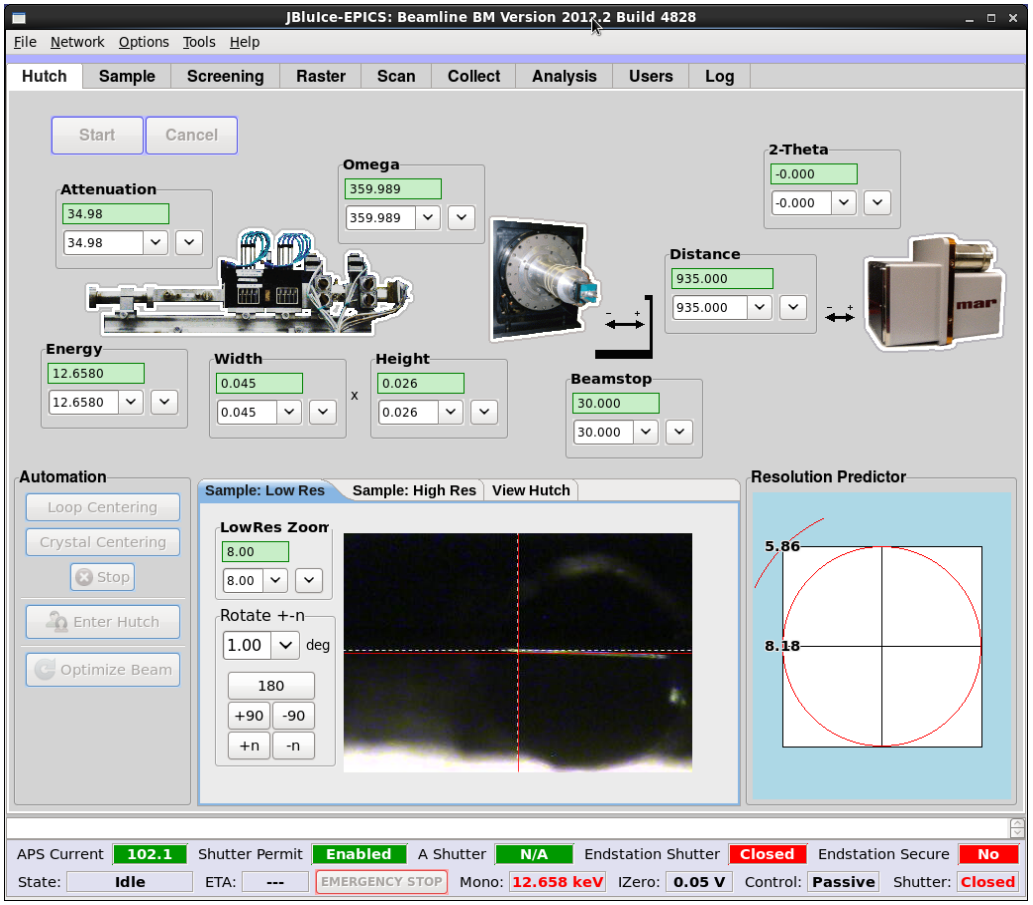
Data collection strategy and Data reduction in JBluice-EPICS

Sudhir Babu Pothineni, Nagarajan Venugopalan, Craig Ogata, Mark Hilgart
GM/CA@APS
TWG Meeting, Aug 16th 2012

Software Highlights of GM/CA@APS

- JBluice-EPICS beamline control software
- Automated screening
- Automated diffraction/fluorescence rastering for invisible crystals and to select best diffracting areas of a visible crystal
- Auto centering on visible crystals
- Automated data collection strategy
- Single site & multiple site data collection (Vector data collection)
- Automated data reduction

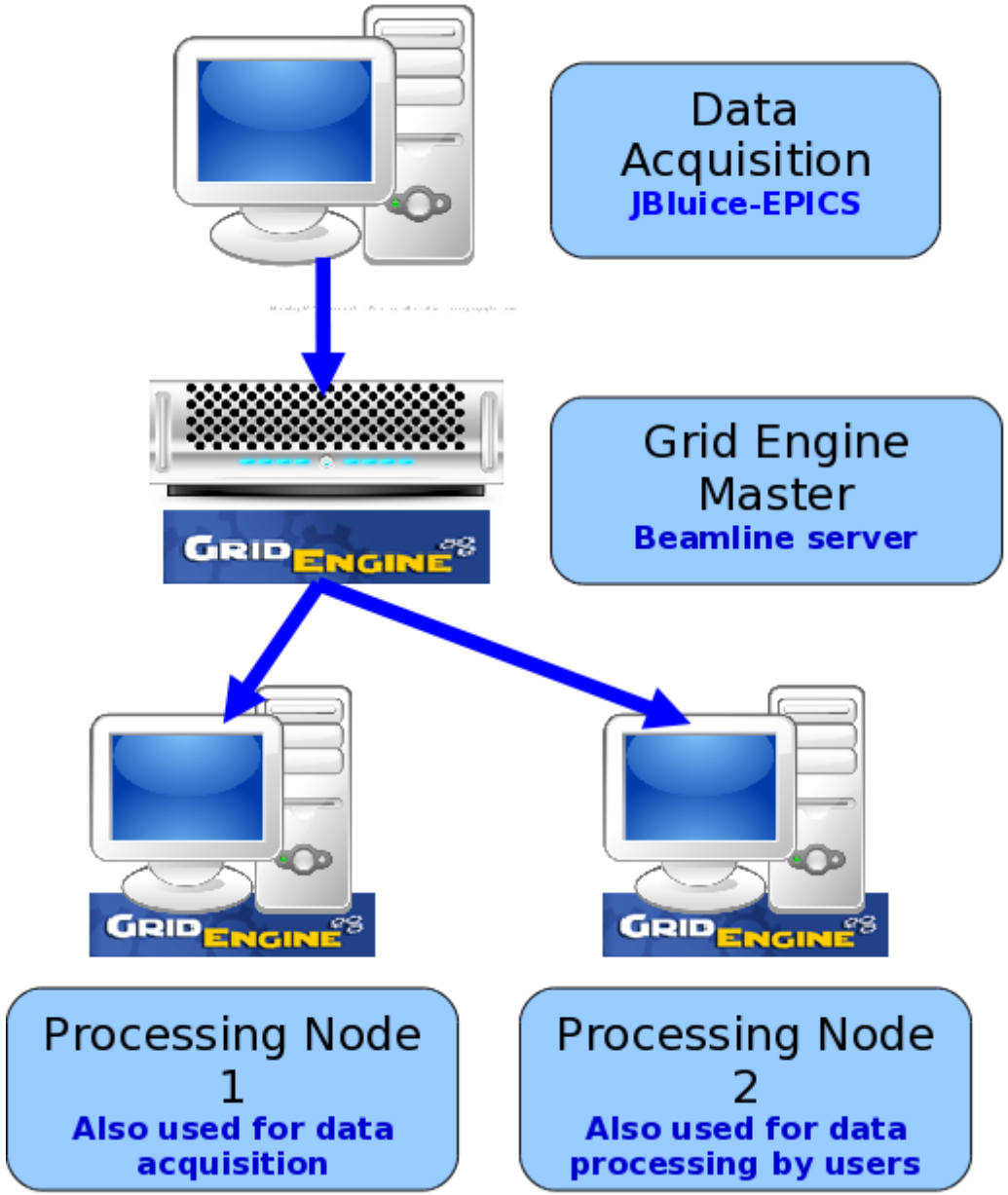


http://www.gmca.anl.gov/jbluice-epics/

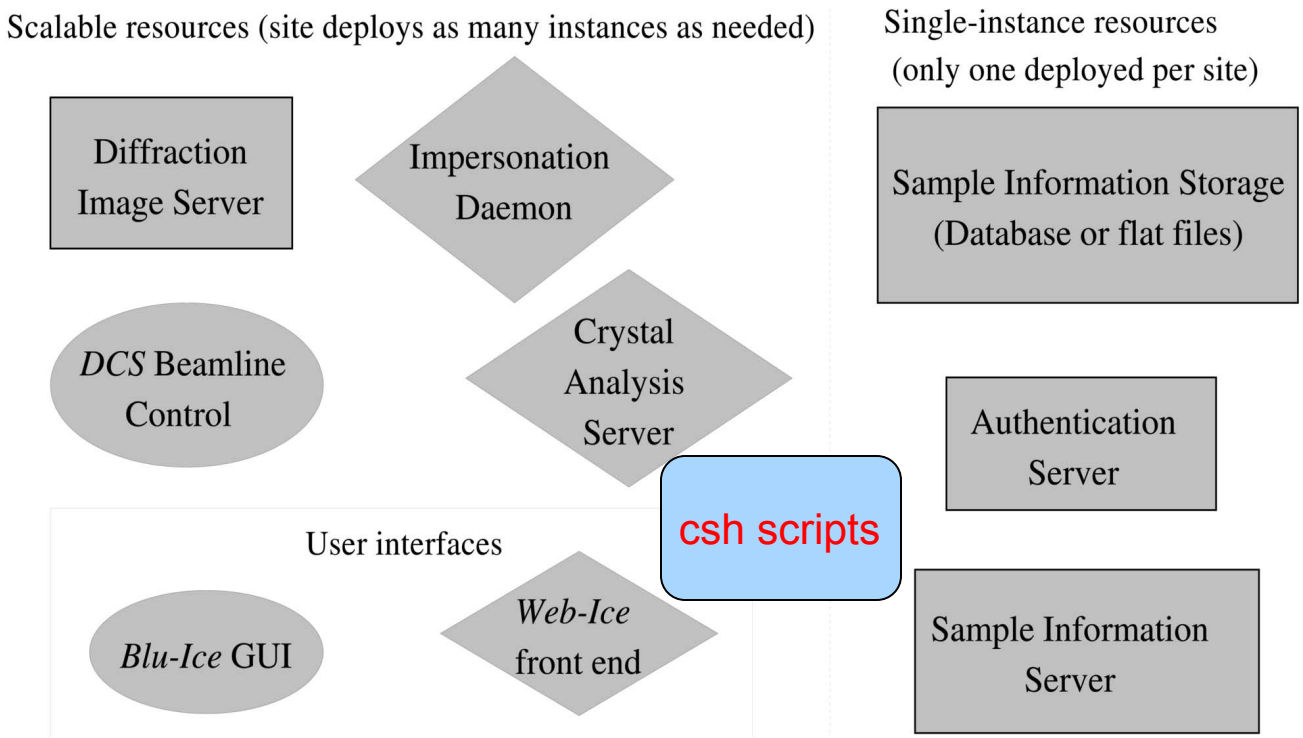
Data Processing in JBluice-EPICS

- Efficient use of available resources.
- JBluice uses Grid Engine for running strategy and data reduction jobs.
- Grid Engine is a open source batch queuing system/job scheduler.
- JBluice submits and monitors the Grid Engine jobs using DRMAA java binding and parses the result log files for quality parameters.
- Scalable to dedicated multiple node cluster.

- SSRL Webice based data collection strategy.
- XDS based data reduction



SSRL WebIce



LABELIT (autoindexing)
MOSFLM (integration)
BEST (strategy)

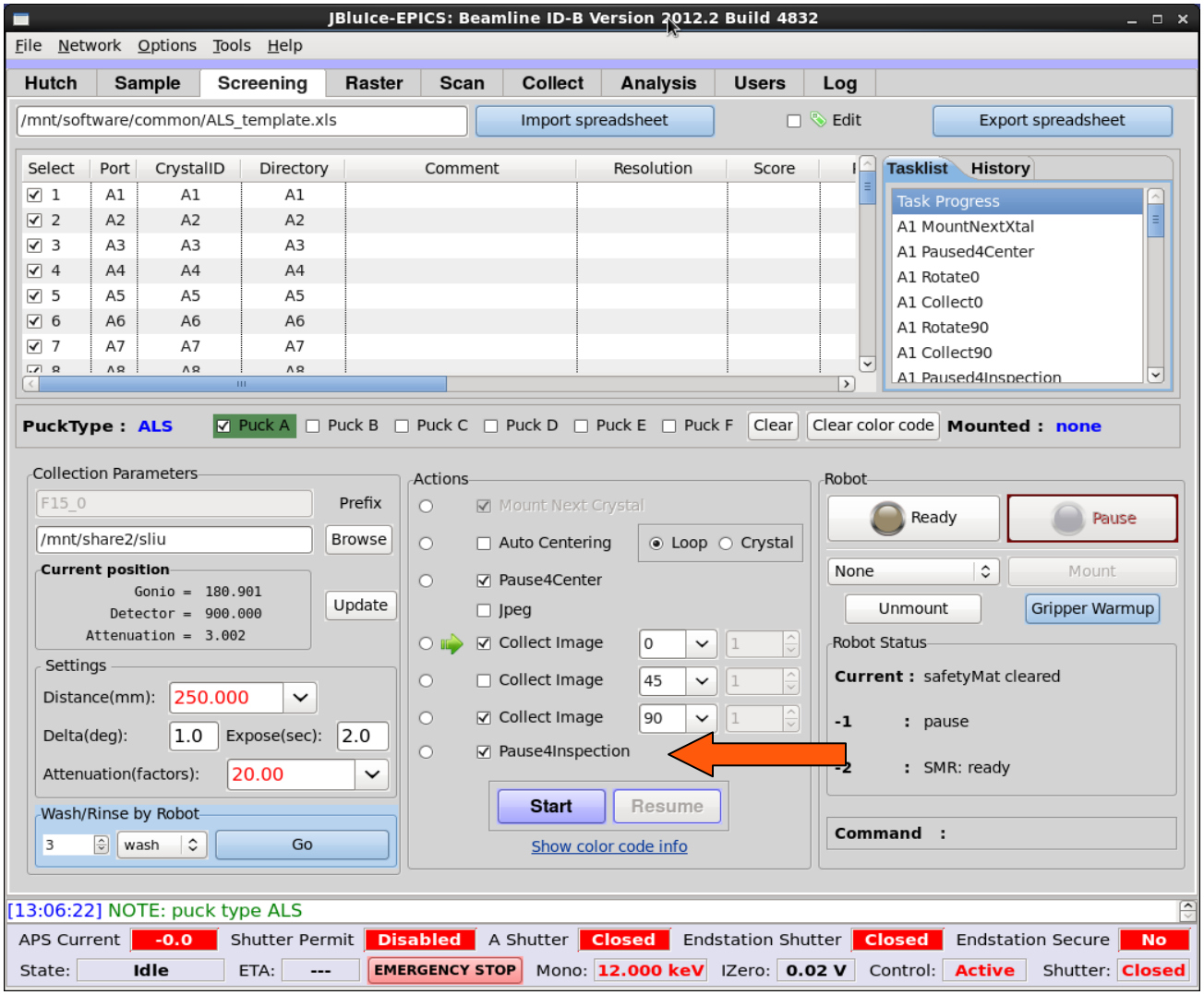
Web-Ice: integrated data collection and analysis for macromolecular crystallography. A. González, P. Moorhead, S.E. McPhillips, J. Song, K. Sharp, J.R. Taylor, P.D. Adams, N.K. Sauter and S.M. Soltis. J. Appl. Cryst. 41, 176-184 (2008).

Porting Webice to JBluice

Webice	JBluice-EPICS
Impersonation daemon, crystal-analysis server	Grid Engine
Sample information server	MySQL
Image server	Image server
JSP web interface	Java swt application interface
SSRL csh scripts for strategy calculations	Modified SSRL csh scripts for strategy calculations

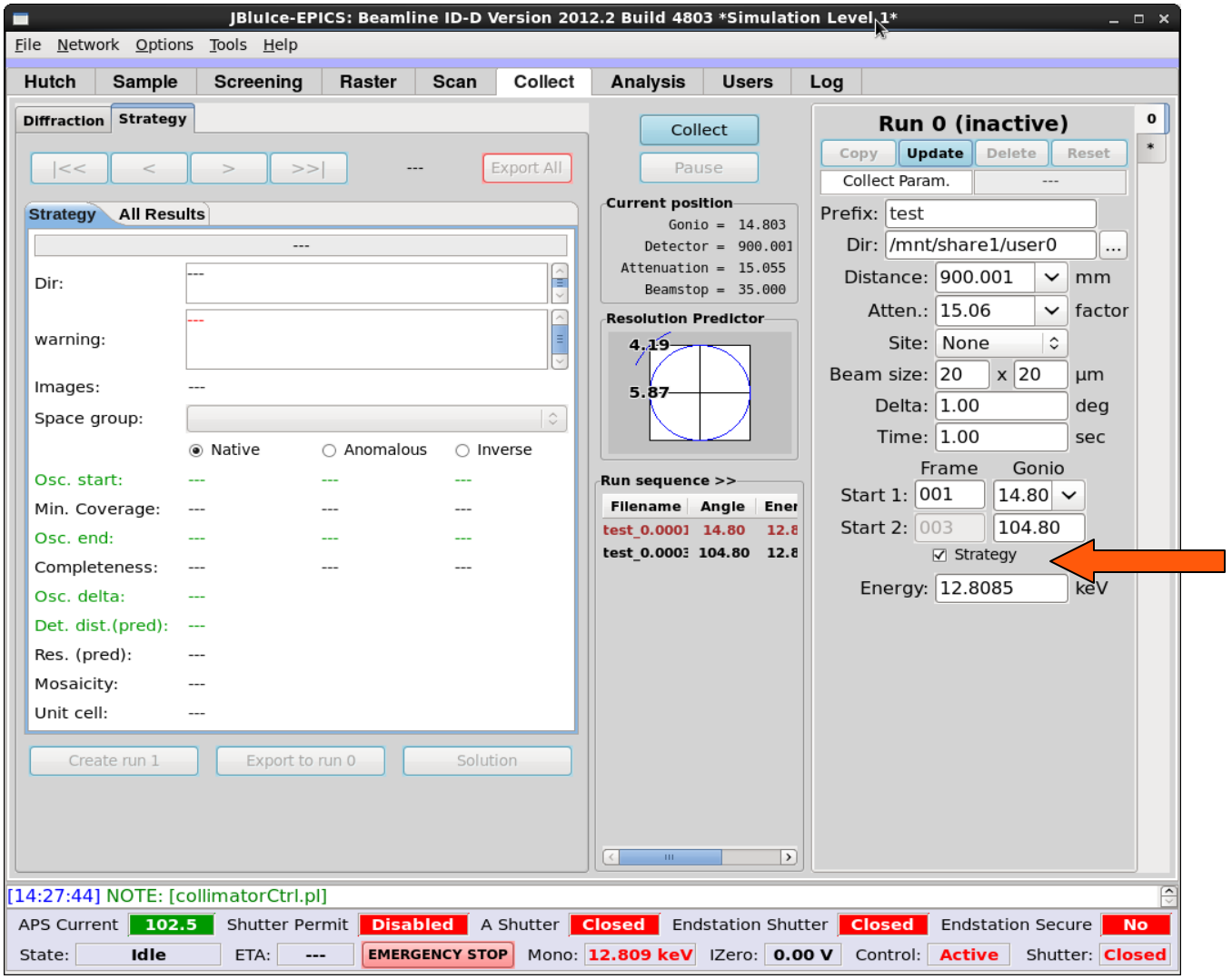
- Maintenance of tomcat web server, authentication server, ssl certificates etc.. are avoided. Parallelization added.

Strategy initiated from screening tab



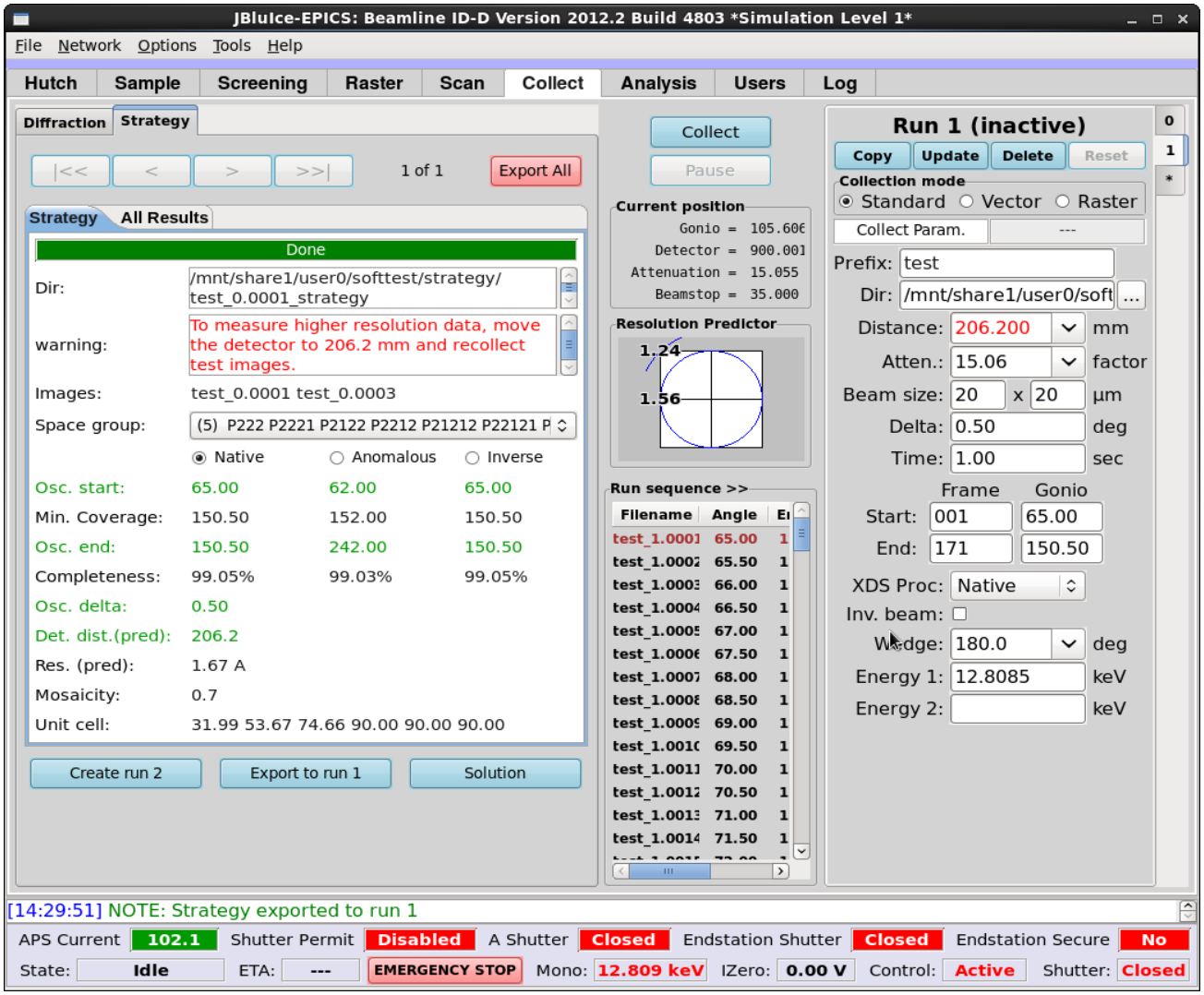
- Strategy calculation initiated automatically after collecting 2 images in screening tab.

Strategy initiated from collect tab



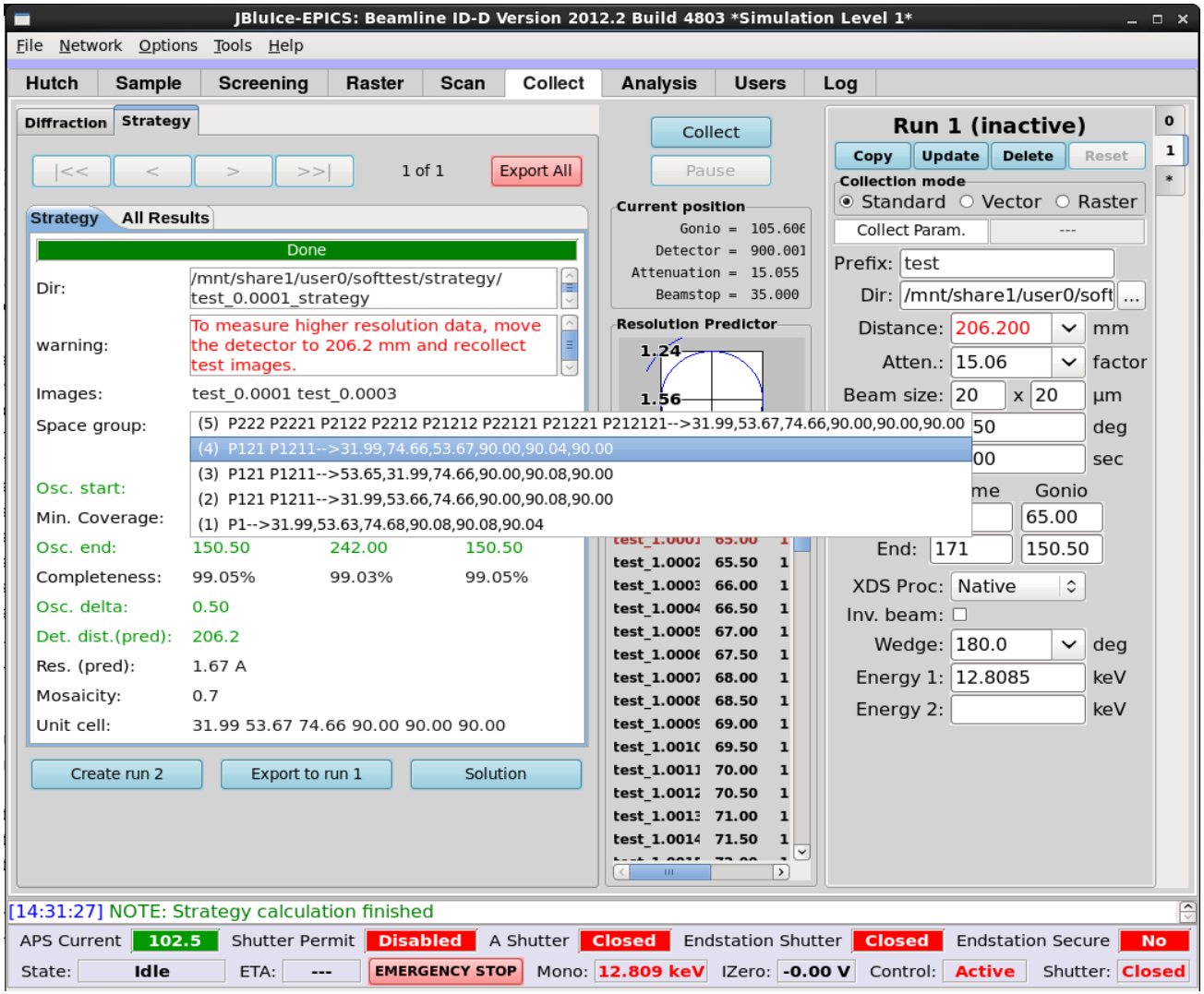
- Strategy can be initiated from collect tab ‘0’ run for manually mounted samples

Strategy parameters



- After strategy calculation parameters are displayed on strategy sub tab

Selecting lower symmetry spacegroup



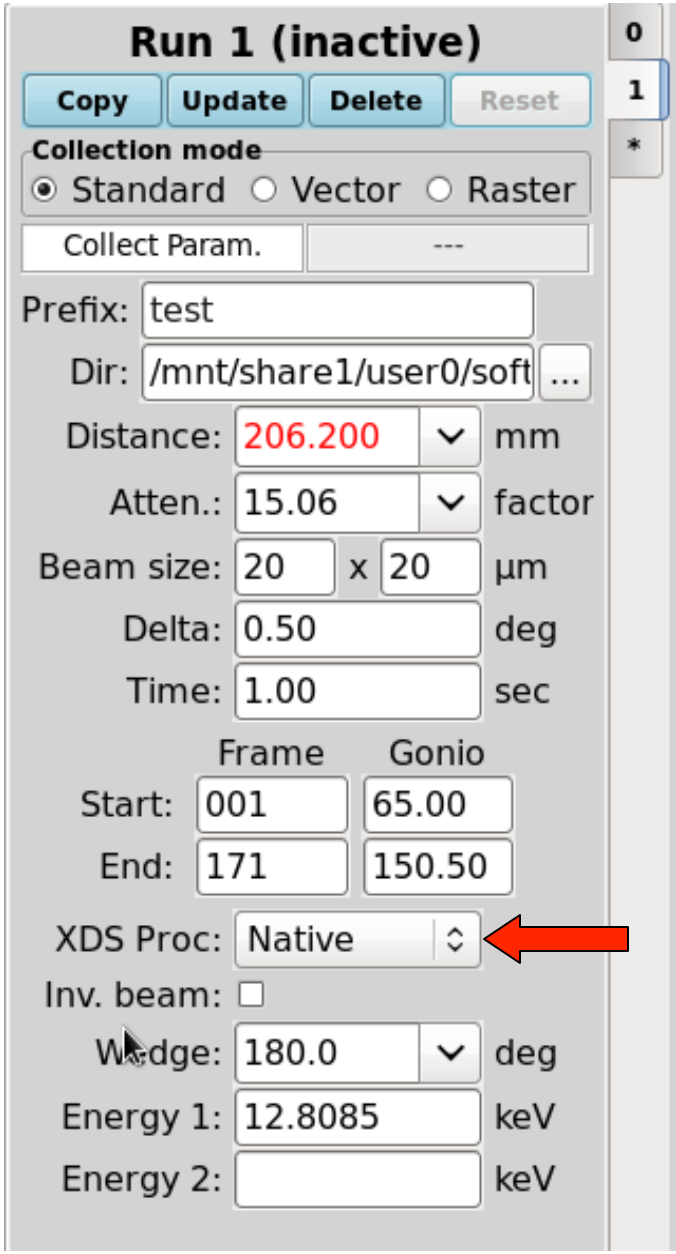
- Lower symmetry spacegroup can be selectable for to re-run strategy

Data Collection Strategy

- List all possible space groups from Labelit solutions.
- Solutions from same crystal symmetry are filtered.
- By default only space groups from highest symmetry are processed initially.
  - They are processed in parallel so it is faster.
- User can choose lower symmetry solution, JBluice will process and display strategy.
- User has an option to choose strategy from BEST/Mosflm.
- Displays Anomalous information, There are two options.
  - Anomalous continuous where osc. range calculated is doubled.
  - Anomalous Inverse where it will use inverse beam with 'Native' osc. range.
- Improved error/warning handling.
- All the strategy results can be saved in Excel file.
- Processing time about 30 sec (if there is index solution).

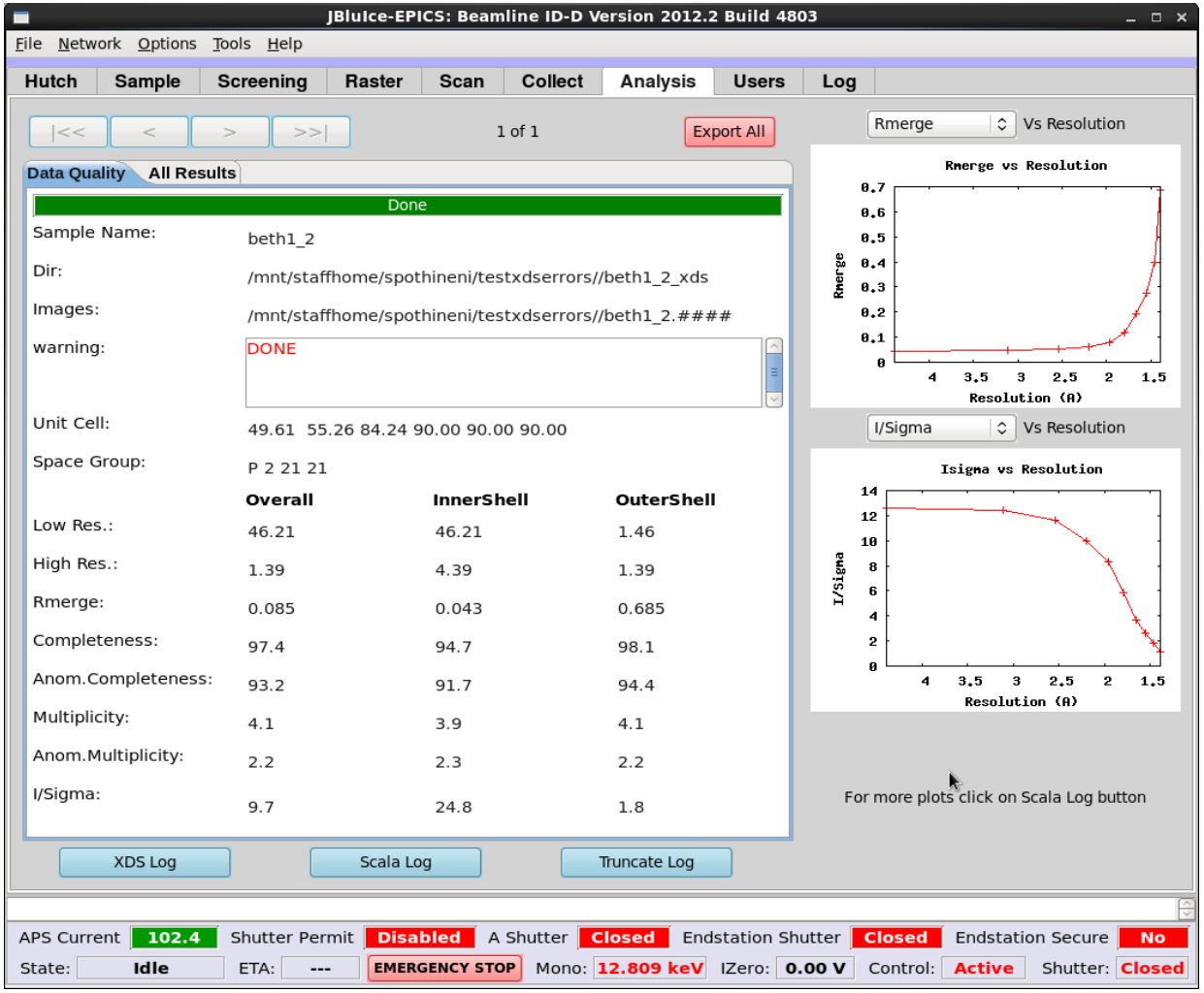
Data Reduction

- There is XDSProc option (None, Native, Anomalous) on each run on collect tab, if checked it will initiate XDS processing through Grid Engine after a dataset is collected.
- JBluice will write XDS input file XDS.INP.
- Runs XDS, POINTLESS, SCALA and TRUNCATE
- Any dataset with more than 10 images collected is automatically reduced.
- XDS runs second time with DEFPIX, INTEGRATE, CORRECT incase of failure with INSUFFICIENT PERCENTAGE of spots.
- XDS can process Anomalous data for SAD & MAD data collection.



Analysis Tab

- Display data quality parameters and plots automatically on Analysis tab.
- Direct access to XDS, SCALA, TRUNCAE logs with the click of a button.
- The scaled mtz file, and other input and intermediate files stores in a subdirectory of data.
- User can change XDS.INP and reprocess the data.



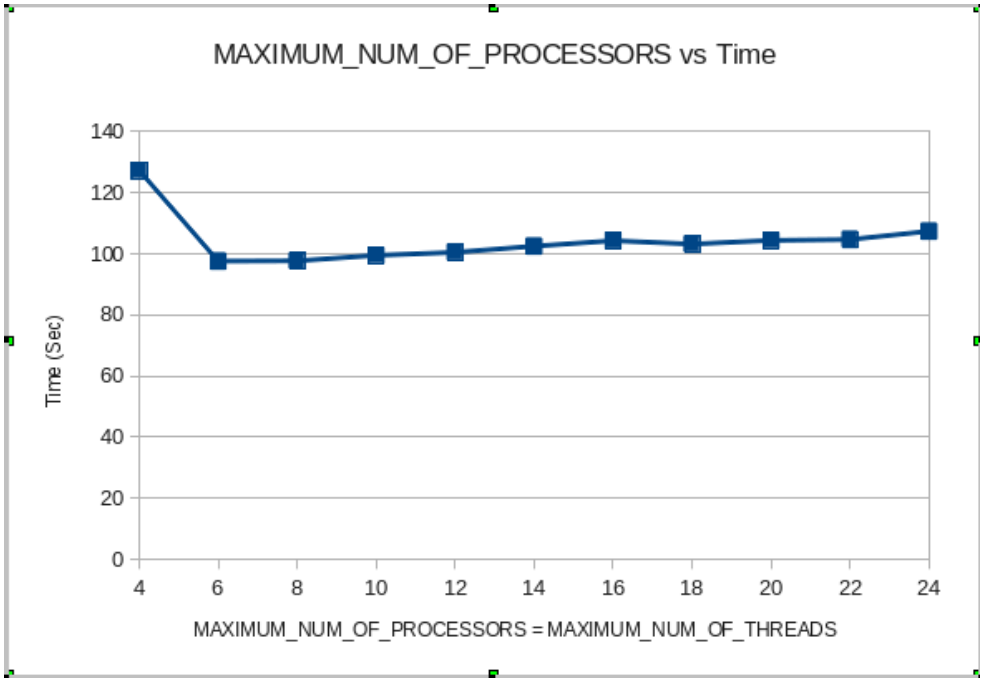
Processing time & specs

Dataset :

360 MAR ccd images
JOB= ALL
SPOT_RANGE= 1 180

MAXIMUM_NUM_PROCESSORS= 6
MAXIMUM_NUM_THREADS= 6
DELTAPHI= 5

Processing time = 97sec.



Since manual processing overlapping with auto processing on same workstations using
MAXIMUM_NUM_PROCESSORS= 4
MAXIMUM_NUM_THREADS= 4

Processing time = 127sec.

Workstation : Intel Xeon with 2 x 6 cores, hyper-threading enabled, total 24 cores.

- limitation : 32 bit.

OMP_STACKSIZE set according to number of cores.

- Ex. 64m for 24 core machine.

Future plans

- Installing dedicated machines for data processing.
- Adopting fast_dp from Diamond Light Source.
    - Error correction and reprocessing
- Running data processing in parallel to data collection.
- Processing vector data with overlaps and inverse beam.
- Processing data collected with raster sites.
- Molecular replacement for structure solution.

Thanks

JBluice-EPICS

Mark Hilgart
Sudhir Babu Pothineni

Controls

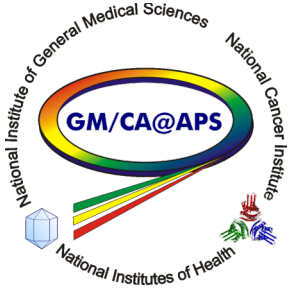
Sergey Stepanov
Oleg Makarov

Crystallography

Craig Ogata
Nagarajan Venugopalan
Ruslan sanishvili (Nukri)
Michael Becker

GM/CA Management

Janet Smith
Robert Fischetti



LABELIT

Nicholas Sauter

BERKELEY LAB
Lawrence Berkeley National Laboratory