

VARIATIONAL-INFERENCE INTEGRATION AT LCLS enables scalable probabilistic intensity estimation for serial X-ray crystallography.

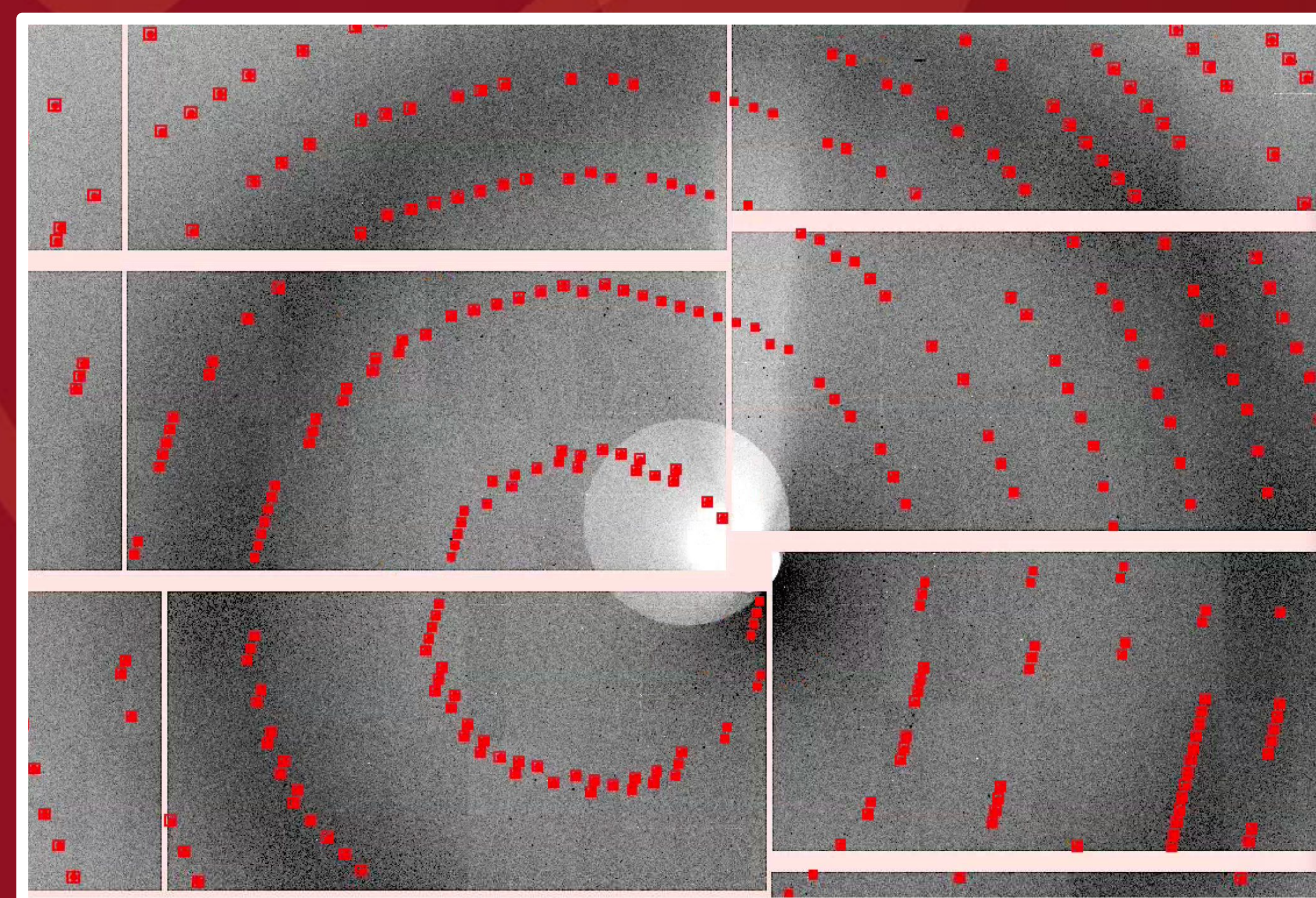
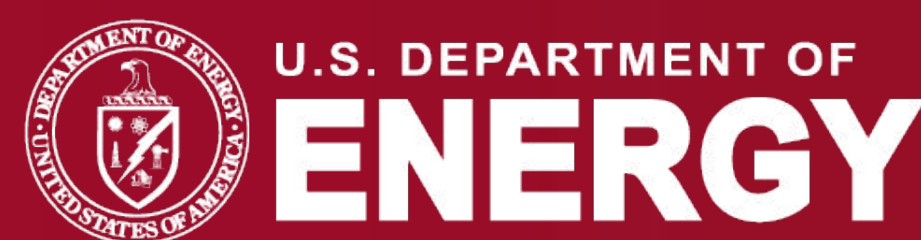
Thao Huynh¹, Luis A. Aldama², Kevin M. Dalton³

Affiliations

¹ College of Engineering, San José State University

² Harvard Biophysics Graduate Program, Harvard University

³ Linac Coherent Light Source, SLAC National Accelerator Laboratory



1. Background & Motivation

- MFX produces challenging FEL diffraction data from many crystals.
- Standard processing uses **summation integration** without explicitly modeling spot shape.
- In conventional crystallography, **profile modeling improves weak-reflection estimates and resolution.**
- Classical profile models are difficult for FEL stills; **variational inference can learn profiles directly from data.**

Motivation: Can learned spot profiles recover weaker, higher-resolution MFX reflections?

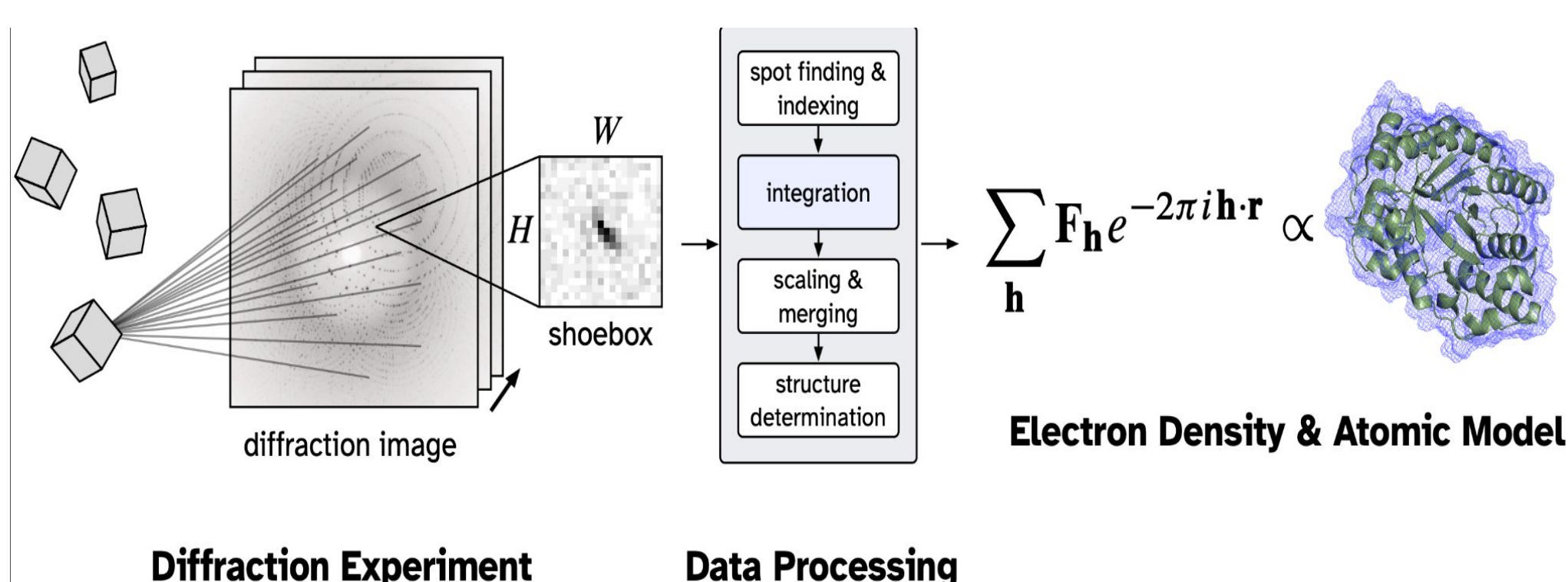


Figure 1. Serial femtosecond crystallography workflow.

2. Research Goal

Evaluate Integrator on real MFX data vs. cctbx.xfel

- Adapt Integrator to **MFX/Jungfrau**
- Test alternative **intensity models**
- Compare **merging** and **refinement**

3. Variational-Inference Method

Integrator models each shoebox probabilistically.

- Learns **intensity, background, and spot profile**
- Outputs **distributions + uncertainty**
- Uses a **Wilson prior** to regularize intensities

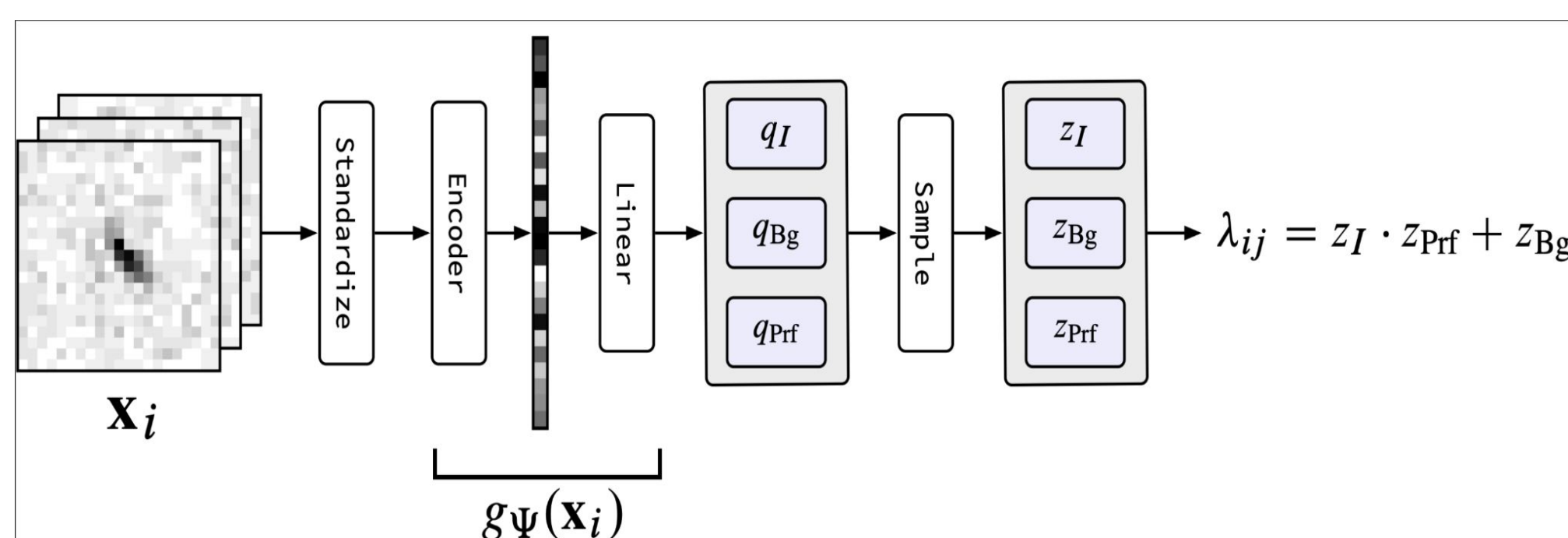


Figure 2. Variational-inference architecture.

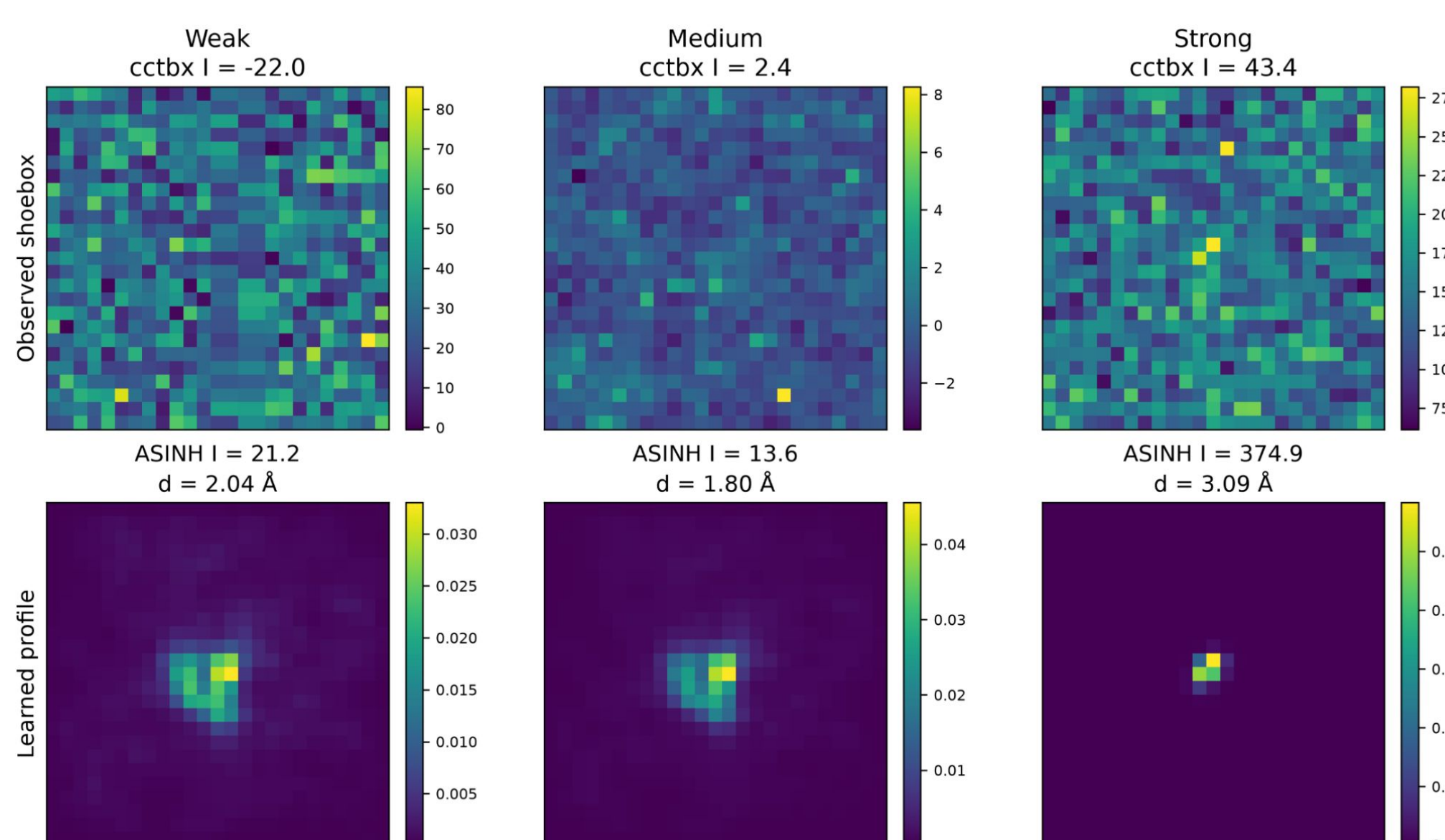


Figure 3. Shoeboxes & Learned Profiles - ASINH.

MFX/INTEGRATOR CODE

Scan to explore →



GitHub [link](#)

4. MFX Workflow / What I Built

Extended Integrator to process large-scale experimental MFX data end-to-end.

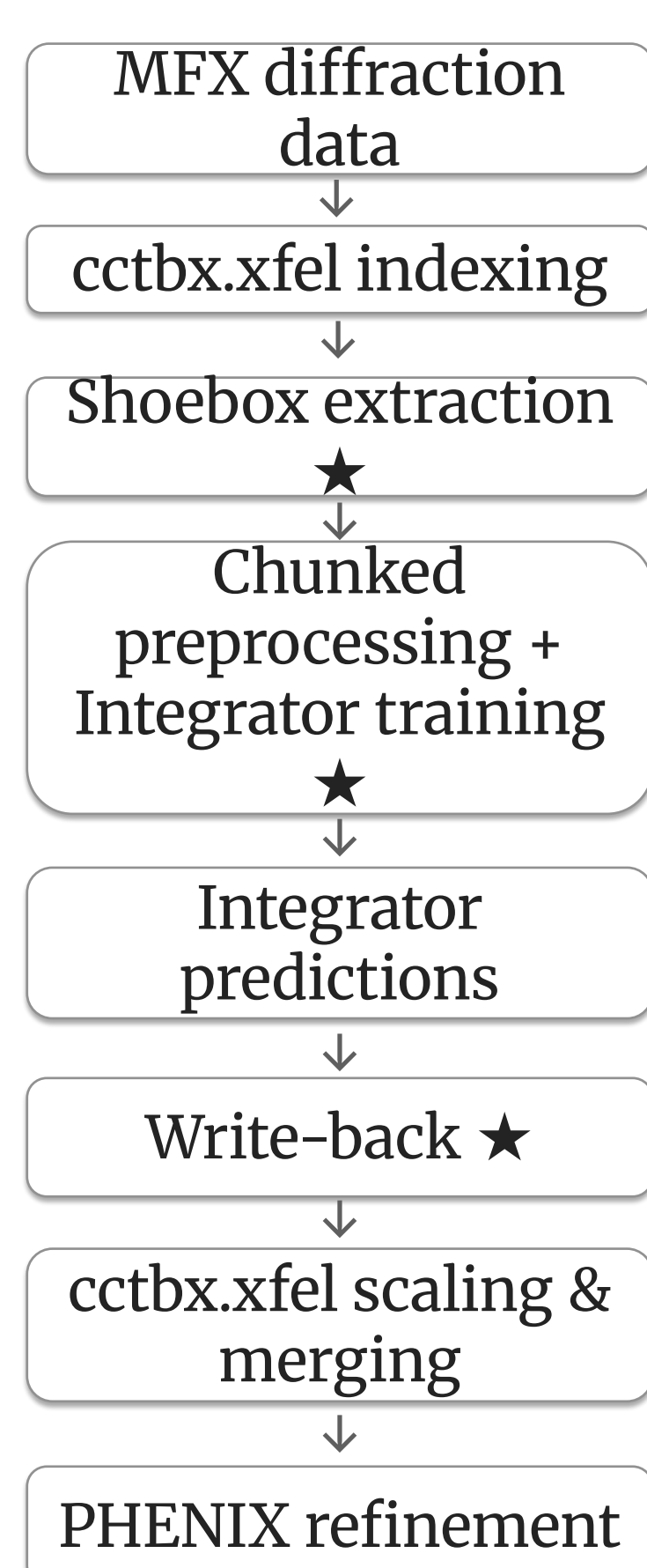


Figure 4. MFX/Integrator workflow.

5. Crystallographic Results

5.1 CC_{1/2} by resolution

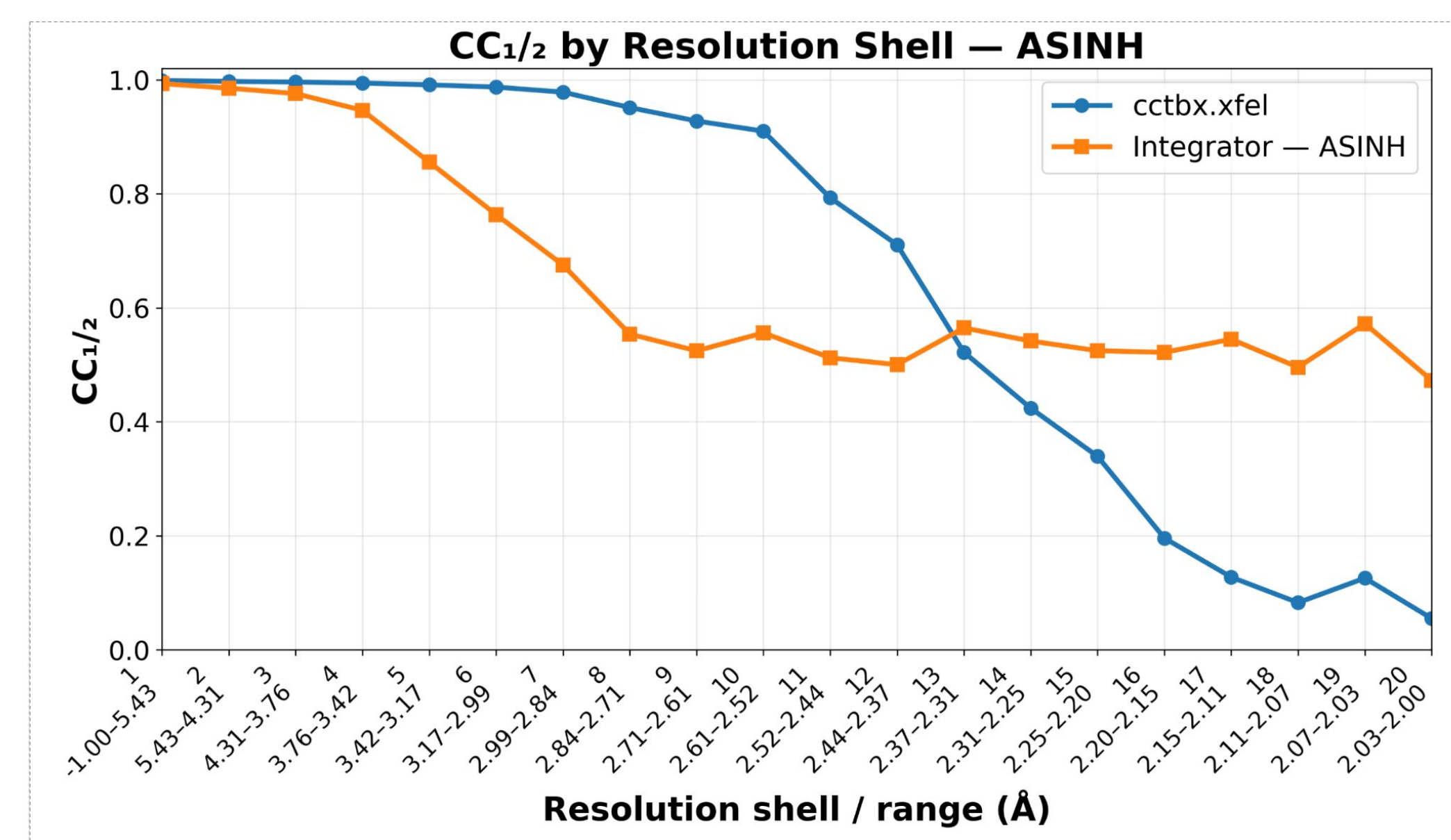
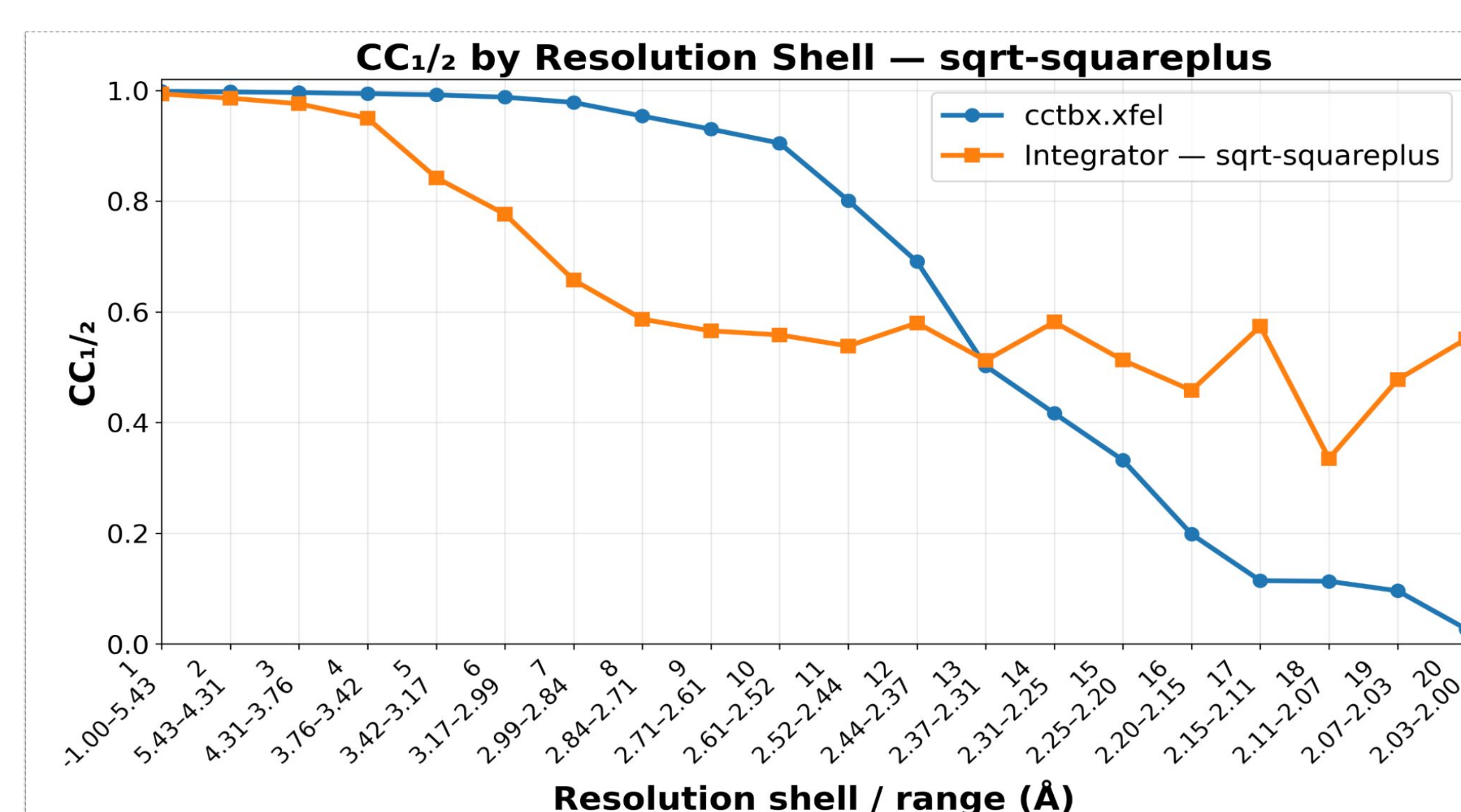


Figure 5. CC_{1/2} vs resolution: cctbx.xfel vs Integrator.

ACKNOWLEDGEMENTS

Grateful to my mentor **Kevin Dalton**, **Luis Aldama**, and the **LCLS Cryptochrome4-SFX** group for their mentorship and support, and to **SLAC** and **SJSU** for making this research experience possible.

5.2 Overall merging statistics

Method	CC _{1/2} ↑	Rint ↓	Rsplitted ↓	N exp.
cctbx (matched)	0.998	6.9%	5.2%	~113.5k
Integrator — ASINH	0.990	15.8%	11.1%	113.8k
Integrator — sqrt	0.990	15.8%	11.1%	113.9k

Table 1. cctbx subsets matched the corresponding Integrator survivor sets.

CC_{1/2} ↑: agreement; Rint/Rsplit ↓: disagreement

6. Structural Refinement

6.1 R-work / R-free

Refinement treatment

- cctbx: French-Wilson on
- Integrator: French-Wilson off

Method	R-work ↓	R-free ↓	Reflections
cctbx — ASINH match	0.1788	0.2132	44,236
Integrator — ASINH	0.2610	0.2768	45,923
cctbx — sqrt match	0.1788	0.2135	44,276
Integrator — sqrt	0.2618	0.2753	45,891

Table 2. Structural refinement statistics.

R-work: fit to refined data · R-free: fit to held-out data · ↓ lower is better

6.2 Protein Structure

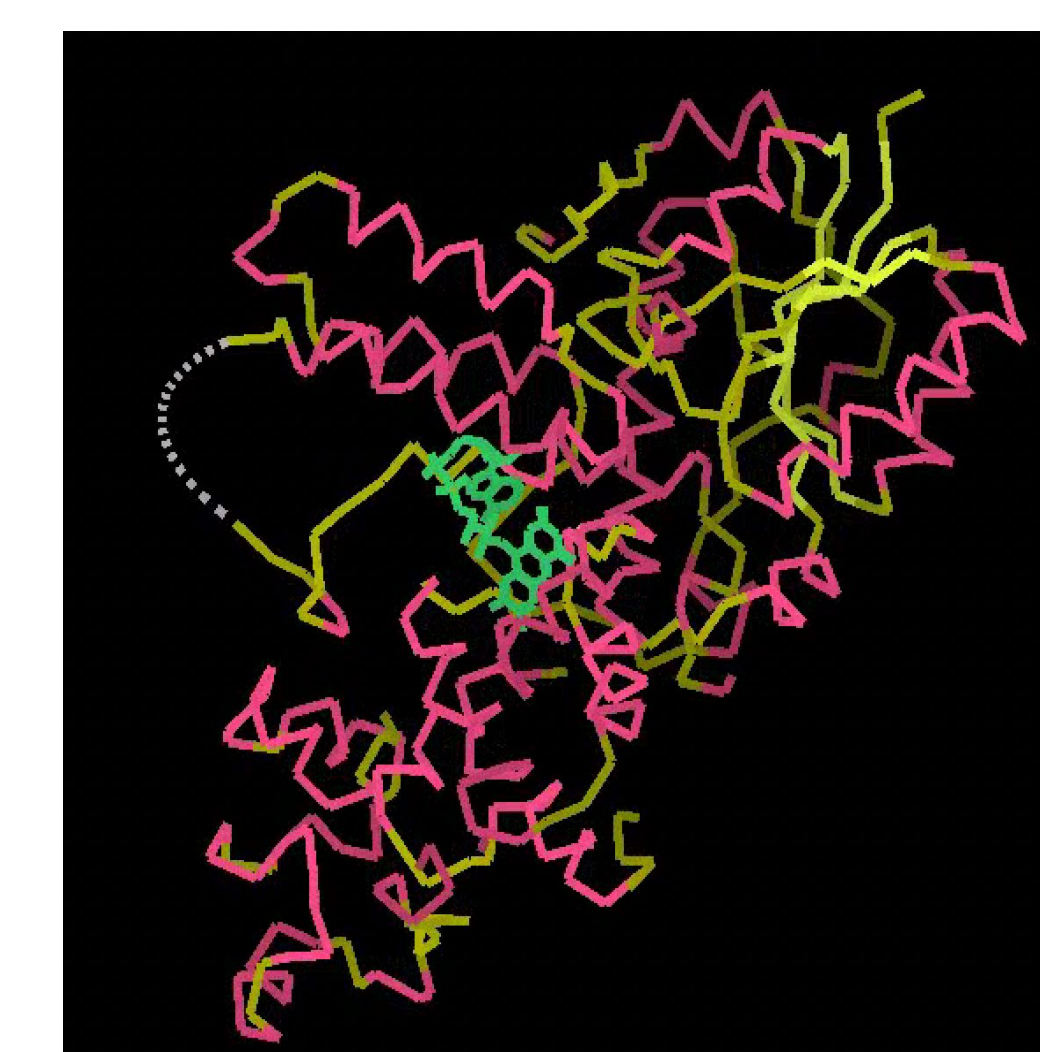


Figure 6. Refined sqrt-squareplus structure in Coot

6.3 Electron Density

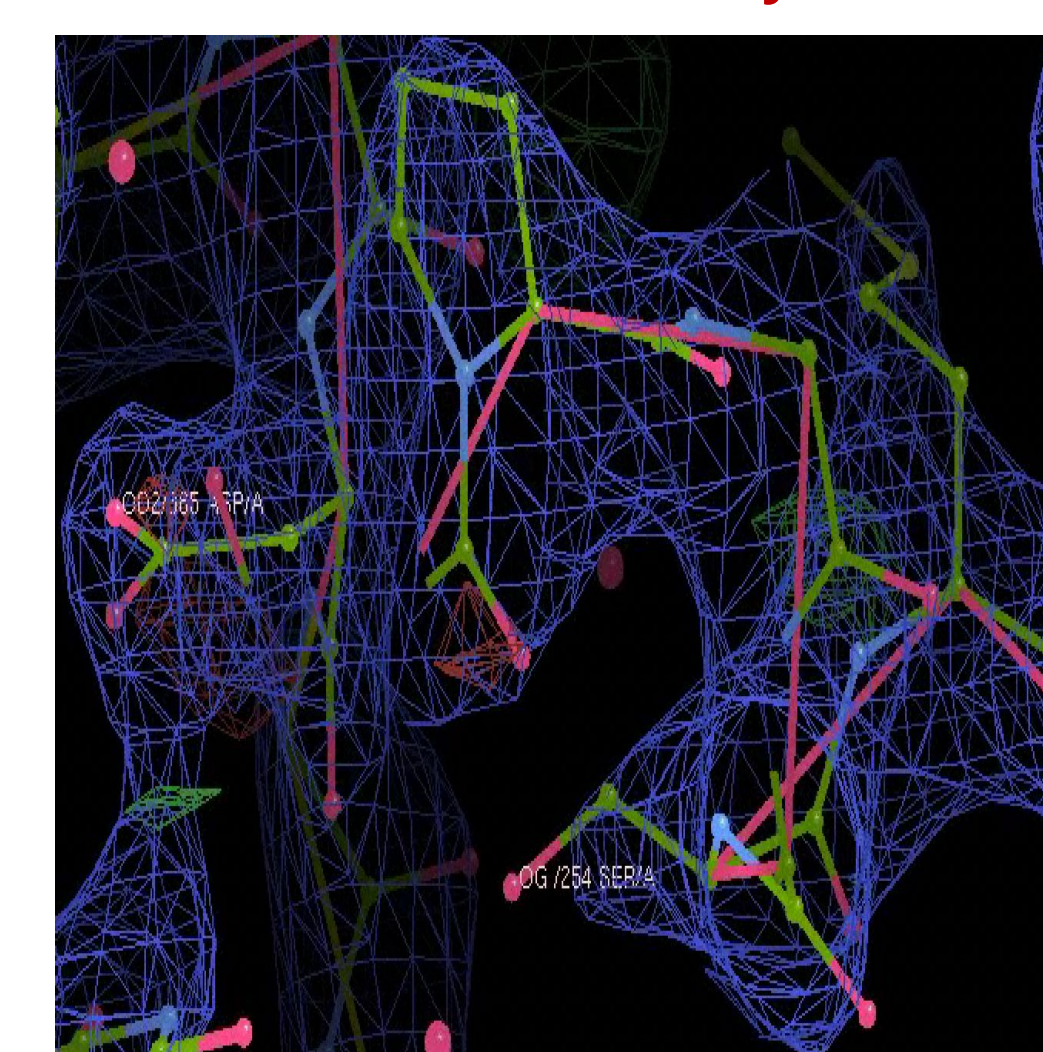


Figure 7. sqrt-squareplus 2mFo-DFc electron density at ~1.6σ in Coot

7. Conclusions & Future Work

7.1 Conclusions

- Established **proof of concept** for end-to-end **variational-inference integration** on real MFX data.
- Scaled to **100M+ reflections** and produced refinable structures.
- cctbx.xfel performs better overall; high-resolution CC_{1/2} requires further validation.

7.2 Future Work

- Train longer with improved B/G initialization.
- Replace global B/G with **per-image B/G**.
- Investigate high-resolution CC_{1/2} behavior.

Scalable · Probabilistic · Refinable

REFERENCES

1. Aldama, L. A., Dalton, K. M., & Hekstra, D. R. *Integrating Macromolecular X-ray Diffraction Data with Variational Inference*.
2. Brewster, A. S. et al. *cctbx.xfel*.
3. Liebschner, D. et al. *PHENIX*.