ganesh C. Sankaran (RENCI) [IEEE ANTS 2024]



Can we identify change using header TCP checksum?

Rare ?

Identified?	Yes	No
Yes	2024	0
No	0	1006

Identification Accuracy

CLOSING REMARKS

International Testbeds for Data-Intensive Science and the Transition to Tbps-800G-400G WANs

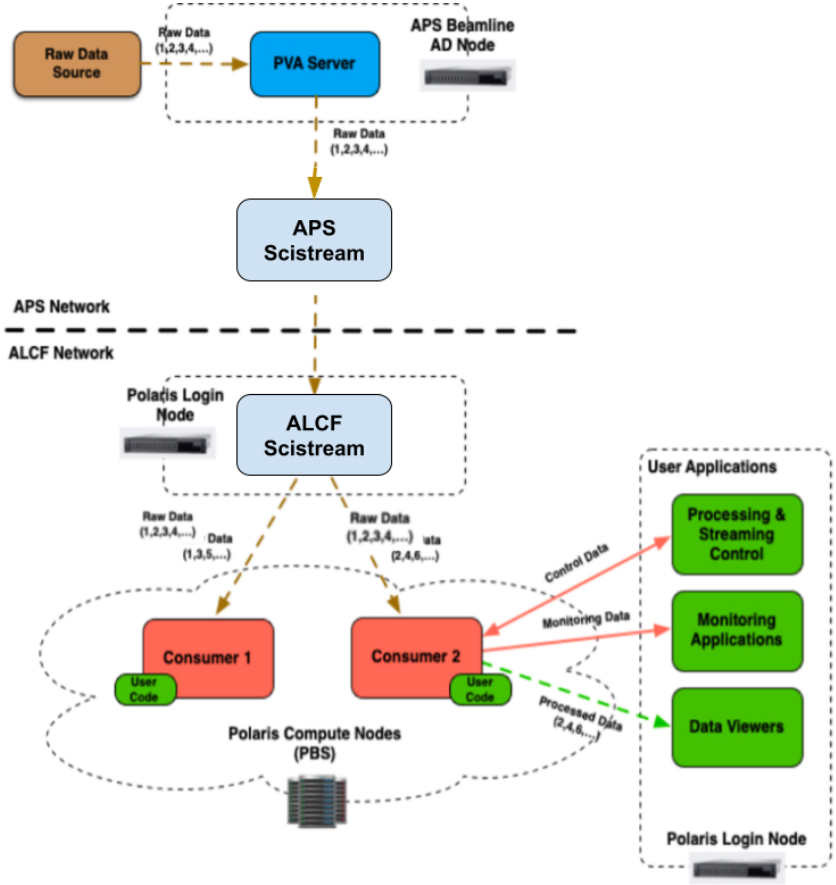
- SciStream is toolkit for enabling applications to **securely** stream data between scientific instruments and HPC
- We have prototyped, benchmarked, and demonstrated SciStream using multiple testbeds using 100Gbps links: **Chameleon Cloud, FABRIC, and SCinet**
- Despite the emergence of powerful AI at the Edge, the need to transition to Tbps-800G-400G WANs to support novel research remains important:
  - Programmable data planes
  - In- and near-network computing
  - Novel network architectures

Argonne NATIONAL LABORATORY

U.S. DEPARTMENT OF ENERGY

email: chungmiranda@anl.gov

EPICS DATA STREAMING FROM APS TO ALCF WITH SCISTREAM

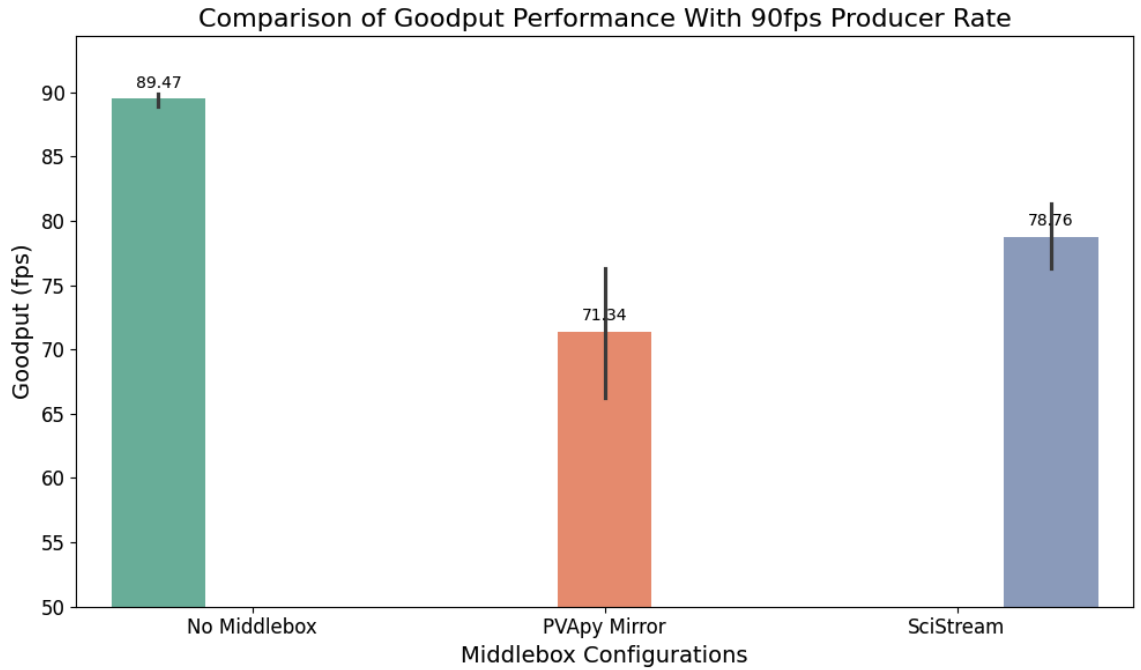


U.S. DEPARTMENT OF ENERGY Argonne National Laboratory is a U.S. Department of Energy laboratory managed by UChicago Argonne, LLC.

Argonne NATIONAL LABORATORY

EPICS DATA STREAMING FROM APS TO ALCF WITH SCISTREAM

Preliminary results



DIII-D

- National Fusion Facility, operated by GA for DOE since late 1980s
- Substantial influence on ITER (The Way), largest fusion experiment
- Make decisions about how to tune their experiment parameters between plasma shots
- Computing model for ITER, which will produce 1PB/day
- SciStream for enabling data streaming between ALCF and GA (In-progress)