Physics Informed Foundation Models For Coherent Diffraction Imaging

Oliver Hoidn, Aashwin Mishra, Matt Seaberg, Apurva Mehta

SLAC NATIONAL ACCELERATOR LABORATORY

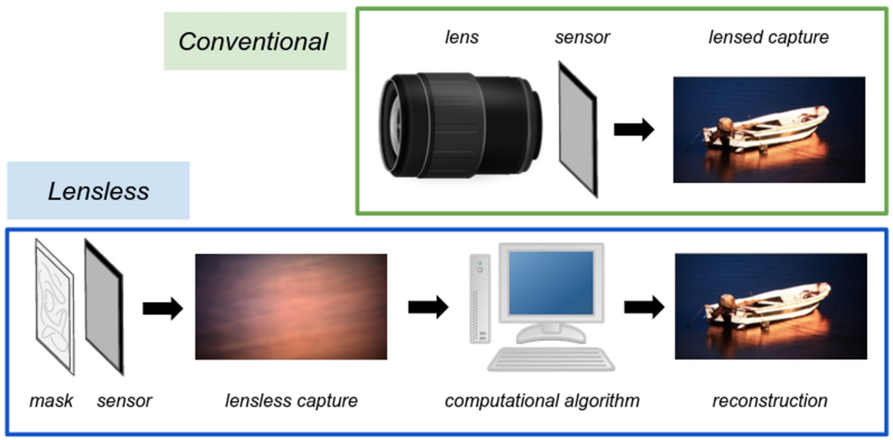
Stanford University | U.S. DEPARTMENT OF ENERGY

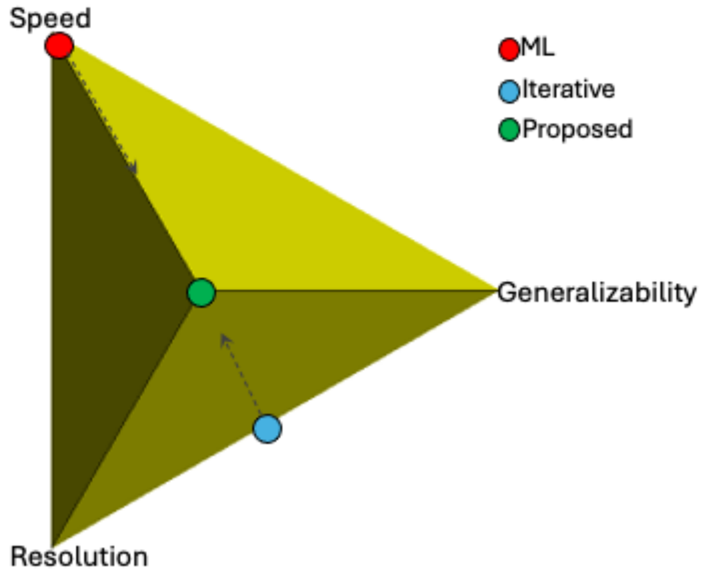
Lensless imaging is compute-bound

- Lensed Imaging: Real time feedback *but* lower resolution.
  Lensless Imaging: Higher resolution *but* no real time feedback.
  Proposed research: high resolution *and* real time feedback, & scaling with LCLS-II-HE.

- Extant iterative approaches are too slow.
  New ML approaches are fast, but unreliable & blurry reconstruction.

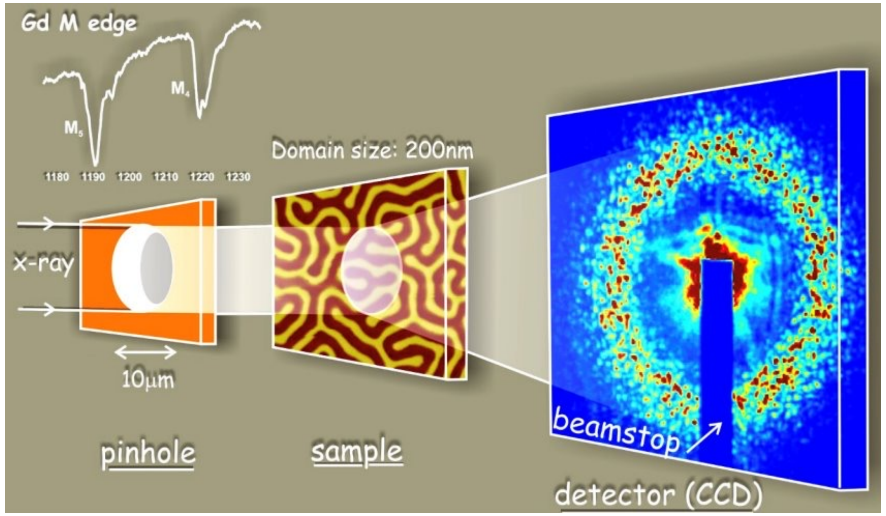
- Hybrid approach: ML speed & physics-based reliability.
  a) Physics Informed: Robust & Generalizable,
  b) Probabilistic: Reliability & Trust.
  c) Physics based Resolution Maximization
  d) Advanced architectures: data scaling & better reconstruction.

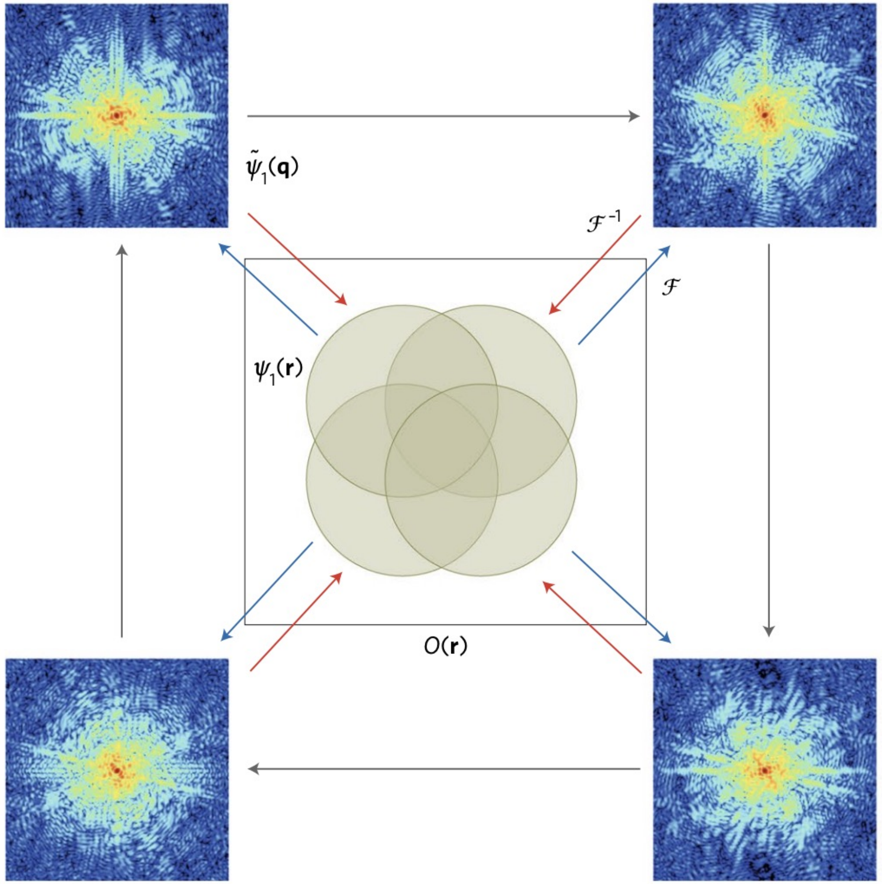


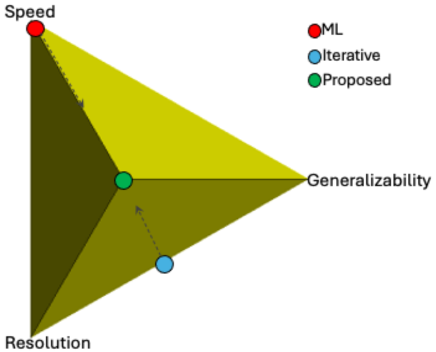


Physics+ML for High-Resolution, Real-Time, Reliable Imaging, scaling with LCLS-II-HE

Scanning Coherent Diffractive Imaging

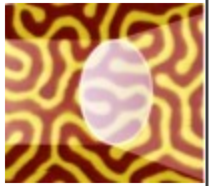


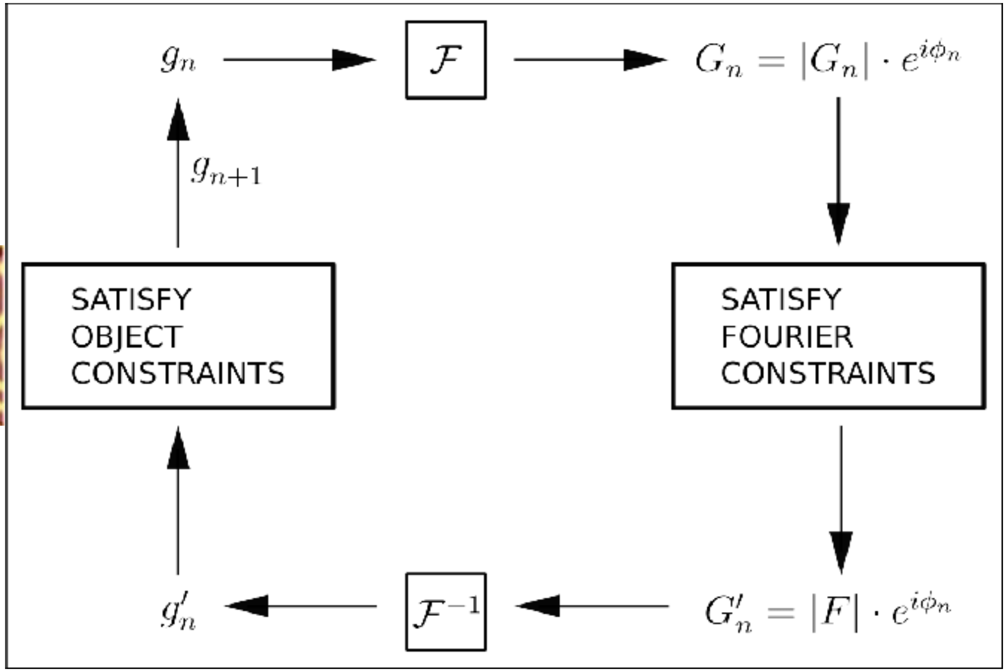


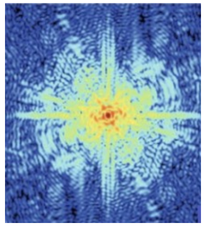


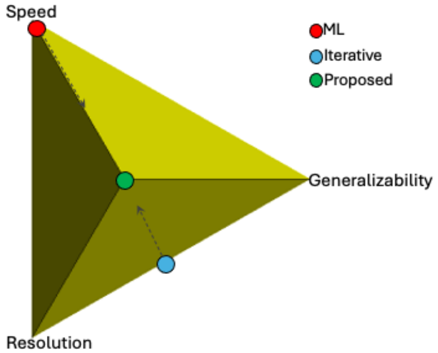
The challenge: iterative phase retrieval is slow

- Reconstruction is constraint-driven:
  - Intensity-matching in reciprocal space
  - Real-space constraints (e.g. shrink-wrapping, overlaps)
- Necessarily iterative and **slow**







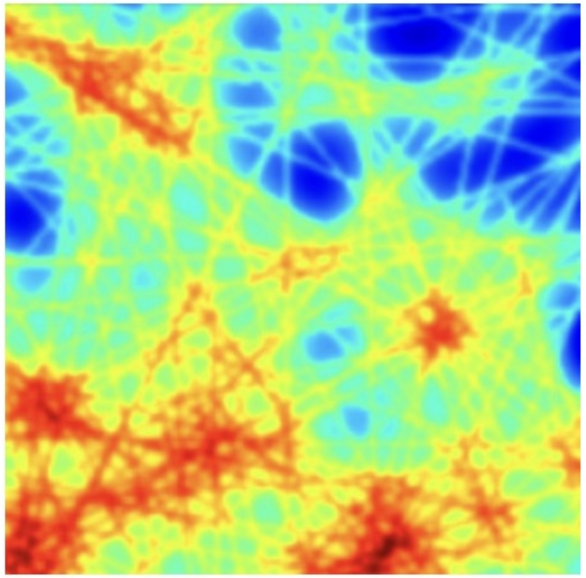


Physics-Informed Neural networks for generalizable reconstruction

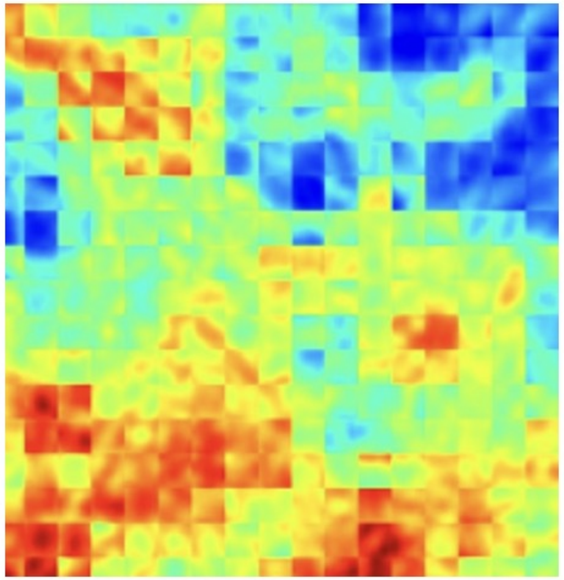
Proof of concept: deep learning-based scanning coherent diffractive imaging with physics-informed neural networks (PINNs) [1]

- Failure modes of prior, non-PINN deep learning approaches:
  - Inefficient training (i.e. poor generalization)
  - Unphysical inverse problem solution → low resolution + artifacts
  - Demanding in infrastructure and (**labeled**) data

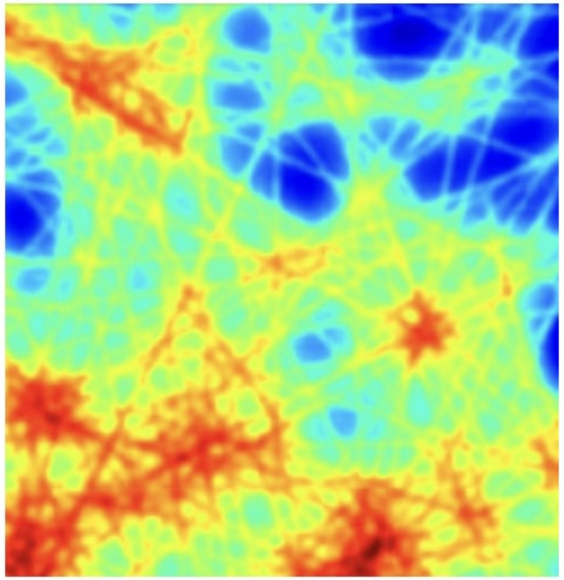
Ground truth



ML (baseline)



ML + physics

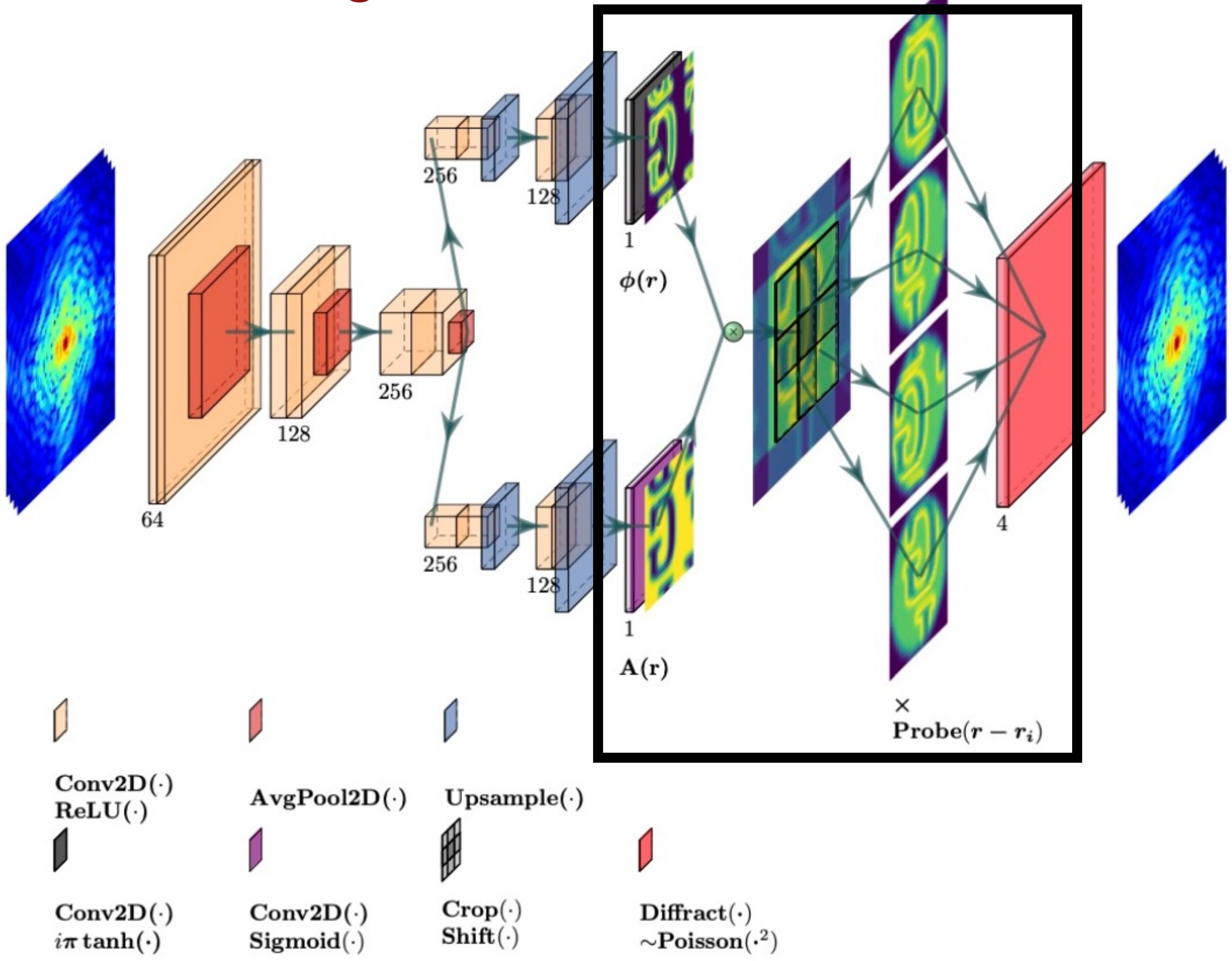


[1] Cherukara et al., doi.org/10.1063/5.0013065
[2] Hoidn et al., doi.org/10.1038/s41598-023-48351-7

Physics + ML gives best of both worlds

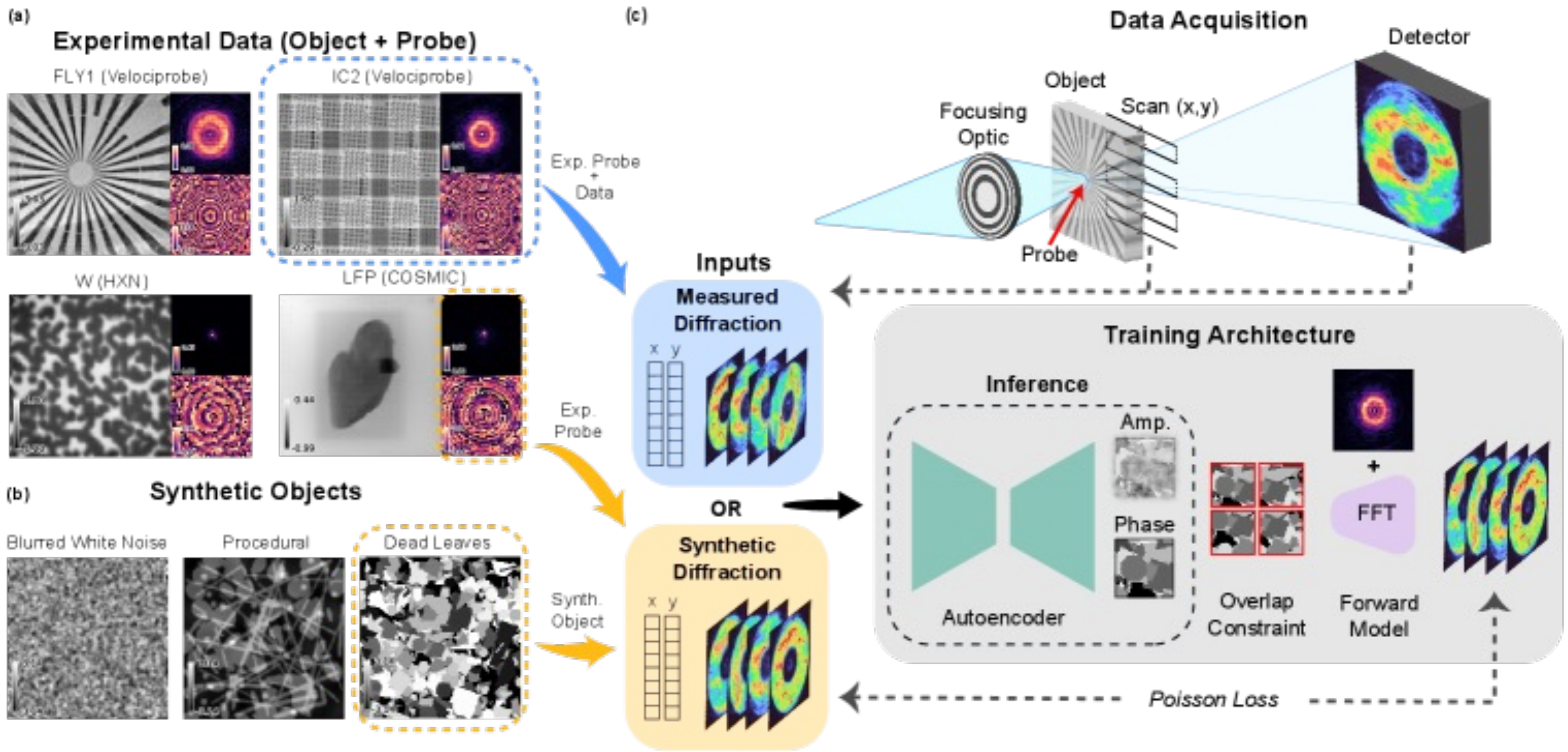
Physics-Informed Neural networks for generalizable reconstruction

- Basic idea: incorporating diffraction physics as architectural constraints
- **Limitations:**
  - 128 x 128 max pixel dimensions → sacrifices *q*-range & resolution
  - Modest model capacity
  - Lack of interpretability and uncertainty quantification



How do we scale this proof of concept?

Experimental results (APS)

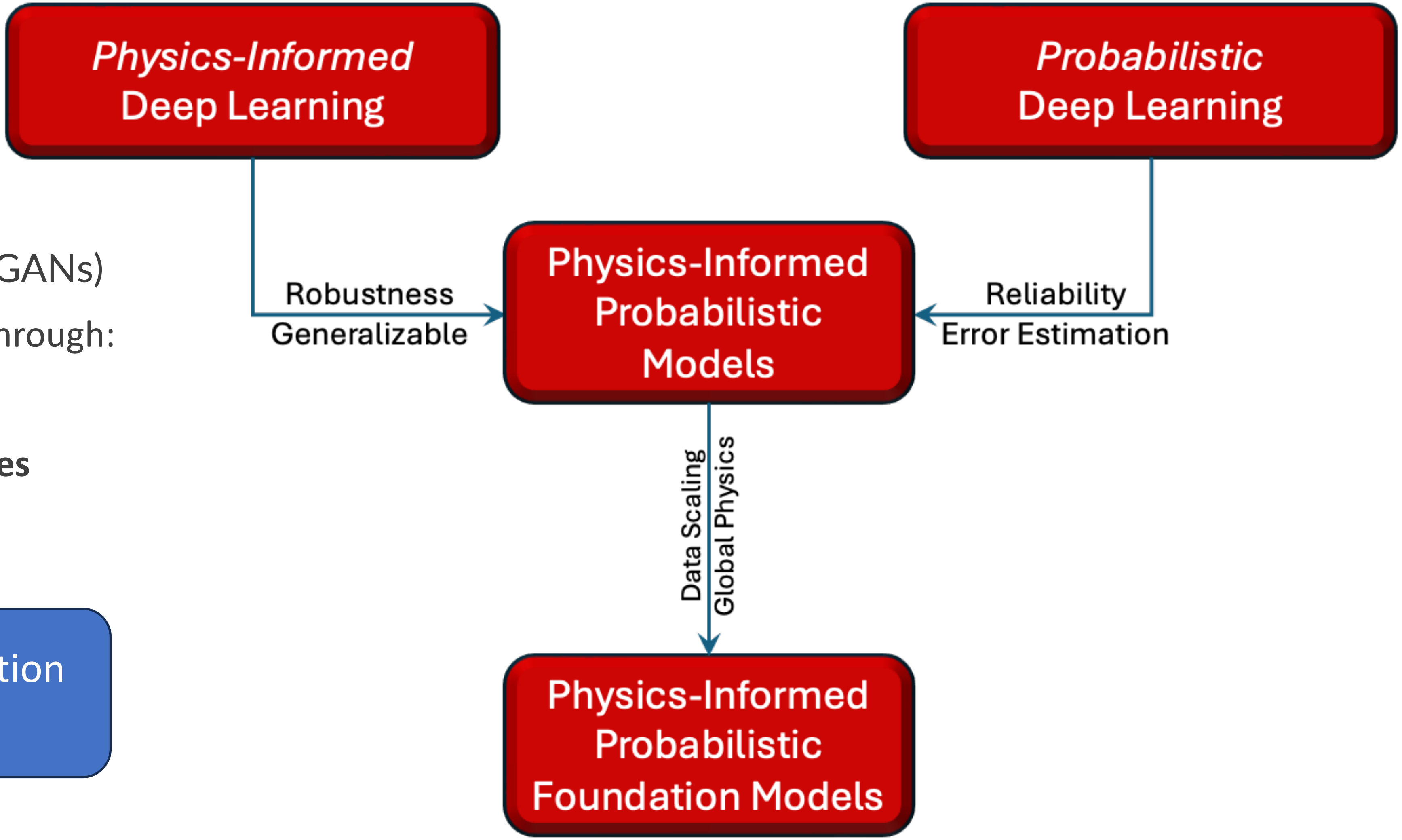


[8] Vong et al. Npj Computational materials, in review

The path forward

- Resolution and dose efficiency improvement through:
  - Noise modeling
  - Architectural scaling (e.g. SRGANs)
- Interpretability and generalization through:
  - Probabilistic machine learning
  - **Vision transformer architectures**

Physics-informed foundation models



Two paths forward

Scaling speed, resolution and generalization

- **Convolutional U-nets:**
  - Ultimate inference speed (100-1000x conventional algorithms) at the expense of generalization
  - 'Easy' path to real-time reconstruction at 30,000 frames per second through parallel inference-time scaling
  - Very rapid training
- **Vision transformers**
  - Ultimate generalization and interpretability at the expense of high data and compute requirements
  - Slow training but potential value as a foundation model for diffractive imaging

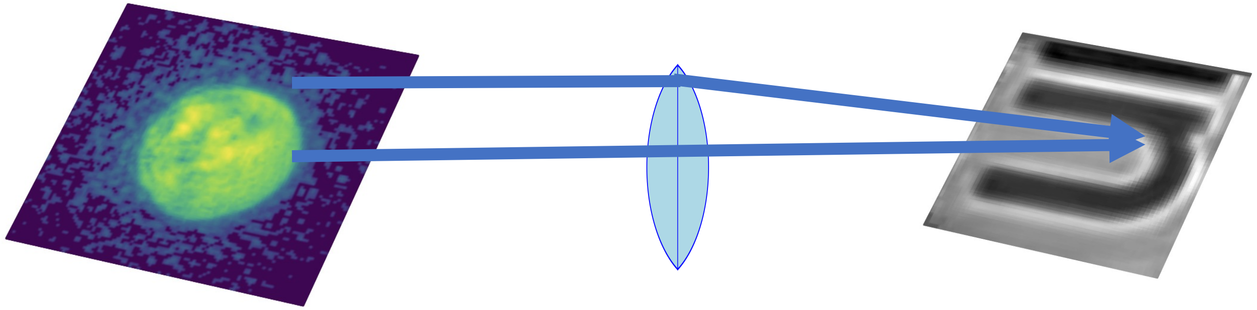
Physics-informed foundation models

Scaling approach: replace CNN backbone by vision transformer (ViT)

- Standard CNN architectures for computer vision:
  - Local receptive fields → poor match to the Fourier transform properties that govern far-field diffraction
  - Limited model capacity → not generalizable enough to reconstruct truly diverse object morphologies

- Transformer architectures:
  - **Global receptive fields** by default
  - Superior generalization [7]

- ViT foundation models would also provide:
  - Interpretable latent space maps
    - Essential for model-driven experiment design
  - Transfer learning opportunities

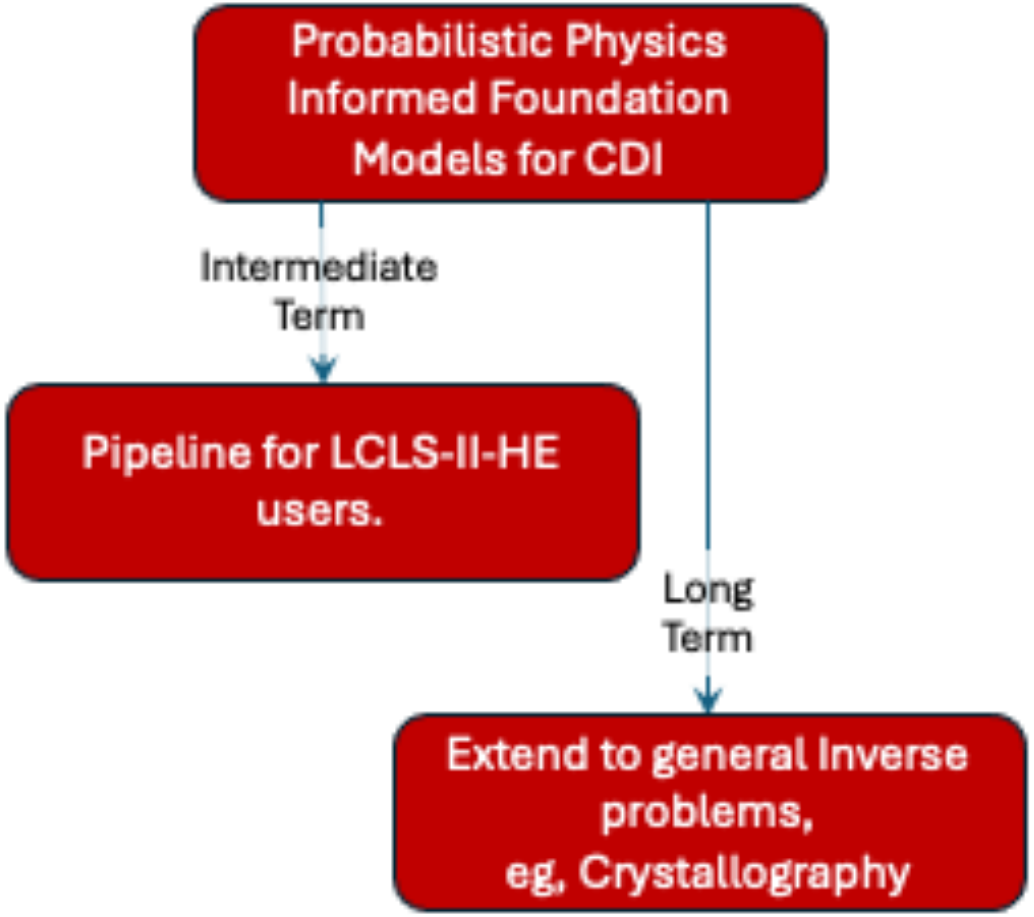


SLAC

Physics-informed ViT: better generalization and interpretable latent space maps

Conclusions and beyond imaging

- Direct ramifications on other imaging techniques, eg crystallography.
- Beyond Imaging: Data compression via latent space embeddings, Active Learning for design of experiments, etc.
- Extend to Physics Informed Foundation Models for general Inverse problems at SLAC.
- Rapid feedback impact on nanoscale characterization capabilities for fields like heterogeneous materials.
    - This capability is being developed right now at SLAC and collaborating institutions (ANL, LBL)



References

[1] Cherukara, M.J., Zhou, T., Nashed, Y., Enfedaque, P., Hexemer, A., Harder, R.J. and Holt, M.V., 2020. AI-enabled high-resolution scanning coherent diffraction imaging. *Applied Physics Letters*, *117*(4).

[2] Hoidn, O., Mishra, A.A. and Mehta, A., 2023. Physics constrained unsupervised deep learning for rapid, high resolution scanning coherent diffraction reconstruction. *Scientific Reports*, *13*(1), p.22789.

[3] Williams, G.J., Quiney, H.M., Dhal, B.B., Tran, C.Q., Nugent, K.A., Peele, A.G., Paterson, D. and De Jonge, M.D., 2006. Fresnel coherent diffractive imaging. *Physical review letters*, *97*(2), p.025506.

[4] Lakshminarayanan, B., Pritzel, A. and Blundell, C., 2017. Simple and scalable predictive uncertainty estimation using deep ensembles. *Advances in neural information processing systems*, *30*.

[5] Blei, D.M., Kucukelbir, A. and McAuliffe, J.D., 2017. Variational inference: A review for statisticians. *Journal of the American statistical Association*, *112*(518), pp.859-877.

[6] Gal, Y. and Ghahramani, Z., 2016, June. Dropout as a bayesian approximation: Representing model uncertainty in deep learning. In *international conference on machine learning* (pp. 1050-1059). PMLR.

[7] He, K., Chen, X., Xie, S., Li, Y., Dollár, P. and Girshick, R., 2022. Masked autoencoders are scalable vision learners. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 16000-16009).