Phase Space Reconstruction using Differentiable Simulations and Neural Networks

CBB Annual Meeting

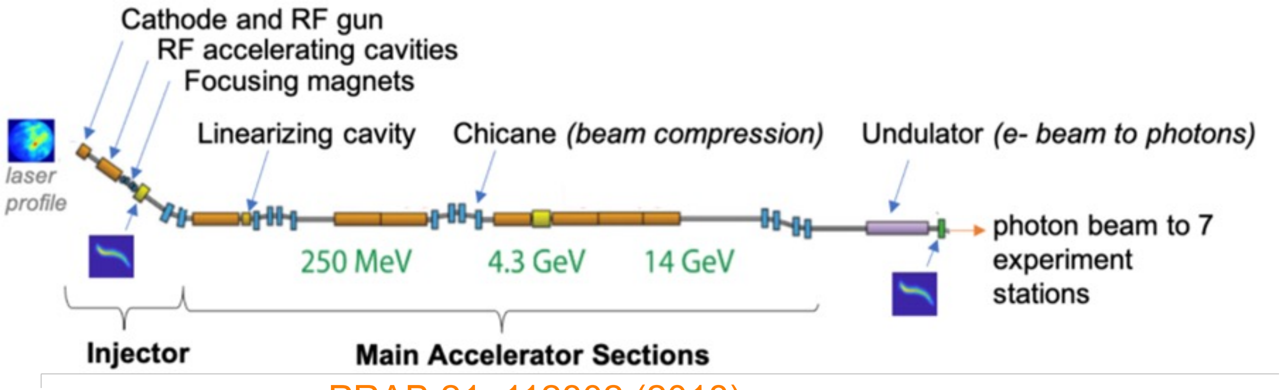
Cornell University - July 20th, 2023

J.P. Gonzalez-Aguilera* *(UChicago)*

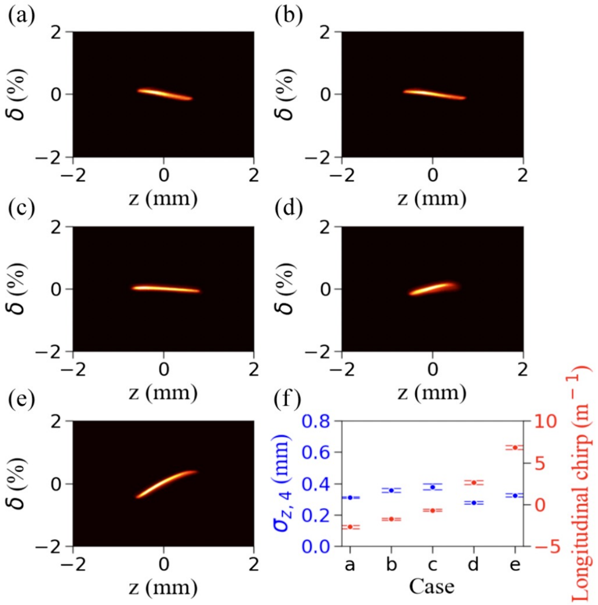
SLAC NATIONAL ACCELERATOR LABORATORY
Argonne NATIONAL LABORATORY
The Center for BRIGHT BEAMS A National Science Foundation Science & Technology Center
THE UNIVERSITY OF CHICAGO
NERSC

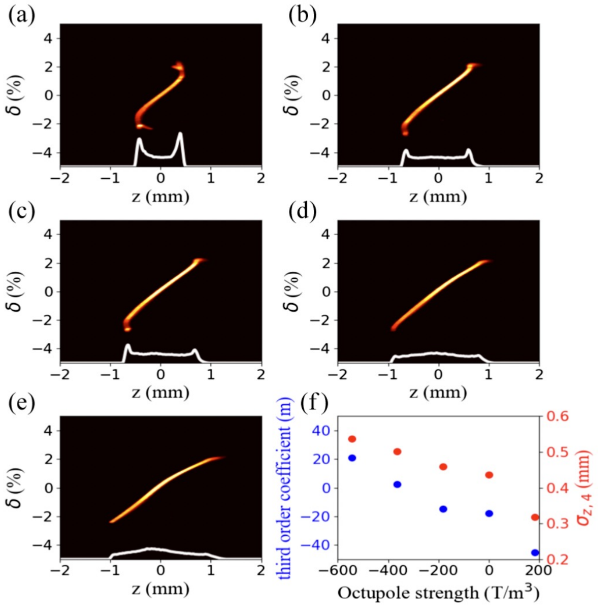
* jpga@uchicago.edu

Manipulating Beams in Phase Space



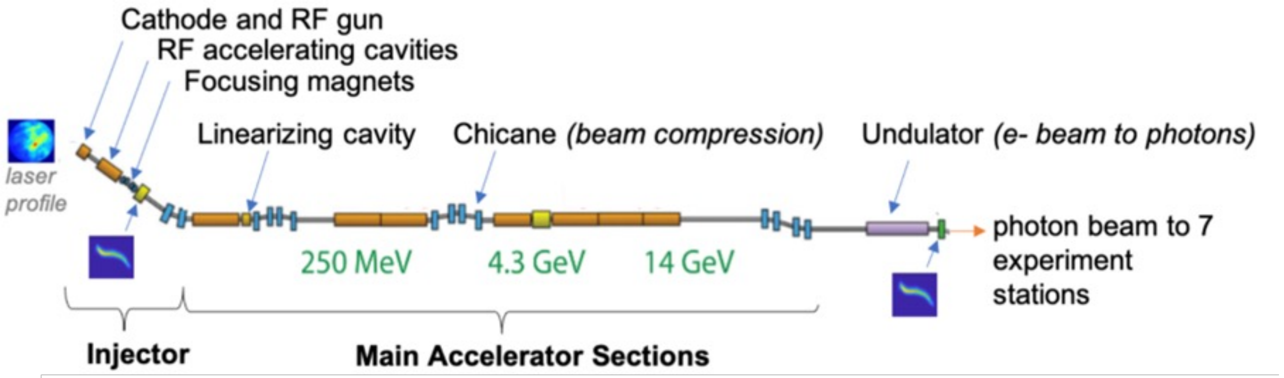
PRAB 21, 112802 (2018)



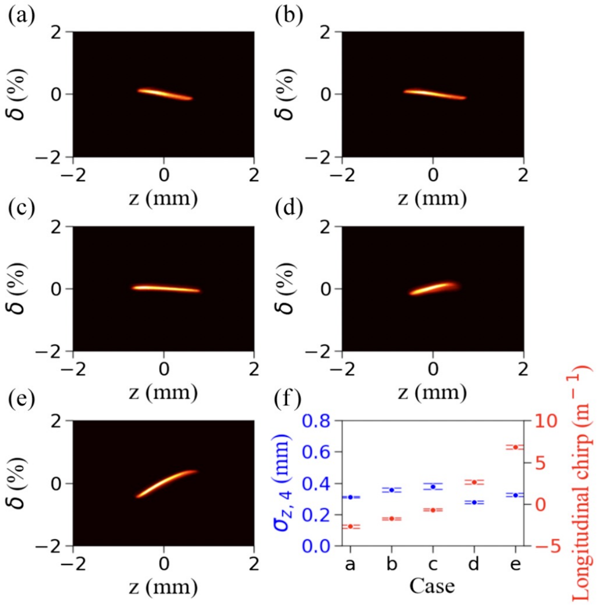


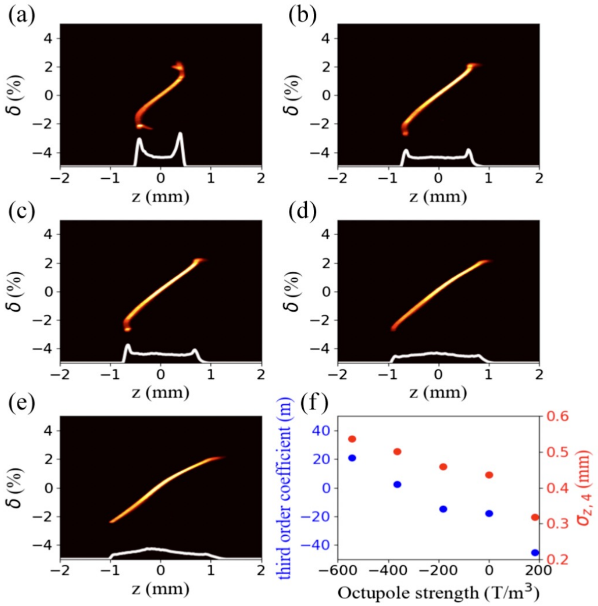
PRL 129, 224801 (2022)

Manipulating Beams in Phase Space



PRAB 21, 112802 (2018)





PRL 129, 224801 (2022)

General Accelerator R&D Program

Accelerator and Beam Physics Roadmap

DOE Accelerator Beam Physics Roadmap Workshop

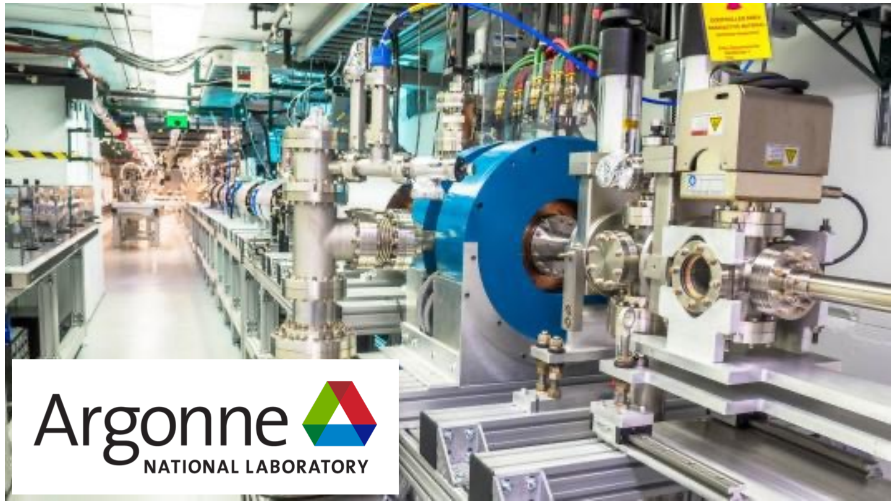
September 6–8, 2022

5 Grand Challenge Three

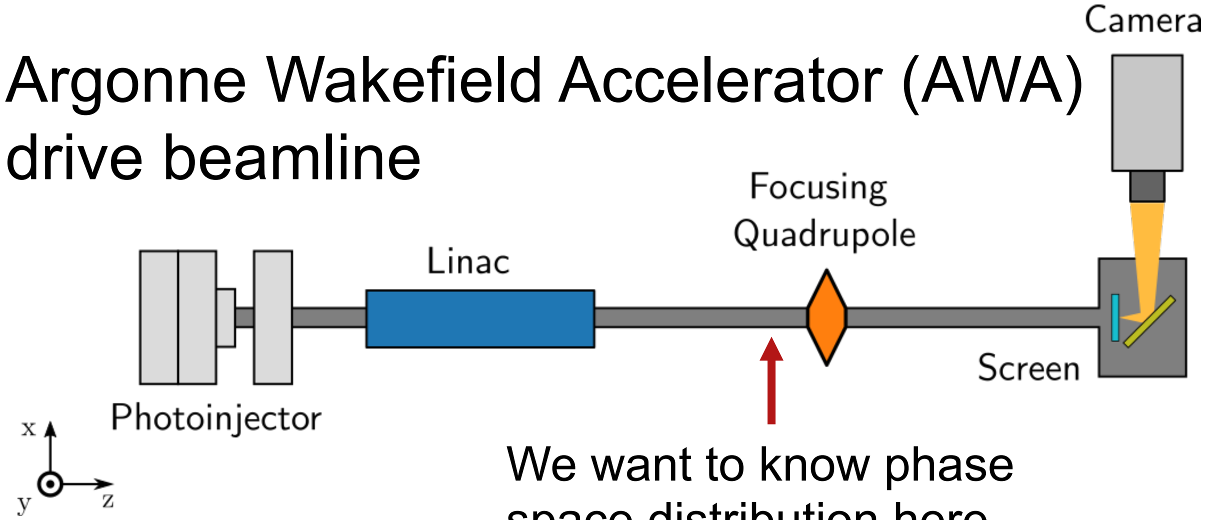
Beam Control: How do we control and diagnose the beam distribution at all scales—from its macroscopic properties down to the level of individual particles?

Detailed measurement of beam phase space distribution is important!

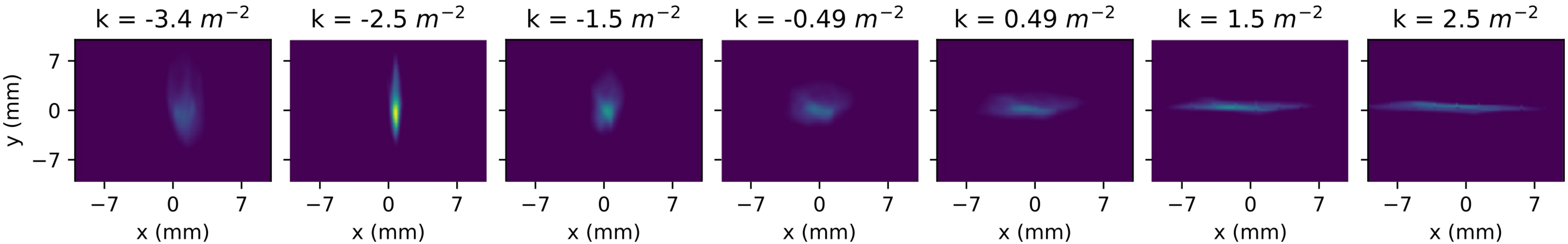
Phase space distribution measurements



Argonne Wakefield Accelerator (AWA) drive beamline



We want to know phase space distribution here

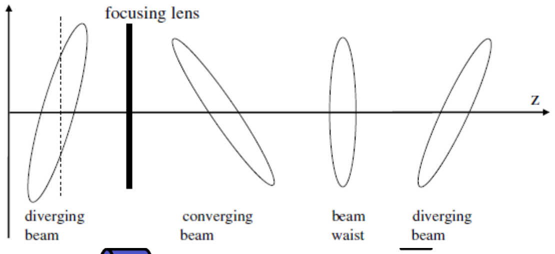


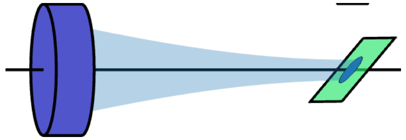
How do I get the most information out of these in an efficient way?

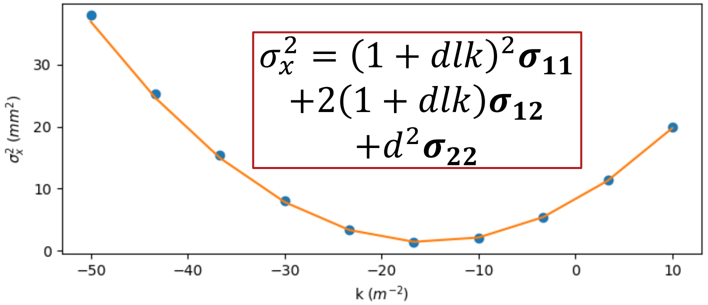
Usual Approaches

Simple quad scan:

- rotate beam by scanning focusing strength
- measure the beam size
- Fit and solve for ε







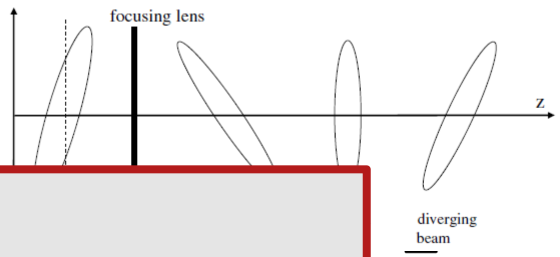
$$\sigma_x^2 = (1+dlk)^2\boldsymbol{\sigma_{11}} + 2(1+dlk)\boldsymbol{\sigma_{12}} + d^2\boldsymbol{\sigma_{22}}$$

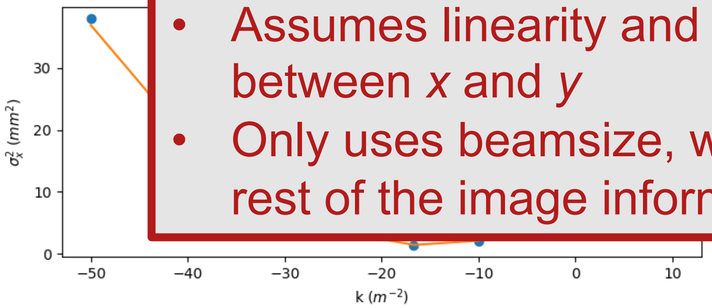
$$\varepsilon = \sqrt{\sigma_{11}\sigma_{22} - \sigma_{12}^2}$$

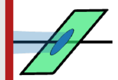
Usual Approaches

Simple quad scan:

- rotate beam by scanning focusing strength
- me
- Fit







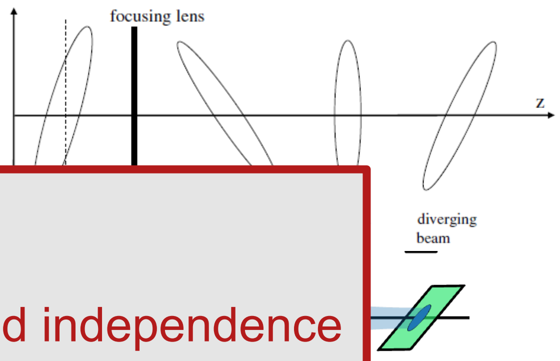
σ_{12}^2

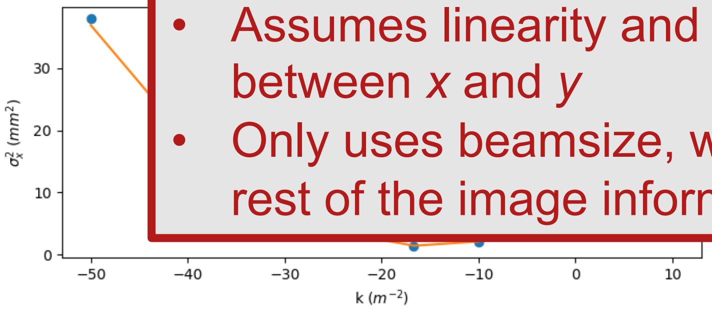
- Easy fit
- Not detailed
- Assumes linearity and independence between *x* and *y*
- Only uses beamsize, wasting the rest of the image information

Usual Approaches

Simple quad scan:

- rotate beam by scanning focusing strength
- me
- Fit

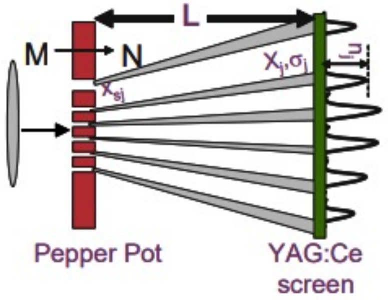




- Easy fit
- Not detailed
- Assumes linearity and independence between *x* and *y*
- Only uses beamsize, wasting the rest of the image information

Specialized diagnostics:

- pepper-pot (single-shot 4D)
- Multi-slit (single-shot 2D)
- Moving slit (multiple measurements)

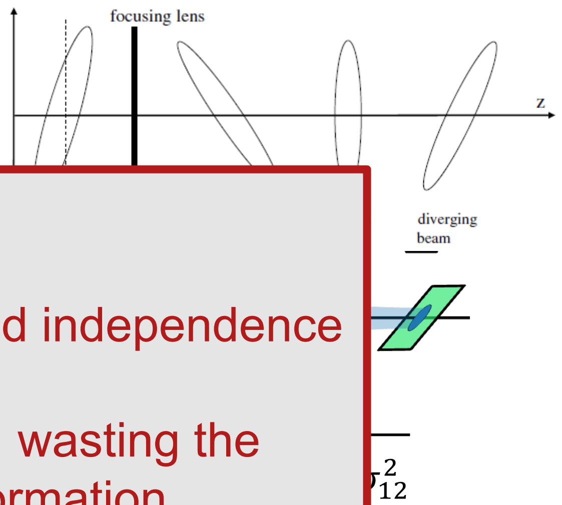


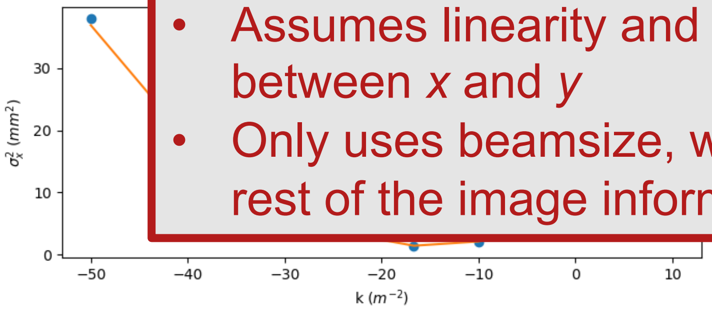
Power. J. et al PAC07, 2007

Usual Approaches

Simple quad scan:

- rotate beam by scanning focusing strength
- me
- Fit





- Easy fit
- Not detailed
- Assumes linearity and independence between *x* and *y*
- Only uses beamsize, wasting the rest of the image information

Specialized diagnostics:

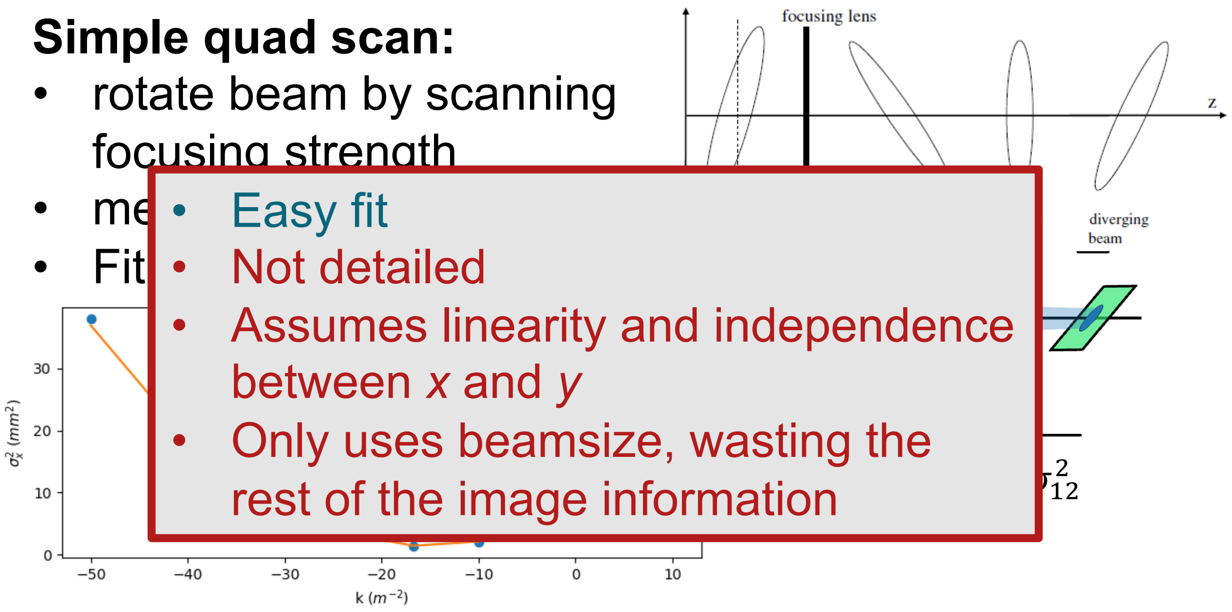
- pepper-pot (single-shot 4D)
- Multi-slit (single-shot 2D)
- Mo

- Fast
- Not as detailed as we would like
- Design considerations for different beam sizes / charges
- Wastes information: only uses beamlets intensities, positions and sizes

Power. J. et al PAC07, 2007

Usual Approaches

Simple quad scan:
- rotate beam by scanning focusing strength
- me…
- Fit…



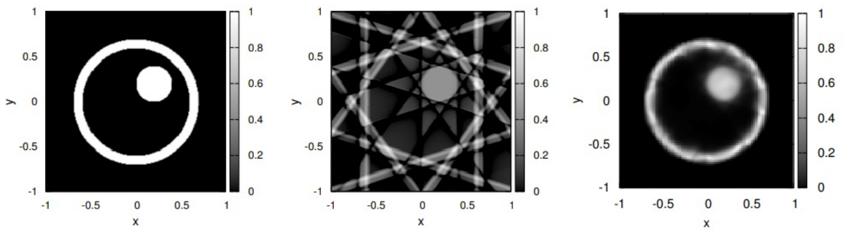
- Easy fit
- Not detailed
- Assumes linearity and independence between *x* and *y*
- Only uses beamsize, wasting the rest of the image information

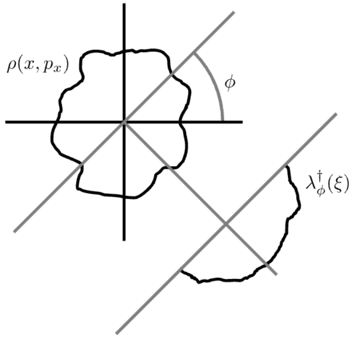
Specialized diagnostics:
- pepper-pot (single-shot 4D)
- Multi-slit (single-shot 2D)
- Mo…

- Fast
- Not as detailed as we would like
- Design considerations for different beam sizes / charges
- Wastes information: only uses beamlets intensities, positions and sizes

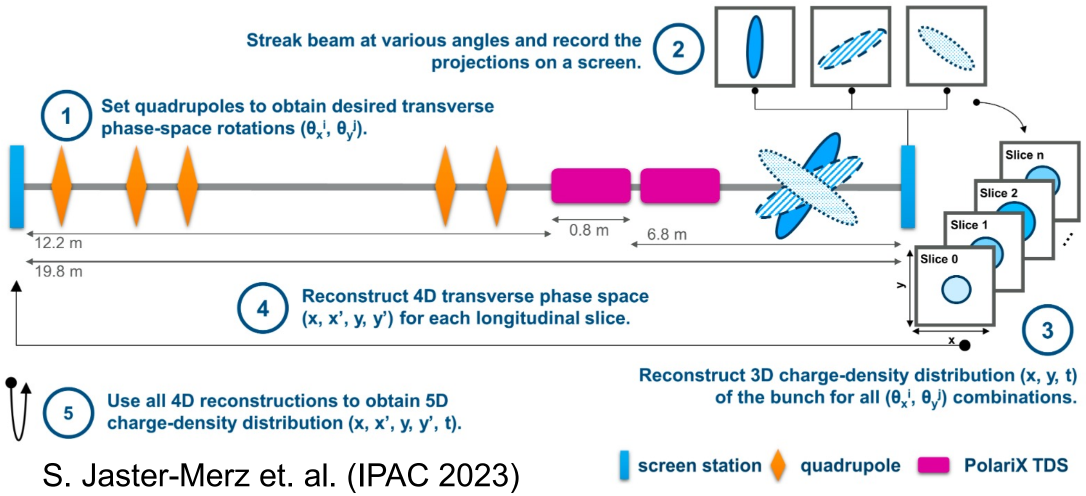
Power. J. et al PAC07, 2007

Advanced tomographic methods:
- Maximum entropy tomography (MENT)
- Algebraic reconstruction (ART, SART)





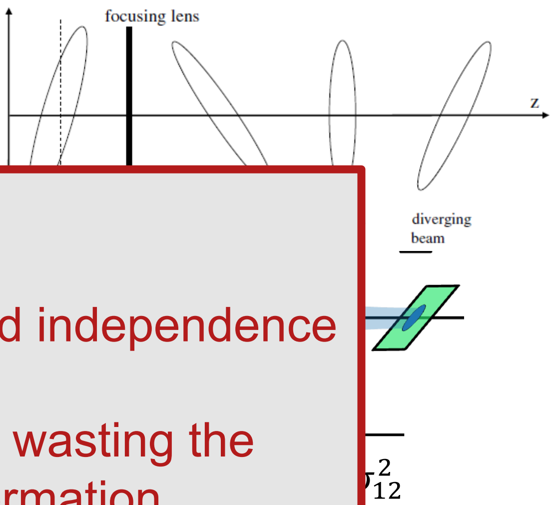
Hock K. and Ibison M., JINST, 2013

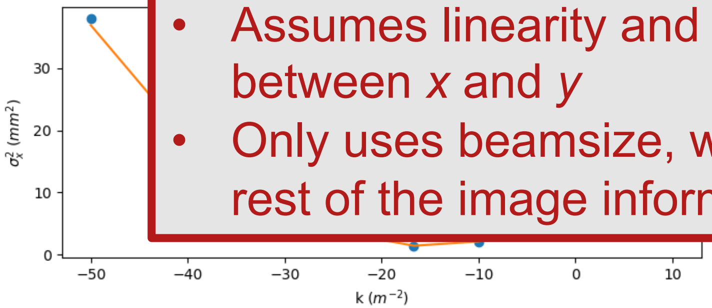


S. Jaster-Merz et. al. (IPAC 2023)

Usual Approaches

Simple quad scan:
- rotate beam by scanning focusing strength
- me
- Fit





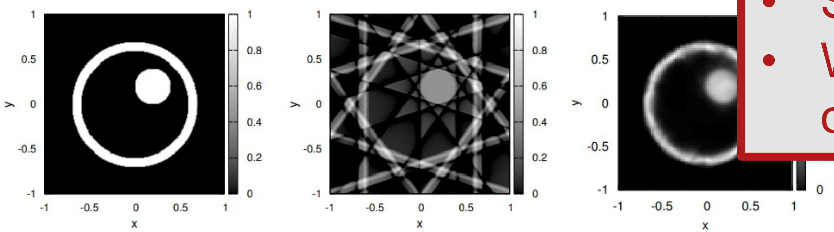
- Easy fit
- Not detailed
- Assumes linearity and independence between *x* and *y*
- Only uses beamsize, wasting the rest of the image information

Specialized diagnostics:
- pepper-pot (single-shot 4D)
- Multi-slit (single-shot 2D)
- Mo

- Fast
- Not as detailed as we would like
- Design considerations for different beam sizes / charges
- Wastes information: only uses beamlets intensities, positions and sizes

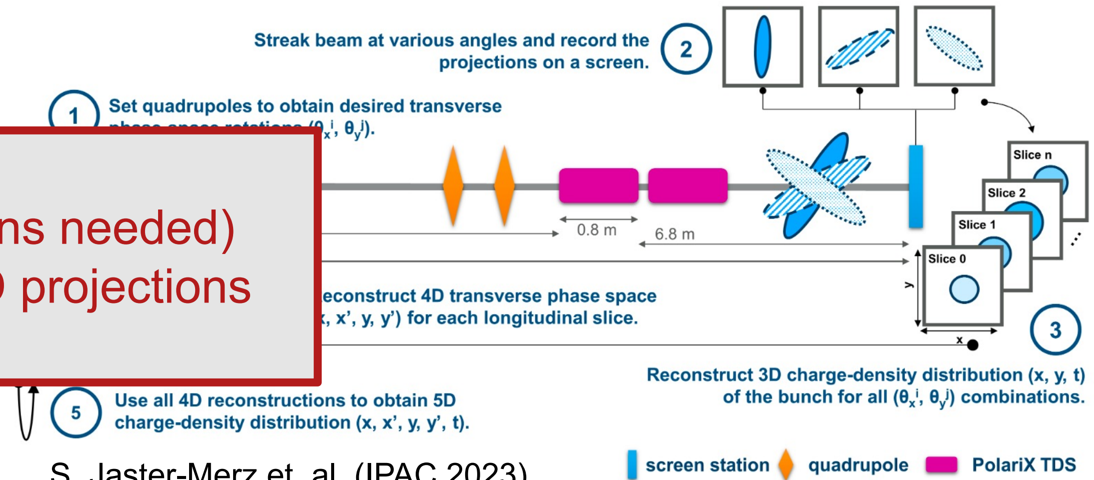
Power. J. et al PAC07, 2007

Advanced tomographic methods:
- Maximum entropy tomography (MENT)
- Algebraic reconstruction (ART, SA



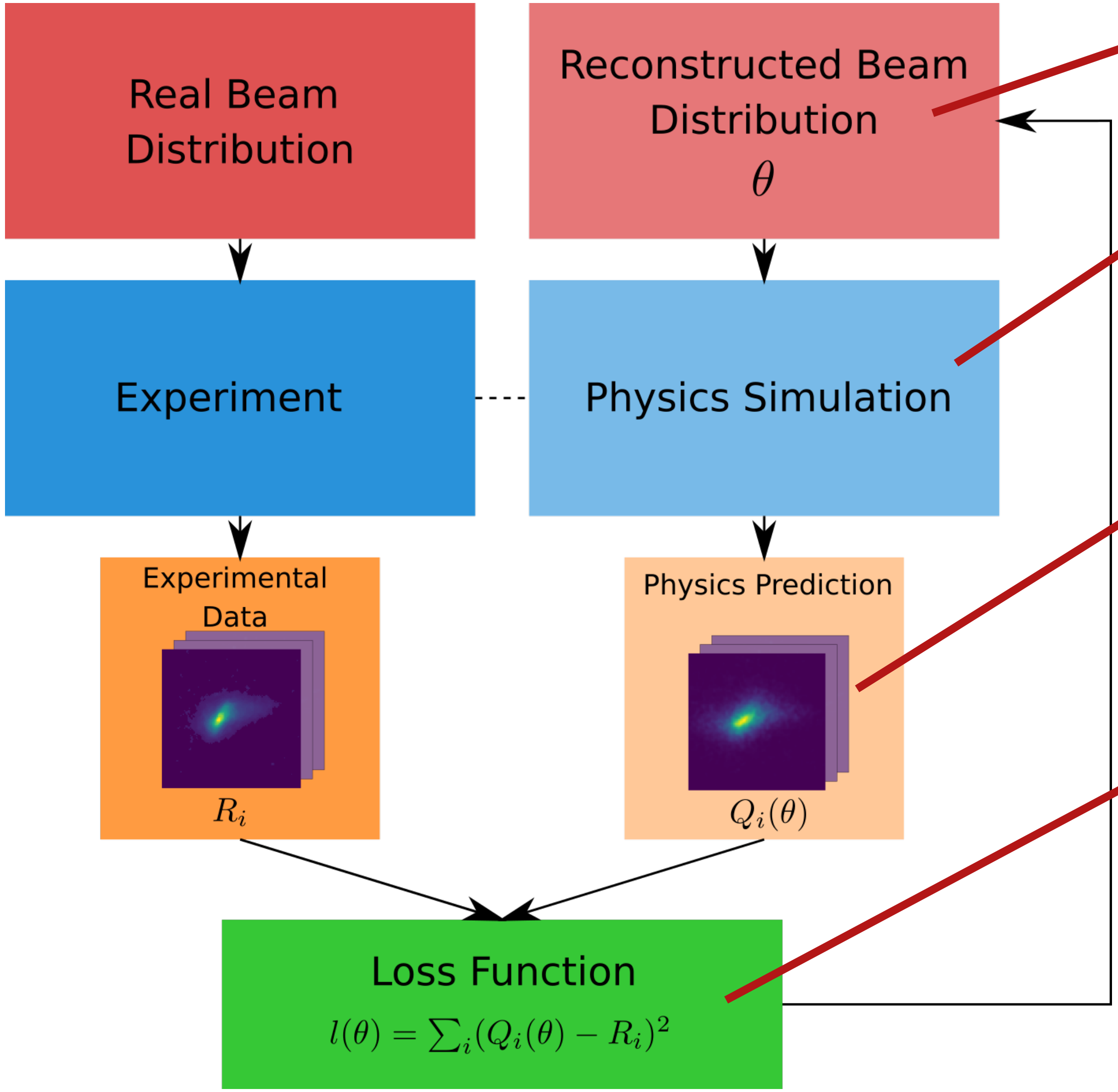
Hock K. and Ibison M., JINST, 2013

- Very detailed
- Slow (many observations needed)
- Wastes information: 1D projections only.



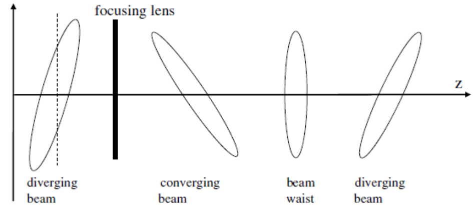
S. Jaster-Merz et. al. (IPAC 2023)

Phase Space Fitting as optimization problem



Simple quad scan:

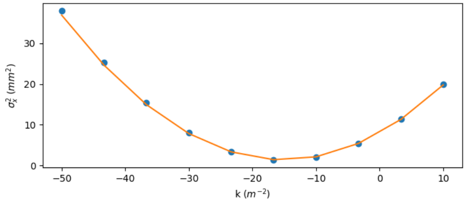
- Beam distribution is assumed to be elliptical. Fully parametrized by σ_{xx}, σ_{xp_x}, $\sigma_{p_xp_x}$
- Assume linear transport of elliptical beam



- Beam sizes from screen downstream

$$\sigma_x^2 = (1 + dlk)^2 \boldsymbol{\sigma_{11}} + 2(1 + dlk)\boldsymbol{\sigma_{12}} + d^2 \boldsymbol{\sigma_{22}}$$

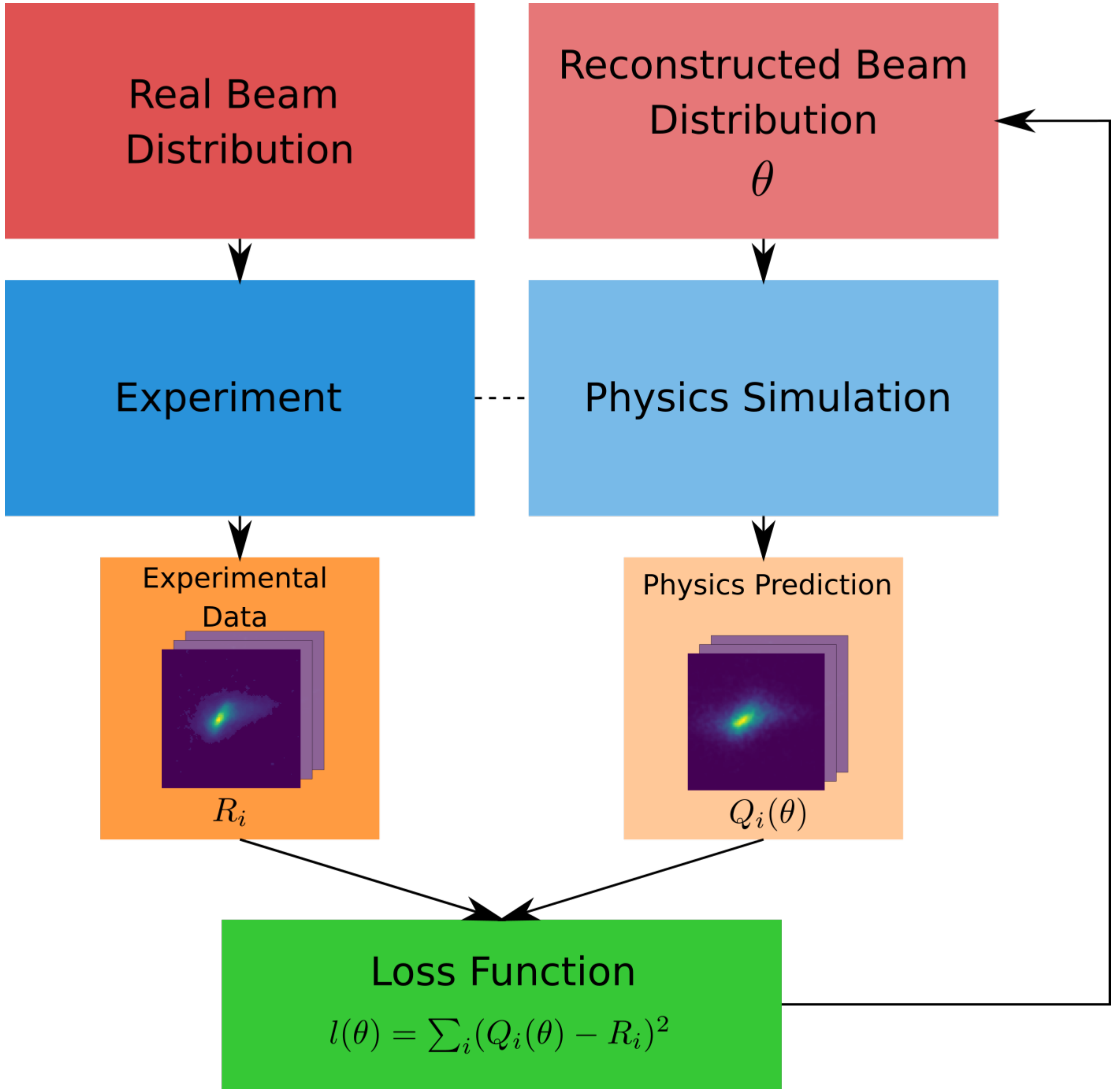
- Error of the quadratic fit



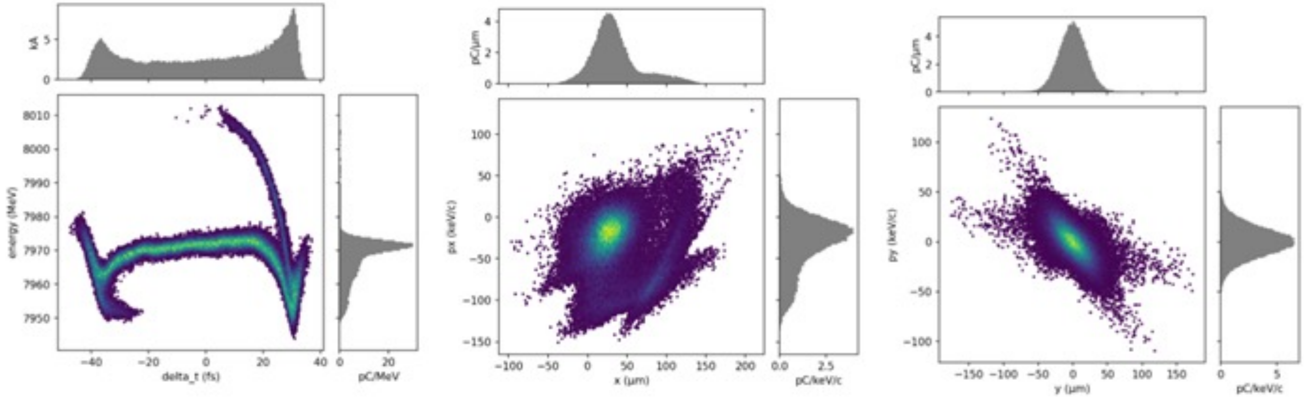
Result:

- Elliptical 2D phase space consistent with beam size measurements.

Phase Space Fitting as optimization problem



We want more detail:

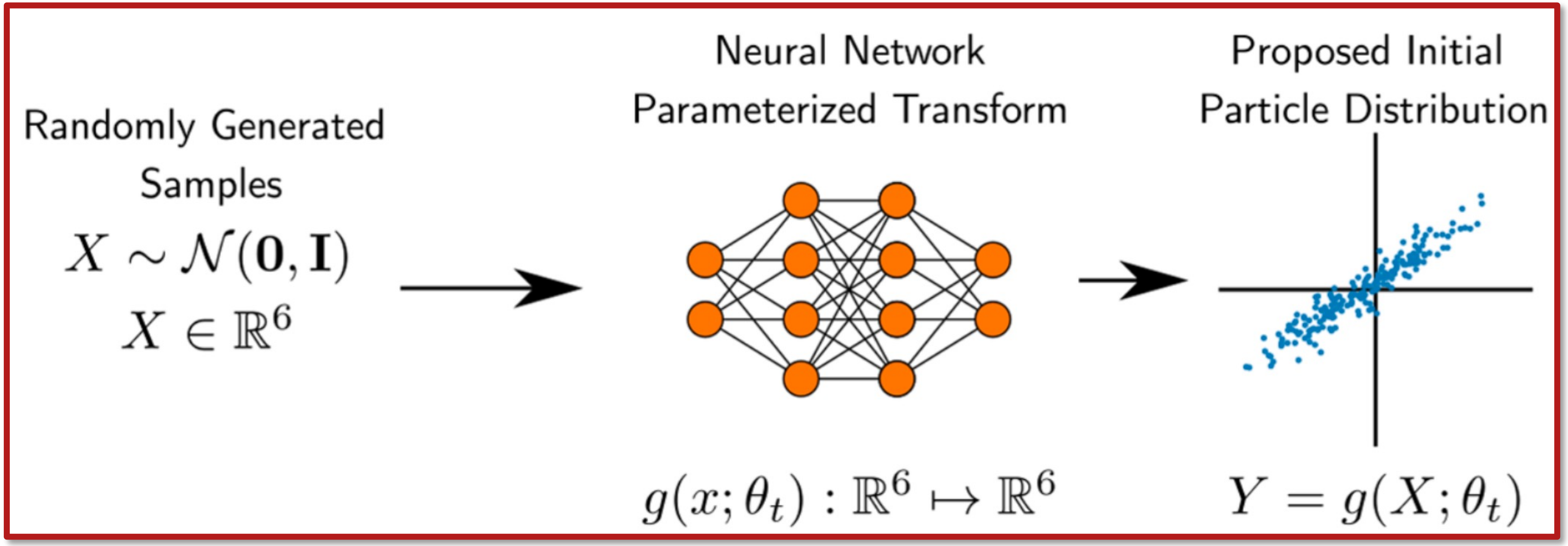


*LCLS

- How do we **parametrize** the beam 6D phase-space distribution in a a **flexible** and **learnable** way?
- How do we run **simulations** that support **optimization** of extremely **high dimensional problems** (~1k parameters)?

Neural Network Parameterization of Beam Distributions

- 6D phase space distribution parametrization that is
  - flexible
  - learnable



Fully connected NN with **~ O(1k) parameters**

Differentiable Simulations (Automatic Differentiation)

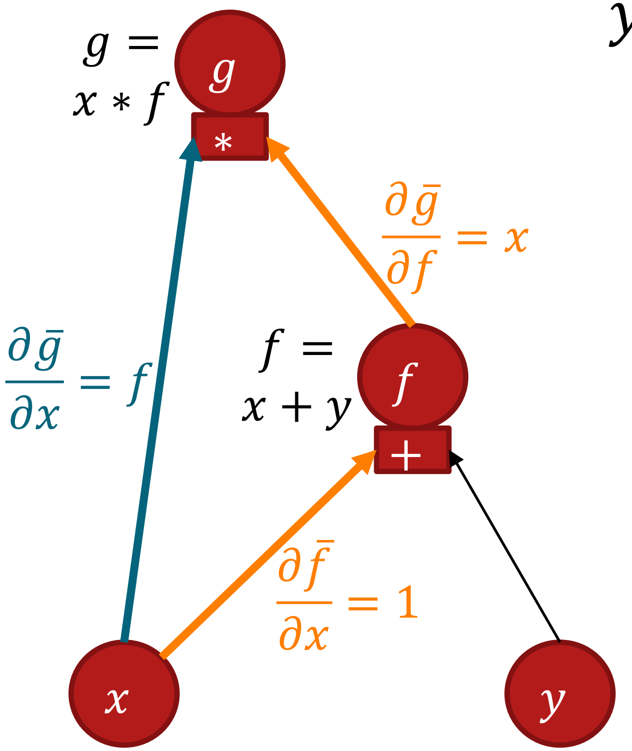
Keep track of derivative information during every calculation step using the chain rule and memory.

Fast and **accurate** high-dimensional gradients

Enables **gradient-based optimization** of model with respect to all free parameters.

Easily optimize models with >10k free parameters.

$$f(x, y) = x + y,$$
$$g(x, f(x, y)) = x * f(x, y),$$
$$x = 3,$$
$$y = 2.$$



$$\frac{\partial g}{\partial x} = \frac{\partial \bar{g}}{\partial x} + \frac{\partial \bar{g}}{\partial f} * \frac{\partial \bar{f}}{\partial x}$$

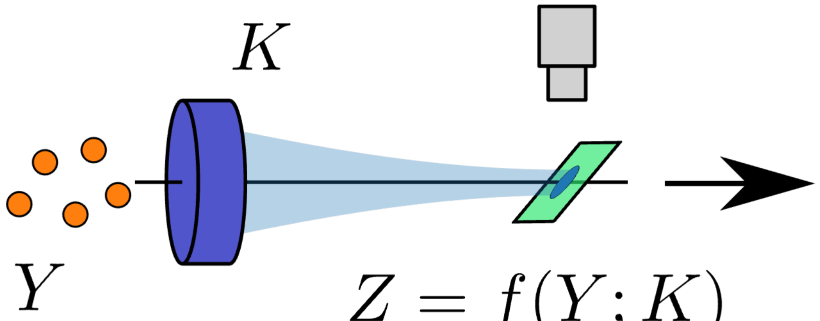
Differentiable Simulations (Automatic Differentiation)

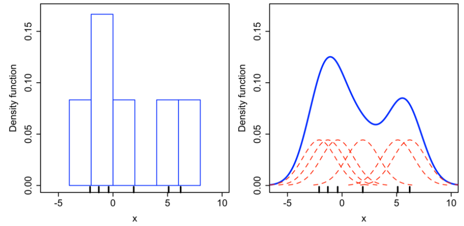
Keep track of derivative information during every calculation step using the chain rule and memory.

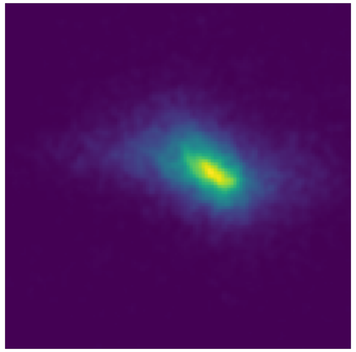
Fast and **accurate** high-dimensional gradients

Enables **gradient-based optimization** of model with respect to all free parameters.

Easily optimize models with >10k free parameters.







$$Z = f(Y; K)$$

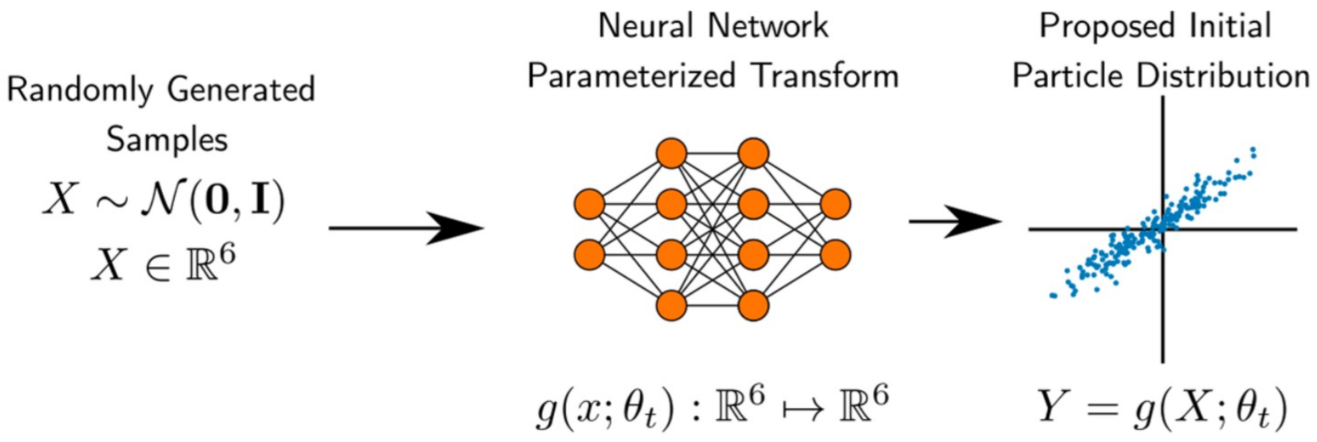
$$Q^{(i,j)} = \mathrm{KDE}(Z)$$

$$\frac{\partial Z}{\partial Y}, \frac{\partial Z}{\partial K}, \frac{\partial \sigma_Z}{\partial K}, \dots$$

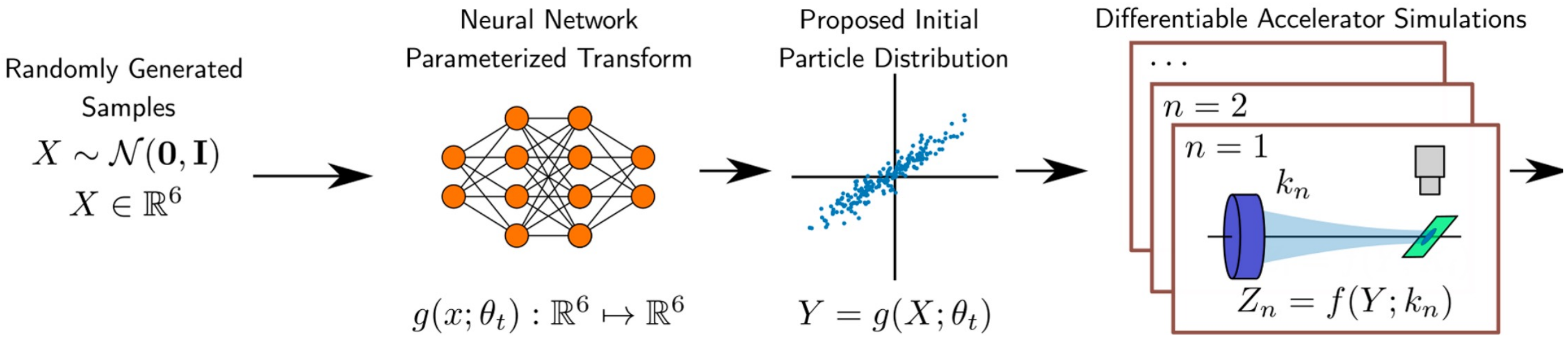
$$\frac{\partial Q^{(i,j)}}{\partial Y}, \frac{\partial Q^{(i,j)}}{\partial K}$$

Poster tomorrow!

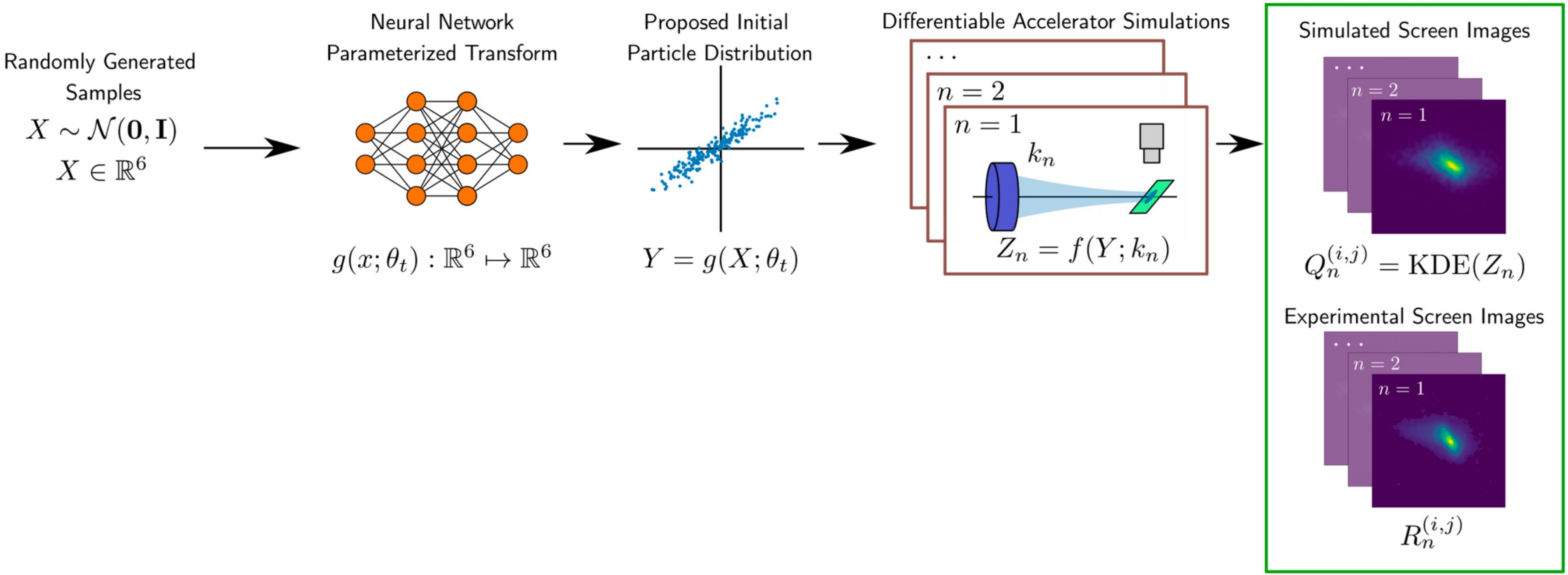
Phase Space Reconstruction Pipeline



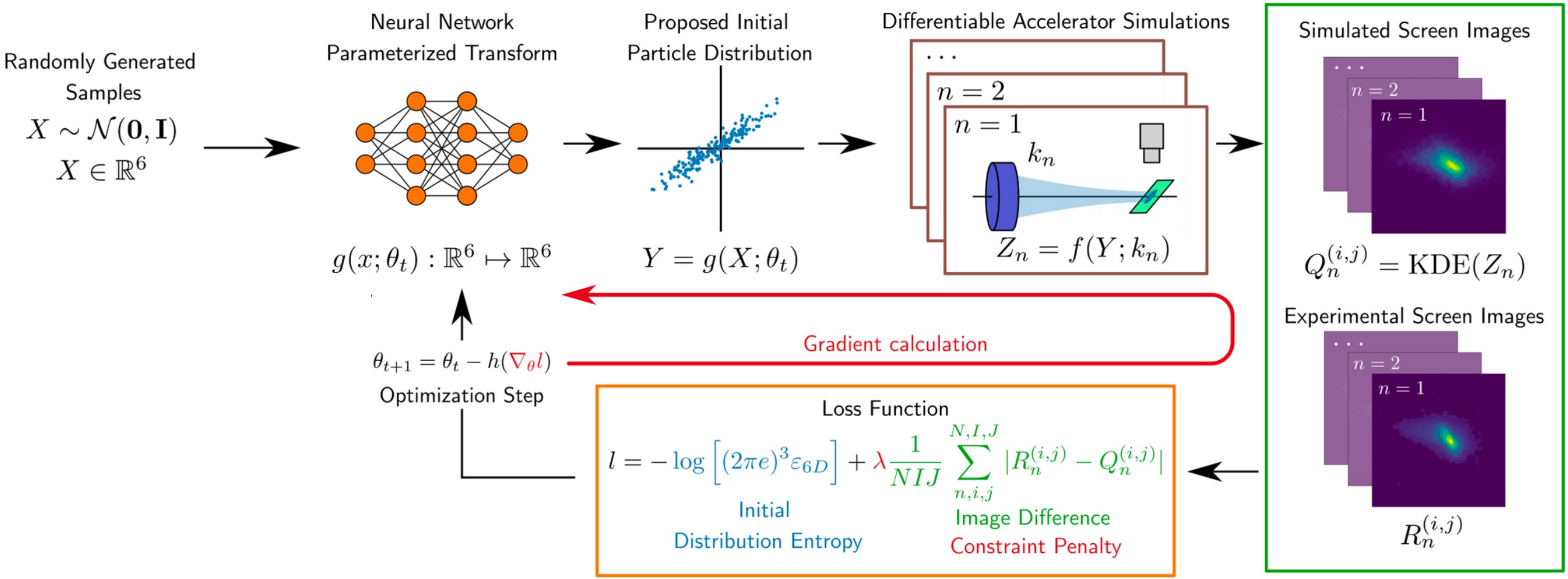
Phase Space Reconstruction Pipeline



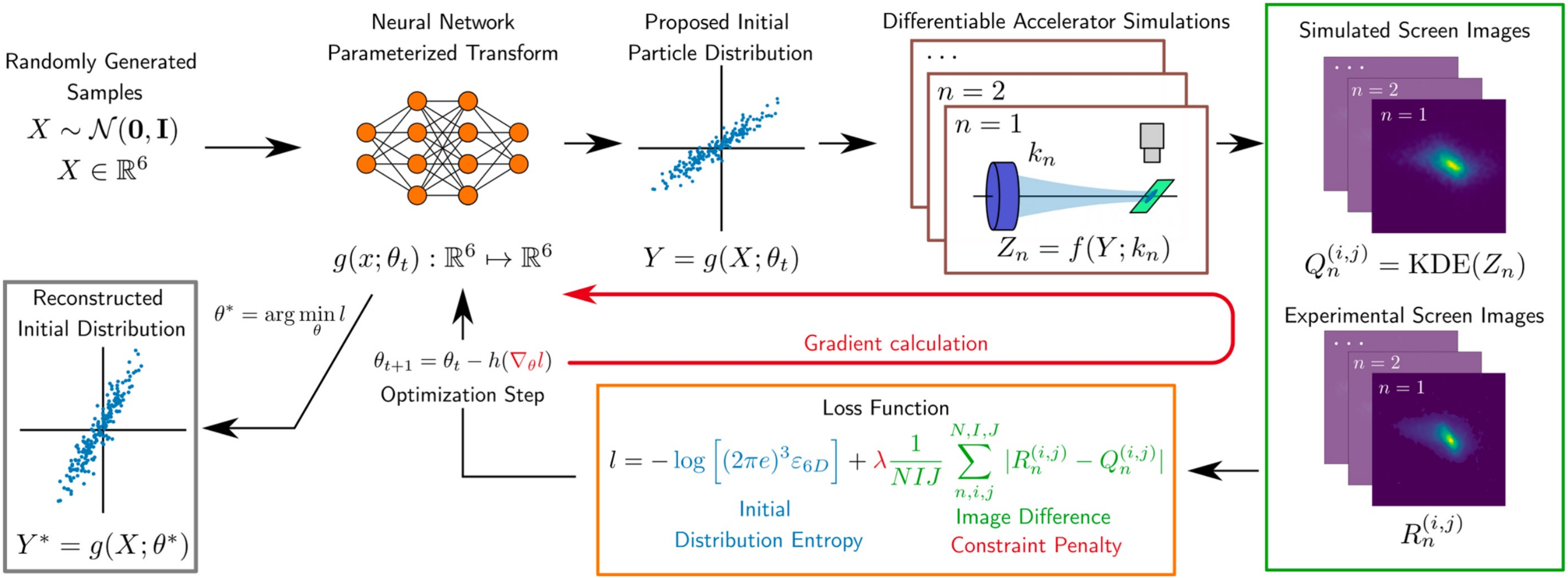
Phase Space Reconstruction Pipeline



Phase Space Reconstruction Pipeline

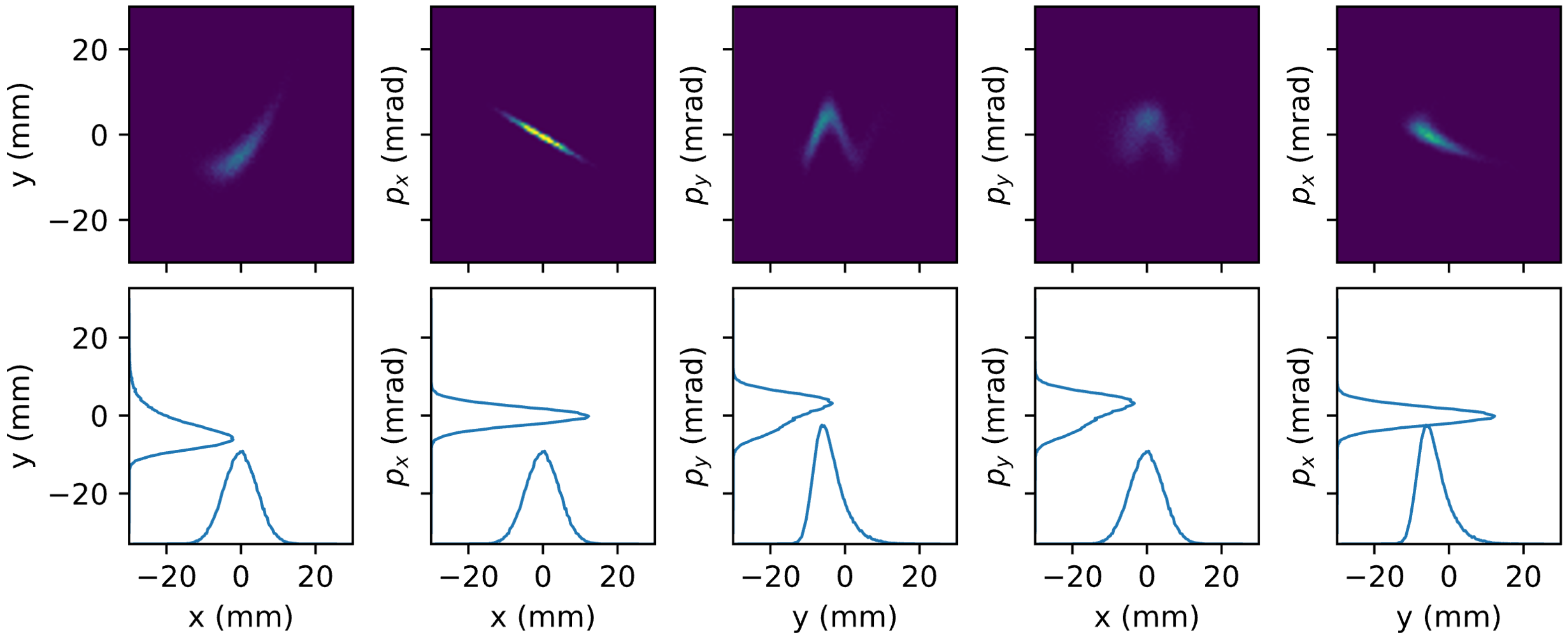


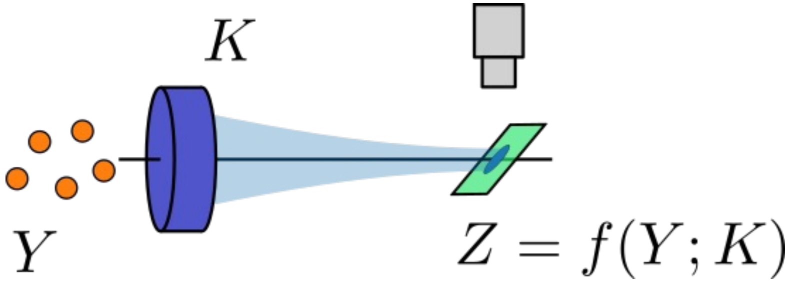
Phase Space Reconstruction Pipeline



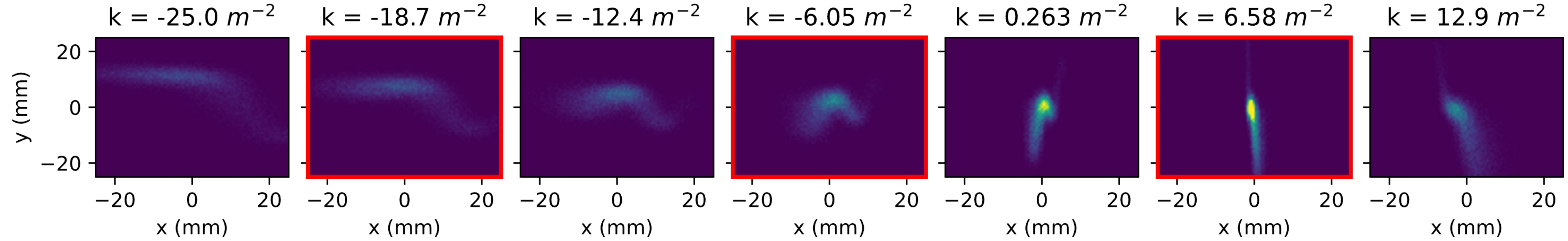
Synthetic Example

Synthetic beam distribution in simulation

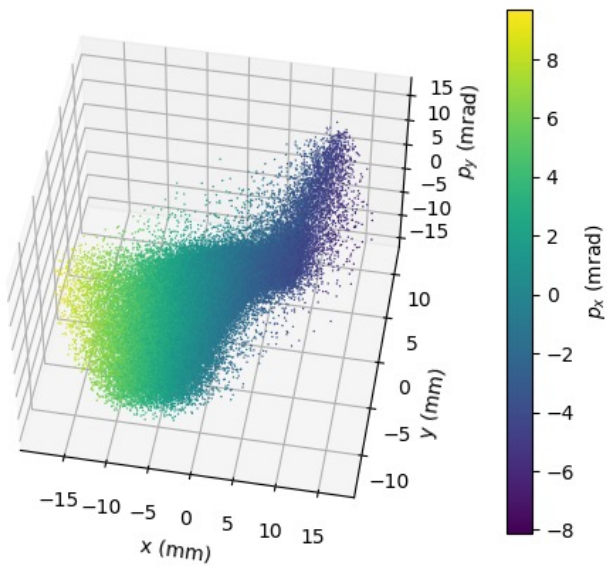


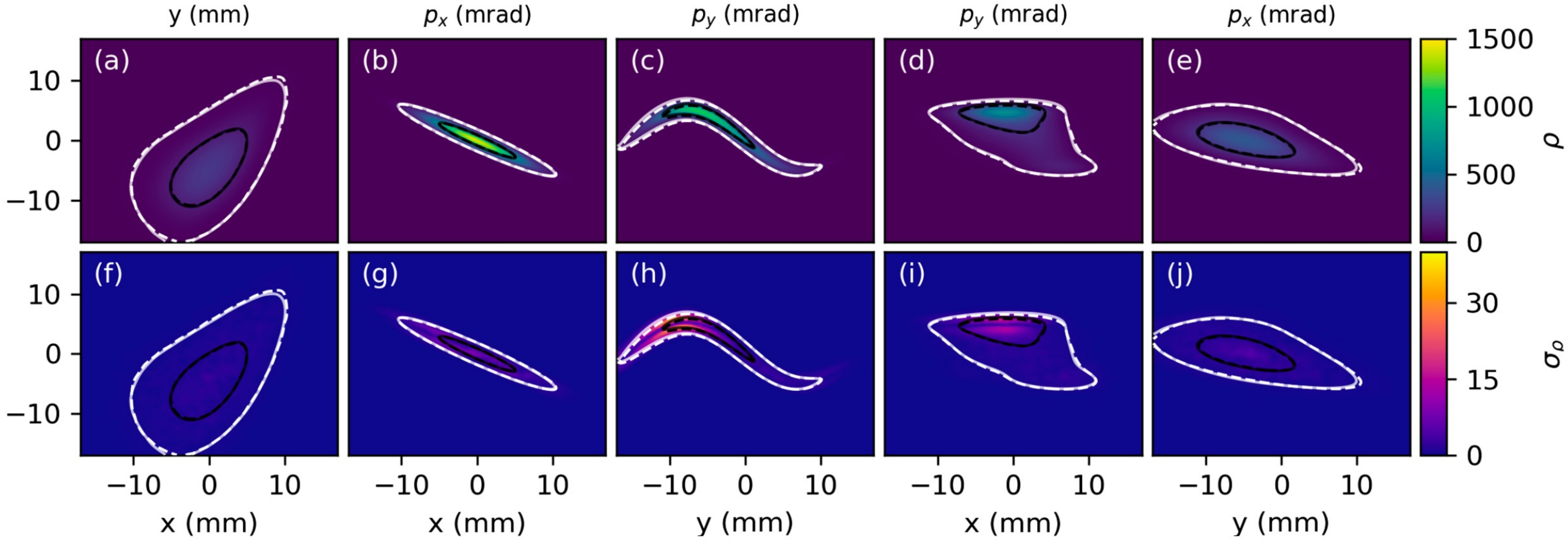


Screen images



Synthetic Example Reconstruction





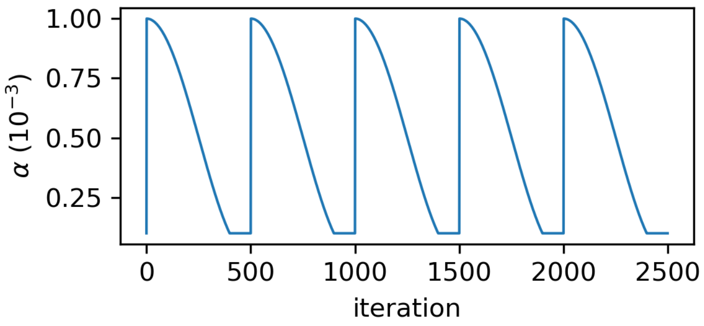
Detailed reconstruction of 4D phase space with only

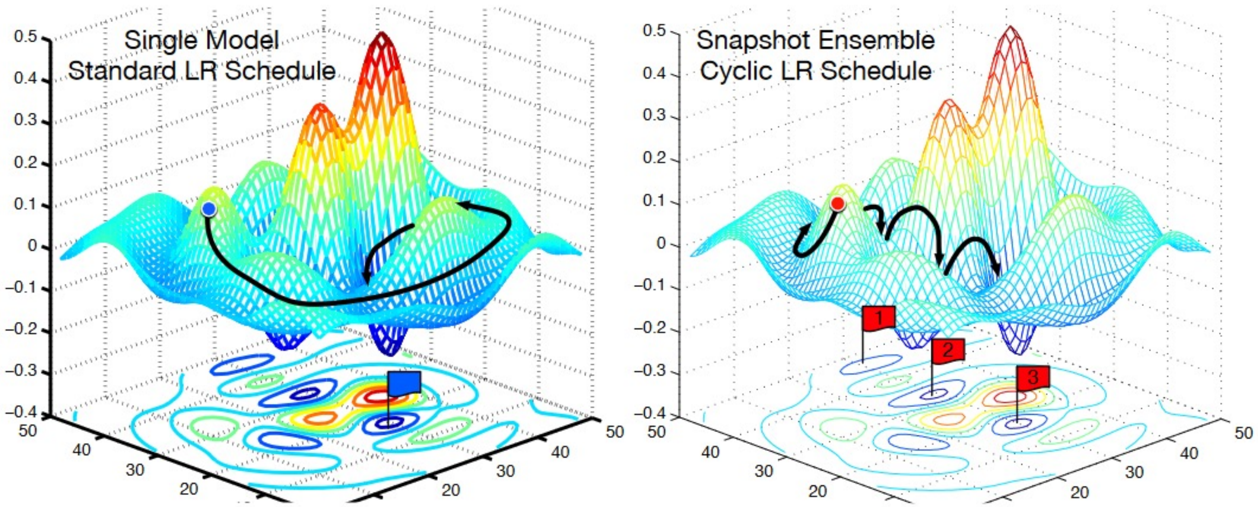
- **a quadrupole and a screen**
- **10 images**

50th percentile ground truth
50th percentile reconstruction
95th percentile ground truth
95th percentile reconstruction

Measuring Model Uncertainty

Create a **snapshot ensemble** to measure uncertainty by cycling the learning rate

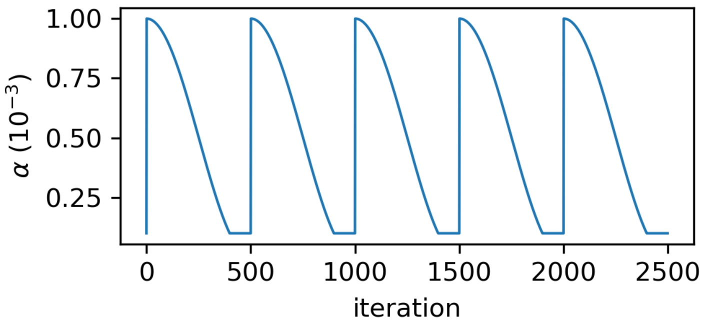


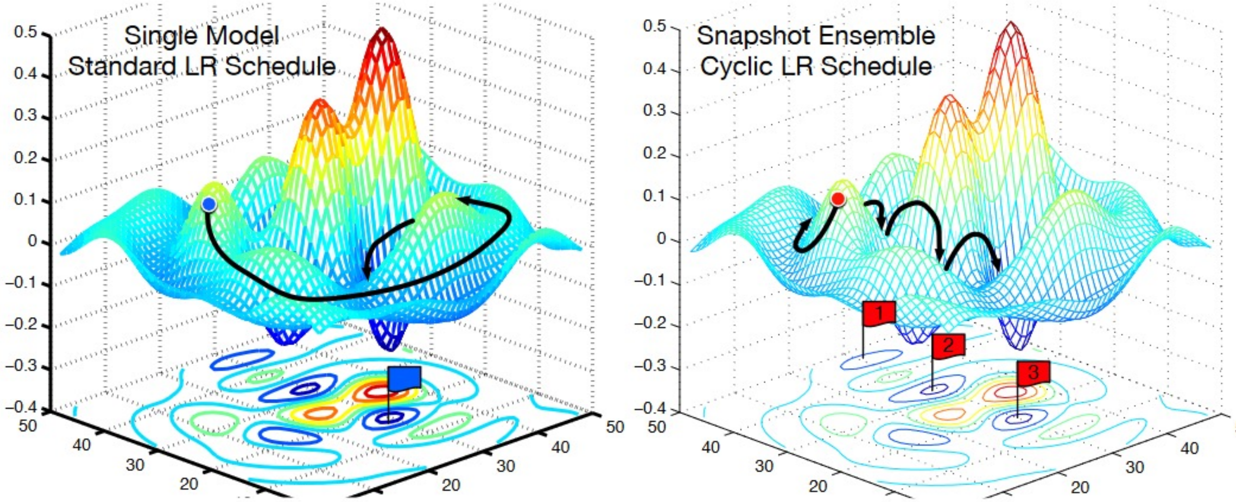


Huang G. et al., ICLR 2017

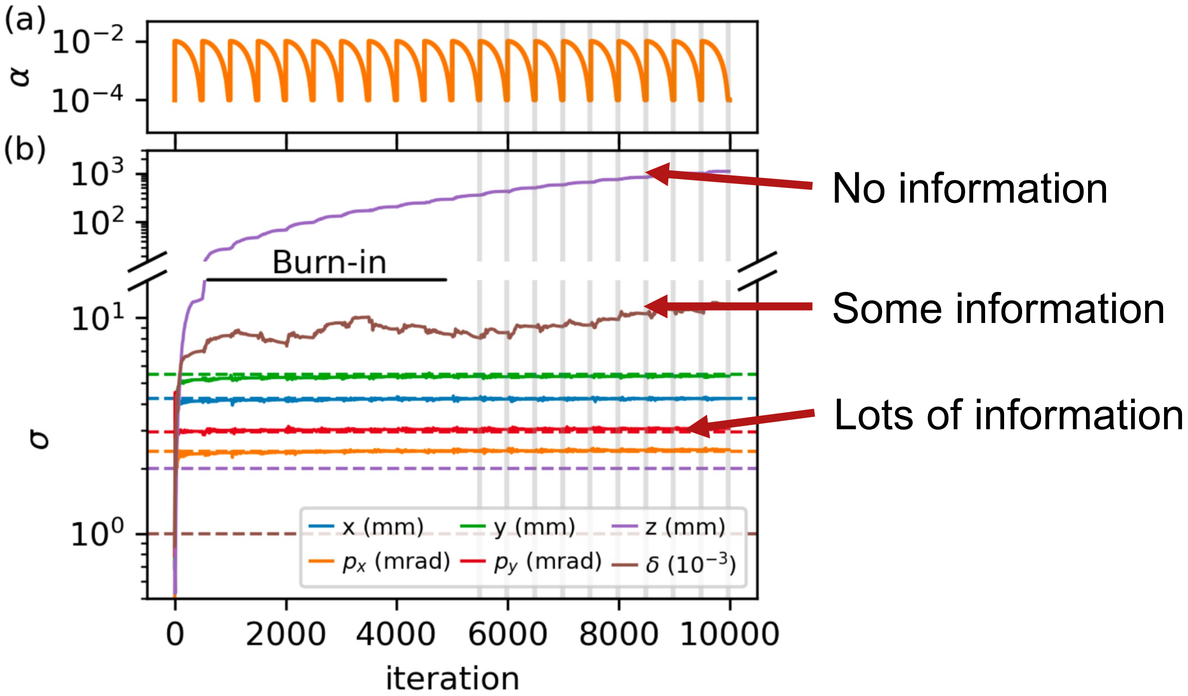
Measuring Model Uncertainty

Create a **snapshot ensemble** to measure uncertainty by cycling the learning rate





Huang G. et al., ICLR 2017

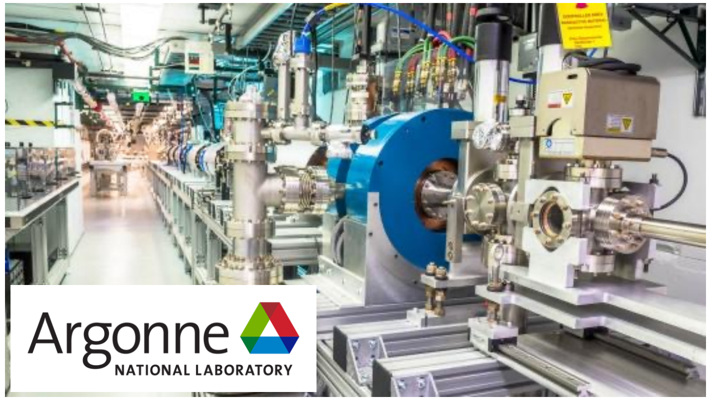


Quadrupole:

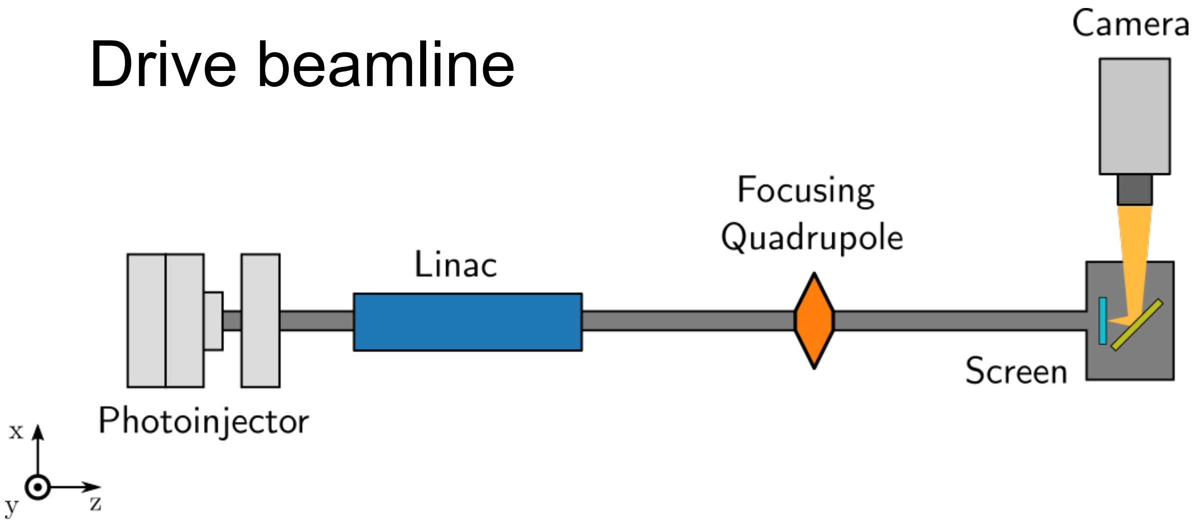
$$H = \frac{p_x^2 + p_y^2}{2(1+p_z)} + \frac{k_1(p_z)}{2}(x^2 - y^2)$$

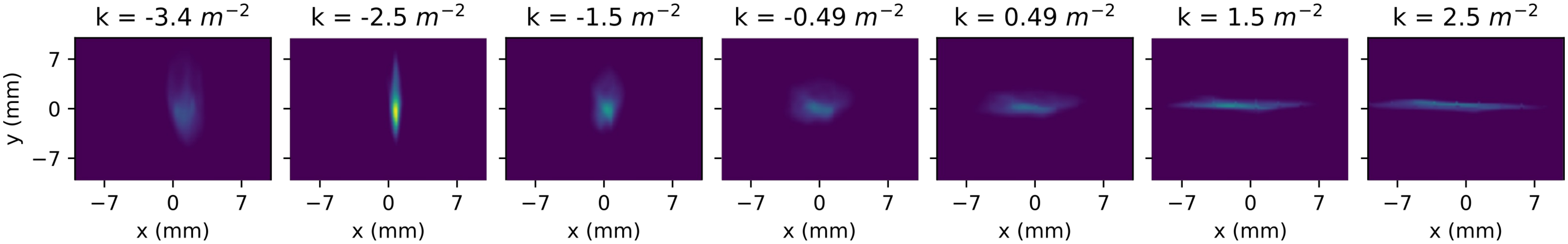
- **Weak dependence on p_z** via chromatic effects
- **No dependence on z**

Tomography Example from AWA

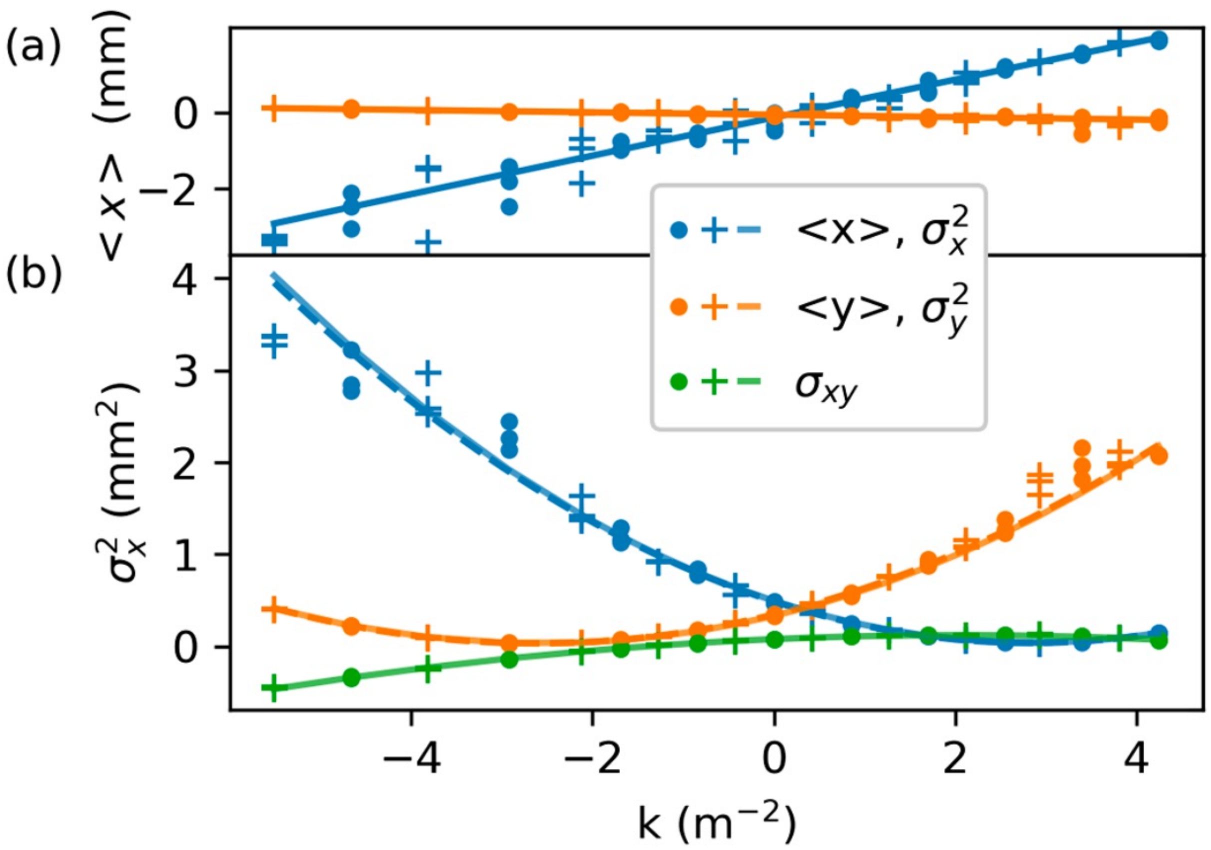


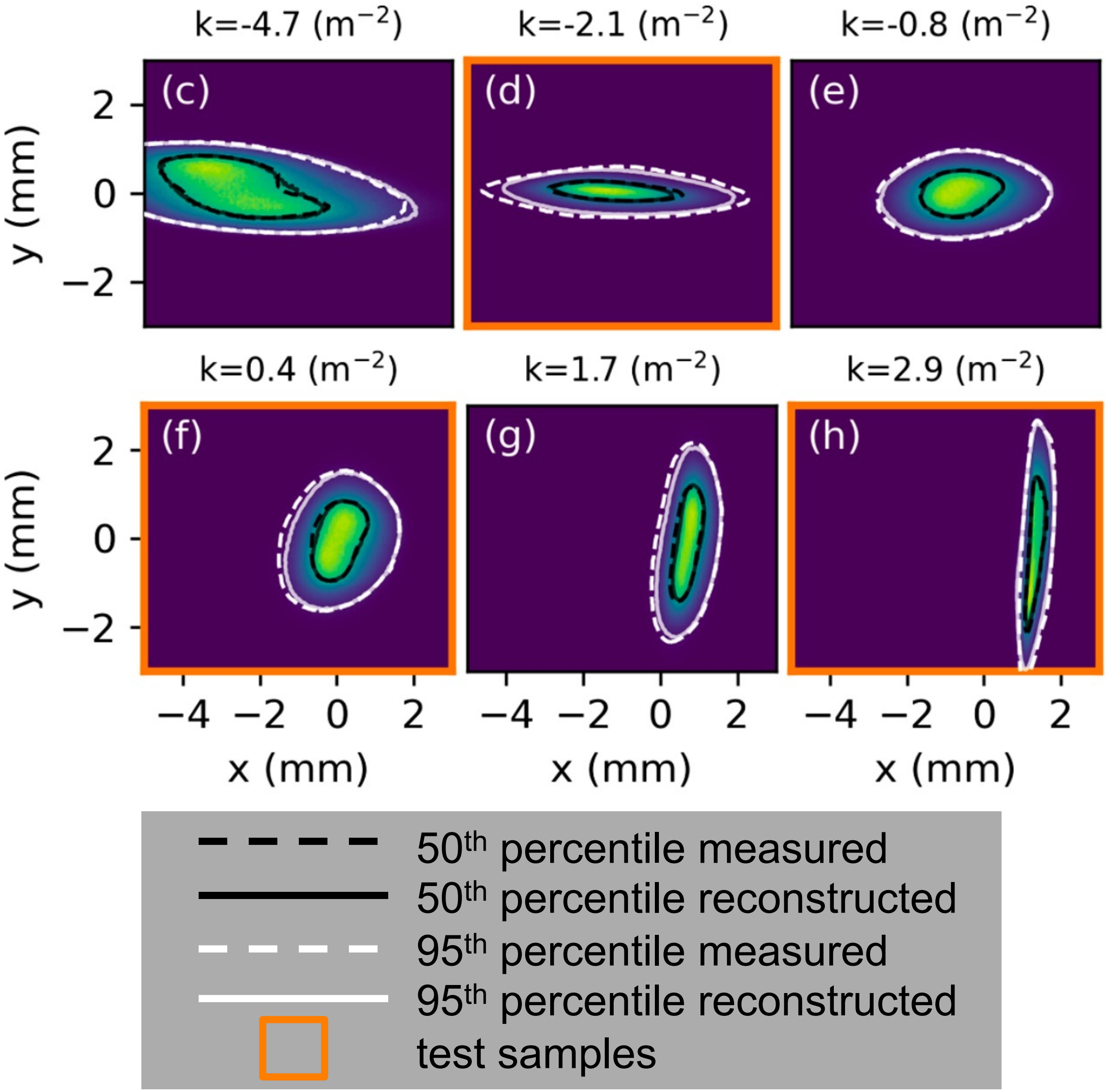
Drive beamline





AWA Reconstruction Results





Detailed reconstruction of 4D phase space in 5 min with only

- **a quadrupole and a screen**
- **10 quad strength, 3 measurements for each**

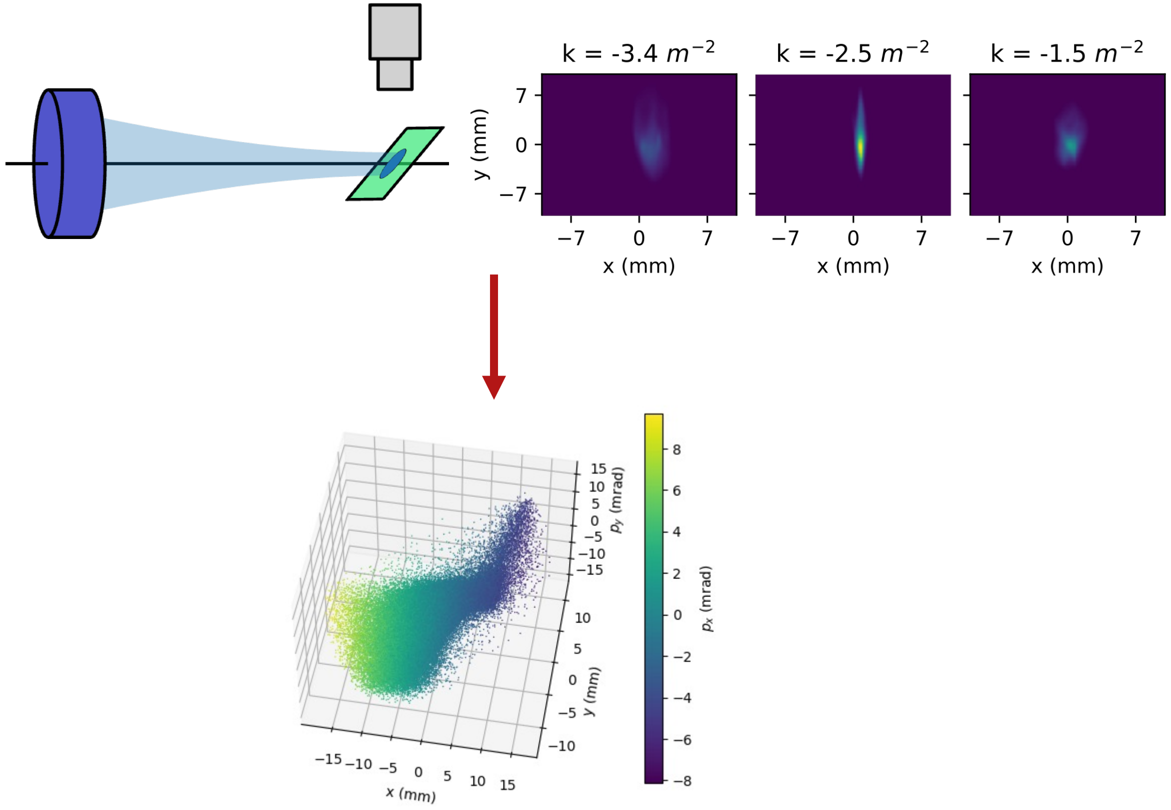
Conclusions

- **4D detailed phase space reconstruction** from **few measurements** and **without special diagnostics**

- Neural Network beam parametrization and differentiable simulations **are not limited by dimensionality**.

- Potentially **extensible to 6D** with the addition of longitudinal diagnostics.

- Can incorporate heterogeneous measurements:
  - More screens, BPMs, …
  - Different types of data



Details: PRL 130, 145001 (2023)

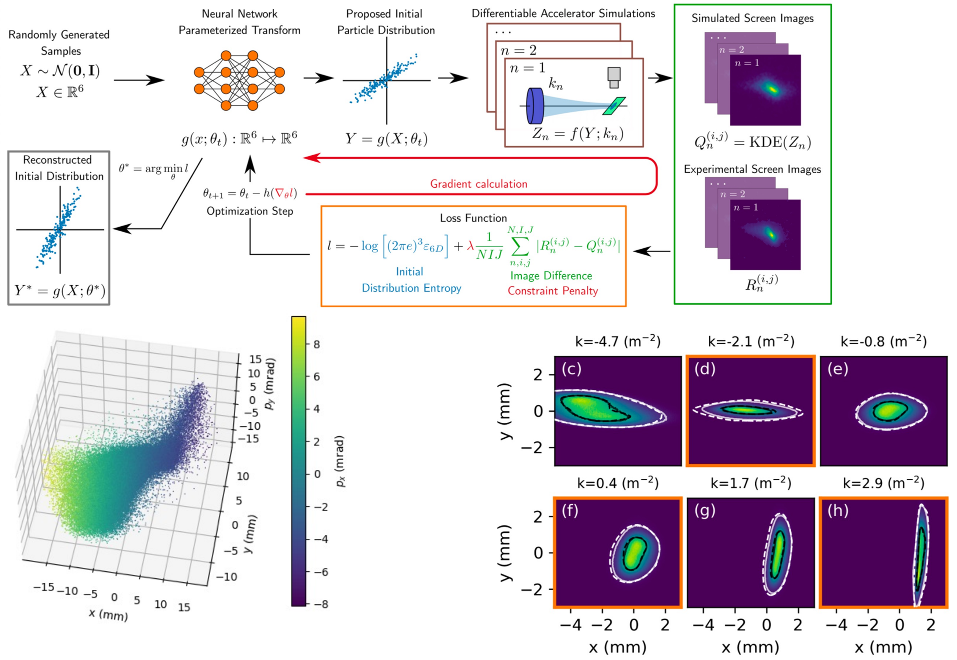
Thanks! Questions?

Phase-Space Reconstruction:

- Ryan Roussel (SLAC)
- Auralee Edelen (SLAC)
- Christopher Mayes (SLAC)
- Daniel Ratner (SLAC)
- Seongyeol Kim (ANL)
- John Power (ANL)
- Eric Wisniewski (ANL)

Differentiable Accelerator Modeling at UChicago:

- Young-Kee Kim
- Chris Pierce
- J.P. Gonzalez-Aguilera



Details: PRL 130, 145001 (2023)

This work was supported by:

- DoE contract No. DE-AC02-76SF00515
- NSF award PHY-1549132, the Center for Bright Beams
- Physical Sciences Division Fellowship, The University of Chicago
- DoE contract No. DE-AC02-05CH11231, NERSC award BES-ERCAP0023724

SLAC NATIONAL ACCELERATOR LABORATORY · Argonne NATIONAL LABORATORY · The Center for BRIGHT BEAMS, A National Science Foundation Science & Technology Center · THE UNIVERSITY OF CHICAGO · NERSC

Backup: Maximum Entropy Loss Function

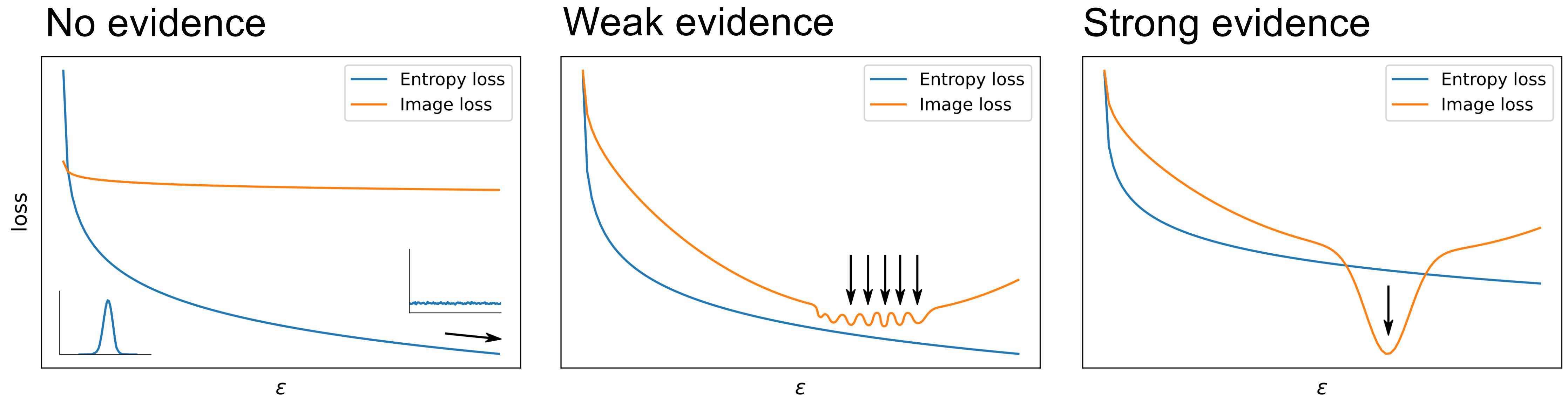
Loss Function

$$l = -\log\left[(2\pi e)^3 \varepsilon_{6D}\right] + \lambda \frac{1}{NIJ}\sum_{n,i,j}^{N,I,J} \left|R_n^{(i,j)} - Q_n^{(i,j)}\right|$$

Initial Distribution Entropy

Image Difference

Constraint Penalty



Backup: Maximum Entropy Tomography (MENT)

Rotate phase space as before, but reconstruct the distribution from 1D projections + **maximize the beam distribution entropy**

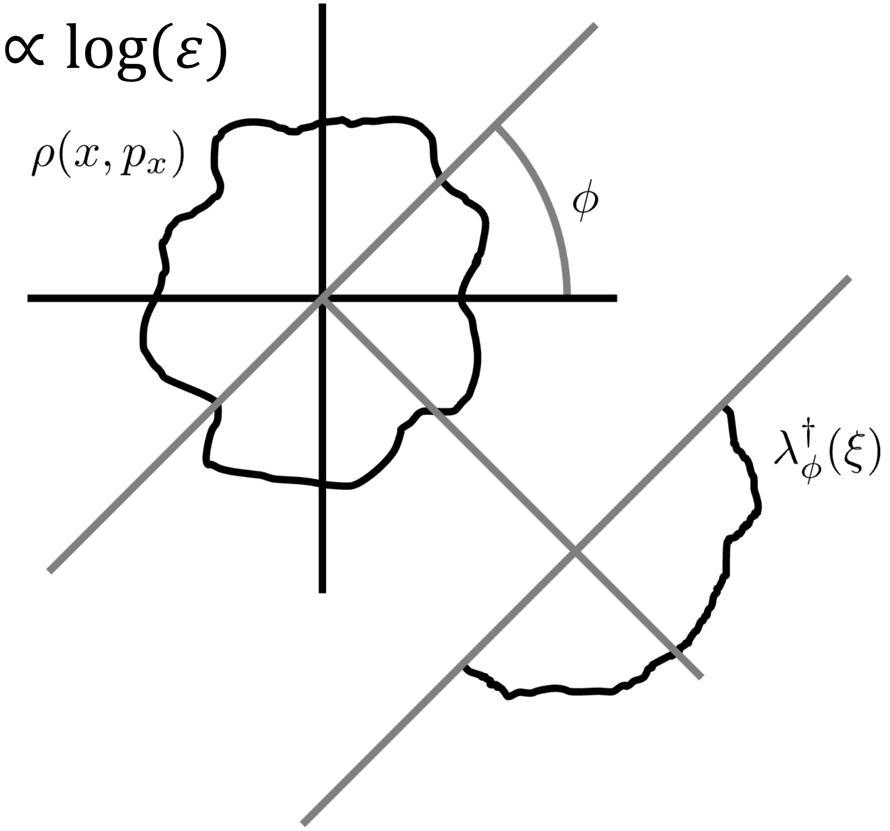
Lagrange multiplier

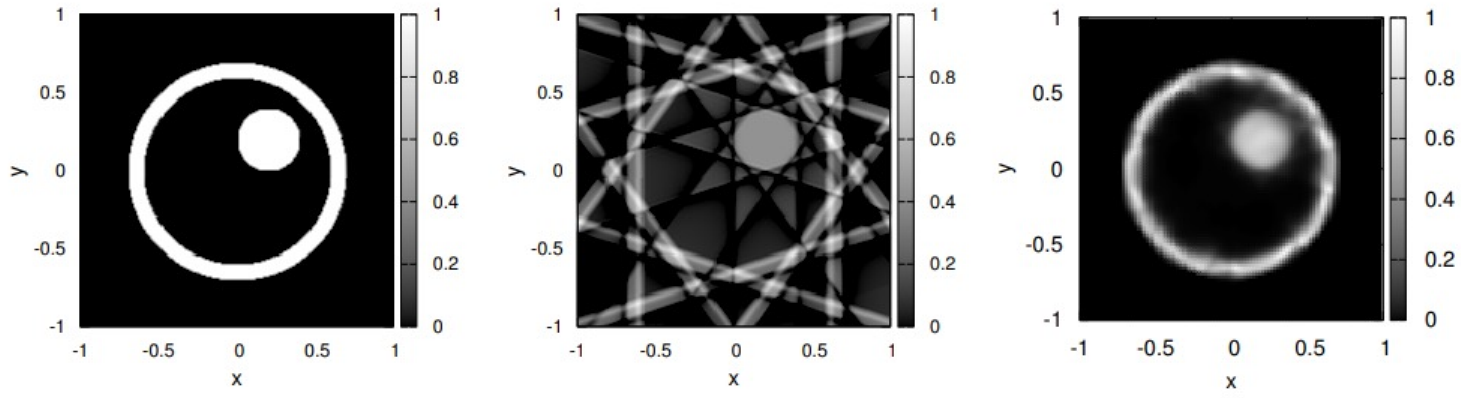
$$\rho^* = \arg\min\{-H(\rho) + \lambda f(\rho)\}$$

Distribution entropy

Discrepancy with measurement

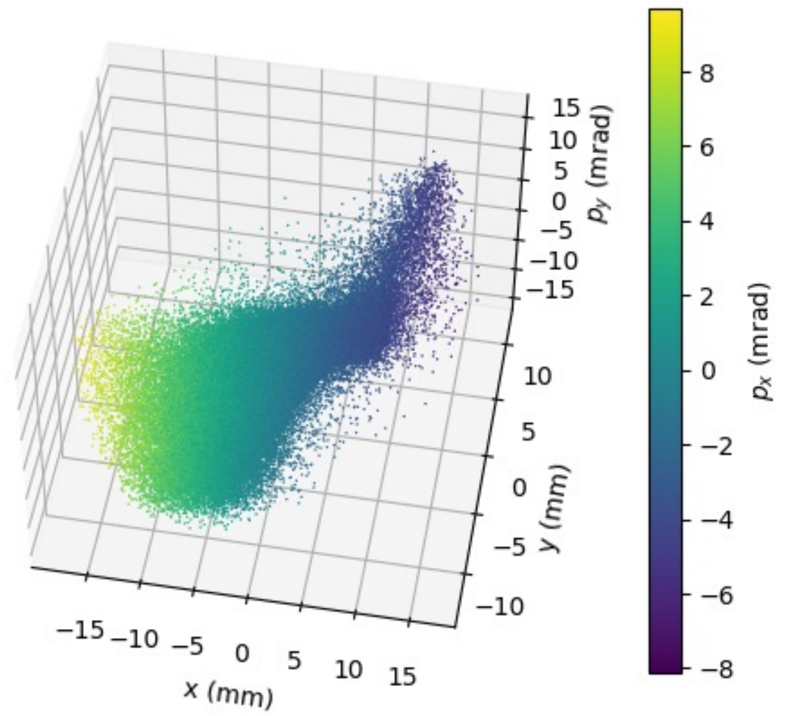
Note: $H \propto \log(\varepsilon)$

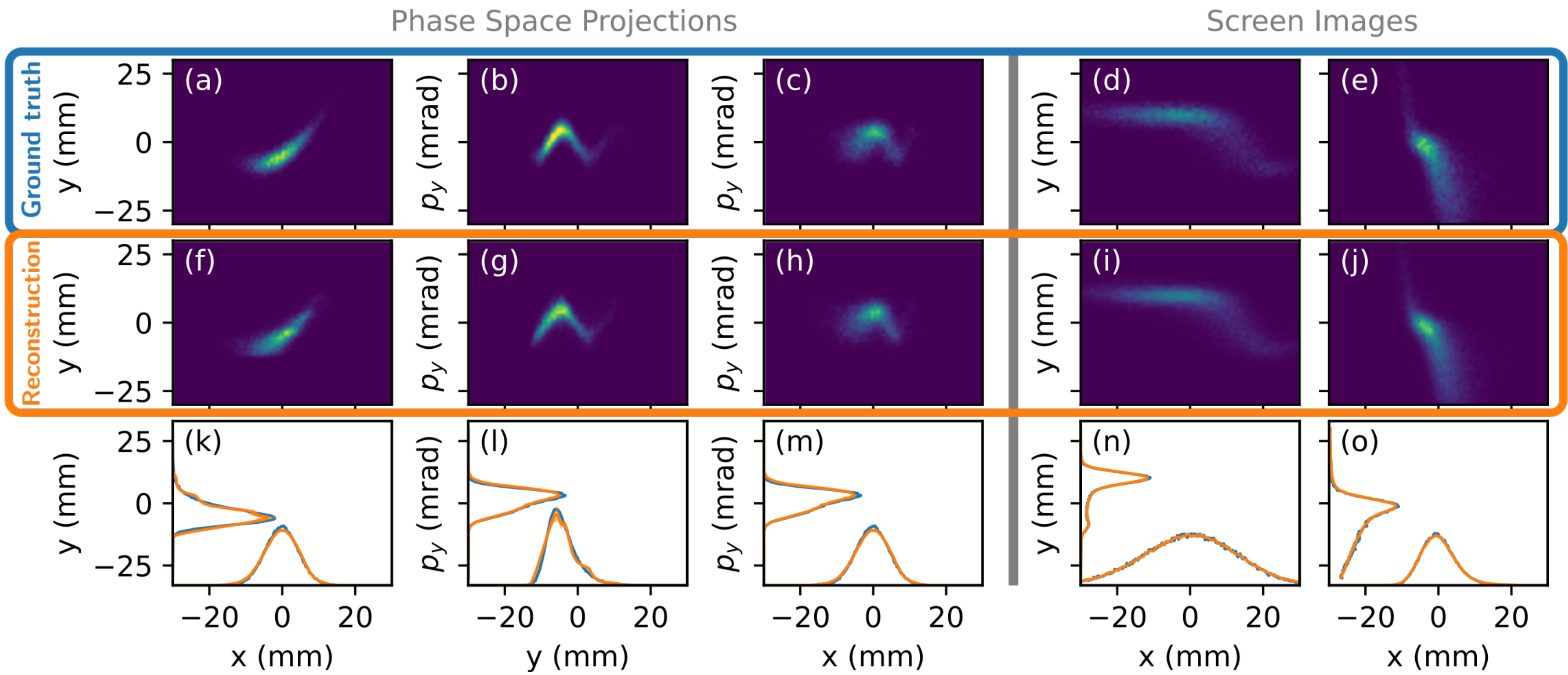




Hock K. and Ibison M., JINST, 2013

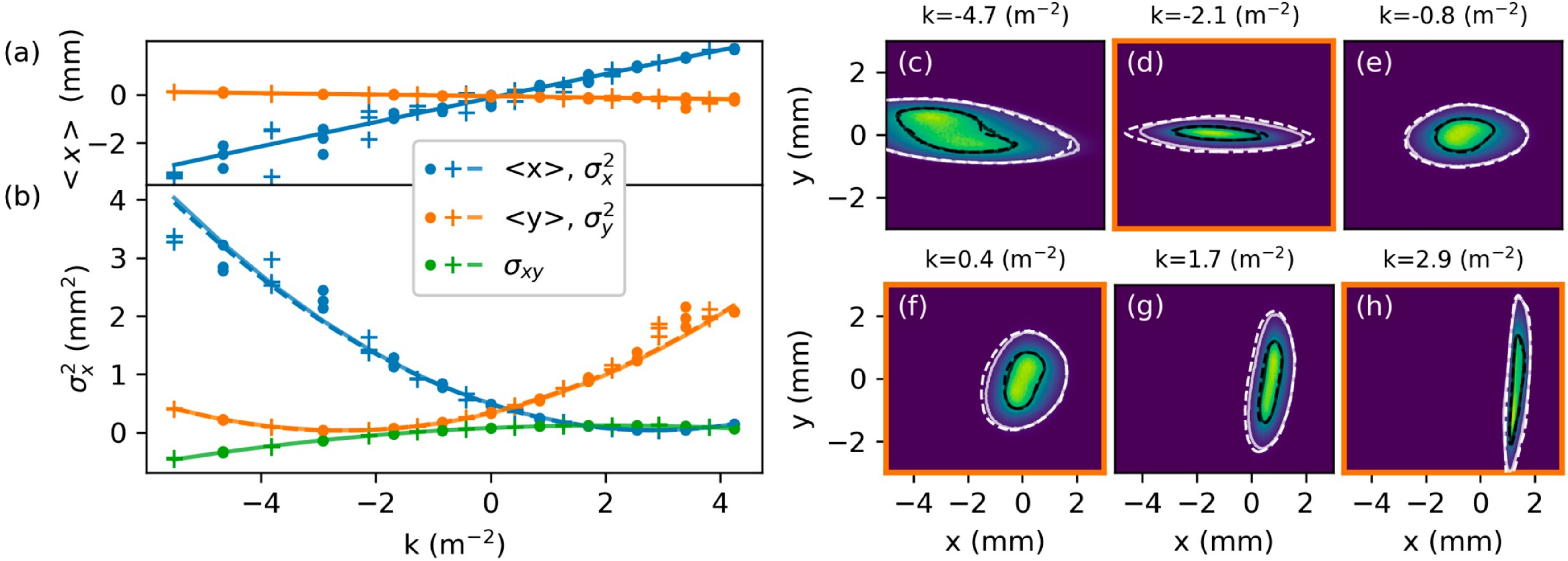
Backup: Synthetic Example Reconstruction





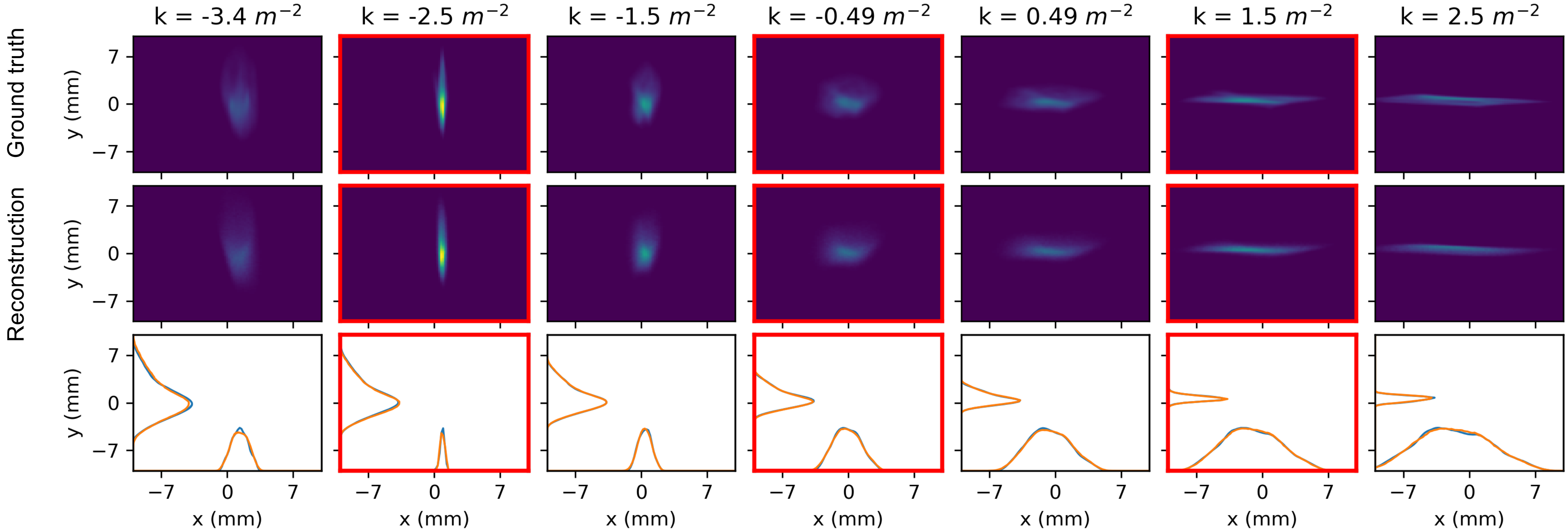
Parameter	Ground truth	rms prediction	Reconstruction	Unit
ε_x	2.00	2.47	2.00 ± 0.01	mm-mrad
ε_y	11.45	14.10	10.84 ± 0.04	mm-mrad
ε_{4D}	18.51	34.83[a]	17.34 ± 0.08	mm^2-mrad^2

Backup: AWA Reconstruction Results



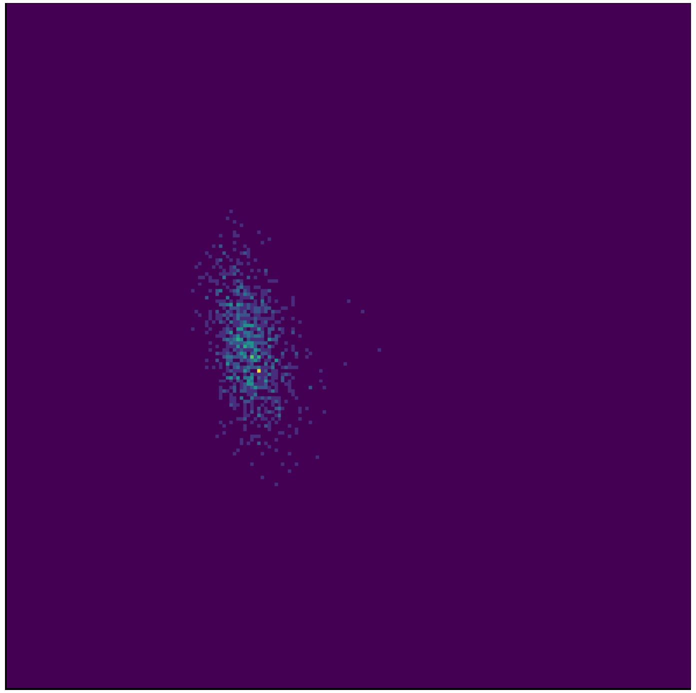
Parameter	rms prediction	Reconstruction	Unit
$\varepsilon_{x,n}$	4.18 ± 0.71	4.23 ± 0.02	mm-mrad
$\varepsilon_{y,n}$	3.65 ± 0.36	3.42 ± 0.02	mm-mrad

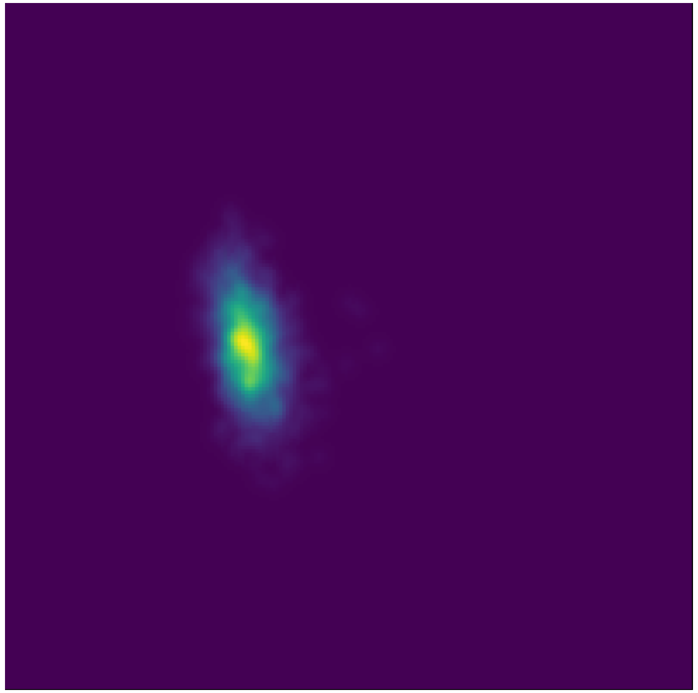
Backup: AWA Reconstruction



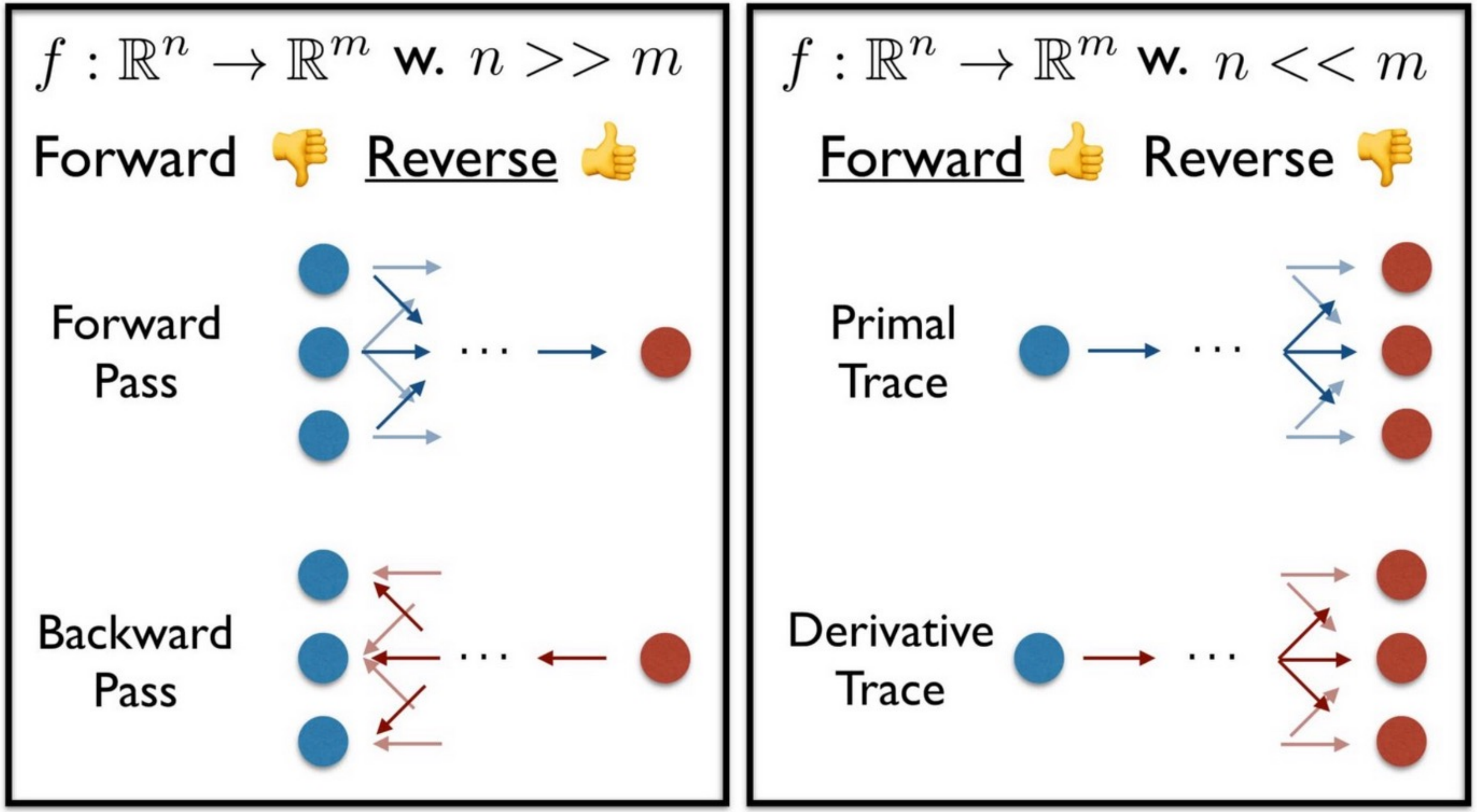
Red border denotes test samples

Backup: Kernel Density Estimation (KDE)



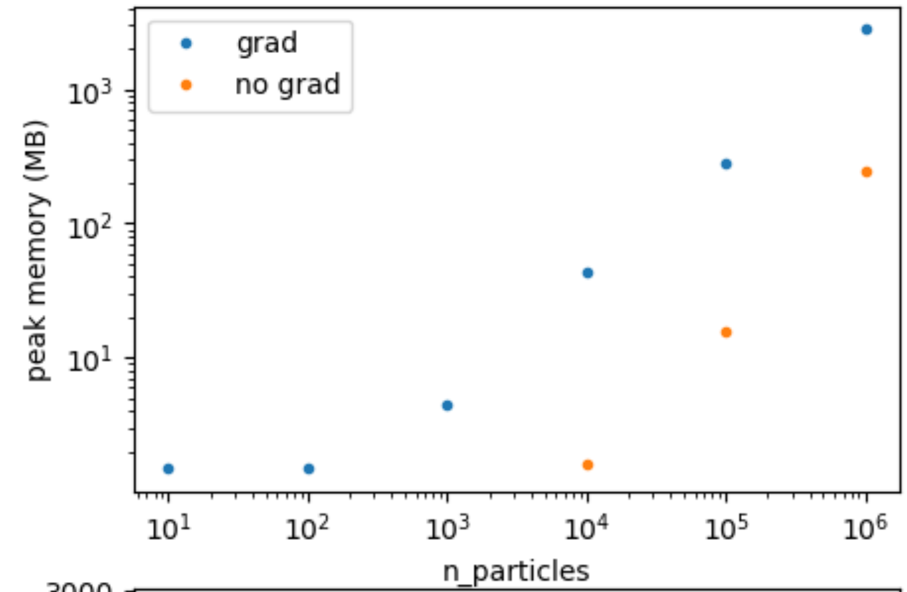


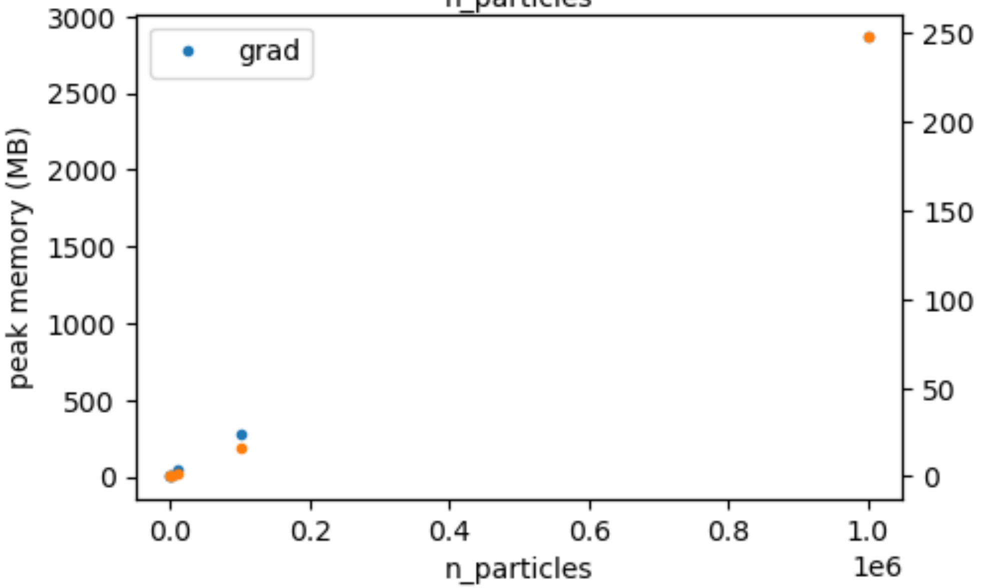
Backup: Reverse vs Forward Autodiff



https://towardsdatascience.com/forward-mode-automatic-differentiation-dual-numbers-8f47351064bf

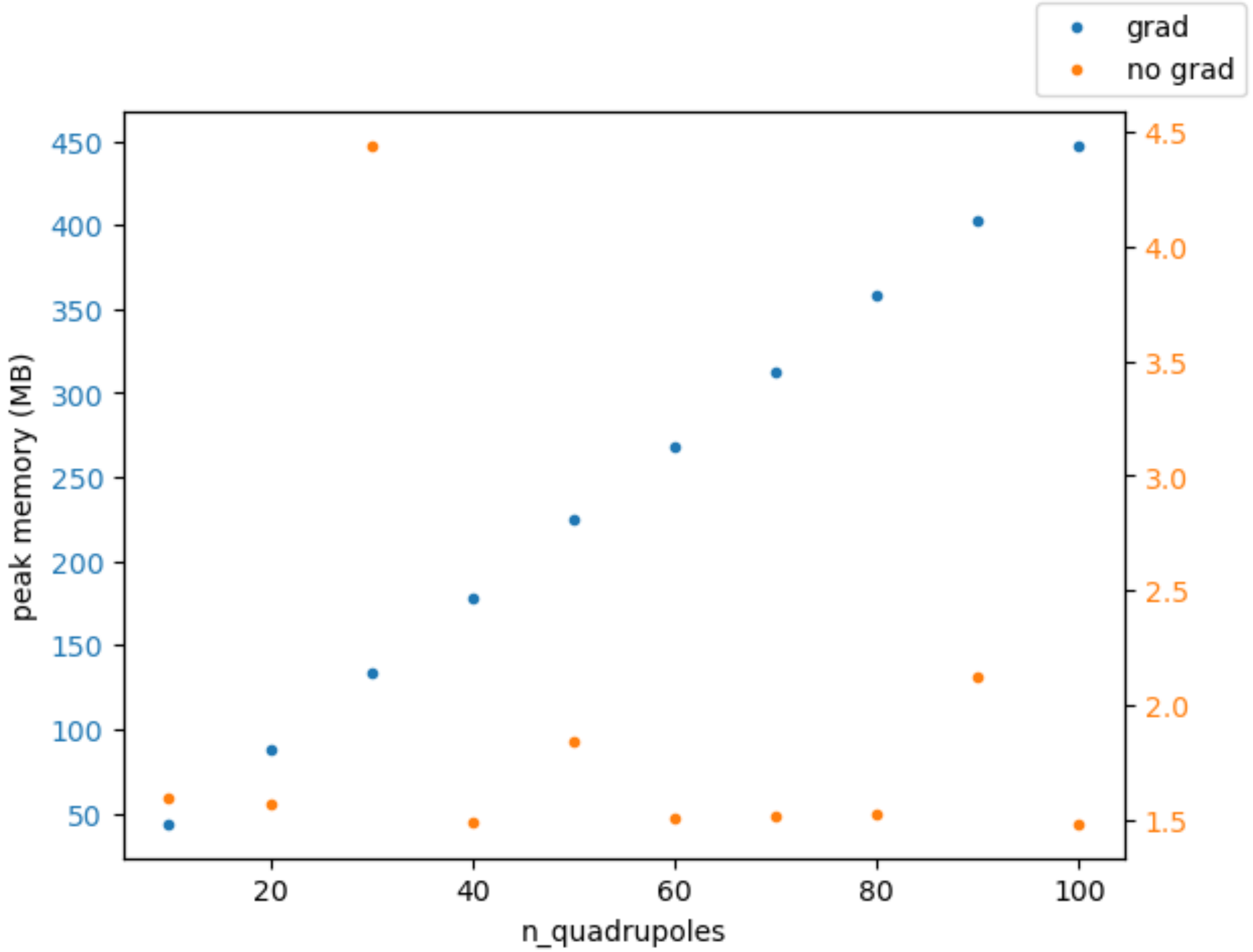
Backup: Memory profiling





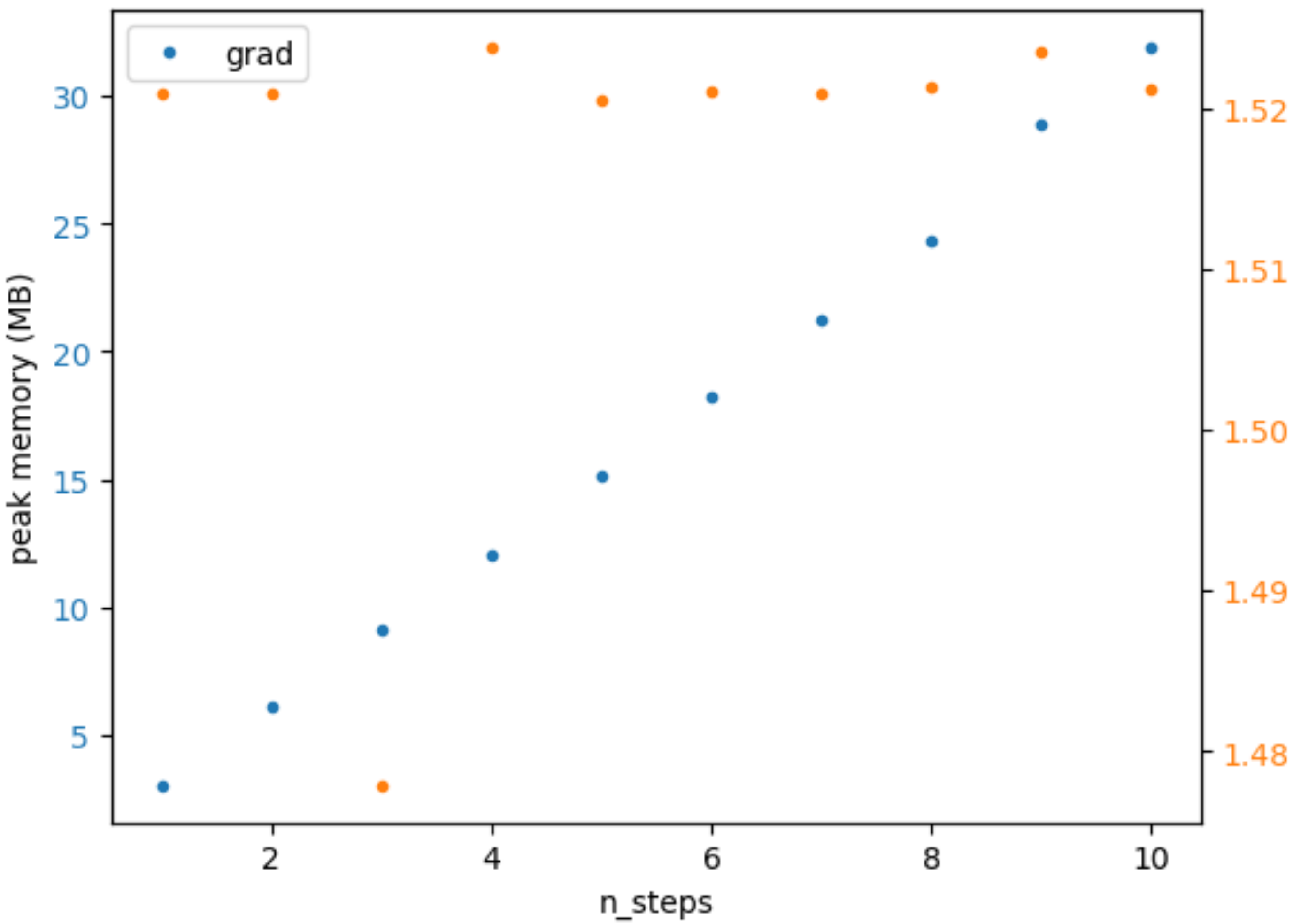
Test 1: 10 quads separated by drifts.
Peak memory vs number of particles

Backup: Memory profiling



Test 2: 10^4 particles
Peak memory vs n quads

Backup: Memory profiling



Test 3: 10^4 particles
Peak memory vs n slices in single quad+drift