mySZE Documentation

Release 0.0.1

Chunhong Yoon

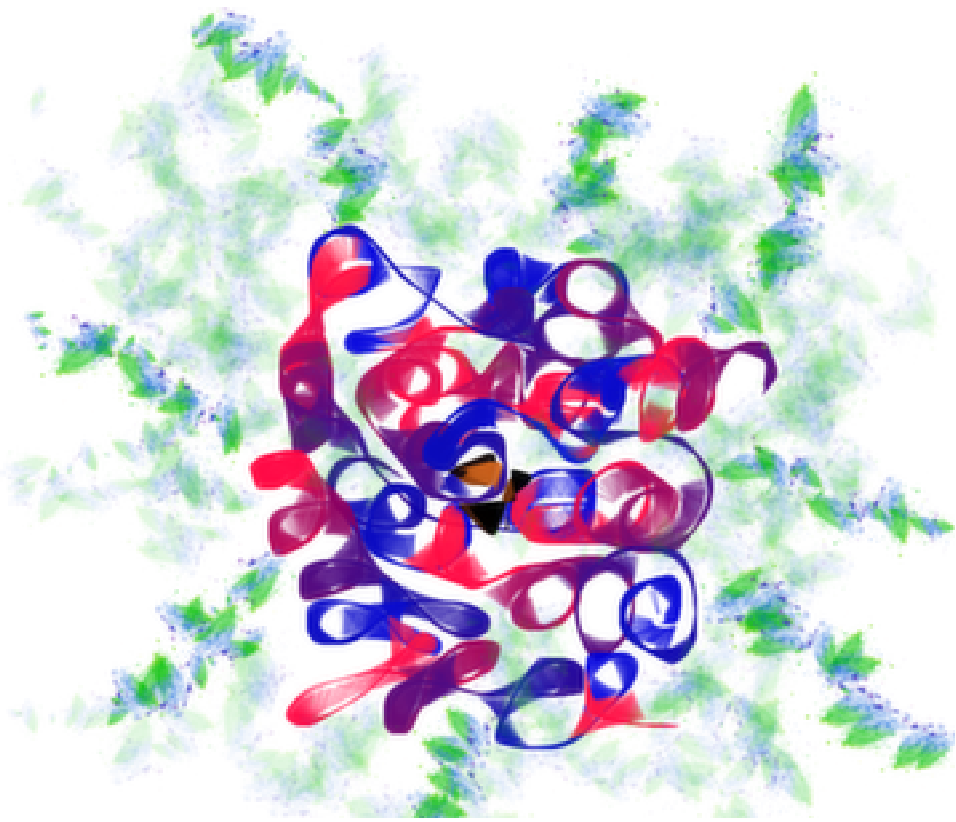
May 20, 2016

Contents

Contents:

CHAPTER 1

Getting started



simS2E is an implementation-agnostic framework that defines the data interfaces between the modules, i.e. Module A must output an hdf5 file that adheres to the simS2E interface which Module B is expecting to read in. Actual implementation of Modules A and B is up to the user. If you understand this concept, then you are ready to use simS2E. You will need to set up the environment for simS2E and if you have a module that you want to plug in, find out what data formats the input and output should be here:

```
http://sims2e.readthedocs.org/en/latest/docs/fel_source_simulation.html
```

If you would like to try running an example simS2E pipeline, then you need Docker installed on your machine.

1.1 System requirements

You will need at least 8GB RAM and 20GB of disk space.

1.2 Installing Docker for example simS2E pipeline

You do NOT need Docker for simS2E. You only need it for running the example simS2E pipeline that we provide. Docker is a new container technology (Think of it as a light-weight virtualbox) that can be run on your choice of OS. Instructions for installation can be found here:

```
https://docs.docker.com/installation/
```

1.3 Setting up the environment for simS2E

You need to clone the simS2E repository from GitHub into your local directory (I will refer to this directory as /host/path):

```
cd /host/path
git clone https://github.com/chuckie82/simS2E.git
```

You now have a copy of all the files/scripts in the sub-directories needed for simS2E:

```
/host/path/simS2E/workflow: Directory for run scripts
/host/path/simS2E/config: Directory for configuration files
/host/path/simS2E/data: Directory for reading/saving data
/host/path/simS2E/packages: Directory for installing software packages
/host/path/simS2E/modules: Directory for module specific scripts
/host/path/simS2E/tmp: Directory for storing temporary files
/host/path/simS2E/docs: Directory for online documentation
```

Go to the packages directory and run setup.sh. This will build all the packages as Docker containers; 1) FAST for FEL source, 2) WPG for optics, 3) pmi_demo for radiation damage to the sample, 4) SingFEL for diffraction patterns, 5) EMC for orientation recovery, and 6) DM for phase retrieval:

```
cd /host/path/simS2E/packages
./setup.sh
```

Now you are ready to run the simulation!!!

1) Let's run the FEL source simulation using FAST:

```
docker run -it -v /host/path/simS2E:/simS2E fast:v0.1 /bin/bash
```

You are now inside the Docker container running bash on Ubuntu v14.04. The FAST package is installed under /home/packages and the simS2E directory is located under /simS2E.

Go to the workflow directory and run the example. This will generate a single FEL pulse:

```
cd /simS2E/workflow
./runFAST
```

This script runs master.sh which in turn runs master_fast.sh. All the simulation configuration is defined in /simS2E/config/config_sim_example. Let's examine the configuration file:

```
nano /simS2E/config/config_sim_example
```

NUM_FELsource_OUT=1 means output 1 instance of the FEL pulse. FELsource=ppFAST means use the FAST package. FAST simulation parameters are defined under ###### ppFAST ######.

When the simulation is complete. Exit the docker container by typing "exit" or Contrl+D. FELsource output hdf5 file will be in /host/path/simS2E/data/sim_example/FELsource. You can examine the hdf5 file by running:

```
h5ls -r /host/path/simS2E/data/sim_example/FELsource/FELsource_out_0000001.h5
```

Note that the output hdf5 names and fields conform to the specifications of the simS2E framework.

2) Let's run the optics simulation using WPG:

```
docker run -it -v /host/path/simS2E:/simS2E wpg:v0.1 /bin/bash
```

You are now inside the Docker container running bash on Ubuntu v14.04. The WPG package is installed under /home/packages and the simS2E directory is located under /simS2E.

Go to the workflow directory and run the example. This will generate the FEL pulse that will hit the sample after propagating through the SPB/SFX beamline optics:

```
cd /simS2E/workflow
./runWPG
```

When the simulation is complete. Exit the docker container by typing "exit" or Contrl+D. WPG output hdf5 file will be in /host/path/simS2E/data/sim_example/prop.

3) Let's run the photon matter interaction simulation using PMI_DEMO:

```
docker run -it -v /host/path/simS2E:/simS2E pmi_demo:v0.1 /bin/bash
```

You are now inside the Docker container running bash on Ubuntu v14.04. The PMI_DEMO package is installed under /home/packages and the simS2E directory is located under /simS2E. Due to the license agreement, the PMI package is not available in this example simulation and demo version is used instead.

Go to the workflow directory and run the example. This will generate the scattering factors of the sample under going radiation damage over time:

```
cd /simS2E/workflow
./runPMI
```

The pdb file that specifies the initial atom positions and scattering factors is stored under /simS2E/data/sim_example/sample/sample.h5. When the simulation is complete, exit the docker container by typing "exit" or Contrl+D. PMI_DEMO output hdf5 file will be in /host/path/simS2E/data/sim_example/pmi.

4) Let's run the diffraction simulation using SingFEL:

```
docker run -it -v /host/path/simS2E:/simS2E singfel:v0.1 /bin/bash
```

You are now inside the Docker container running bash on Ubuntu v14.04. The SingFEL package is installed under /home/packages and the simS2E directory is located under /simS2E.

Go to the workflow directory and run the example. This will generate the diffraction patterns of the sample under going radiation damage over time:

```
cd /simS2E/workflow
./runSingFEL
```

Let's open the simulation configuration file again in /simS2E/config/config_sim_example. NUM_DIFFR_OUT=100 means generate 100 time evolution diffraction patterns. In order to run a meaningful simulation, try increasing this number to 50,000. DIFFR=singfel means use the SingFEL package. SingFEL parameters are defined under ######

SingFEL ######. When the simulation is complete, exit the docker container by typing "exit" or Contrl+D. SingFEL output hdf5 file will be in /host/path/simS2E/data/sim_example/pmi.

You can examine the hdf5 file by running:

```
cd /host/path/s2eDocs/modules/diffr
python diagnostic_singfel.py /host/path/simS2E/data/sim_example
```

You should observe two matplotlib plots: 1) photon field and 2) photon count. You may need to install h5py, matplotlib and numpy to run this script.

5) Let's run the orientation recovery simulation using EMC:

```
docker run -it -v /host/path/simS2E:/simS2E emc:v0.1 /bin/bash
```

You are now inside the Docker container running bash on Ubuntu v14.04. The EMC package is installed under /home/packages and the simS2E directory is located under /simS2E.

Go to the workflow directory and run the example. This will generate the 3D diffraction volume after orientation recovery. Note that EMC may take many hours to converge to a solution. On my Linux box, it takes about a day:

```
cd /simS2E/workflow
./runEMC
```

Let's open the simulation configuration file again in /simS2E/config/config_sim_example. ORIENT=EMC specifies the EMC algorithm for orientation recovery. The EMC parameters are defined under ###### EMC ######. When the simulation is complete, exit the docker container by typing "exit" or Contrl+D. EMC output hdf5 file will be in /host/path/simS2E/data/sim_example/orient.

6) Let's run the phase retrieval simulation using DM:

```
docker run -it -v /host/path/simS2E:/simS2E dm:v0.1 /bin/bash
```

DM ###### You are now inside the Docker container running bash on Ubuntu v14.04. The DM package is installed under /home/packages and the simS2E directory is located under /simS2E.

Go to the workflow directory and run the example. This will generate the 3D electron density.:

```
cd /simS2E/workflow
./runDM
```

Let's open the simulation configuration file again in /simS2E/config/config_sim_example. PHASE=DM specifies the Difference Map algorithm for phase retrieval. The DM parameters are defined under ###### DM ######. When the simulation is complete. Exit the docker container by typing "exit" or Contrl+D. DM output hdf5 file will be in /host/path/simS2E/data/sim_example/phase.

1.4 Setting up Sphinx for documenting simS2E simulation

You need clone the simS2E repository from GitHub:

```
git clone https://github.com/chuckie82/start-to-end.git
```

The index.rst is the master ReST for your project.

You may already have sphinx installed – you can check by doing:

```
python -c 'import sphinx'
```

If that fails install the latest version with:

```
> sudo easy_install -U Sphinx
```

Let's see if we can build our html:

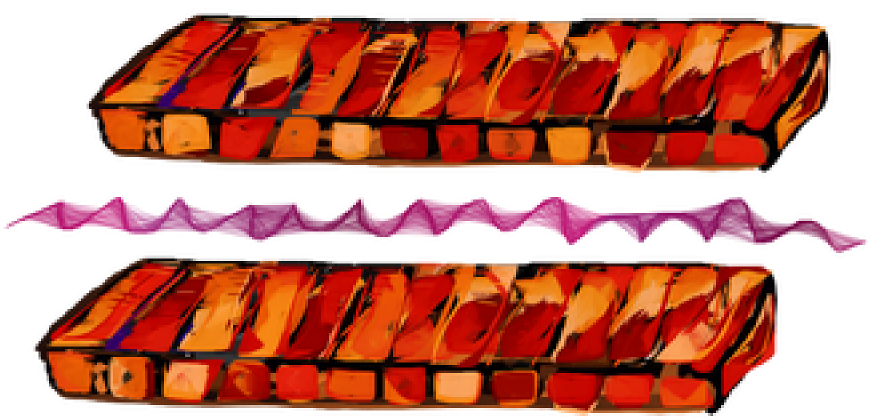
```
make html
```

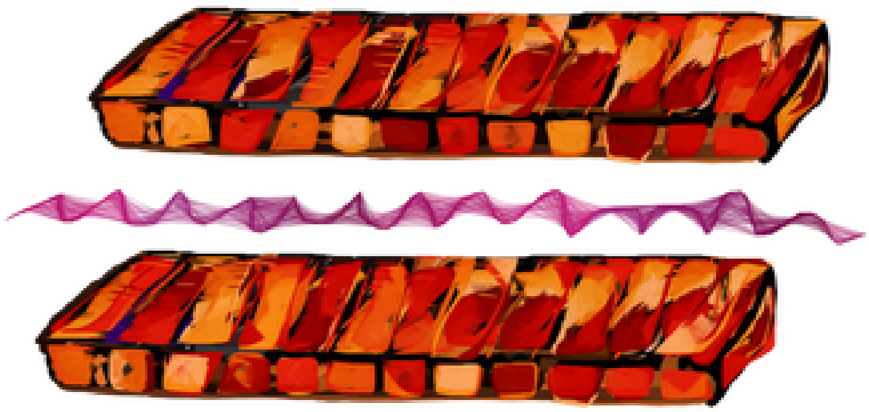
If you now open your favorite internet browser and type `_build/html/index.html`, you should see the documentation website.

To update the document on the web, just push your changes:

```
git add *.rst
git commit -m "Update all documents"
git push -u origin master
```

That's it! Now you are ready to





FEL source simulation

2.1 Introduction

Documentation for FEL source simulation can be found on this page.

2.2 Data access

Data in archive can be exported using web browser. Initial FEL source can be downloaded from here:

FEL source web site

with authentication (xfel/desy account)

FEL source web site

If you use this dataset, please acknowledge *[SALDIN99]*.

2.3 Output data description

The output data is expected in hdf5 format, and the glossary can be found below. FEL source module is responsible for writing out in the format specified below.

2.3.1 FELsource_out_<7 digit ID>.h5 (Output HDF glossary)

Field name	Description
data/	
data/arrEhor	Complex EM field written in 4D array, horizontal polarization
data/arrEver	Complex EM field written in 4D array, vertical polarization
params/	Parameters for wavefront propagation
params/Mesh/nSlices	Numbers of points vs photon energy/time for the pulse
params/Mesh/nx	Numbers of points, horizontal
params/Mesh/ny	Numbers of points, vertical
params/Mesh/sliceMax	Max value of time [s] or energy [ev] for pulse (fragment)
params/Mesh/sliceMin	Min value of time [s] or energy [ev] for pulse (fragment)
params/Mesh/xMax	Maximum of horizontal range

Table 2.1 – continued from previous page

Field name	Description
params/Mesh/xMin	Minimum of horizontal range
params/Mesh/yMax	Maximum of vertical range
params/Mesh/yMin	Minimum of vertical range
params/Mesh/zCoord	Longitudinal position, for FEL output data - length of active undulator
params/Rx	Instantaneous horizontal wavefront radius
params/Ry	Instantaneous vertical wavefront radius
params/dRx	Error of wavefront horizontal radius
params/dRy	Error of wavefront vertical radius
params/nval	complex electric field nval==2
params/photonEnergy	Average photon energy
params/wDomain	Wavefront in time or frequency (photon energy) domain
params/wEFieldUnit	Electric field units, {sqrt(W/mm^2) (time domain), arbitrary}
params/wFloatType	Electric field numerical type
params/wSpace	R-space or Q-space wavefront presentation
params/xCentre	Horizontal transverse coordinates of wavefront instant 'source center'
params/yCentre	Vertical transverse coordinates of wavefront instant 'source center'
history/parent/info/	**Information about input data**
history/parent/info/ contact	Contact Information
history/parent/info/ data_description	Description of FEL data
history/parent/info/ method_description	Method description
history/parent/info/ package_version	Package version
misc/	**Complimentary information**
history/parent/misc/ FAST2XY.DAT	FELsource_params_FAST2XY.txt used for post-processing FAST output
history/parent/misc/ angular_distribution	radial distribution of far field divergence
history/parent/misc/ spot_size	near field transverse FEL beam size (FWHM)
history/parent/misc/ gain_curve	gain curve, dependence of FEL pulse energy (column 2) from number of working point (c
history/parent/misc/nzc	number of working point defines active undulator length
history/parent/misc/ temporal_struct	FEL pulse temporal structure, instantaneous power P(tau)
version	**hdf5 format version**

2.4 Diagnostic (diagnostic_felsrc.py)

Fig.1. Pulse irradiance XY map (number of photons per pixel), the title contains size of the pixel;

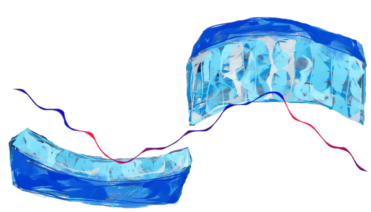


Fig.2. Pulse time structure, the title contains the pulse energy value.

CHAPTER 3

Propagation, including optics

3.1 Input data description

The input data is expected in hdf5 format, and the glossary can be found in the link below.

3.1.1 prop_out_<7 digit ID>.h5 (Output HDF glossary)

Field name	Description
data/	
data/arrEhor	Complex EM field written in 4D array, horizontal polarization
data/arrEver	Complex EM field written in 4D array, vertical polarization
params/	Parameters for wavefront propagation
params/Mesh/nSlices	Numbers of points vs photon energy/time for the pulse
params/Mesh/nx	Numbers of points, horizontal
params/Mesh/ny	Numbers of points, vertical
params/Mesh/qxMax	Maximum of horizontal frequency (If params/wSpace is Q-space)
params/Mesh/qxMin	Minimum of horizontal frequency (If params/wSpace is Q-space)
params/Mesh/qyMax	Maximum of vertical frequency (If params/wSpace is Q-space)
params/Mesh/qyMin	Minimum of vertical frequency (If params/wSpace is Q-space)
params/Mesh/sliceMax	Max value of time [s] or energy [ev] for pulse (fragment)
params/Mesh/sliceMin	Min value of time [s] or energy [ev] for pulse (fragment)
params/Mesh/xMax	Maximum of horizontal range (If params/wSpace is R-space)
params/xMin	Minimum of horizontal range (If params/wSpace is R-space)
params/yMax	Maximum of vertical range (If params/wSpace is R-space)
params/yMin	Minimum of vertical range (If params/wSpace is R-space)
params/zCoord	Longitudinal position, for FEL output data - length of active undulator
params/beamline/printout	(add description)
params/Rx	Instantaneous horizontal wavefront radius
params/Ry	Instantaneous vertical wavefront radius
params/dRx	Error of wavefront horizontal radius
params/dRy	Error of wavefront horizontal radius
params/nval	complex electric field nval==2
params/photonEnergy	Average photon energy
params/wDomain	Wavefront in time or frequency (photon energy) domain
params/wEFieldUnit	Electric field units, sqrt(Phot/s/0.1%BW/mm^2), sqrt(W/mm^2) for time domain, sqrt(J/eV/mm^2) for fr
params/wFloatType	Electric field numerical type

Table 3.1 – continued from previous page

Field name	Description
params/wSpace	R-space or Q-space wavefront presentation
params/xCentre	Horizontal transverse coordinates of wavefront instant 'source center'
params/yCentre	Vertical transverse coordinates of wavefront instant 'source center'
info/	
info/package_version	Package version
info/contact	Contact details of author
info/data_description	Short description of what the data is
info/method_description	Short description of what method was used to generate the data
history	**Information about input data**
misc/	**Complimentary information**
misc/xFWHM	FWHM belong x-axis
misc/yFWHM	FWHM belong y-axis
version	**hdf5 format version**

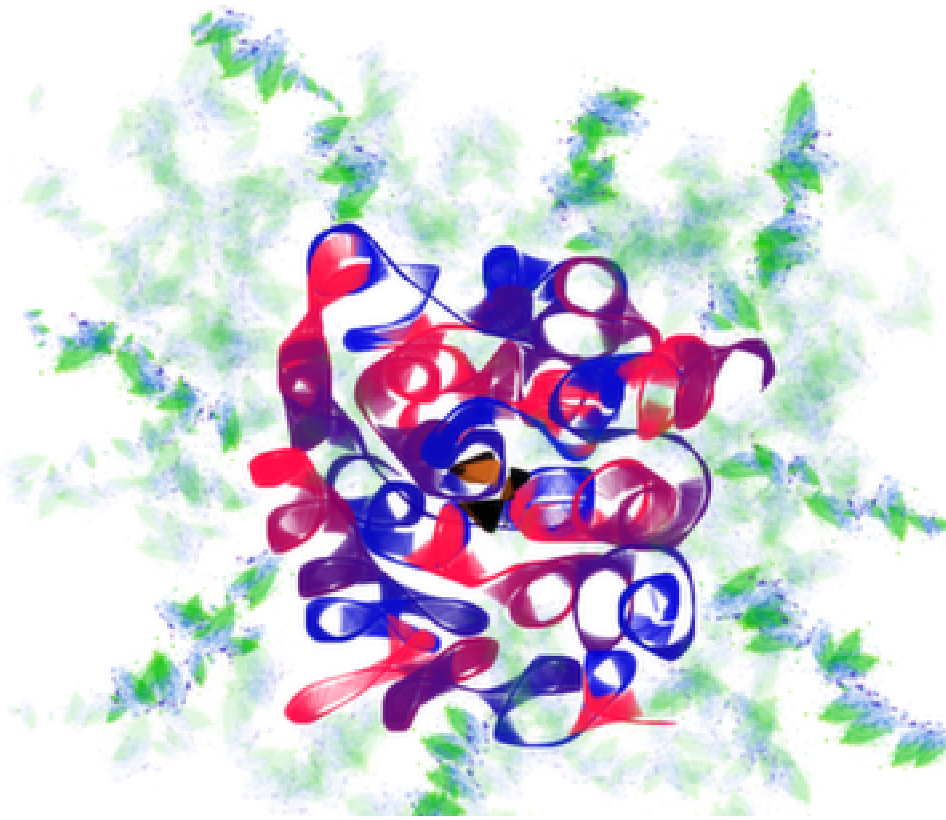
3.2 Diagnostic (diagnostic.py)

Fig.1. Pulse irradiance XY map (number of photons per pixel), the title contains size of the pixel;

Fig.2. Plot of pulse time structure before and after propagating, the title contains the propagated pulse energy value.

3.3 About WPG

WPG, WaveProperGator is an interactive simulation framework for coherent X-ray wavefront propagation. WPG provides intuitive interface to the SRW library. The application examples oriented on European XFEL design parameters.



Online documentation page

CHAPTER 4

Photon Matter Interaction

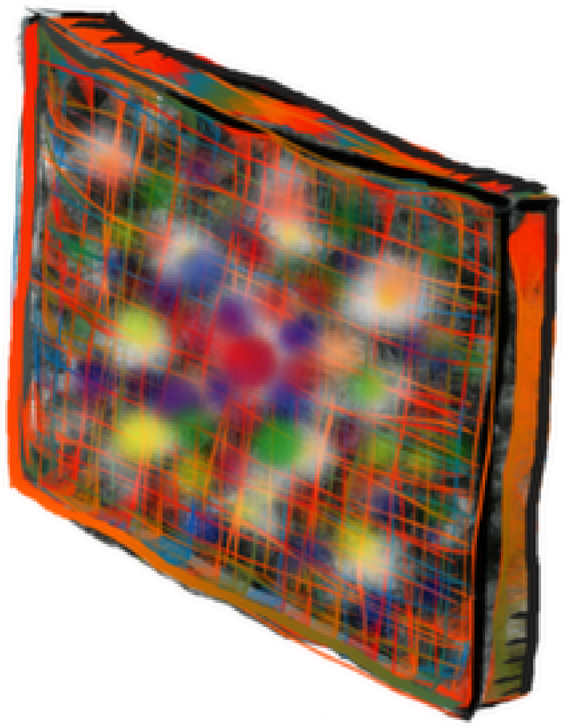
4.1 Input/Output data description

The input/output data is expected in hdf5 format, and the glossary can be found below. Photon matter interaction module is responsible for reading in and writing out in the format specified below.

4.1.1 pmi_out_<7 digit ID>.h5 (Output HDF glossary)

Field name	Description	Data type	Units
data/			
data/snp_<7 digit index>/ff	Atomic form factor in 2D array (number of unique ions x length of atomFormFactorQ)	Float	
data/snp_<7 digit index>/halfQ	Reciprocal space spanned by the atomic form factor in 1D array (number of samples of reciprocal q)	Float	1/Å
data/snp_<7 digit index>/Nph	Numbers of photons in the beam	Int	ph
data/snp_<7 digit index>/r	Atomic position in real space in 2D array (number of ions x 3D coordinates {x,y,z})	Float	
data/snp_<7 digit index>/T	List of unique ID numbers given to each atomFormFactor in 1D array (number of unique ions)	Int	
data/snp_<7 digit index>/Z	List of atomType present at atomPosition in 1D array (number of ions x number of frames)	Int	
data/snp_<7 digit index>/xyz	List of indices of ff for each atom in Z	Int	
data/snp_<7 digit index>/Sq_halfQ	Reciprocal space spanned by the Compton scattering in 1D array (number of samples of reciprocal q)	Float	1/Å
data/snp_<7 digit index>/Sq_bound	Compton scattering by bound electrons in 1D array (length of Sq_Q)	Float	
data/snp_<7 digit index>/Sq_free	Compton scattering by free electrons in 1D array (length of Sq_Q)	Float	
history	**Information about input data**		
history/parent/detail	Details of the parent inclusing /data, /info, /misc, /params		
history/parent/parent	Iteratively list parent modules		
info/			
info/package_version	Package version		
info/contact	Contact details of author		
info/data_description	Short description of what the data is		
info/method_description	Short description of what method was used to generate the data		
misc/	**Miscellaneous information**		
params/	**Parameters used to run the module**		
version	**hdf5 format version**	Float	0.1

4.2 Python script for HDF



Script on bitbucket

Coherent Diffraction

5.1 Input/Output data description

The input/output data is expected in hdf5 format, and the glossary can be found below. Coherent diffraction module is responsible for reading in and writing out in the format specified below.

5.1.1 diffr_params_SingFEL (Input Parameter glossary)

Field name	Description	DataType
–input_dir	Input directory where pmi_out files are stored	String
–output_dir	Output directory where diffr_out files will be stored	String
–config_file	Full path and filename of this file	String
-b	Experimental beam file	String
-g	Experimental geometry file	String
–uniformRotation	Rotations are selected uniformly in given rotation space	Int
–calculateCompton	Calculate Compton scattering in diffraction pattern	Int
–sliceInterval	Interval to calculates diffraction	Int
–numSlices	Number of time slices to use for calculating diffraction	Int
–pmiStartID	Start ID of PMI trajectory	Int
–pmiEndID	End ID of PMI trajectory	Int
–dpID	Diffraction pattern index for current pmiID	Int
–numDP	Number of diffraction patterns to generate per pmiID	Int
–USE_GPU	Options to use GPU (1) or not (0)	Int
version	SingFEL version	0.1

5.1.2 diffr_out_<7 digit ID>.h5 (Output HDF glossary)

Field name	Description	Data type	Units
data/			
data/data	Diffraction pattern in 2D matrix	Float	
data/diffr	Diffracted intensity before Poisson noise (Optional)	Float	
data/angle	Additional rotation applied to the rotated pmi_out position. Initial rotation angle can be found in pmi_out/data/angle. Active right handed rotations applied in quaternion.	Float	
history/	Information about input data		
history/parent/detail	Details of the parent including /data, /info, /misc, /params /data should be soft-linked with a relative path		
history/parent/parent	Iteratively list parent modules		
info/	Information		
info/package_version	Package name and version	String	
info/contact	Contact details of author	String	
info/data_description	Short description of what the data is	String	
info/method_description	Short description of what method was used to generate the data	String	
misc/	Miscellaneous information		
params/	Parameters used for coherent diffraction		
params/geom/detectorDist	Detector distance from point of interaction	Float	m
params/geom/pixelWidth	Pixel width	Float	m
params/geom/pixelHeight	Pixel height	Float	m
params/geom/mask	Mask of a diffraction pattern to indicate	Int	
	pixel ON (1) or OFF (0) in 2D array		
params/beam/photonEnergy	Photon energy	Float	eV
params/beam/photons	Number of photons in the beam	Int	ph
params/beam/focusArea	Beam focus area	Float	
params/info	Input for Coherent diffraction program	String	
version	hdf5 format version	Float	0.1

5.2 Diagnostic

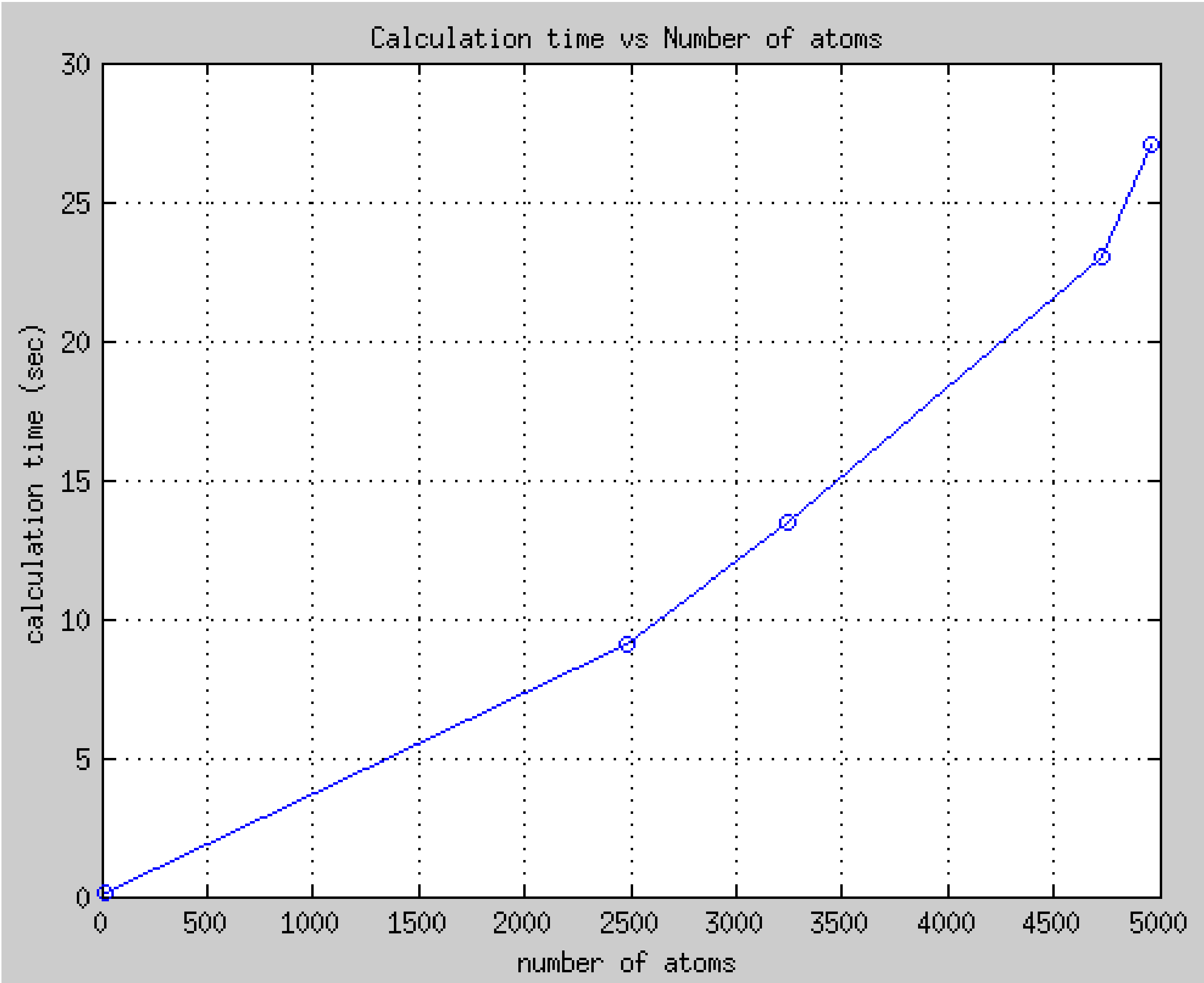
Python script displays /data/data and /data/diffr at completion of the module execution.

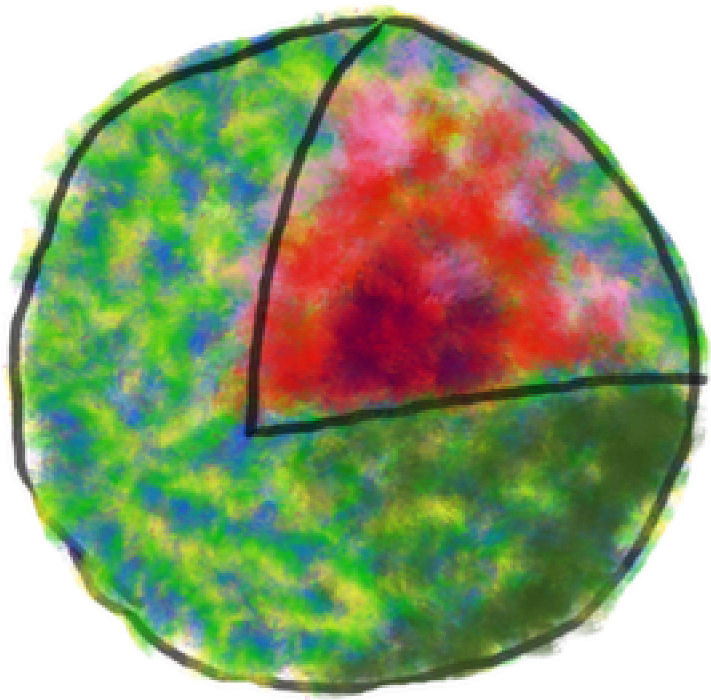
5.3 Scaling behaviour of SingFEL

Calculation time using single processor vs number of atoms is non-linear, perhaps quadratic.

Detector number of pixels: 131x131

Benzoic acid: 15 atoms Chignolin: 2484 atoms 2YBE: 3240 atoms 2NIP: 4735 atoms 4AS4: 4963 atoms


Calculation time vs Number of atoms
calculation time (sec)
number of atoms
0
5
10
15
20
25
30
0
500
1000
1500
2000
2500
3000
3500
4000
4500
5000



Orientation Determination

6.1 Input/Output data description

Test

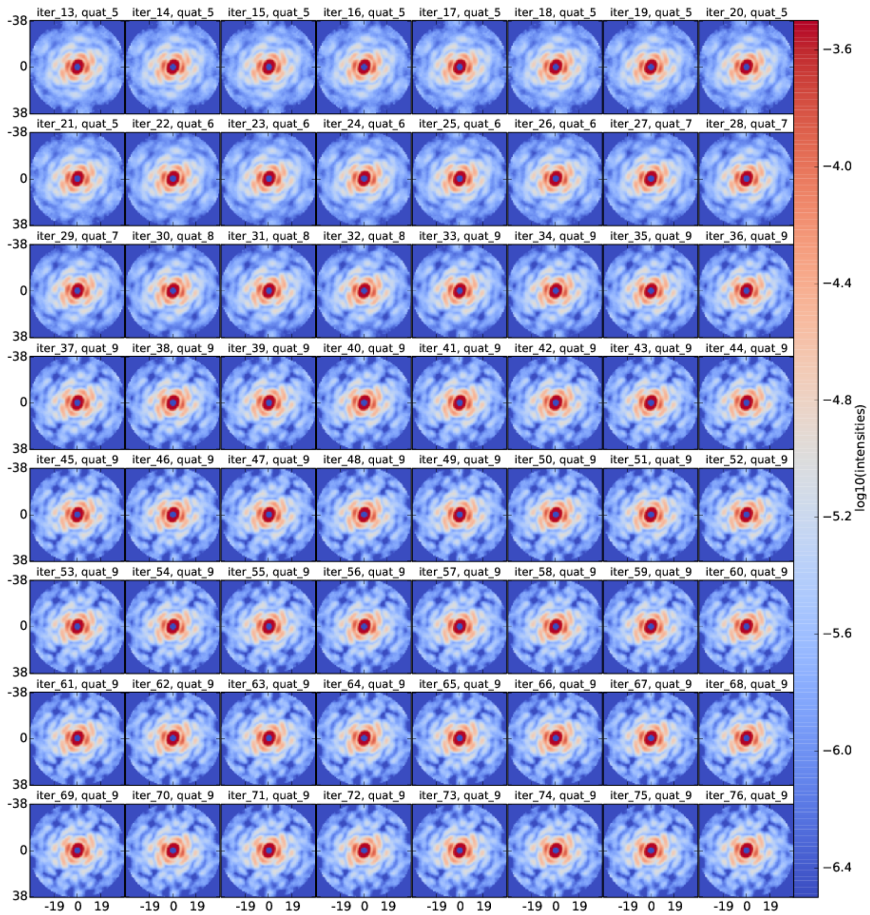
The input/output data is expected in hdf5 format, and the glossary can be found below. Orientation determination module is responsible for reading in and writing out in the format specified below.

6.1.1 orient_out_<7 digit ID>.h5 (Output HDF glossary)

Field name	Description	Data type	Units
data/			
data/data	Diffraction volume in 3D array (dimX x dimY x dimZ)	Float	
data/angle	Most likely orientation	Float	
data/center	center of diffraction volume x,y,z	Int	pixels
params/	Parameters used for coherent diffraction		
params/info	Input for orientation determination program and version	String	
history/	Information about input data		
info/	Information		
misc/	Miscellaneous information		
version	hdf5 format version	Float	0.1

6.2 Diagnostics of reconstructed 3D diffraction volume

EMC reconstruction parameters..


iter_13, quat_5
iter_14, quat_5
iter_15, quat_5
iter_16, quat_5
iter_17, quat_5
iter_18, quat_5
iter_19, quat_5
iter_20, quat_5
iter_21, quat_5
iter_22, quat_6
iter_23, quat_6
iter_24, quat_6
iter_25, quat_6
iter_26, quat_6
iter_27, quat_7
iter_28, quat_7
iter_29, quat_7
iter_30, quat_8
iter_31, quat_8
iter_32, quat_8
iter_33, quat_9
iter_34, quat_9
iter_35, quat_9
iter_36, quat_9
iter_37, quat_9
iter_38, quat_9
iter_39, quat_9
iter_40, quat_9
iter_41, quat_9
iter_42, quat_9
iter_43, quat_9
iter_44, quat_9
iter_45, quat_9
iter_46, quat_9
iter_47, quat_9
iter_48, quat_9
iter_49, quat_9
iter_50, quat_9
iter_51, quat_9
iter_52, quat_9
iter_53, quat_9
iter_54, quat_9
iter_55, quat_9
iter_56, quat_9
iter_57, quat_9
iter_58, quat_9
iter_59, quat_9
iter_60, quat_9
iter_61, quat_9
iter_62, quat_9
iter_63, quat_9
iter_64, quat_9
iter_65, quat_9
iter_66, quat_9
iter_67, quat_9
iter_68, quat_9
iter_69, quat_9
iter_70, quat_9
iter_71, quat_9
iter_72, quat_9
iter_73, quat_9
iter_74, quat_9
iter_75, quat_9
iter_76, quat_9
-38
0
38
-19 0 19
−3.6
−4.0
−4.4
−4.8
−5.2
−5.6
−6.0
−6.4
log10(intensities)

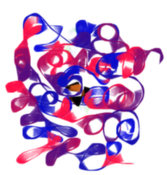
Phasing

7.1 Input/Output data description

The input/output data is expected in hdf5 format, and the glossary can be found below. Phasing module is responsible for reading in and writing out in the format specified below.

7.1.1 phase_out_<7 digit ID>.h5 (Output HDF glossary)

Field name	Description	Data type	Units
data/			
data/electronDensity	Recovered electron density volume in 3D array (dimX x dimY x dimZ)	Float	
params/			
params/info	Input for phasing program and version	String	
history/	Information about input data		
info/	Information		
misc/	Miscellaneous information		
version	hdf5 format version	Float	0.1



CHAPTER 8

Indices and tables

- genindex
- modindex
- search

Bibliography

[SALDIN99] 5. (a) Saldin, E. A. Schneidmiller, and M. V. Yurkov. *Nucl. Instrum. and Methods*, A(429):233, 1999.