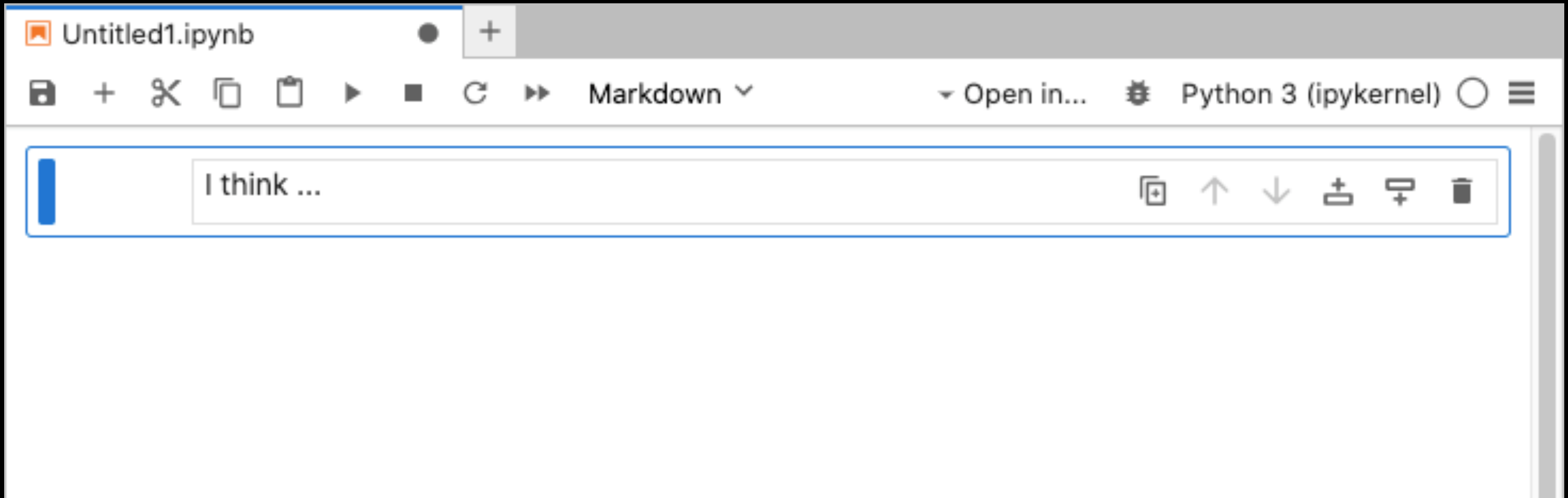
A Duality between Experiments and Tools

ILLINOIS TECH

Nik Sultana

KNIT10 – Mar 11-12, 2025

NSF CNS-2346499
URA #24-S-23


Untitled1.ipynb
Markdown
Open in...
Python 3 (ipykernel)
I think ...

I think

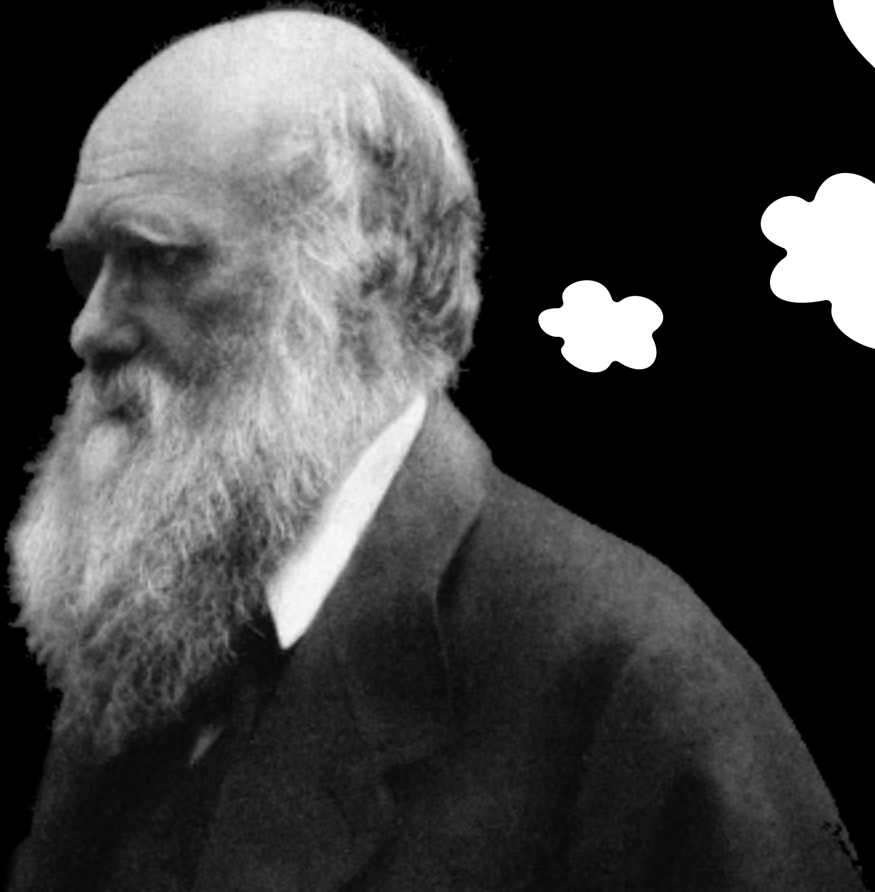
Case must be that one generation then should be as many living as now. To do this & to have many species in same genus (as is) requires extinction.

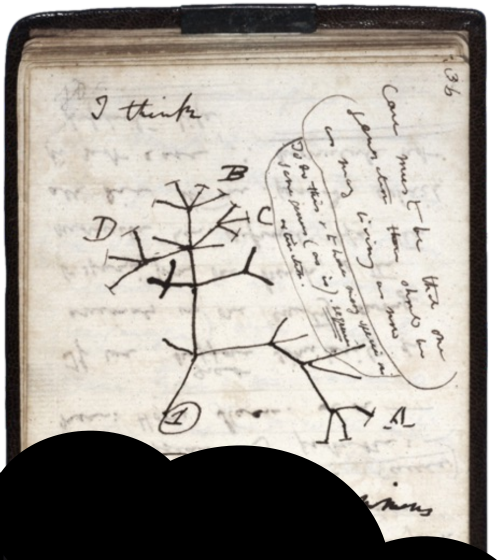
Thus between A & B. immens gap of relation. C & B. the finest gradation, B & D rather greater distinction Thus genera would be formed. — bearing relation

I think

Case must be that one generation then should be as many living as now. To do this & to have many species in same genus (as is) requires extinction.

Thus between A & B immense gap of relation. C & B the finest gradation, B & D rather greater distinction. Thus genera would be formed. — bearing relation




I think

Tab. OPTICKS.

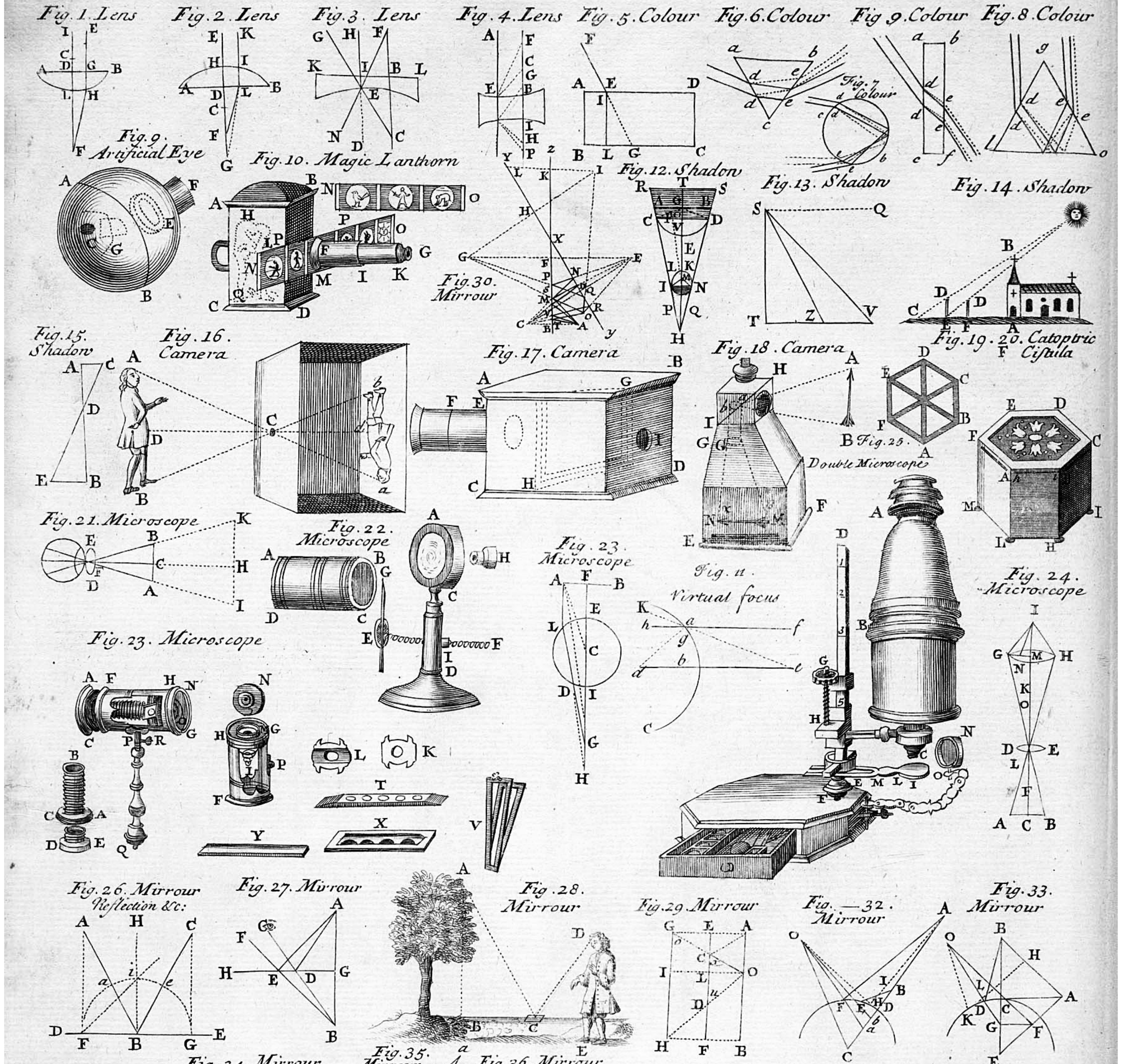
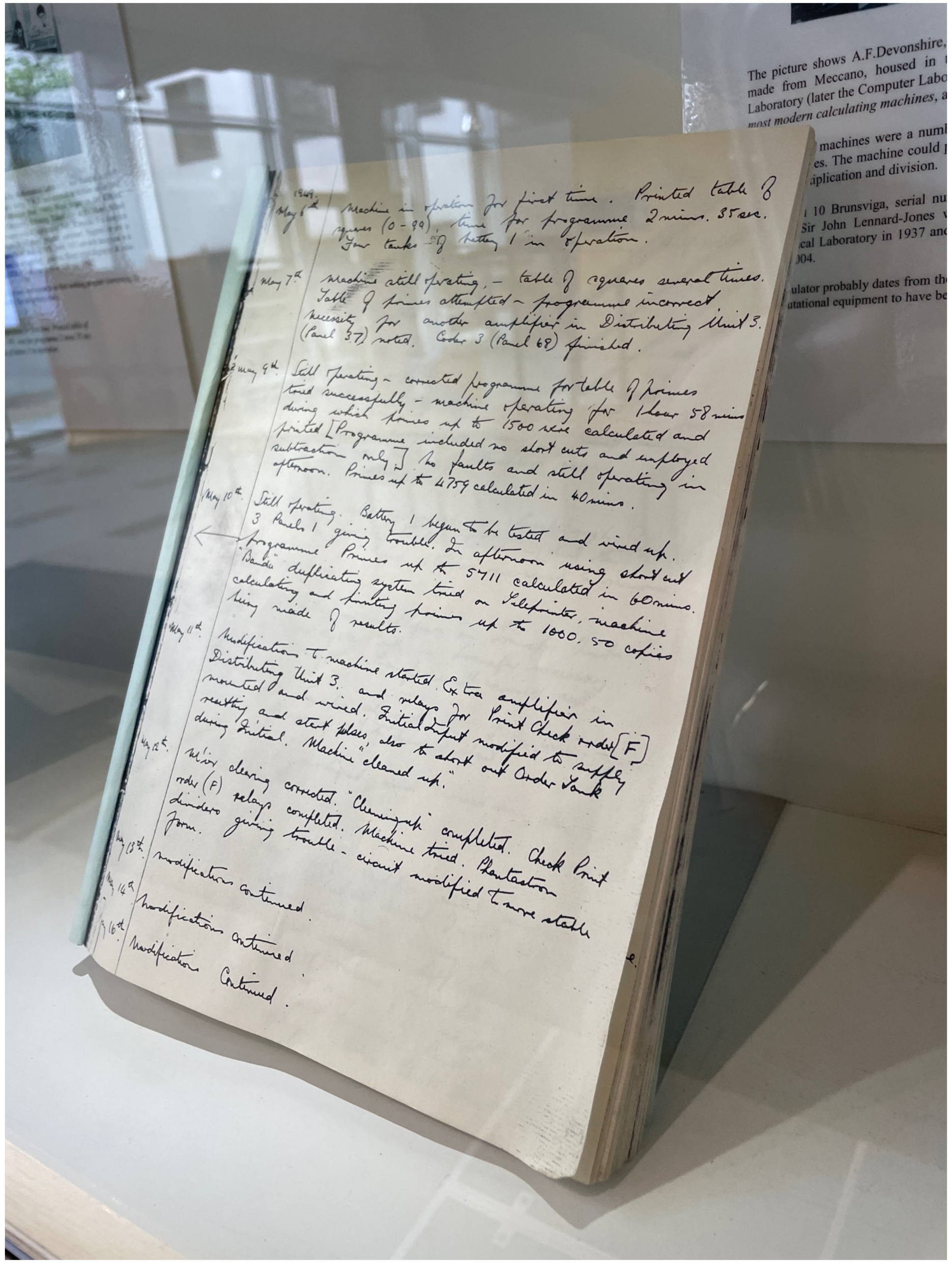


Fig. 1. Lens
Fig. 2. Lens
Fig. 3. Lens
Fig. 4. Lens
Fig. 5. Colour
Fig. 6. Colour
Fig. 9. Colour
Fig. 8. Colour
Fig. 7 Colour
Fig. 9. Artificial Eye
Fig. 10. Magic Lanthorn
Fig. 12. Shadow
Fig. 13. Shadow
Fig. 14. Shadow
Fig. 30. Mirrour
Fig. 15. Shadow
Fig. 16. Camera
Fig. 17. Camera
Fig. 18. Camera
Fig. 19. 20. Catoptric Cistula
Fig. 25
Double Microscope
Fig. 21. Microscope
Fig. 22. Microscope
Fig. 23. Microscope
Fig. 11.
Virtual focus
Fig. 24. Microscope
Fig. 23. Microscope
Fig. 26. Mirrour Reflection &c:
Fig. 27. Mirrour
Fig. 28. Mirrour
Fig. 29. Mirrour
Fig. 32. Mirrour
Fig. 33. Mirrour
Fig. 35.

Tool-building

- Includes instruments, machines, components.
- Developing techniques and knowledge.
- Developing workforce.
- Cottage industry, leading to ecosystem.


Main Injector Particle Accelerator Tunnel
Recycler Particle Storage Ring
Full-scale Model
Main Injector Particle Accelerator



1949.

May 6th. Machine in operation for first time. Printed table of squares (0 – 99), time for programme 2 mins. 35 sec. Four tanks of battery 1 in operation.

May 7th. Machine still operating. – table of squares several times. Table of primes attempted – programme incorrect. Necessity for another amplifier in Distributing Unit 3. (Panel 37) noted. Coder 3 (Panel 68) finished.

May 9th. Still operating – corrected programme for table of primes tried successfully – machine operating for 1 hour 58 mins during which primes up to 1500 were calculated and printed [Programme included no short cuts and employed subtraction only.] No faults and still operating in afternoon. Primes up to 4759 calculated in 40 mins.

May 10th. Still operating. Battery 1 begun to be tested and wired up. 3 Panels 1 giving trouble. In afternoon using short cut "programme" Primes up to 5711 calculated in 60 mins. "Banda" duplicating system tried on Teleprinter, machine calculating and printing primes up to 1000. 50 copies being made of results.

May 11th. Modification to machine started. Extra amplifier in Distributing Unit 3. and relays for Print Check order [F] mounted and wired. Initial Input modified to supply resetting and start pulses, also to short out Order Tank during Initial. Machine "cleaned up."

May 12th. M/c clearing corrected. "Clean-up" completed. Check Print order (F) relays completed. Machine tried. Phantastron dividers giving trouble – circuit modified to more stable form.

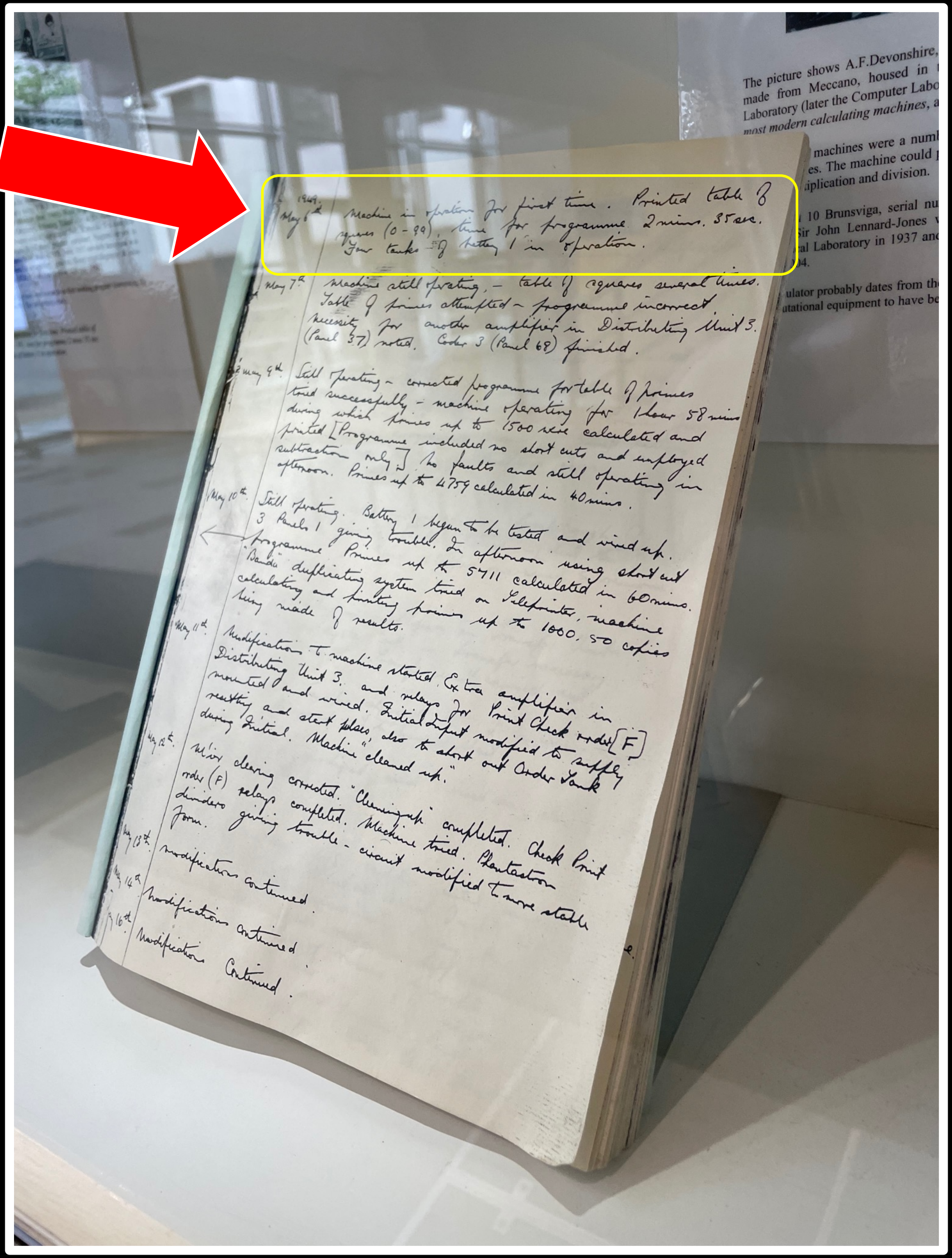
May 13th. Modifications continued.

May 14th. Modifications continued.

May 16th. Modifications continued.

May 6th 1949

Machine in operation for first time.
Printed table of squares (0-99).
Time for programme 2 mins 35 secs.
Four tanks of battery 1 in operation.

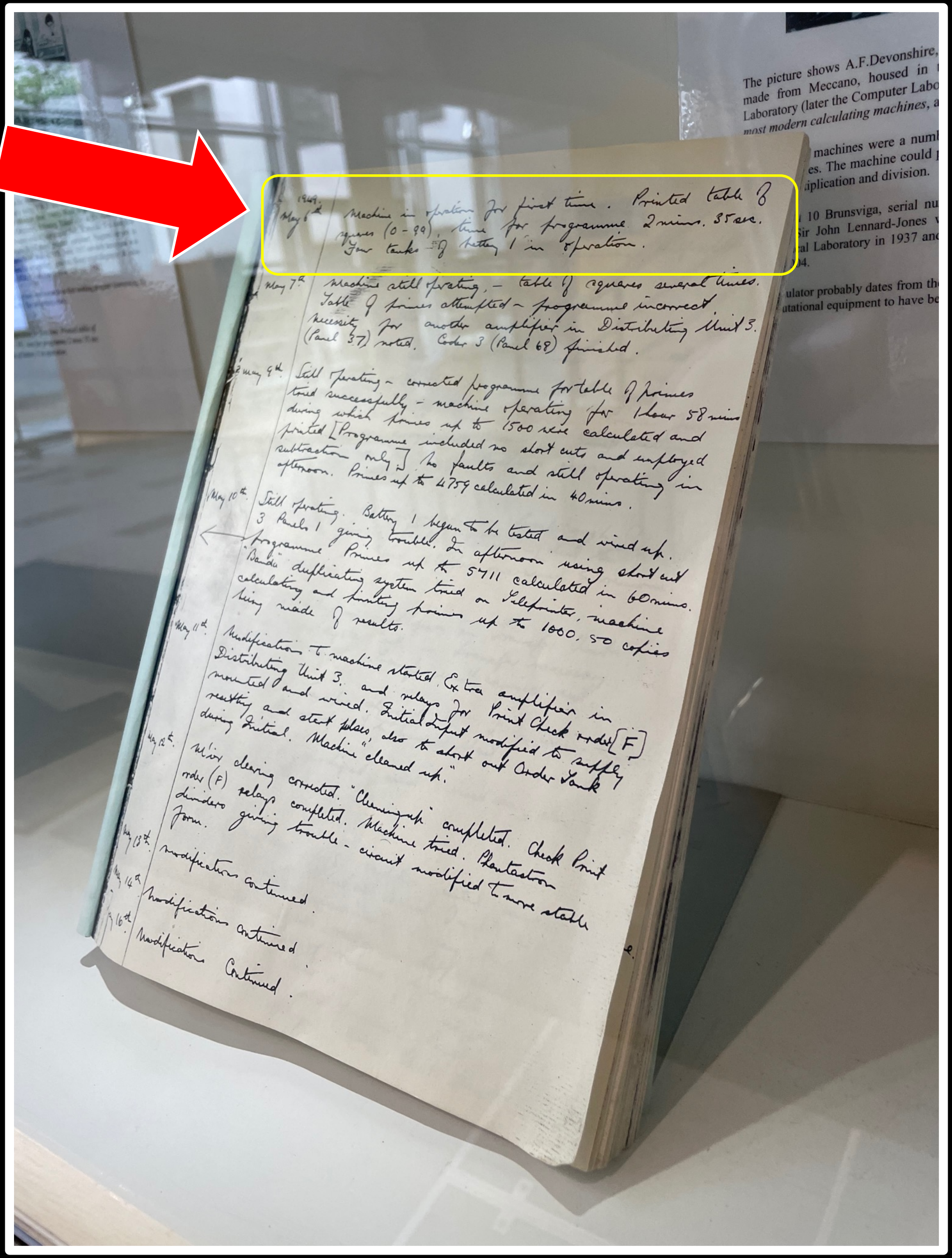


May 6th 1949

Machine in operation for first time.
Printed table of squares (0-99).
Time for programme 2 mins 35 secs.
Four tanks of battery 1 in operation.

```
nsultana$ cal May 1949
      May 1949
Su Mo Tu We Th Fr Sa
 1  2  3  4  5  6  7
 8  9 10 11 12 13 14
15 16 17 18 19 20 21
22 23 24 25 26 27 28
29 30 31

nsultana$
```

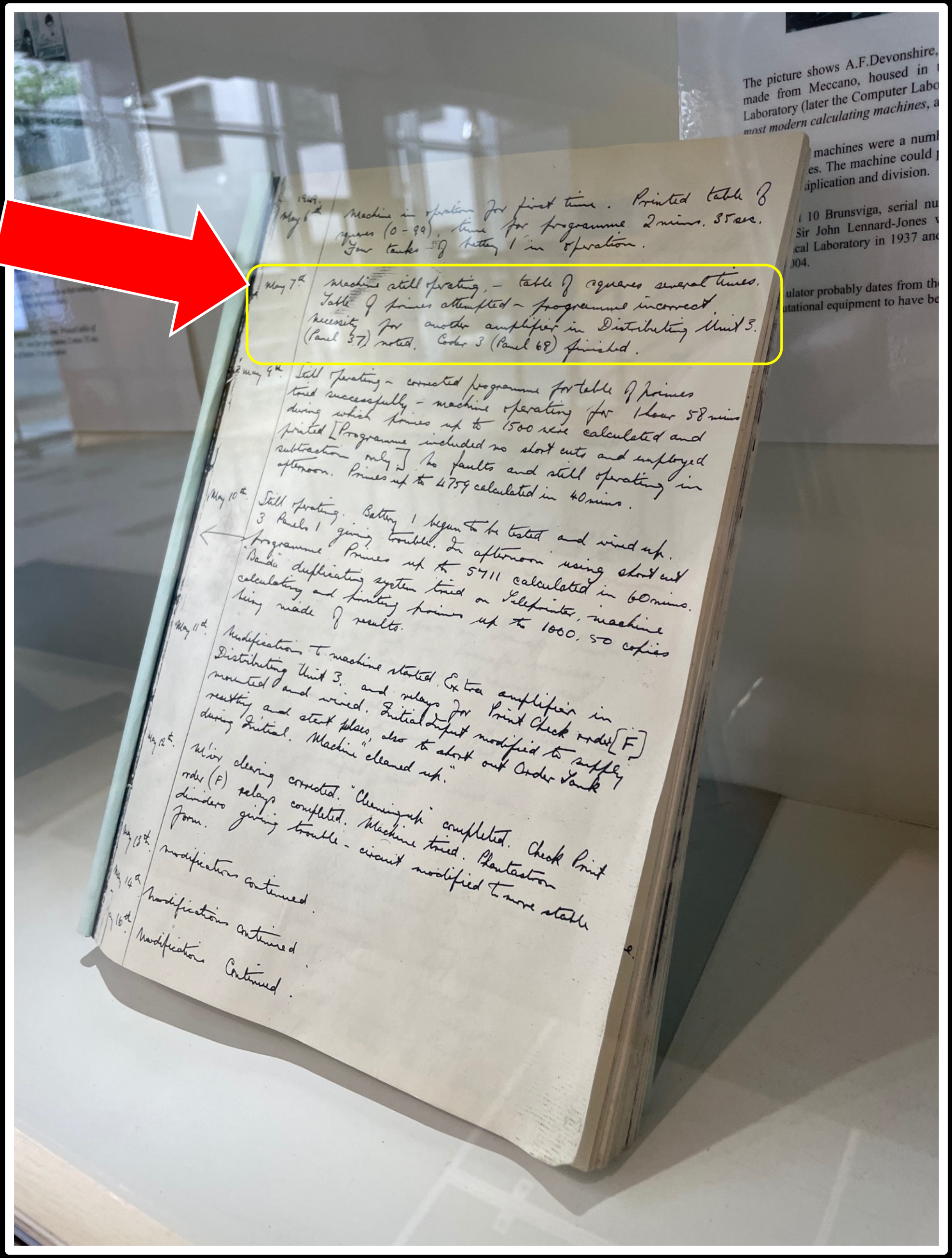


May 6th 1949

Machine in operation for first time.
Printed table of squares (0-99).
Time for programme 2 mins 35 secs.
Four tanks of battery 1 in operation.

May 7th 1949

Machine still operating – table of squares several times.
Table of primes attempted – programme incorrect.
Necessity for another amplifier in Distributing Unit 3
(Panel 37) noted. Coder 3 (Panel 68) finished.



May 6th 1949

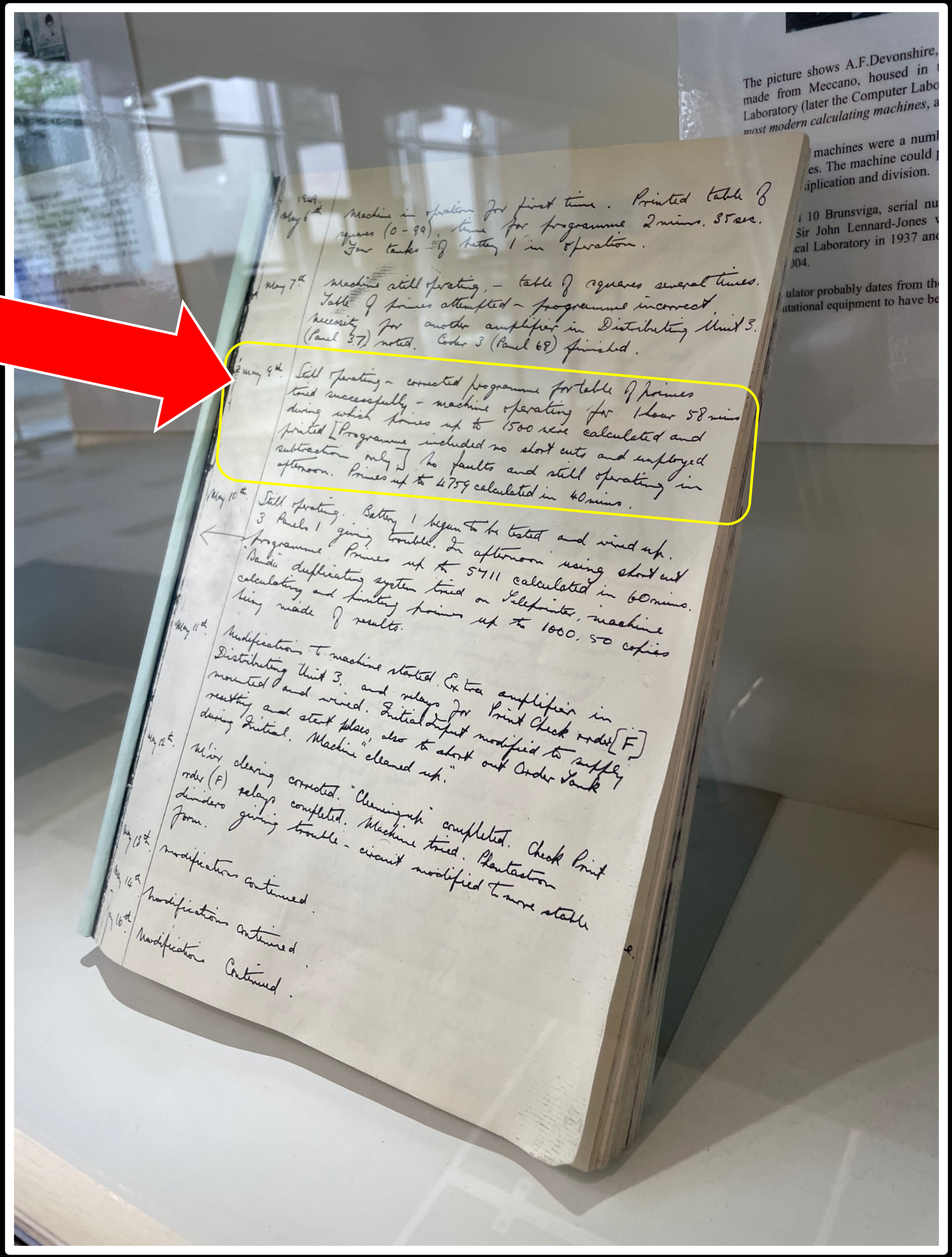
Machine in operation for first time.
Printed table of squares (0-99).
Time for programme 2 mins 35 secs.
Four tanks of battery 1 in operation.

May 7th 1949

Machine still operating – table of squares several times.
Table of primes attempted – programme incorrect.
Necessity for another amplifier in Distributing Unit 3 (Panel 37) noted. Coder 3 (Panel 68) finished.

May 9th 1949

Still operating – correct programme for table of primes tried successfully – machine operating for 1 hour 58 minutes during which primes up to 1500 were calculated and printed [Programme included no short cuts and employed subtraction Only.] No faults and still operating in the afternoon. Primes up to 4749 calculated in 40 minutes.



Idea

Idea

↑

Observations

Idea

Observations

Experiments

Idea

↑

Observations

↑

Experiments

↑

Resources

Idea

Observations

Experiments

Resources

Idea

↑

Observations

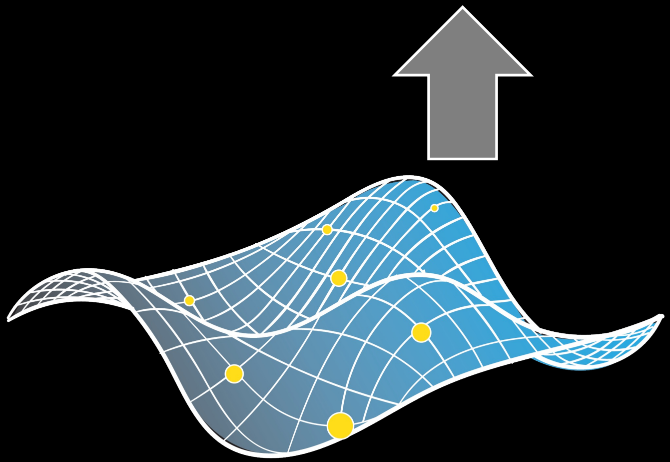
↑

Experiments

↑

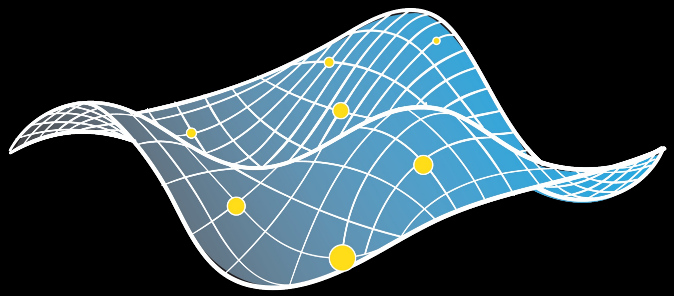
Infrastructure + Tools

Experiments



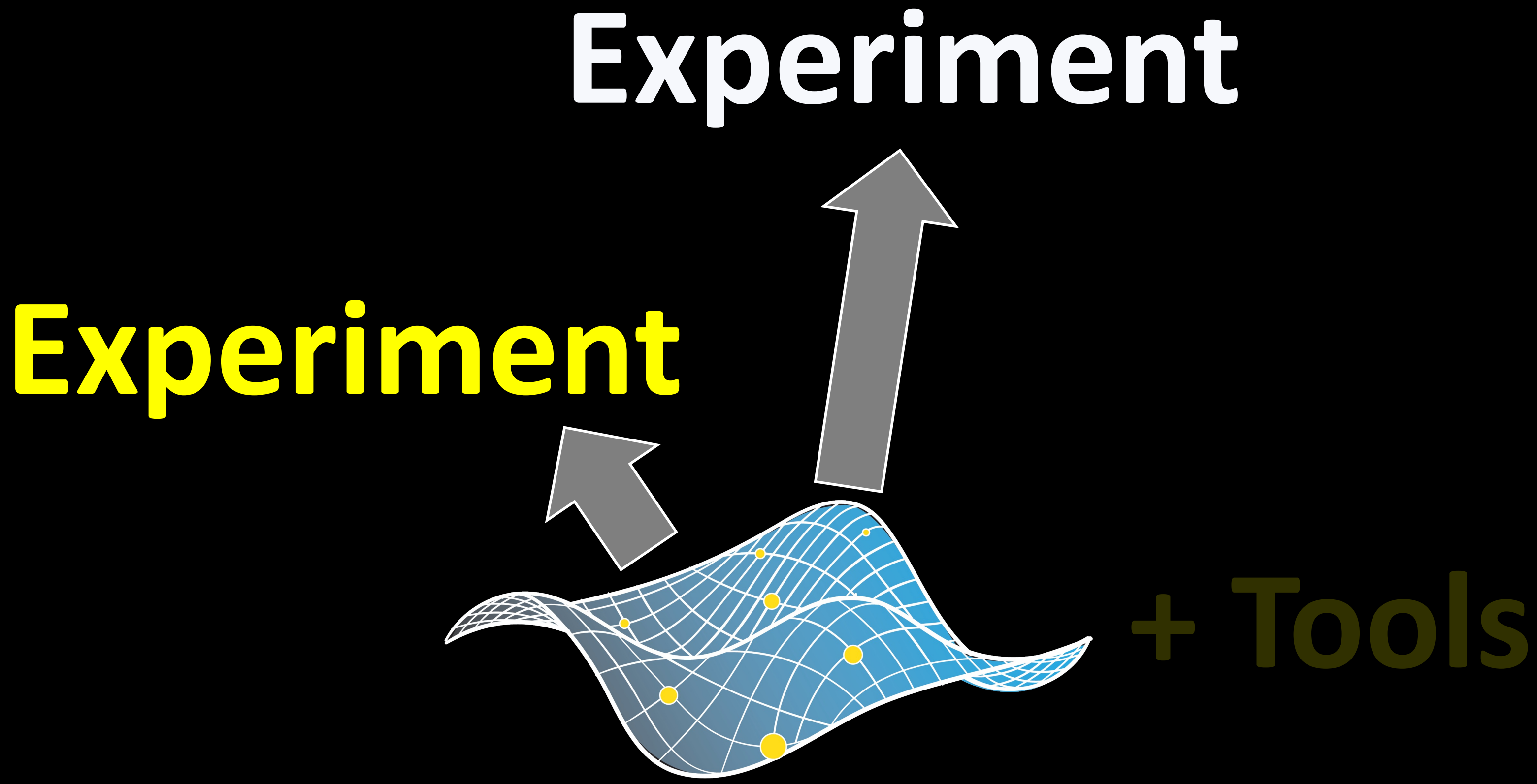
+ Tools

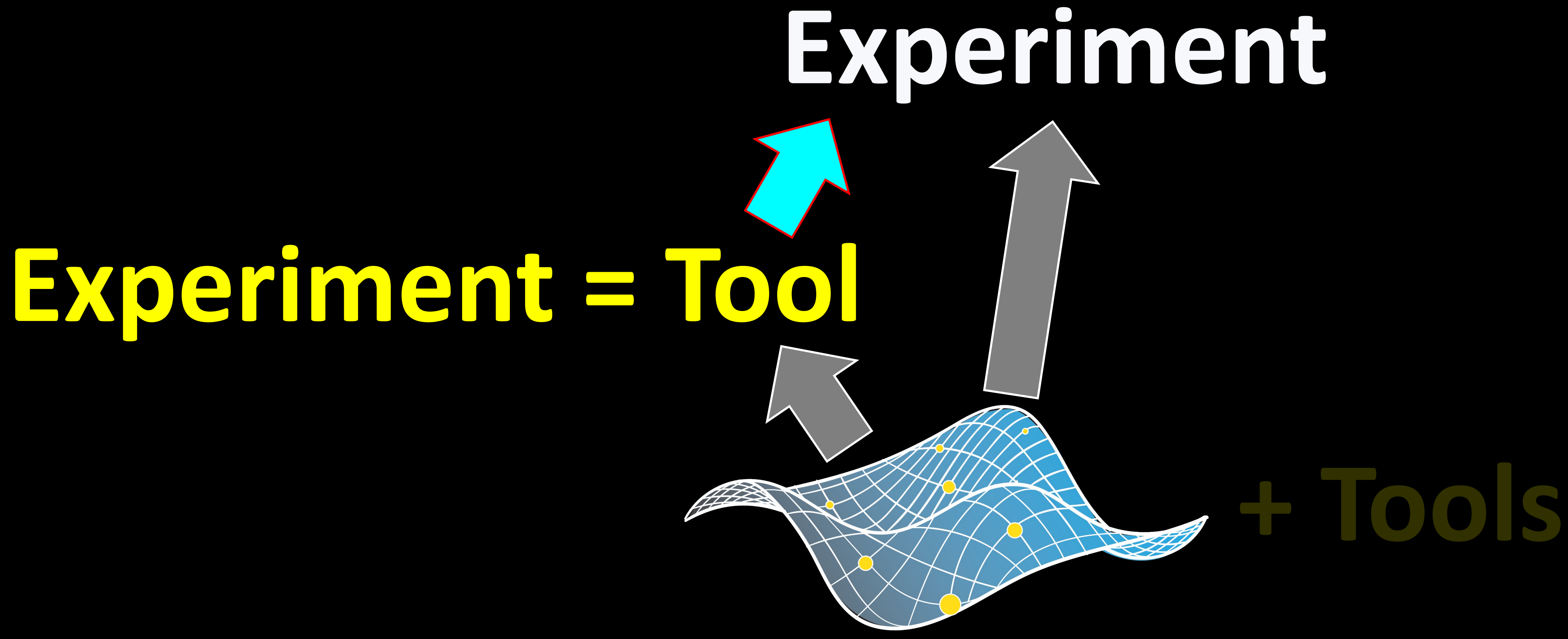
Experiments



+ Tools

{iperf, scapy,
DPDK, …}


Experiment
Experiment
+ Tools


Experiment
Experiment = Tool
+ Tools

A Duality between Experiments and Tools

Tools as Experiments

Experiments as Tools

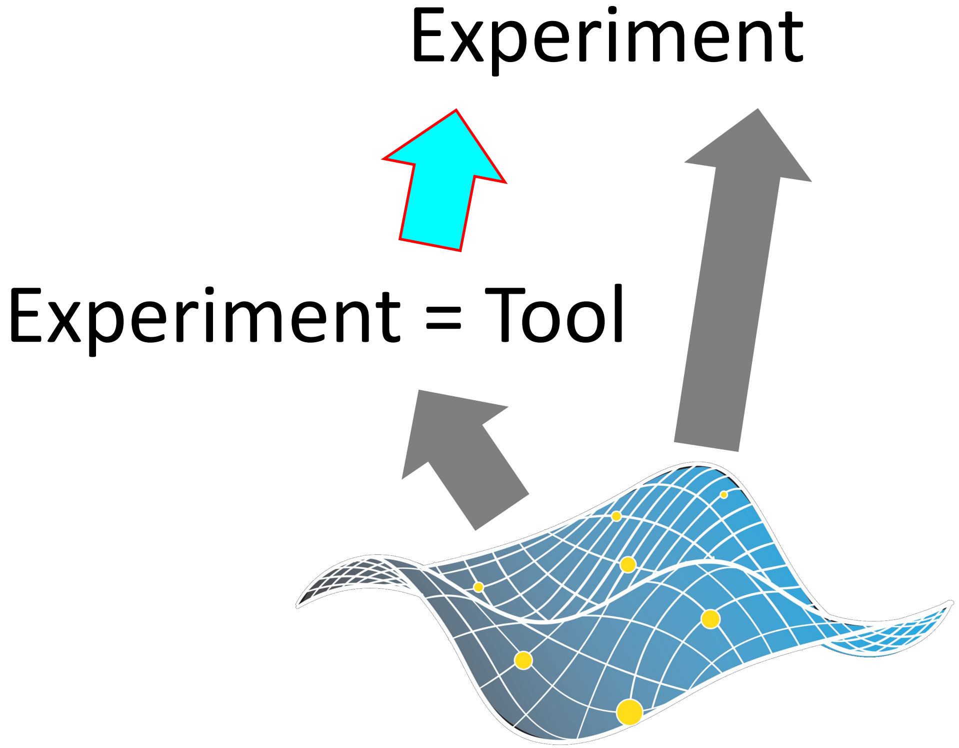
~~Tool-building~~ Experiment-building

- Includes instruments, ~~machines~~ experiments, components.
- Developing techniques and knowledge.
- Developing workforce.
- Cottage industry, leading to ecosystem.

Examples: Supporting other experiments

- **Network debugging**
  Also being used for teaching.

- **Network profiling**
  Producing experiment-level and testbed-level profiles.



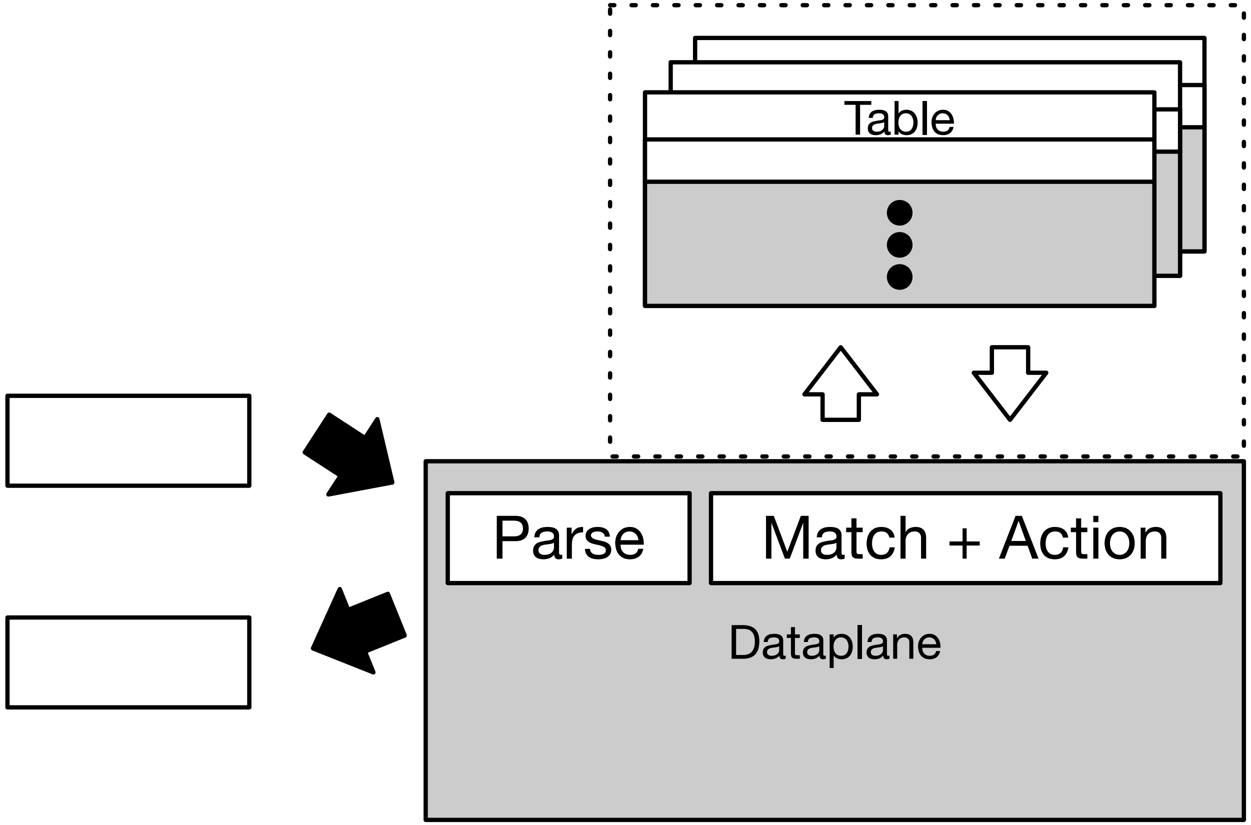
Network Debugging

- Goal is providing facilities for:
  - Self-reported configuration, observed configuration, provenance reasoning.
  - Mutation of experiments.
- Relevance to FABRIC: [Do attend Alexander’s tutorial tomorrow]
  debuggability,
  diagnosticability and
  reproducibility of Software-Defined Networking (SDN) experiments.
- CREASE project: https://crease.cs.iit.edu
  Causal **RE**asoning and **A**ttestation for **S**cientific **E**xperimentation

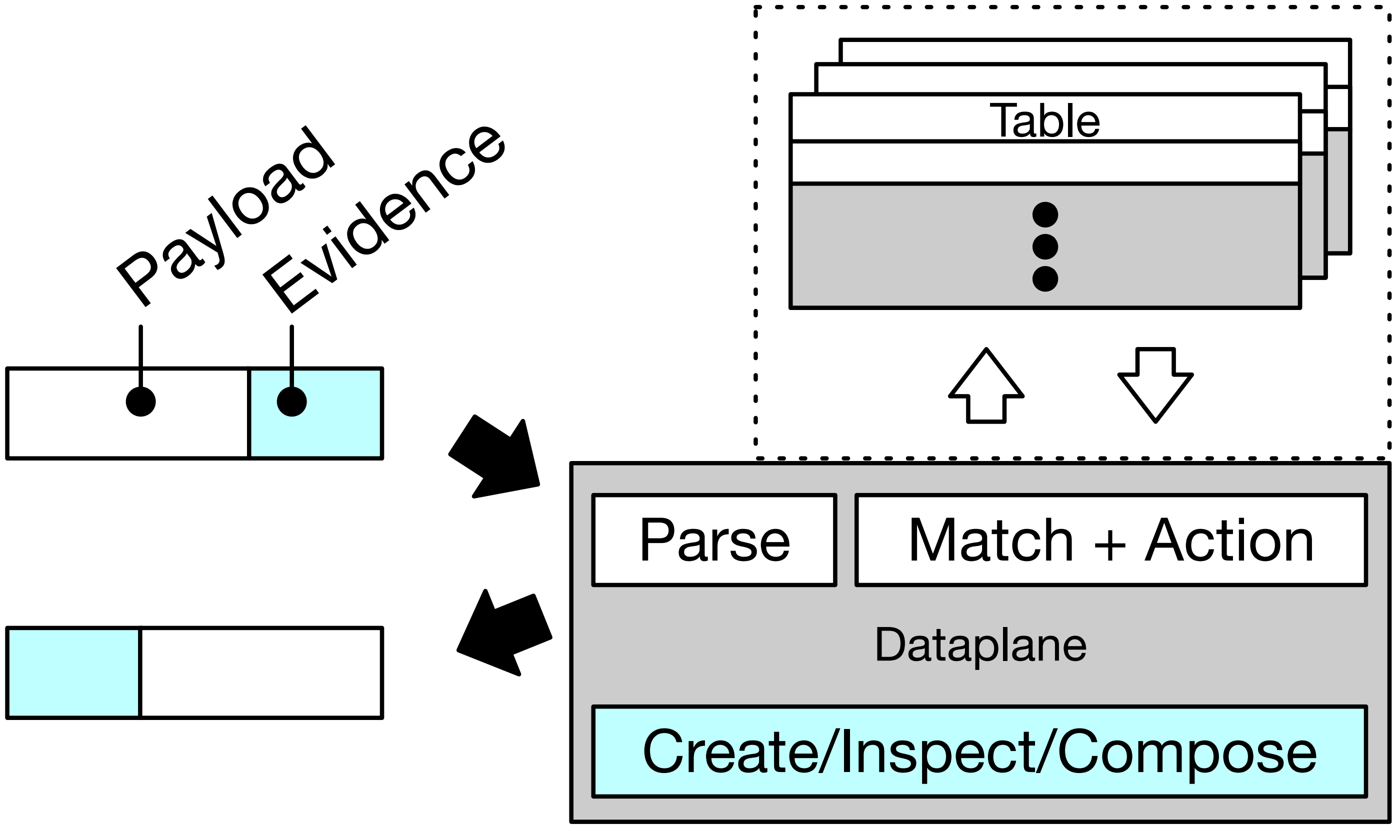
CREASE Background

- Origins in security: Early detection of misconfigurations and APTs, mitigating attack surface from programmability.
- Remote Attestation and Provenance Reasoning
- Motivations:
  Transparency and
  Accountability
- How to adapt these ideas in CREASE?

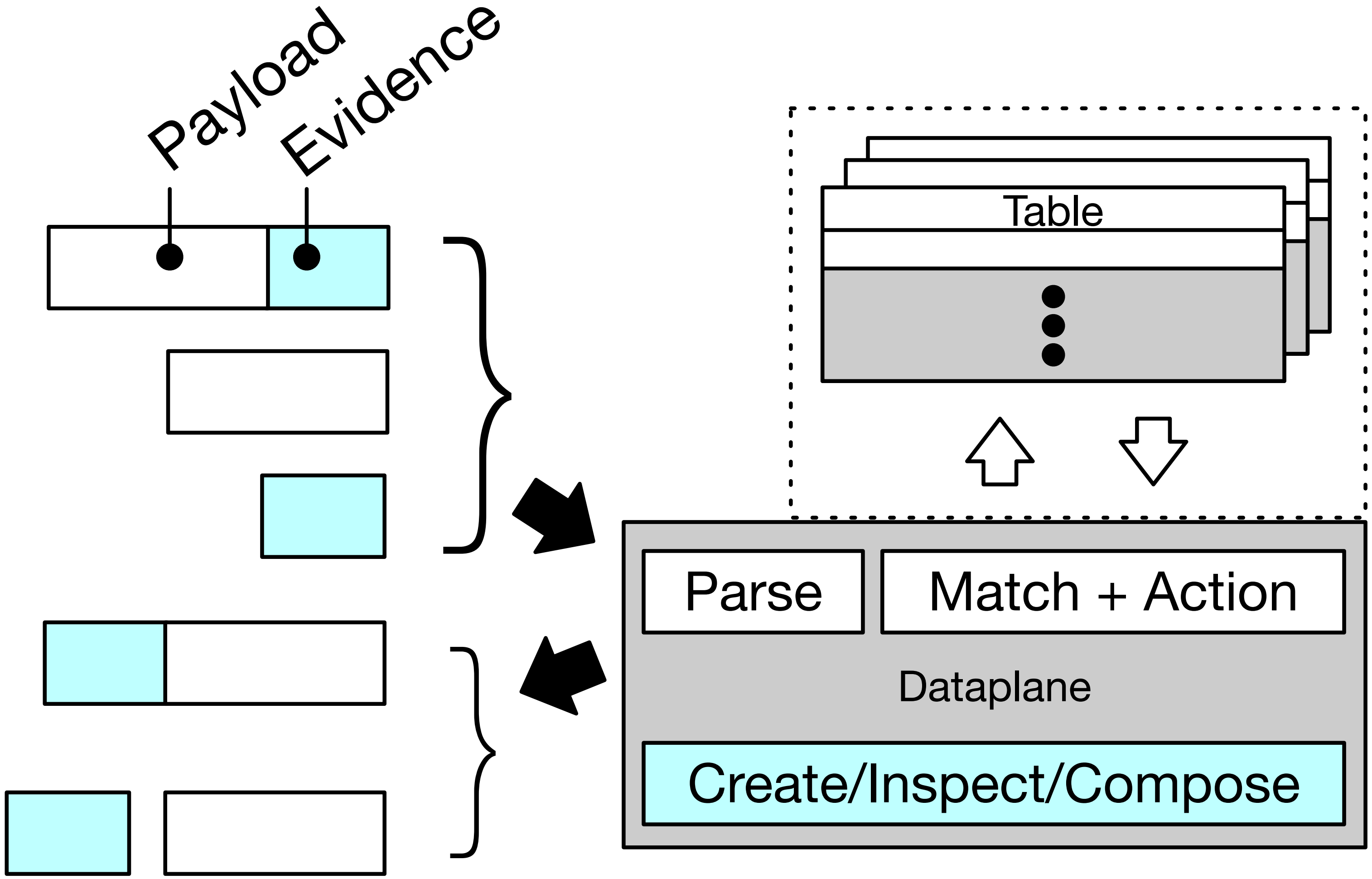
How CREASE works



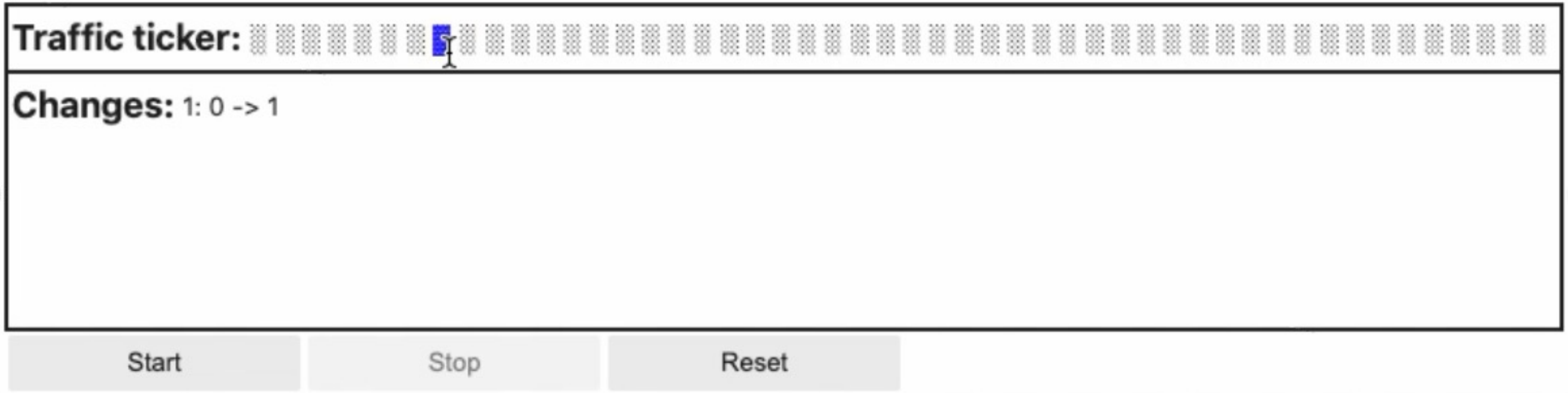
How CREASE works



How CREASE works



See the full Webinar!



Research

Shape-shifting Elephants: Multi-modal Transport for Integrated Research Infrastructure

Nik Sultana
Illinois Institute of Technology
USA

Yatish Kumar
ESnet
USA

Chin Guok
ESnet
USA

James B Kowalkowski
Fermilab
USA

Michael H L S Wang
Fermilab
USA

ABSTRACT

Data Acquisition (DAQ) workloads form an important class of scientific network traffic that by its nature (1) flows across different research infrastructure, including remote instruments and supercomputer clusters, (2) has ever-increasing throughput demands, and (3) has ever-increasing integration demands—for example, observations at one instrument could trigger a reconfiguration of another instrument. Today's DAQ transfers rely on UDP and (heavily tuned) TCP, but this is driven by convenience rather than suitability. The mismatch between Internet transport protocols and scientific workloads becomes more stark with the steady increase in link capacities, data generation, and integration across research infrastructure.

This position paper argues the importance of developing specialized transport protocols for DAQ workloads. It proposes a new transport feature for this kind of elephant flow: *multi-modality* involves the network actively configuring the transport protocol to change how DAQ flows are processed across different underlying networks that connect scientific research infrastructure. Multi-modality is a layering violation that is proposed as a pragmatic technique for DAQ transport protocol design. It takes advantage of programmable network hardware that is increasingly being deployed in scientific research infrastructure. The paper presents an initial evaluation through a pilot study that includes a Tofino2 switch and Alveo FPGA cards, and using data from a particle detector.

KEYWORDS

DAQ Workloads,Transport Protocol,Scientific Networking

ACM Reference Format:

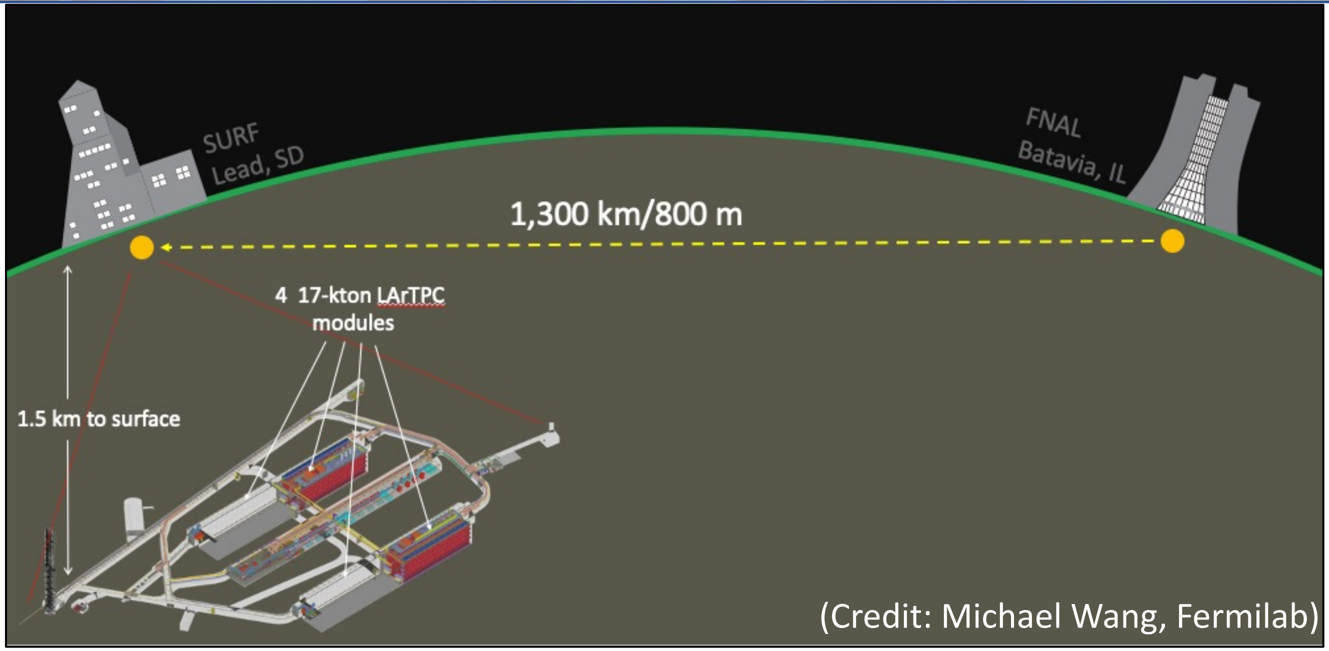
Nik Sultana, Yatish Kumar, Chin Guok, James B Kowalkowski, and Michael H L S Wang. 2024. Shape-shifting Elephants: Multi-modal Transport for Integrated Research Infrastructure. In *The 23rd ACM Workshop on Hot Topics in Networks (HOTNETS '24), November 18–19, 2024, Irvine, CA, USA.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3696348.3696855

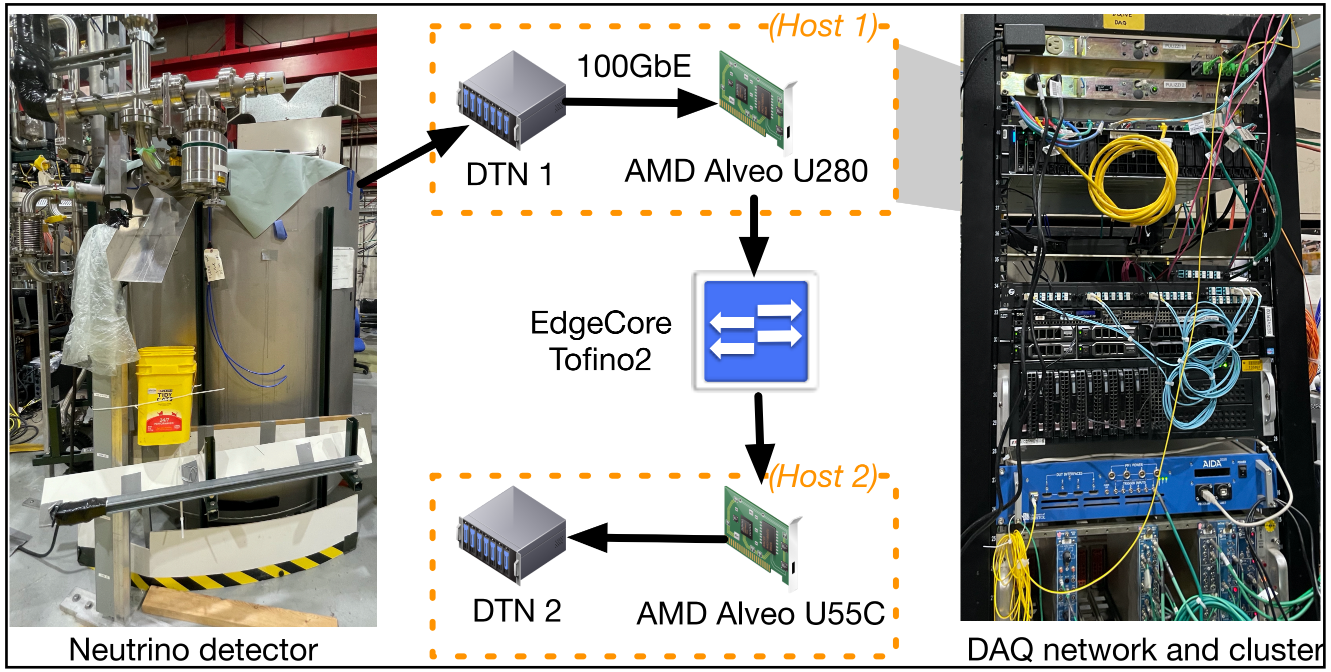
1 INTRODUCTION

As instruments become more precise, they produce more data. We can only process a fraction of data from large instruments. For example, the Large Hadron Collider (LHC) *generates* data at more than 600Tbps but around 40Tbps is currently *acquired* [52]—that is, read out of the instrument. Table 1 lists examples of recent and under-development experiments, and their *data acquisition* (DAQ) rates.

Large instruments—such as those in Table 1—have a *DAQ network* that connects the instruments' sensors to a small downstream processing facility. From there, data is transferred over other networks to reach large-scale processing facilities. Often, the DAQ network is an Ethernet built using commodity equipment. Traffic consists of elephant flows with a regular shape (size and arrival rate)—detailed further in §2.

In addition to producing more data, research infrastructure





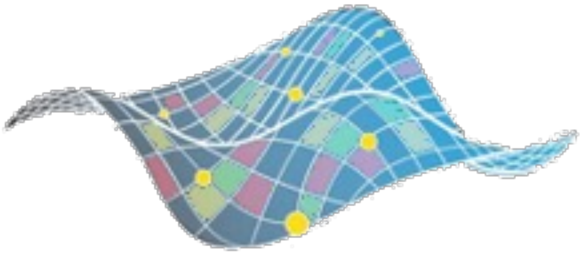
Teaching

- A key skill: problem solving.
- Learning fundamentals,
  but not reasoning from first principles all the time.
- Need tools to support us to learn problem solving,
  help gather information, organize it,
  and deduce from it.

Network Profiling

- Informally: what's on the network?
- Traffic composition over time:
  - **Header types**: What protocols are being used.
  - **Encapsulation patterns**: How is the network being used.
  - **Flows**
    - Number of packets
    - Packet sizes
    - Inter-packet delays
    - Other details – e.g., some/all flows contain frequent RSTs.
  - **Relative utilization**: should some types of traffic be prioritized?
  - Indicators of misconfiguration and compromise.

Goals of the Patchwork project



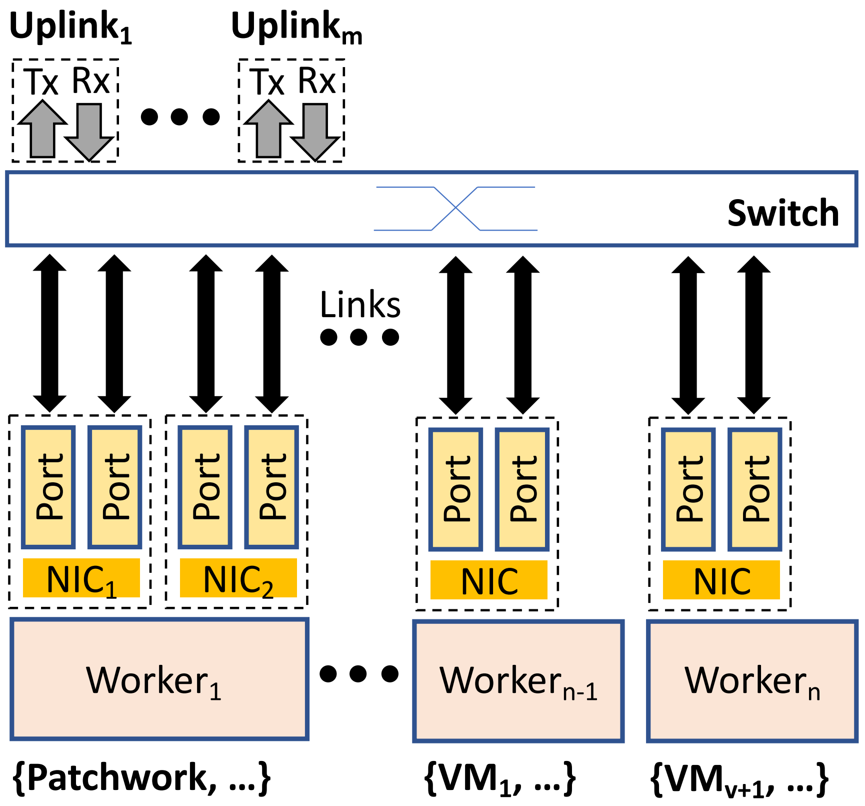
1) Providing a network profiler for FABRIC.
   Developing user-provided infrastructure for shared, federated testbeds.

   Two usage modes:
   1. Individual experiment profiling. [See Vaneshi's demo later today]
   2. Testbed-wide profiling.

2) Developing an initial profile of the entire FABRIC network.
   Ongoing work: building the profile periodically or on demand.

Outcomes

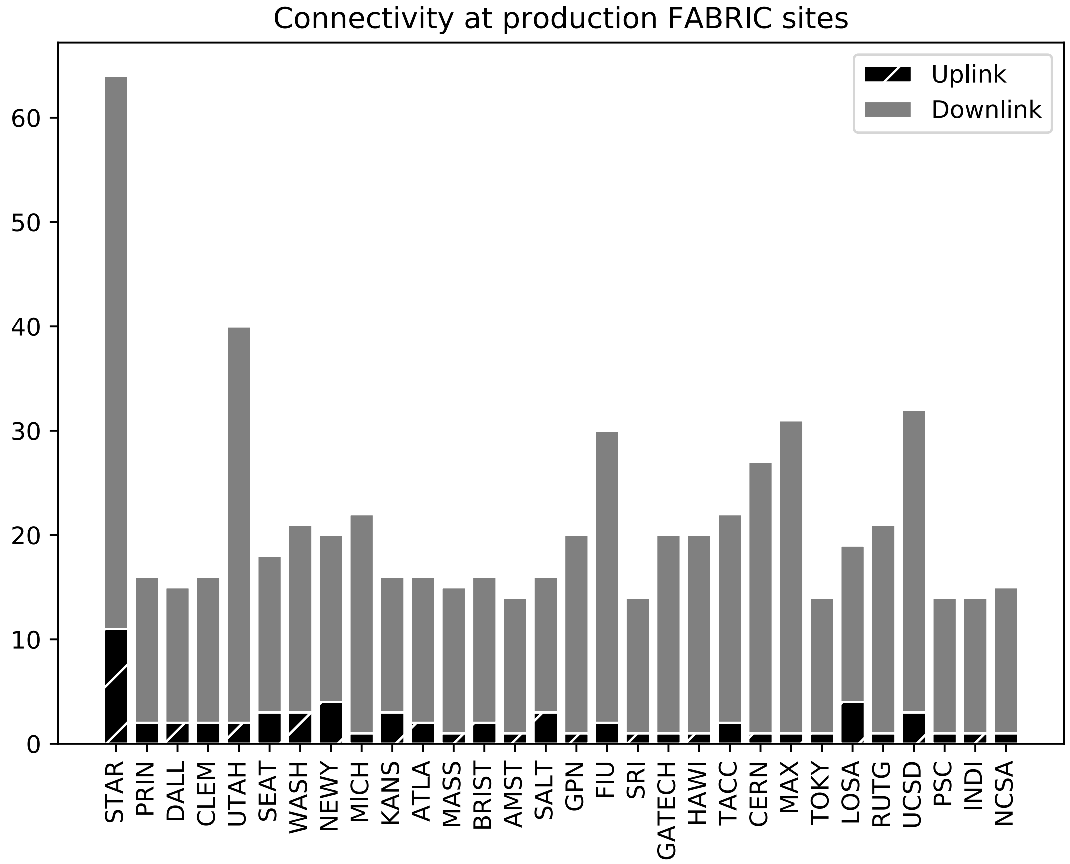
- Resource and Infrastructure Study
- FABRIC Network Traffic Profile
- Patchwork system:
  - End-to-end automation
  - Loss detection: Tx+Rx >= (Mirror)Rx
  - Cycling ports to get more coverage.
    - Mitigates disparity between switch ports and mirrors.
    - Ranking ports by activity.
  - Offloading to Alveos.
    - Filter, Truncation, Editing+Anonymisation.
      + custom **DPDK** application for capture serialisation.
    - Software can capture up to ~8.5Gbps.
    - Hardware can capture up to 100Gbps.

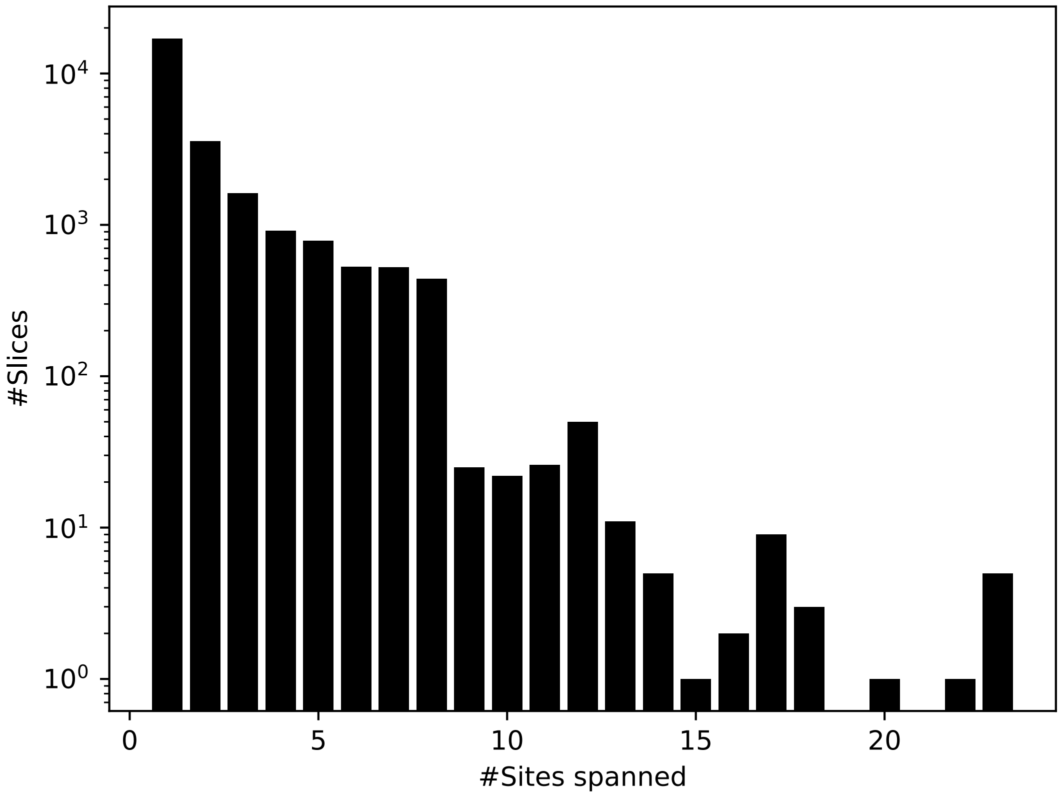


Resource and Infrastructure Study

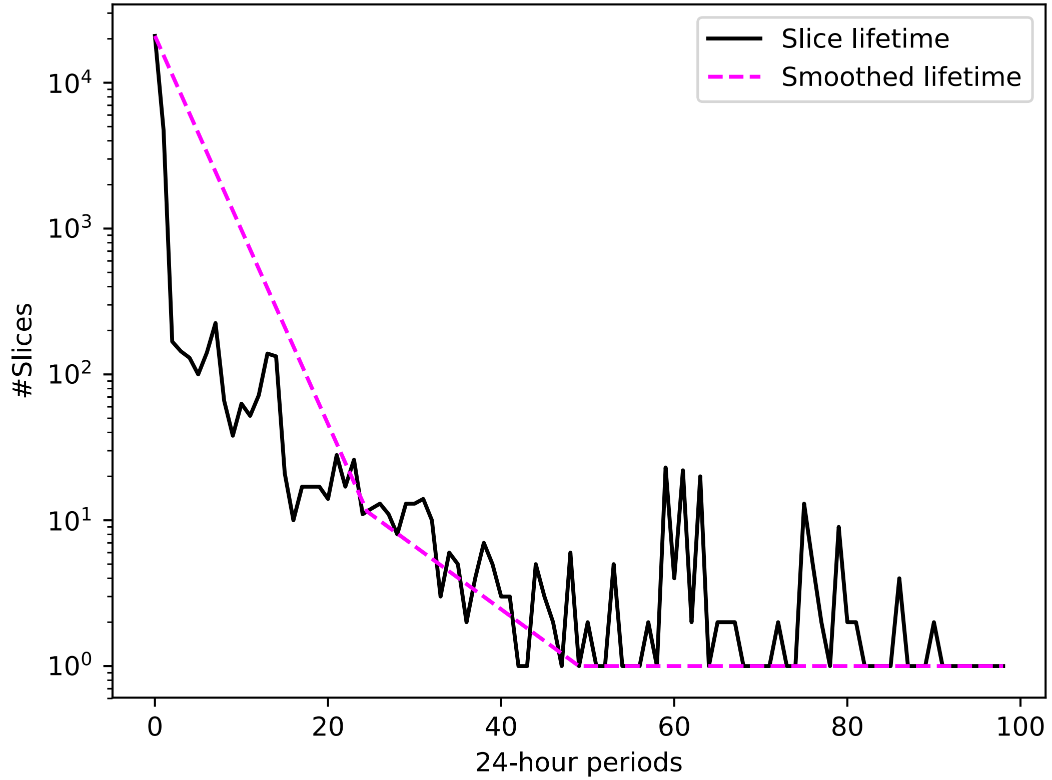
- **Which ports should we focus on?**
  Uplinks? Mostly uplinks? Both uplinks and downlinks?
  How do we choose?
- **When** should we sample?
  How are "busy testbed" and "busy network" related?
- **How long** should we sample for?
- **Which sites** should we sample?
- **What data rates** should we support?

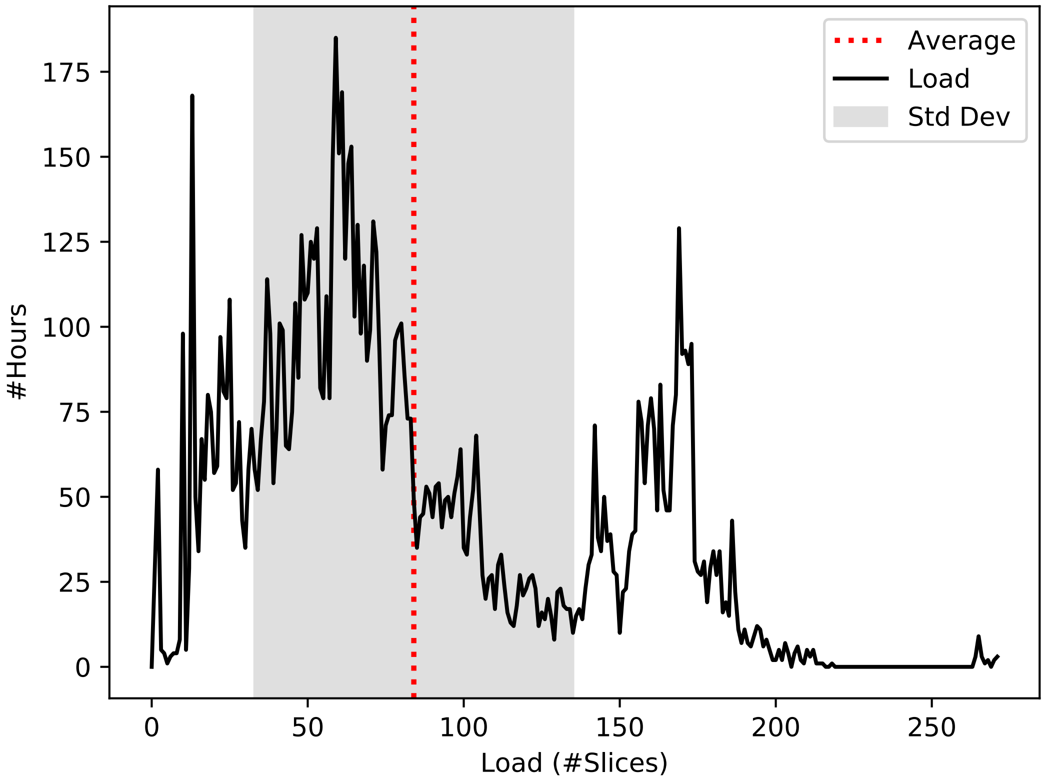
Uplinks and their Use



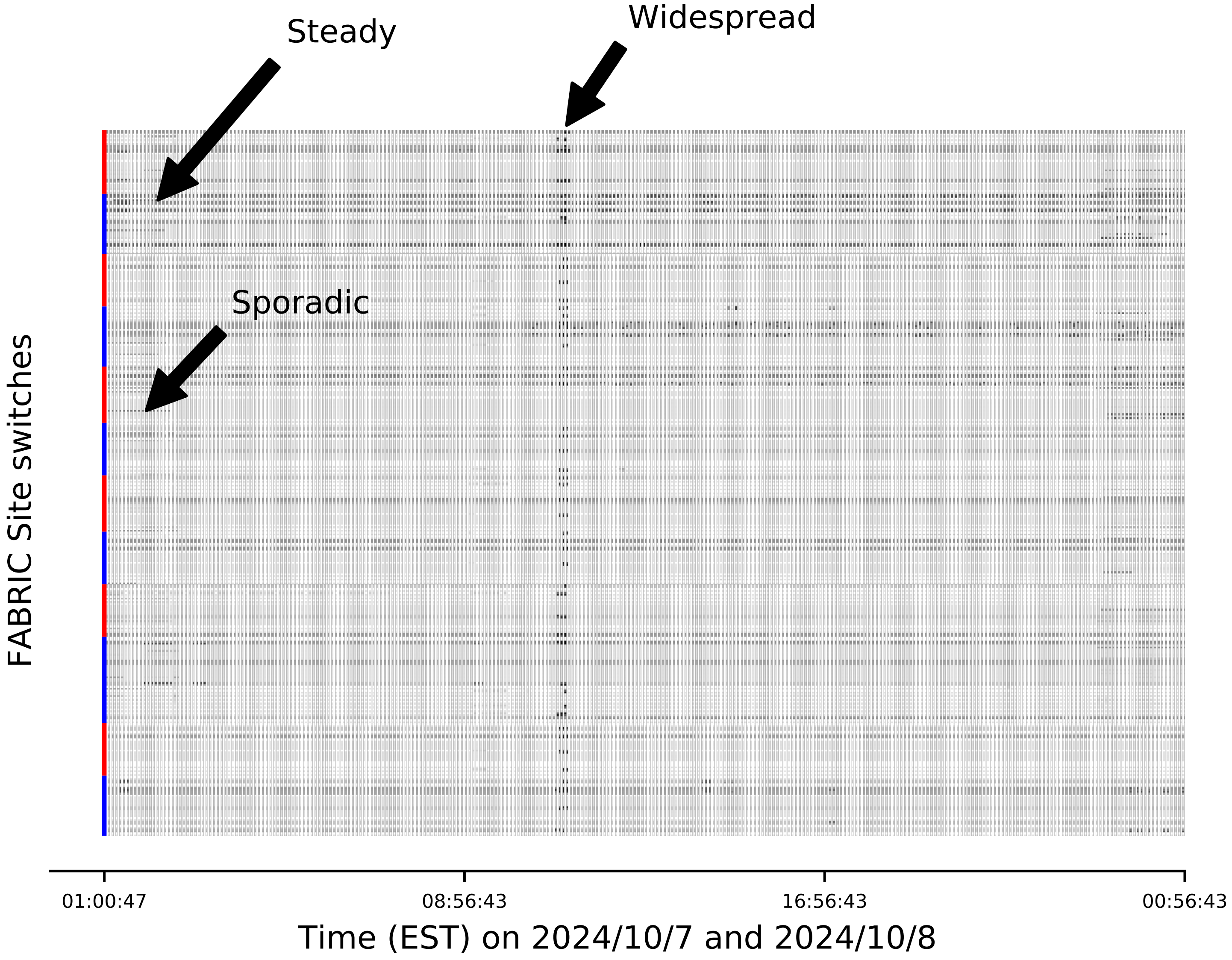


Slice characteristics (FABRIC-wide)

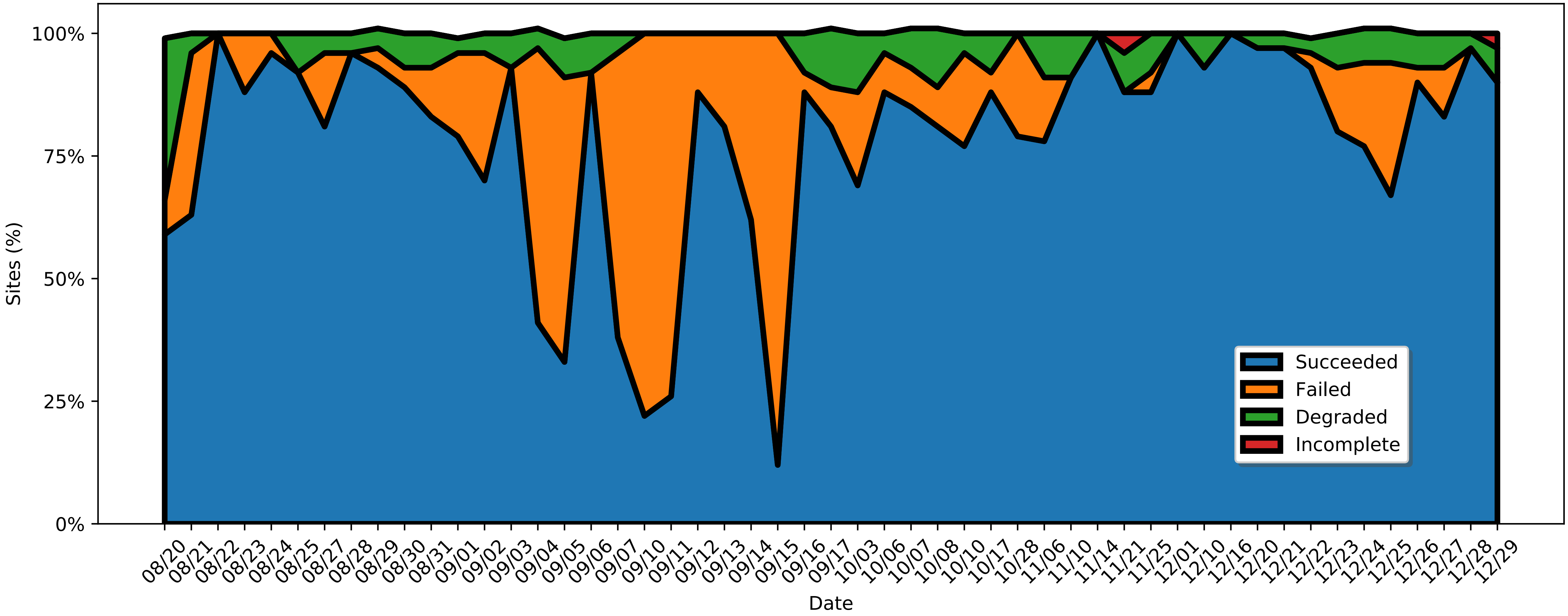




FABRIC Network Utilization



Patchwork is both a Tool and Experiment



FABRIC's Network Profile

- **Customisable! Gathering, analysis, visualization…**
- Granularity
  - Space: Global vs Site-level
  - Time: hour of day (other options: day of week, month)

- Headers/protocols
- Header diversity
- Frame sizes
- # sampled flows
- Size (in bytes) of sampled flows
- Frame size distribution in sampled flows
- Encapsulation depth and patterns

Food for thought

- Scale enables new features.
  FABRIC helps realize a duality between Experiments and Tools.

- How can we better work together?
  Testbeds as living communities of researchers.
  Actionable opportunities for collaboration through shared resources.

- Challenges: Complexity and Change
  Continuity and Reproducibility

- Starting points:
  - What would you like to know from Patchwork?
  - What additional CREASE features would help you?
  - PhD topics

Thank you!

- **Team members at Illinois Tech,** particularly contributors to Patchwork and CREASE: Hyunsuk Bang, Sean Cummings, Vaneshi Ramdhony, Bjoern Sagstad, Nishanth Shyamkumar, Prajwal Somendyapanahalli Venkateshmurthy, Alexander Wolosewicz. And many past team members at Illinois Tech and students who took my courses.
- Our co-authors and collaborators across projects related to this talk: Komal Thareja, Mert Cevik and Paul Ruth at **RENCI;** Charles Carpenter, Yongwook Song, Zongming Fei and Jim Griffioen at **UKY;** Tom Lehman at **FABRIC;** Xi Yang, Dale Carder, Stacey Sheldon, Jonathan Sewter, Peter Bengough, Yatish Kumar, and Chin Guok at **ESnet;** Ilya Baldin at **JLAB;** Anita Nikolich at **UIUC;** James Kowalkowski and Michael Wang at **Fermilab;** Vinod Yegneswaran, Deborah Shands, Ashish Gehani and Phil Porras at **SRI;** Dhiraj Saharia and Benjamin Ujcich at **Georgetown;** Adam Petz and Perry Alexander at **UKansas;** Gordon Brebner and Chris Neely at **Xilinx/AMD;** Joe Mambretti at **StarLight/ICAIR/Northwestern;** and **OTS** at **Illinois Tech** (Jim Tufts, Babar Kamran, Adrian Bucurica, Ibukun Oyewole, and Sejal Vaishnav).

Find out more during KNIT10

Vaneshi Ramdhony: **Network Profiling and Traffic Analysis for FABRIC users**

- Presentation today at 3:15pm (in Bellflower, Student Engagement Session)
- Demo today at 5pm (in room Trillium A)

Alexander Wolosewicz: **A Network Debugger for FABRIC Experiments**

- Demo today at 5pm (in room Trillium A)
- Tutorial tomorrow at 2:35pm (in room Redbud)