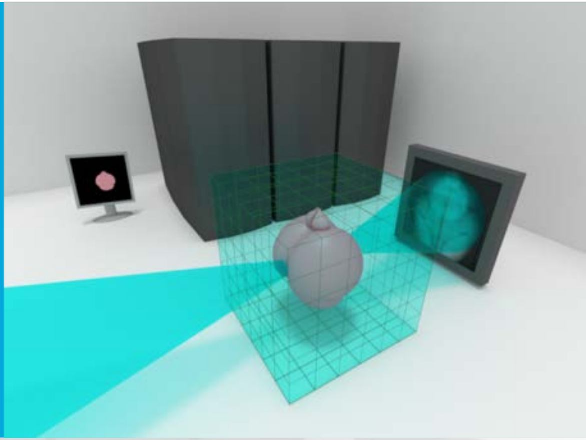
Three dimensions, two microscopes, one code

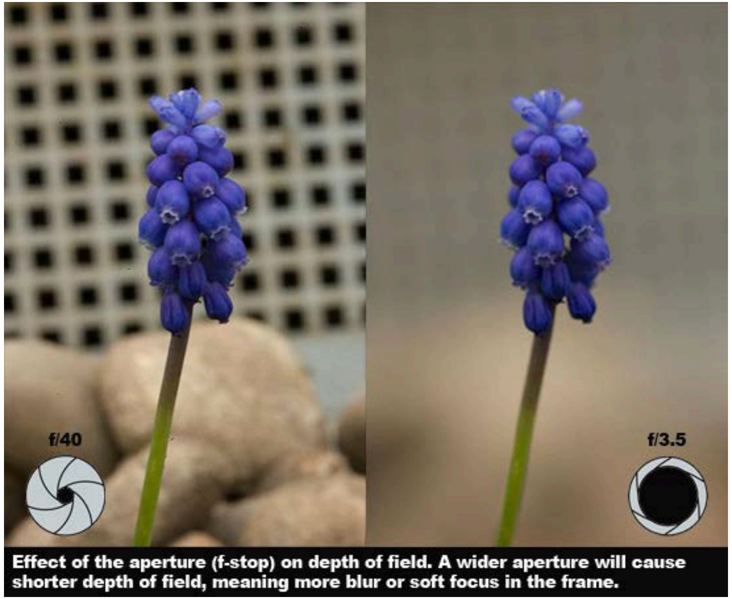
Automatic differentiation for x-ray nanotomography beyond the depth of focus limit



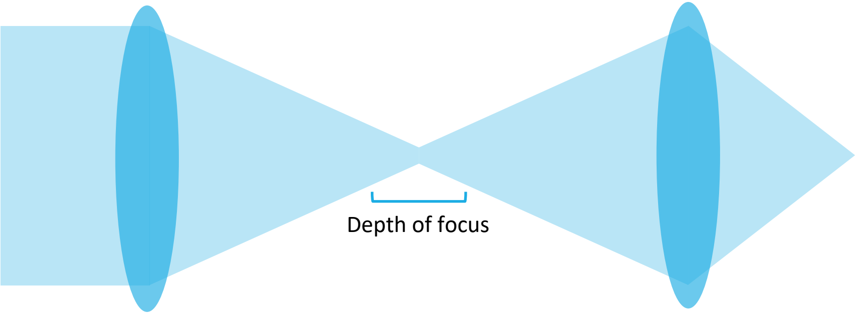
Ming Du
Postdoctoral Appointee
Argonne National Laboratory

October 2, 2019

Depth of focus: what is it



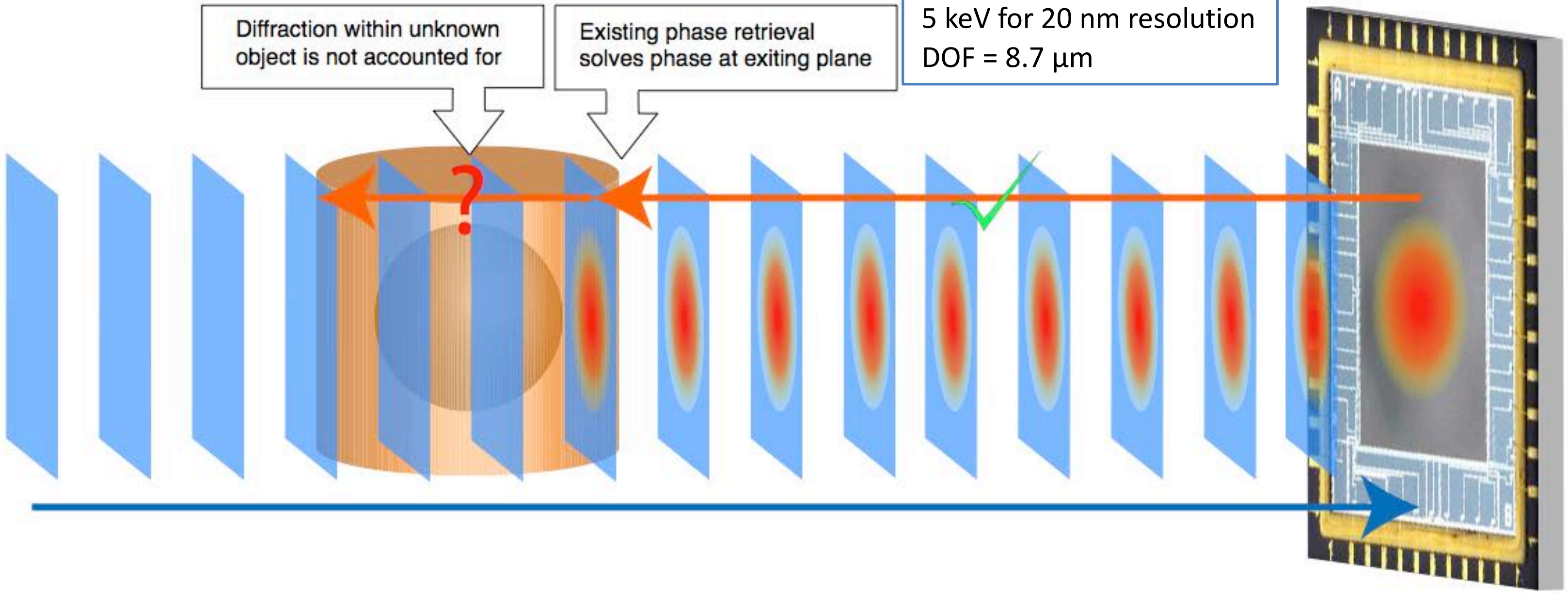
http://www.elementsofcinema.com/cinematography/depth-of-field.html



In-sample diffraction must be accounted for when t > DOF

$$\mathrm{DOF} = \frac{2}{0.61^2}\frac{\delta_r^2}{\lambda} \simeq 5.4\delta_r\frac{\delta_r}{\lambda}$$ (5.2 prefactor as in Tsai et al., 2016)

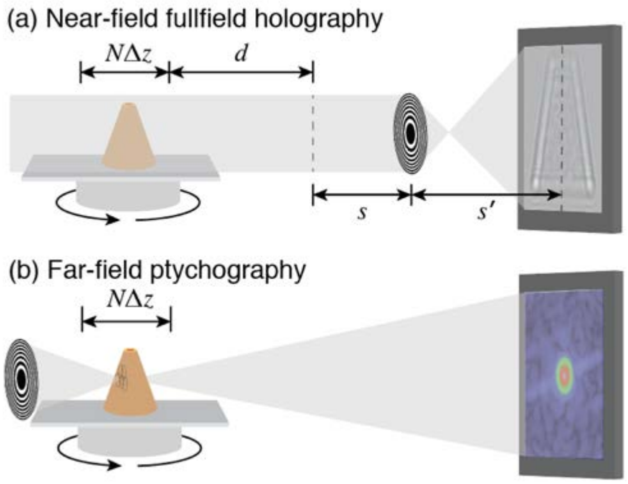
5 keV for 20 nm resolution
DOF = 8.7 µm



[1] E. H. R. Tsai, I. Usov, A. Diaz, A. Menzel, and M. Guizar-Sicairos, "X-ray ptychography with extended depth of field," Opt Express 24, 29089–20 (2016).

The forward model

$$L = \frac{1}{N_\theta N_p N_k} \sum_{\theta,k} \left\| \boxed{\left| f(\boldsymbol{x}, \theta, k, \Delta z, d) \right|} - \sqrt{\boldsymbol{y}_{\theta,k}} \right\|^2$$

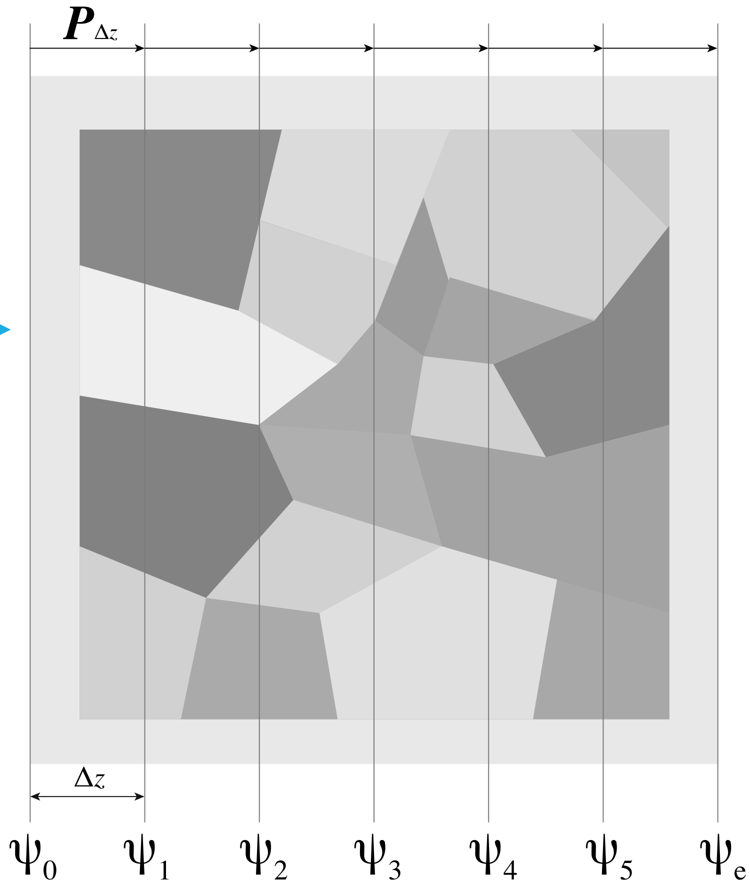


$$f(\boldsymbol{x}) = \bar{\boldsymbol{y}}(\boldsymbol{x}) = \boldsymbol{P}_d \boldsymbol{M}_{\boldsymbol{x}} \psi_0$$

$$f(\boldsymbol{x}) = \bar{\boldsymbol{y}}(\boldsymbol{x}) = \boldsymbol{P}_\infty \boldsymbol{M}_{\boldsymbol{x}} \psi_0$$

Multislice propagation

Fresnel (near-field) propagation



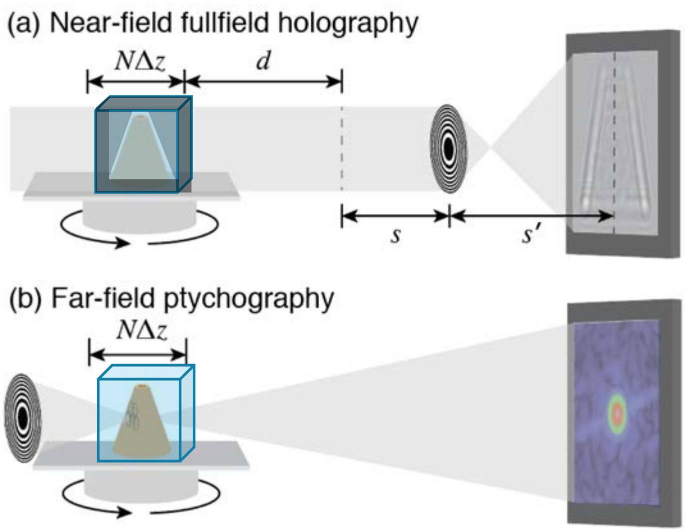
[1] J. M. Cowley and A. F. Moodie, "The scattering of electrons by atoms and crystals. I. A new theoretical approach," Acta Cryst (1957). Q10, 609-619 [doi:10.1107/S0365110X57002194] 10, 1–11 (1957).

Just something else…

$$L = \frac{1}{N_\theta N_p N_k} \sum_{\theta,k} \left\| |f(\boldsymbol{x},\theta,k,\Delta z,d)| - \sqrt{\boldsymbol{y}_{\theta,k}} \right\|^2 + \alpha_\delta |\boldsymbol{x}_\delta|_1 + \alpha_\beta |\boldsymbol{x}_\beta|_1 + \gamma \mathrm{TV}(\boldsymbol{x}_\delta)$$

$$\text{subject to } x_w = 0 \text{ for } x_w \notin \Theta \text{ and } x_w \geq 0 \text{ for } x_w \in \Theta.$$

L1 norm	• Object sparsity • Noise and artifact suppression
Total variation	• Object gradient sparsity • Noise and artifact suppression
Non-negativity	• Solution stabilization • Works as long as one avoids anomalous dispersion at an absorption edge
Finite support and shrink-wrap	• For fullfield holography only • Initialized by thresholding conventional reconstruction results • Shrunk by taking out low-value voxels per several iterations



[1] Tibshirani, R. (2011). *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *73*(3), 273–282.
[2] Horn, R. A. and Johnson, C. R. "Norms for Vectors and Matrices." Ch. 5 in Matrix Analysis. Cambridge, England: Cambridge University Press, 1990.
[3] Sidky, E. Y., & Pan, X. (2008). Physics in Medicine and Biology, 53(17), 4777–4807.
[4] Marchesini, S., He, H., Chapman, H. N., Hau-Riege, S. P., Noy, A., Howells, M. R., et al. (2003). Physical Review B, 68(14), 843–4.

It's all about gradient

$$L = \frac{1}{N_\theta N_p N_k} \sum_{\theta,k} \left\| |f(\boldsymbol{x}, \theta, k, \Delta z, d)| - \sqrt{\boldsymbol{y}_{\theta,k}} \right\|^2 + \alpha_\delta |\boldsymbol{x}_\delta|_1 + \alpha_\beta |\boldsymbol{x}_\beta|_1 + \gamma \mathrm{TV}(\boldsymbol{x}_\delta)$$

subject to $x_w = 0$ for $x_w \notin \Theta$ and $x_w \geq 0$ for $x_w \in \Theta$.

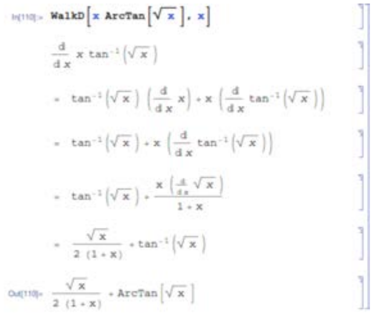
$$\bar{\boldsymbol{x}} = \underset{\boldsymbol{x}}{\operatorname{argmin}}\{L(\boldsymbol{x})\}$$

$$\nabla_{\boldsymbol{x}} L = ?$$

$$
\begin{aligned}
L &= c \sum_{\theta,k} \left\| |f(\boldsymbol{x},\theta,k,\Delta z,d)| - \sqrt{y_{\theta,k}} \right\|^2 \\
\nabla_{\boldsymbol{x}} L &= 2c \sum_{\theta,k} \left(|f(\boldsymbol{x},\theta,k,\Delta z,d)| - \sqrt{y_{\theta,k}}\right)^{\mathrm{T}} \nabla_{\boldsymbol{x}} f \\
g_{\theta,k} &= f(\boldsymbol{x},\theta,k,\Delta z,d) = \boldsymbol{P}_d \boldsymbol{M}_{\boldsymbol{x},\theta,\Delta z} \psi_{0,k} \\
\nabla_{\boldsymbol{x}} g_{\theta,k} &= \boldsymbol{P}_d \frac{d\boldsymbol{M}_{\boldsymbol{x},\theta,\Delta z}}{d\boldsymbol{x}} \psi_{0,k} \\
\boldsymbol{M}_{\boldsymbol{x},\theta,\Delta z} &= \prod_j^J \boldsymbol{P}_{\Delta z} \boldsymbol{A}_{\boldsymbol{x},\theta,j} \\
&= \prod_j^J \boldsymbol{P}_{\Delta z} \exp[\mathrm{diag}(\boldsymbol{S}_j \boldsymbol{R}_\theta \boldsymbol{x})] \\
\frac{d\boldsymbol{M}_{\boldsymbol{x},\theta,\Delta z}}{d\boldsymbol{x}} &= \sum_j^J \Bigg\{ \prod_{j+1}^J \left\{ \boldsymbol{P}_{\Delta z} \exp[\mathrm{diag}(\boldsymbol{S}_j \boldsymbol{R}_\theta \boldsymbol{x})] \right\} \\
&\quad \left\{ \boldsymbol{P}_{\Delta z} \exp[\mathrm{diag}(\boldsymbol{S}_j \boldsymbol{R}_\theta \boldsymbol{x})] \mathrm{diag}(\mathbf{S}_j \mathbf{R}_\theta \right\} \prod_0^{j-1} \left\{ \mathbf{P}_{\Delta z} \exp[\mathrm{diag}(\mathbf{S}_j \mathbf{R}_\theta \boldsymbol{x})] \right\} \Bigg\}
\end{aligned}
$$

Automatic differentiation: more than machine learning

$$\frac{df}{dx} = \frac{f(x+\epsilon) - f(x-\epsilon)}{2\epsilon}$$


In[110]:= WalkD[x ArcTan[√x], x]

$\frac{d}{dx} x \tan^{-1}(\sqrt{x})$

$= \tan^{-1}(\sqrt{x}) \left(\frac{d}{dx} x\right) + x \left(\frac{d}{dx} \tan^{-1}(\sqrt{x})\right)$

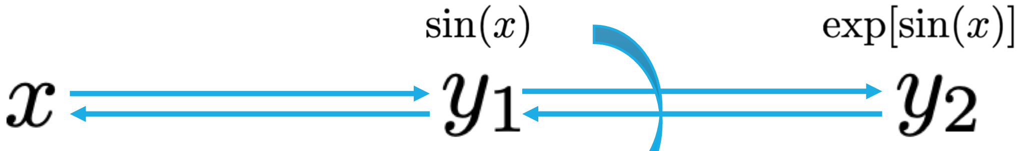
$= \tan^{-1}(\sqrt{x}) + x \left(\frac{d}{dx} \tan^{-1}(\sqrt{x})\right)$

$= \tan^{-1}(\sqrt{x}) + \frac{x \left(\frac{d}{dx} \sqrt{x}\right)}{1+x}$

$= \frac{\sqrt{x}}{2(1+x)} + \tan^{-1}(\sqrt{x})$

Out[110]= $\frac{\sqrt{x}}{2(1+x)}$ + ArcTan[√x]

$$f(x) = y_2[y_1(x)]$$
$$y_1(w) = \sin(w)$$
$$y_2(w) = \exp(w)$$



$$\frac{dy_1}{dx} = \cos(x) \qquad \frac{dy_2}{dy_1} = \exp(y_1)$$

$$\frac{df}{dx} = \frac{dy_2}{dy_1}\frac{dy_1}{dx}$$

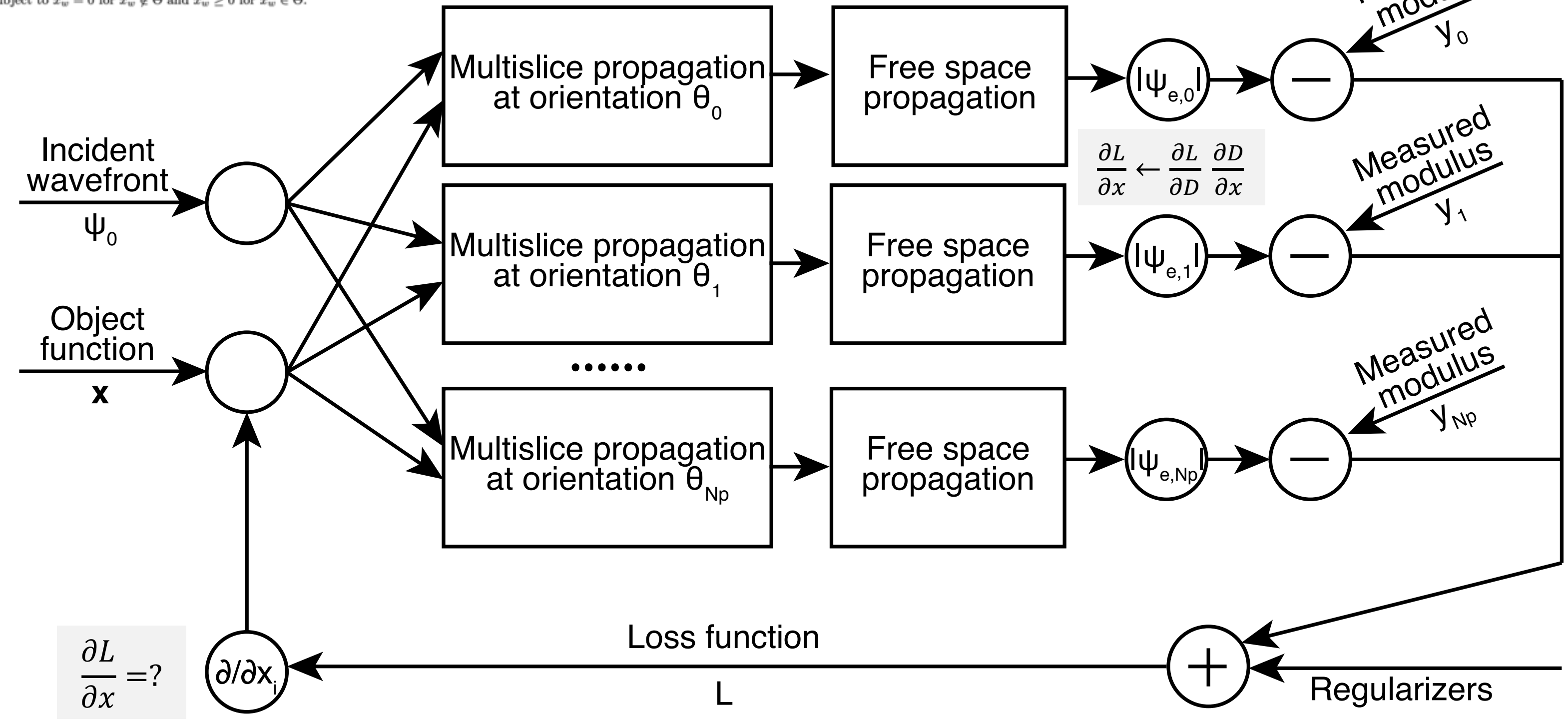
Automatic differentiation: more than machine learning

$$L = \frac{1}{N_\theta N_p N_k} \sum_{\theta,k} \underbrace{\left\| |f(\boldsymbol{x}, \theta, k, \Delta z, d)|^2 - \boldsymbol{y} \right\|^2}_{D} + \alpha_\delta |\boldsymbol{x}_\delta|_1 + \alpha_\beta |\boldsymbol{x}_\beta|_1 + \gamma \mathrm{TV}(\boldsymbol{x}_\delta)$$

subject to $x_w = 0$ for $x_w \notin \Theta$ and $x_w \geq 0$ for $x_w \in \Theta$.

$$\frac{\partial P[M(x)]}{\partial x} \leftarrow \frac{\partial P[M(x)]}{\partial M(x)} \frac{\partial M(x)}{\partial x}$$

$$\frac{\partial D}{\partial x} \leftarrow \frac{\partial D}{\partial P[M(x)]} \frac{\partial P[M(x)]}{\partial x}$$



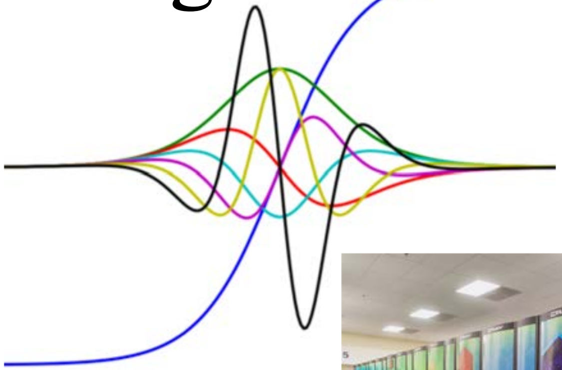
Automatic differentiation: more than machine learning

TensorFlow

(Has been applied for ptychography by Nashed *et al.*)

HOROVOD

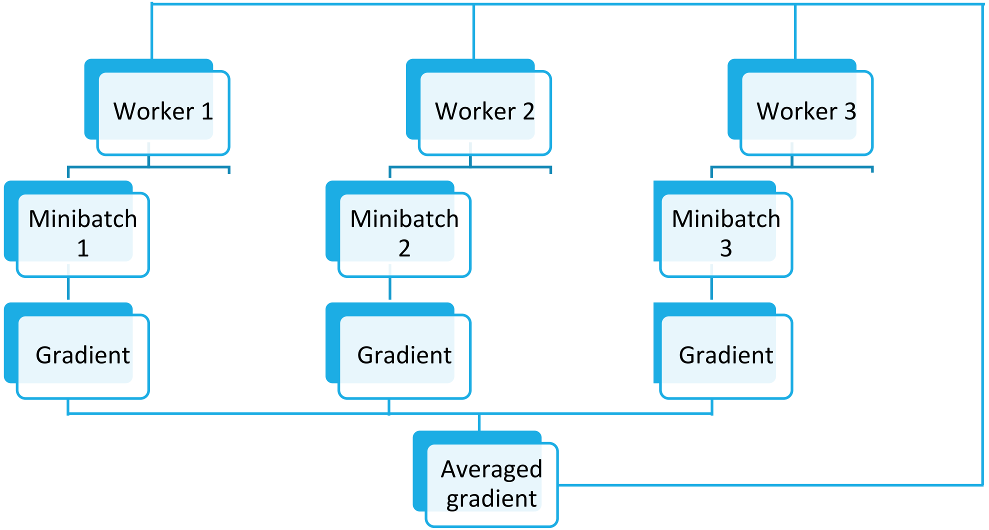
Autograd



[1] M. Abadi *et al.*, "Tensorflow: Large-scale machine learning on heterogeneous distributed systems," arXiv cs.DC, arXiv:1603.04467 (2016).
[2] D. Maclaurin, "Modeling, Inference and Optimization with Composable Differentiable Procedures," (PhD thesis), Harvard University (2014).
[3] Nashed, Y. S. G., Peterka, T., Deng, J., & Jacobsen, C. (2017). Procedia Computer Science, 108, 404–414.

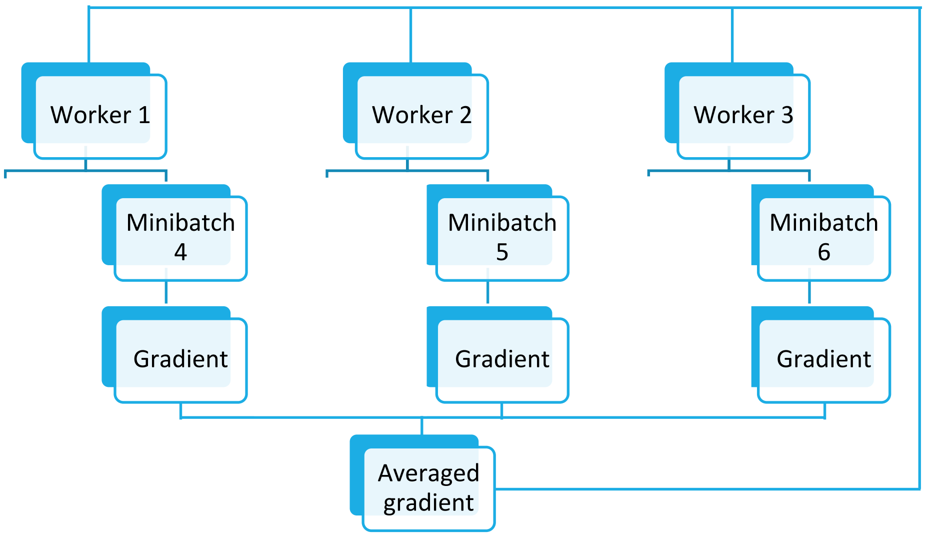
Automatic differentiation: more than machine learning

Divide projection data into minibatches to fit in memory



Automatic differentiation: more than machine learning

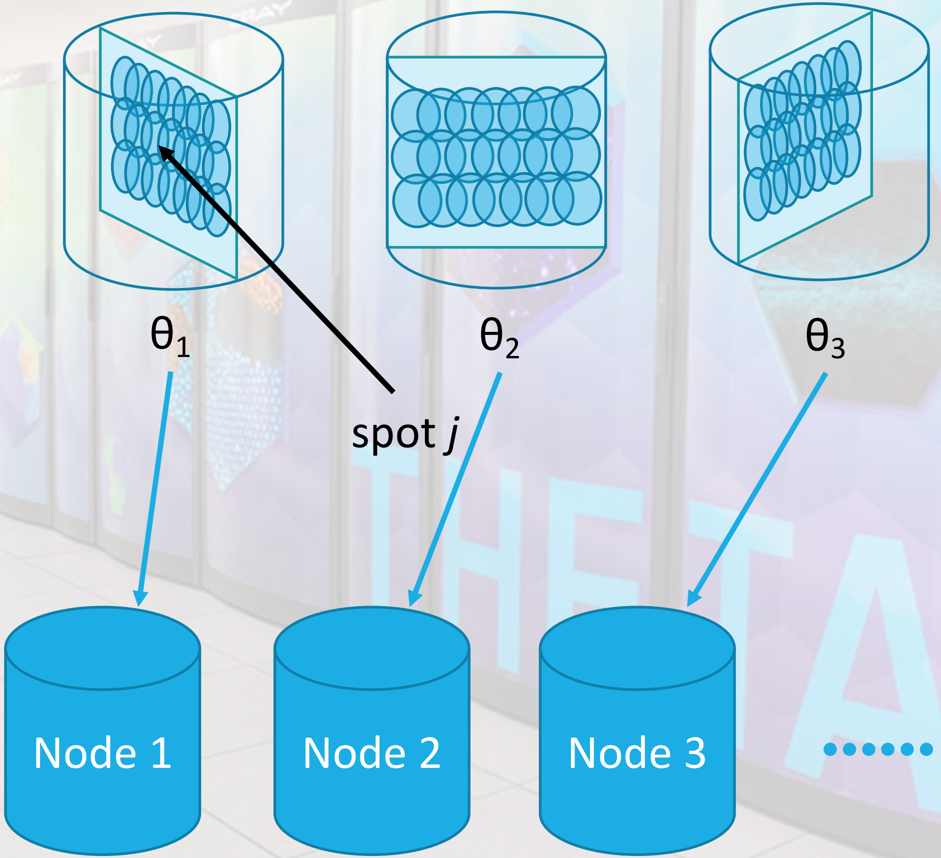
Divide projection data into minibatches to fit in memory



Automatic differentiation: more than machine learning


Taylor Childers

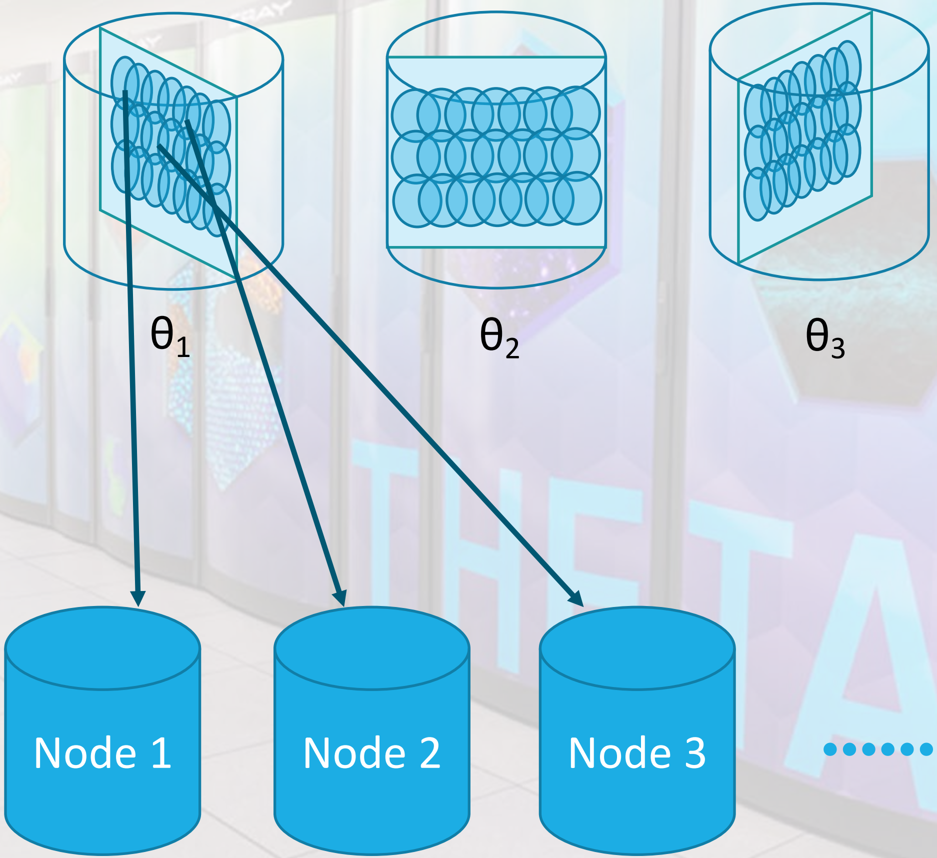

Corey Adams



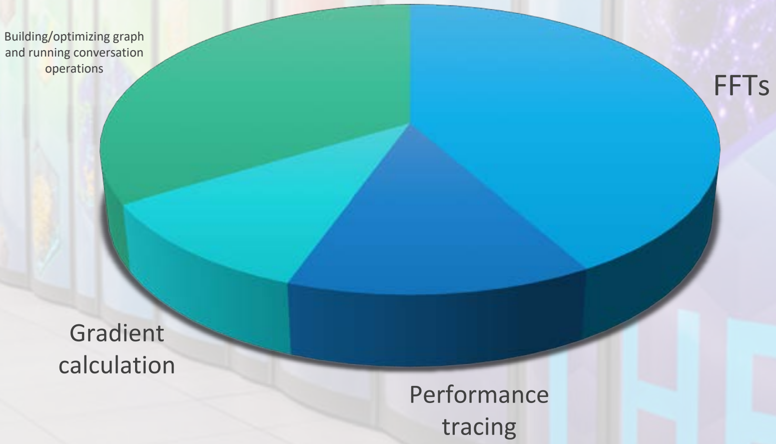
Automatic differentiation: more than machine learning

Taylor Childers

Corey Adams



Automatic differentiation: more than machine learning



TensorFlow

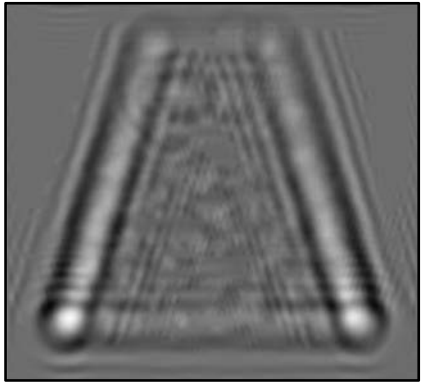
Autograd

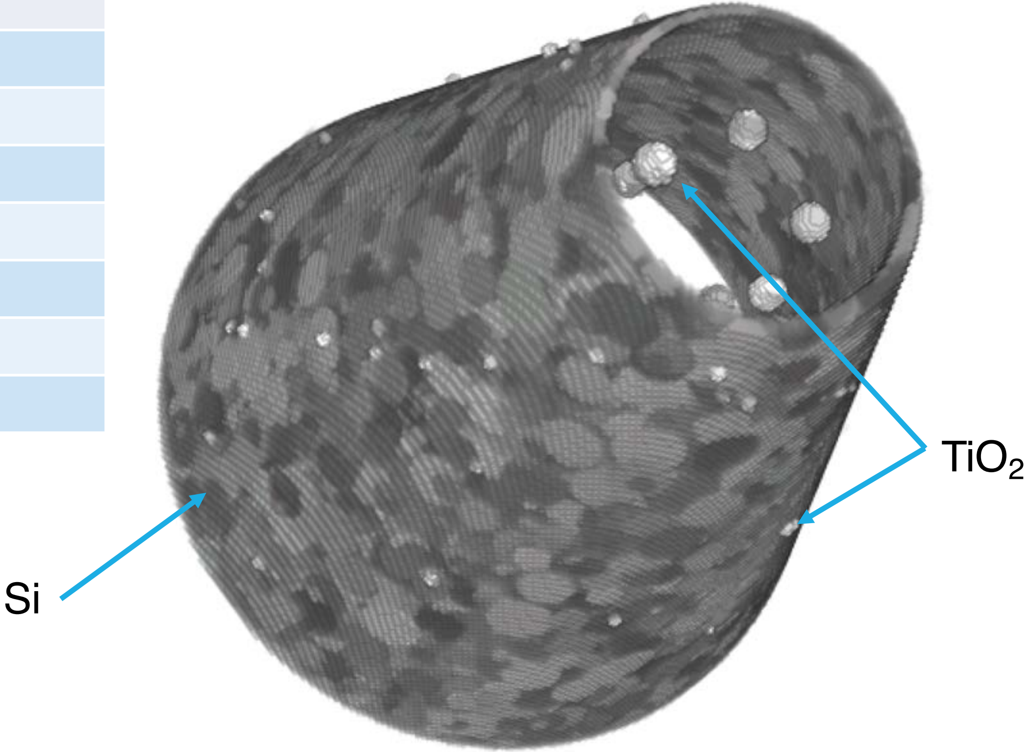
MPI4py

For FFT-heavy applications on KNL nodes

Test case 1: a designed sample

Pixel size (nm)	1
Energy (eV)	5000
Depth of focus (nm)	21.77
Largest sample thickness (nm)	200
Sample-detector distance (nm)	1000
Object grid size	256^3
# of projections	500
# of diffraction spots in ptychography	23×23

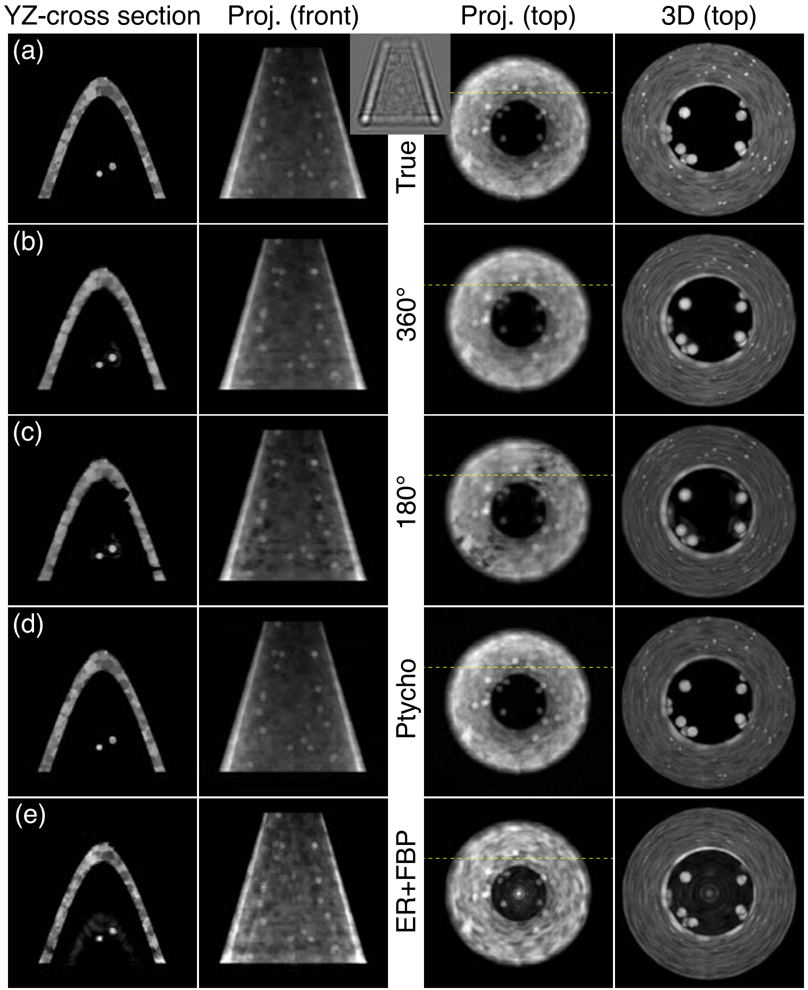




Test case 1: a designed sample

Pixel size (nm)	1
Energy (eV)	5000
Depth of focus (nm)	21.77
Largest sample thickness (nm)	200
Sample-detector distance (nm)	1000
Object grid size	256^3
# of projections	500
# of diffraction spots in ptychography	23×23
Platform	Cooley
Fullfield # of threads	4
Fullfield time	5 h
Ptychography # of threads	20
Ptychography time	45.9 h

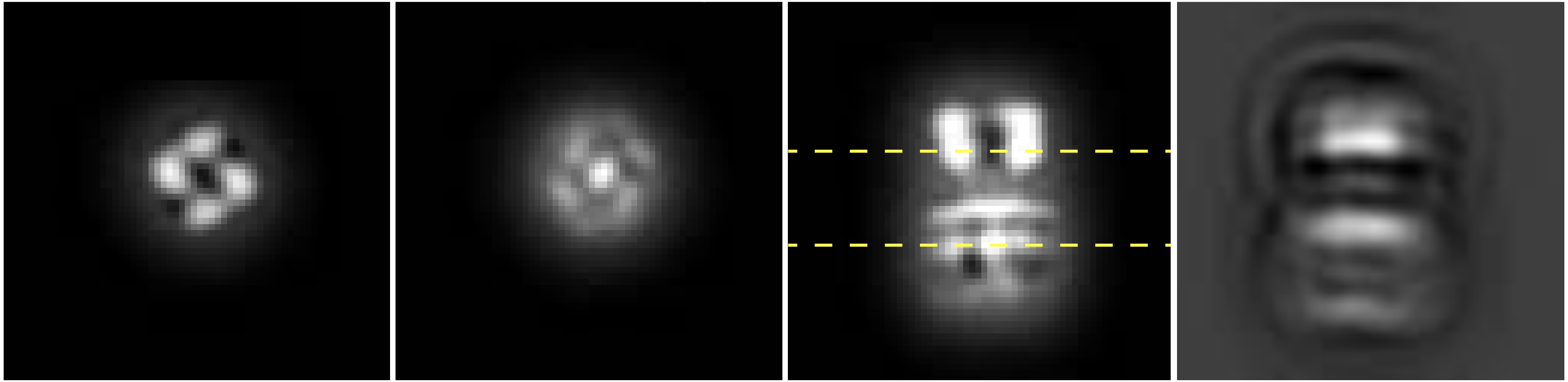
True: simulated object
360°: full-filed reconstruction with 360° projection data
180°: full-filed reconstruction with 180° projection data
Ptycho: ptychography reconstruction with 360° projection data
ER+FBP: reconstruction with conventional CDI and tomography



Test case 2: a semi-experimental protein molecule

Human adhesin complex originally acquired using EM; data retrieved from EM databank.

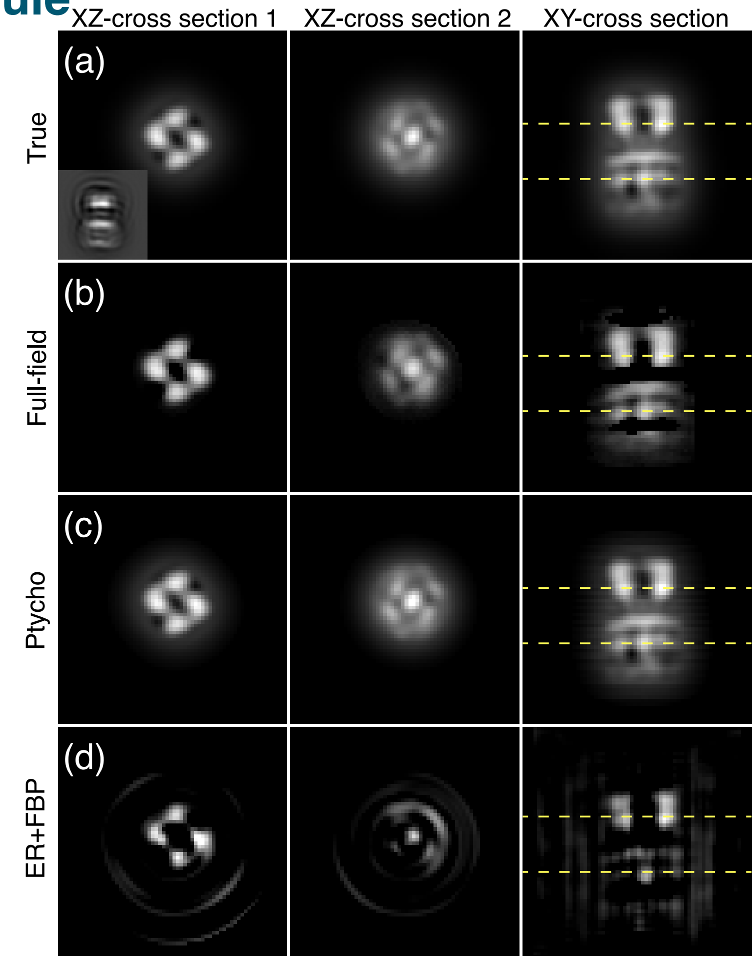
Pixel size (nm)	0.67
Energy (eV)	800
Depth of focus (nm)	1.56
Largest sample thickness (nm)	30
Object grid size	64^3
Num. of projections	500
# of diffraction spots in ptychography	23×23



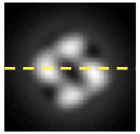
Test case 2: a semi-experimental protein molecule

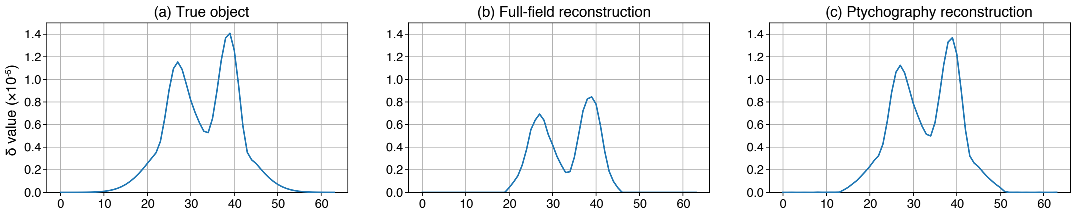
Pixel size (nm)	0.67
Energy (eV)	800
Depth of focus (nm)	1.56
Largest sample thickness (nm)	30
Object grid size	64^3
Num. of projections	500
# of diffraction spots in ptycho.	23×23

Fullfield platform	Workstation (GPU)
Fullfield # of threads	4
Fullfield time	0.15 h
Ptychography platform	Cooley
Ptycho. # of threads	20
Ptychography time	1.45 h

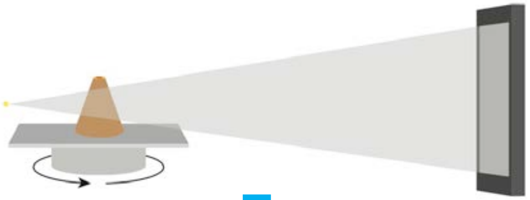


Full-field reconstruction throws away halos, while ptychography preserves it.

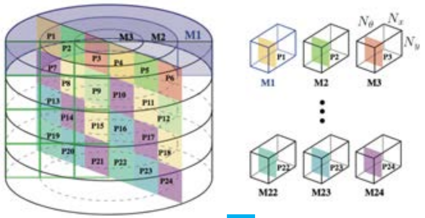




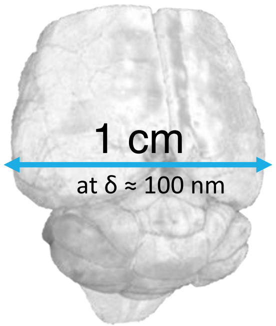
Future perspectives

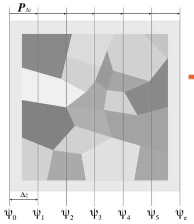


Point projection



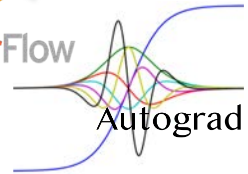
Tomosaic

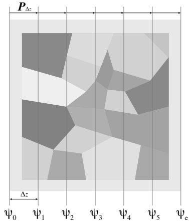


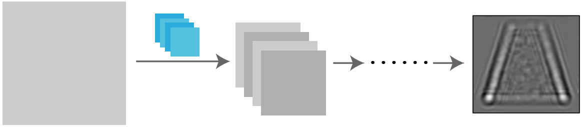


TensorFlow

Beyond DOF

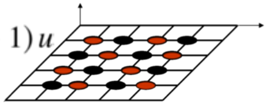






Convolutional neural network

$$\frac{\partial u}{\partial z} = \frac{1}{2ikn} \Delta_{\perp} u - ik(n-1)u$$



O(N)

Finite difference propagation

Acknowledgement

Thank you

Q & A