

LCLS Realtime Analysis Needs at NERSC

Christopher O'Grady, LCLS Data Systems

July 8, 2020

Linac Coherent Light Source

... the world's first "hard x-ray" laser



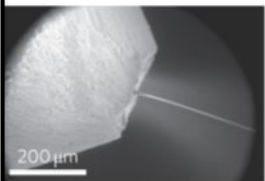
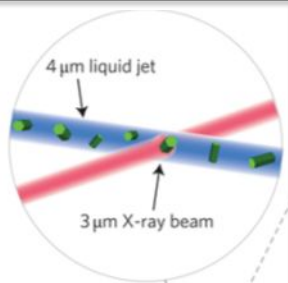
LCLS operates 24 hours/day with 95% beam availability, 120 Hz now, 1Mhz in 2021

- ~\$1B facility runs 24/7
- **1MHz, 20GB/s** in 2021: requires supercomputers.
- **Computing needs change from “minute to minute”**
- Experiments change significantly multiple times per week
- Realtime data analysis feedback is critical for running experiment
 - ~1s latency for subset of data (before data reaches disk)
 - Few-minute latency for all data (from disk)
- I am here to discuss the **few-minute latency (from disk)** which I will **(loosely) call “realtime”**

LCLS Nanocrystallography Example



Experiment Description



Gas dynamic virtual nozzle



LCLS beam

- Individual nanocrystals are injected into the focused LCLS pulses
- Diffraction patterns are collected on a pulse-by-pulse basis
- Crystal concentration dictates “hit” rate

Multi-megapixel detector

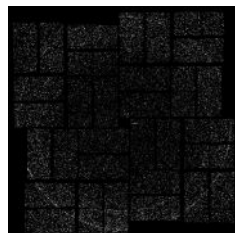


60 GB/s

1 TB/s

8 kHz in 2024 (4 MP)
40 kHz in 2027 (16 MP)

X-ray diffraction image



6 GB/s

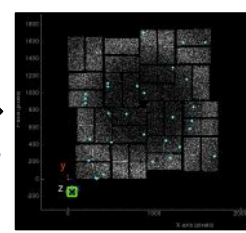
100 GB/s

Data Reduction

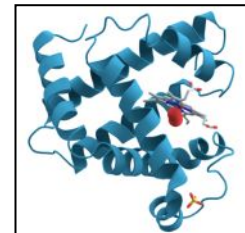
- Remove “no hits”
- >10x reduction

3 TFlops
16 TFlops

Intensity map from multiple pulses



Interpretation of system structure / dynamics



Data Analysis

- Bragg peak finding
- Index / orient patterns
- Average
- 3D intensity map
- Reconstruction

4 PFlops
20 PFlops

Data reduction mitigates storage, networking, and processing requirements

Current NERSC Possibilities (my best understanding)



- **Reservations**

- Need >1 day advance notice? While useful, LCLS is too dynamic: e.g. accelerator or expt breaks, or job takes longer than expected

- **“Realtime” QOS**

- Dedicated resources that are idle when not being used. Inefficient, but very useful for smaller users.
- LCLS has been approved for 20 nodes

- **Flex queue**

- Jobs that can checkpoint (e.g. density functional theory codes like VASP, Quantum Espresso...)
- Used by NERSC to chop big jobs in small pieces to “fill in the cracks”

- **DMTCP** (<https://www.nersc.gov/assets/Uploads/Checkpoint-Restart-20191106.pdf>)

- A work in progress by Zhengji Zhao and others

- “Realtime QOS” is inefficient, so not an option for larger efforts like LCLS
- I’ve been told “suspended jobs” (remain in memory/swap) is not an option at NERSC
- My **best guess**:
 - **Flex queue is closest**: NERSC system is already preempting checkpointable jobs, which receive a discount
 - **Expand flex-queue idea: a “high-priority queue”** where LCLS pays a premium to be able to preempt flex-queue jobs that can checkpoint (VASP, Quantum, Espresso, DMTCP?)