SACLA and the K Computer

A. Hori (RIKEN AICS)

A. Tokuhisa (RIKEN AICS)
T. Kameyama (RIKEN AICS)
K. Okada (JASRI/RIKEN RSC)
M. Yamaga (JASRI/RIKEN RSC)
Y. Joti (JASRI/RIKEN RSC)
M. Yabashi (RIKEN RSC)
Y. Ishikawa (Univ. Tokyo / RIKEN AICS)

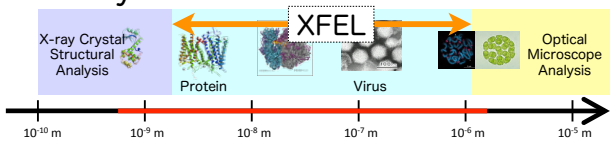
K. Yoshinaga (RIKEN AICS)
J. ARAI (Univ. Tokyo -> NTT)
T. Sugimoto (JASRI)
R. Tanaka (JASRI/RIKEN RSC)
T. Hatsui (RIKEN RSC)
Y. Sugita (RIKEN AICS)
N. Go (JAEA)

SACLA and the K

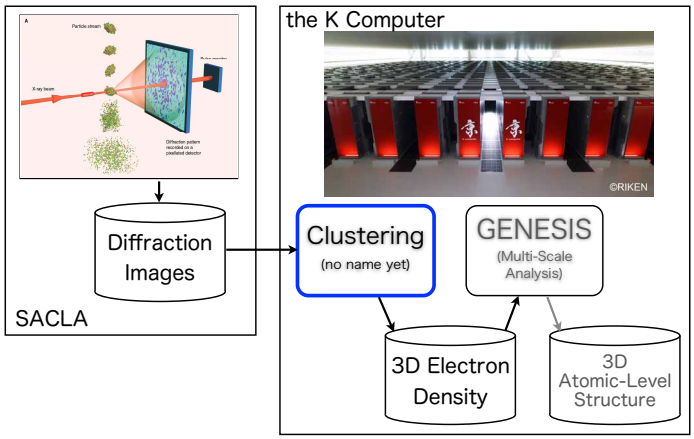
- Data Acquisition: SACLA (XFEL Facility)
  XFEL: X-ray Free Electron Laser
- Data Processing: the K computer

WAN (80Km)

- To Analyze 3D Structure of Particles

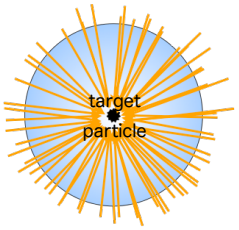


Entire Data Flow



Clustering Algorithm

- Orientations of target particles are uncontrollable
  - Correlation (FFT) is the clue of viewpoint
    => Clustering
  - All-to-all computation is needed $O(N^2)$!

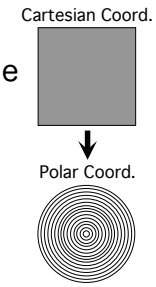


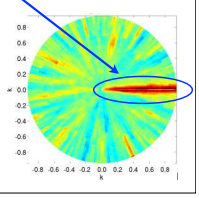
- Diffraction image contains quantum noise
  - 100 images must be averaged.

➡ 1 million images must be shot !!

Rotation Handling

- 2-D Cartesian coordinate image is converted to Polar coordinate image (coaxial rings)
- Each ring is FFTed
- Two images are compared (correlation), ring by ring
- Finally, resulting "a correlation line" indicating they are close enough





High-speed classification of coherent X-ray diffraction patterns on the K computer for high-resolution single biomolecule imaging (A. Tokuhisa, et al.), In Journal of Synchrotron Radiation, volume 20, 2013.

Fact Sheet

- Data Acquisition (SACLA)
  - max. 30 images/sec (depending on the kind of target particle)
  - 20MB/image (3Å resolution) (depending on resolution)
  - 1 million images for one particle analysis
    - quantum noise and all possible orientations

➡ yielding 20TB in 10 hours !

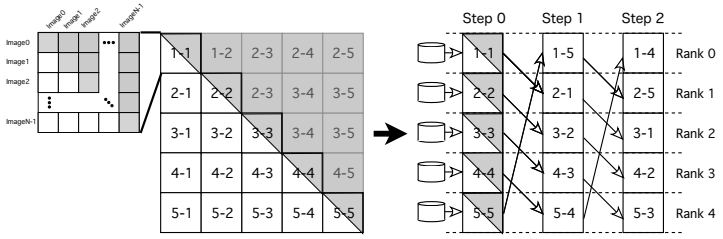
- Data Transfer (from SACLA to "K")
  - *Gfarm* copy tool (*gfpcopy*) takes 20 hours to copy 20TB data
- Data Processing ("K")
  - $O(N^2)$, but can be reduced to $O(K \cdot N)$
    K is the number of clustering groups

OPT-1: All-to-all Processing

- Load must be balanced
- Each data file must be read *only once*

➡ Read data are passed to neighbors at every step

- #steps is *(N+1)/2* (*N* is #chunks or #nodes)

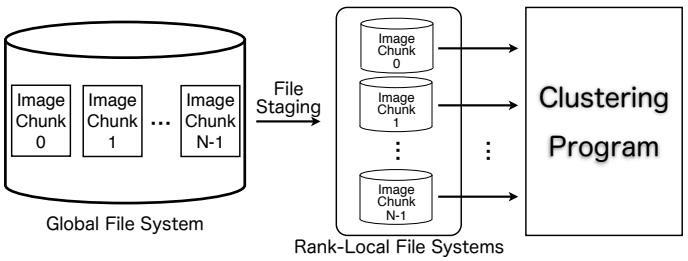


OPT-2: File Distribution

- Parallel file system can be a bottleneck
- Maximizing parallelism up to #nodes

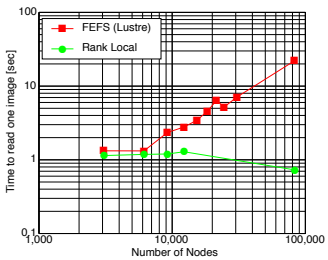
➡ Utilizing the K file system

- File staging system, and
- "Rank-local" file system



Performance Numbers

- Computation time
  - 5Å 0.06 Sec/2images
  - 3Å 0.45 Sec/2images
- Communication time
  - Hidden
- Read Time →



- Computation time dominates ("K", 80K nodes)
  - 5Å 1 hour
  - 3Å 24 hours

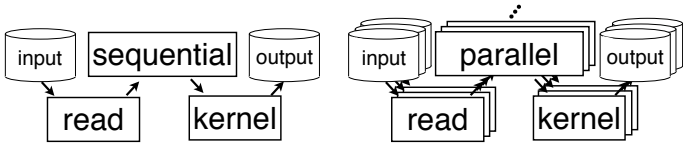
Current Work

- Generalization to handle *any* all-to-all data processing
  (e.g. Transmission Electron Microscope (TEM))

➡ Decoupling kernel code and parallelizing (MPI) code

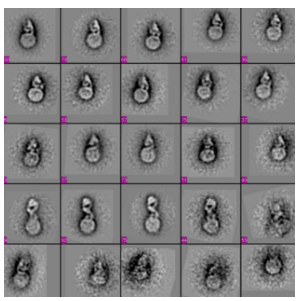
- User develops sequential programs
  - file read and sequential kernel code
- Easy-to-develop, easy-to-debug, easy-to-port, and programing language free

```
$ ./nonameyet ./read-prg params ./kernel-prg params
```



Future Work

- 4-D Analysis (3-D + Time)
  - to analyze time-variant particles,
  - requiring more number (x10 ?) of images to be shot and analyzed.



F. Ye, *et al.*, *JCB*, **188**, 157 (2010)

- Exa-Machine
  - Finer Resolution
  - 4-D Analysis

Summary responding to Jack's e-mail

- Workflow
  - *All-to-all data processing* can be a workflow
- Architecture
  - Fast sequential (+ OpenMP) FFT (3.3 B/F)
- Virtual Data Facility
  - Silent error can be handled as quantum noise (?)
- Software Stack
  - *Decoupling Architecture*
- Data Mini-apps
- Data Stewardship
- Data Life Cycle
  - Scientists never want to discard experimental data
  - Facilities (SACLA and K) are so big and hard to relocate
  - File staging and node(rank)-local disks (no-need of PFS ?)