Sparsity in Optical Imaging: Bandwidth Extrapolation, Phase Retrieval, and Nonlinearities

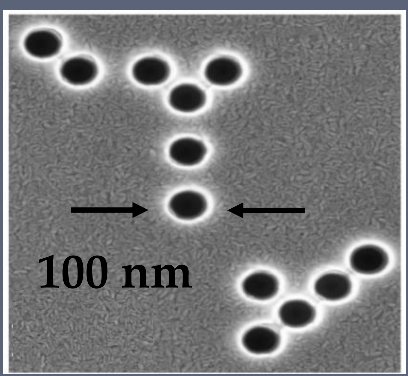
Yonina Eldar
Collaboration with the groups of Prof. Moti Segev and Prof. Oren Cohen
Technion – Israel Institute of Technology

Subwavelength Imaging

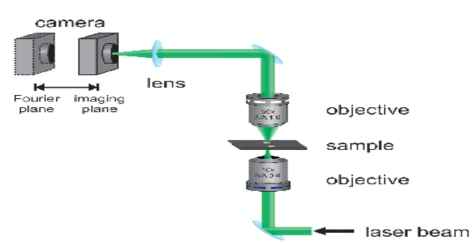
Diffraction limit: Even a perfect optical imaging system has a resolution limit determined by the wavelength λ

- The smallest observable detail is larger than ~ $\lambda/2$
- This results in image smearing

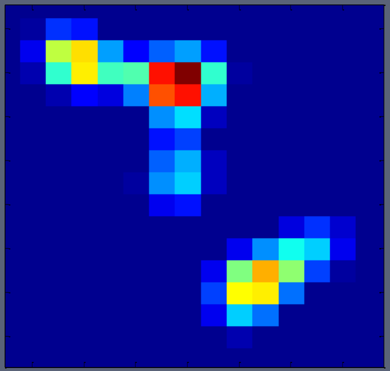
λ=514nm



Nano-holes
as seen in
electronic microscope



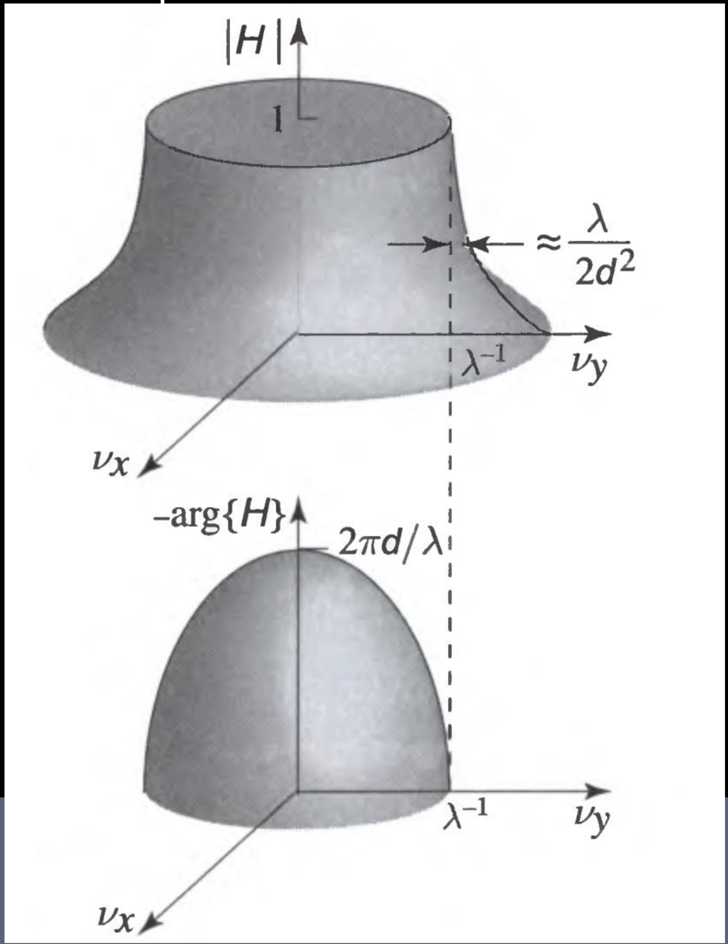
Sketch of an optical microscope: the physics of EM waves acts as an ideal low-pass filter



Blurred image
seen in
optical microscope

Optical Cut-off for High Spatial Frequencies

Free space transfer function



field propagation (z = 0 ⟶ z > 0)

$$\Psi(x,y,z) = FT^{-1}\left\{ FT\{\Psi(x,y,z=0)\} \underbrace{e^{iz\sqrt{\left(\frac{2\pi}{\lambda}\right)^2 - \left(k_x^2+k_y^2\right)}}}_{H(k_x,k_y)} \right\}$$

$$\sqrt{k_x^2 + k_y^2} < \frac{2\pi}{\lambda}$$ **propagating waves**

$$\sqrt{k_x^2 + k_y^2} > \frac{2\pi}{\lambda}$$ **evanescent waves**

Free space acts as a lowpass filter

Hardware Solutions – Near Field

NSOM

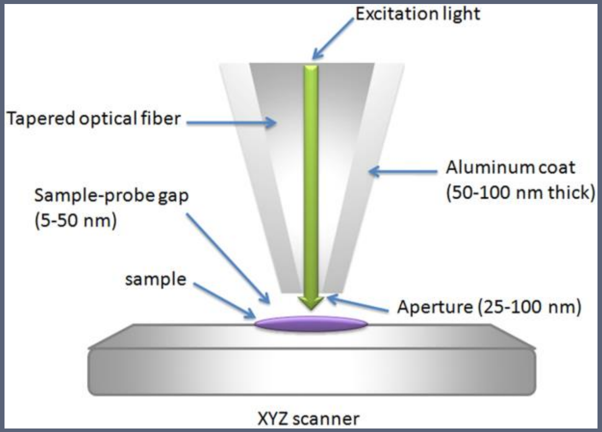
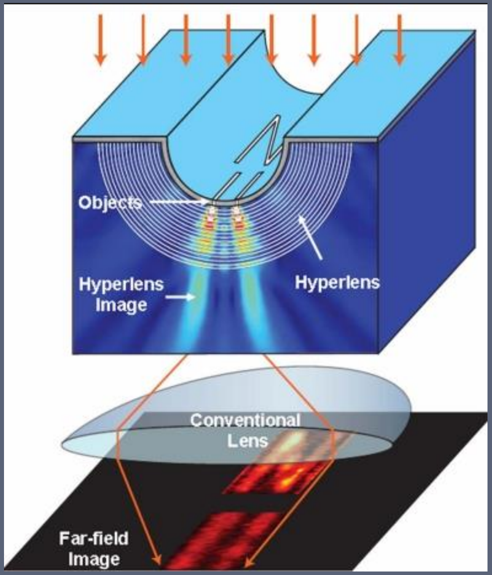


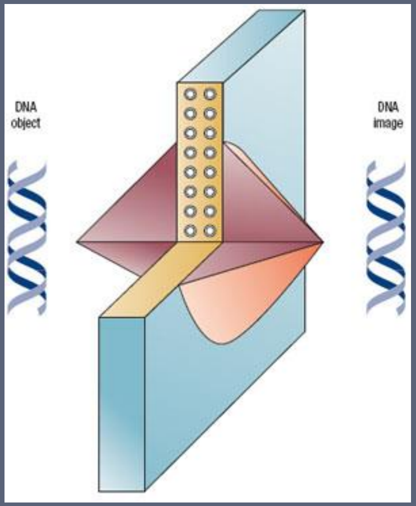
Image from
D. J. Müller et al., Pharmacol. Rev. 60,43 (2008).

E. H. Synge, Phil. Mag. 6, 356 (1928).
E. A. Ash et al., Nature 237, 510 (1972).
A. Lewis et al., Ultramicroscopy 13, 227 (1984).
E. Betzig et al., Science 251, 1468 (1991).

Superlens / Hyperlens



Z. Jacob et al., Opt. Exp. 14, 8247 (2006).
A. Salandrino et al., Phys. Rev. B 74, 075103 (2006).
Z. Liu et al., Science 315, 1686 (2007).



V. G. Veselago, Soviet Phys. Uspheki 10, 509 (1968).
J. B. Pendry, Phys. Rev. Lett. 85, 3966 (2000).
N. Fang et al., Science 308, 534 (2005).
I. I. Smolyaninov et al., Science 315, 1699 (2007).

Algorithmic Subwavelength Imaging

- Can we recover the subwavelength data algorithmiclly from far field data?
- Mathematically this amounts to recovering a signal from its low pass data
- Standard regularization methods have not been able to yield satisfactory results

``All methods for extrapolating bandwidth beyond the diffraction limit are known to be extremely sensitive to both:

- noise in the measured data and
- the accuracy of the assumed a priori knowledge."

J. W. Goodman, Introduction to Fourier Optics

Subwavelength real-time optical imaging remains a major challenge

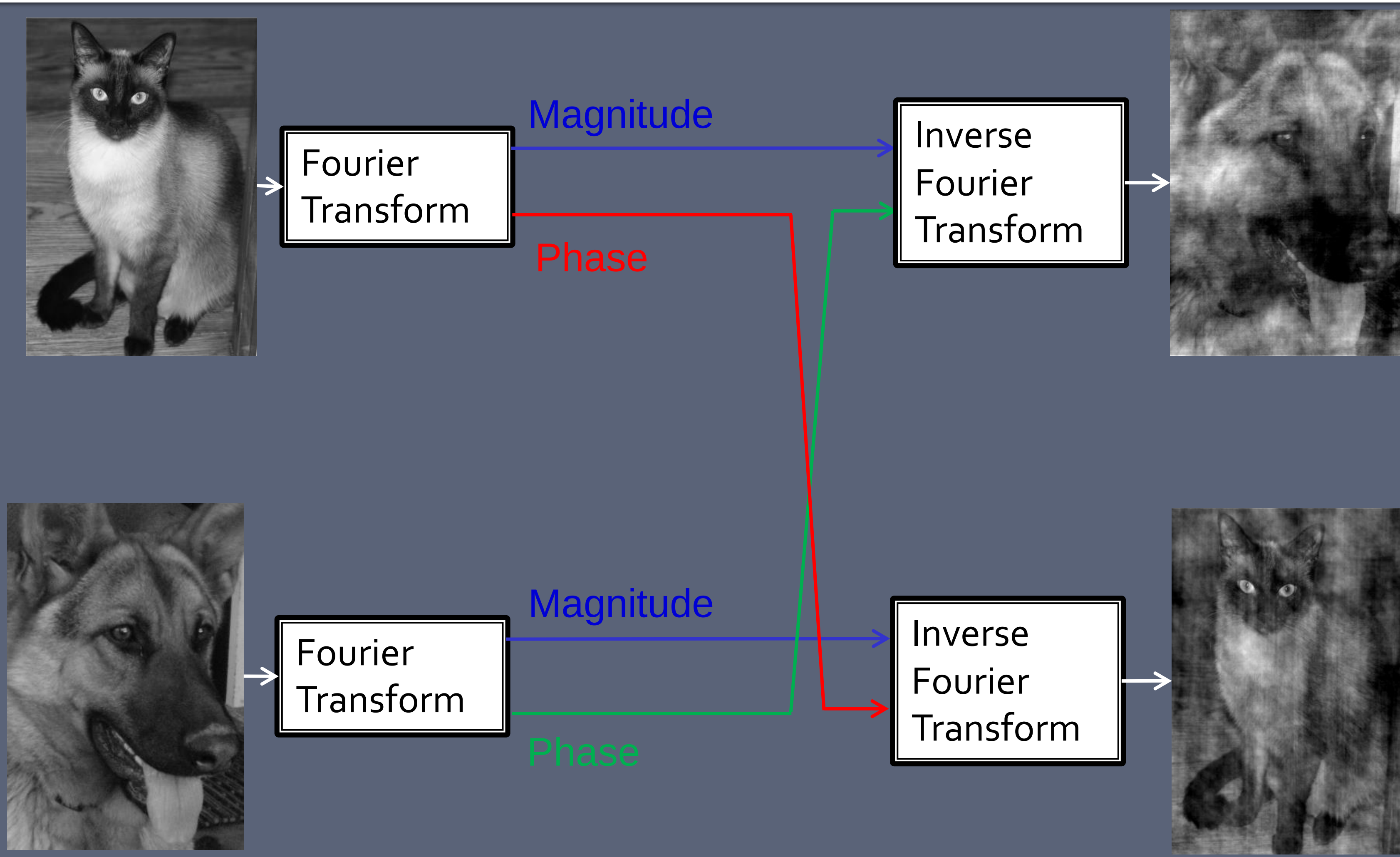
Phase Retrieval:
Recover a signal from its Fourier magnitude

$$x[n] \longrightarrow \boxed{\text{Fourier + Absolute value}} \longrightarrow y[k] = |X[k]|^2$$

- Arises in many fields: microscopy, crystallography, astronomy, optical imaging, and more
- Given an optical image illuminated by coherent light, in the far field we obtain the Fourier transform of the image
- Optical measurement devices measure the photon flux, which is proportional to the magnitude squared of the field
- Phase retrieval can allow direct recovery of the image
- Phase importance in imaging: Hayes, Lim, Oppenheim 80

Arises naturally in many problems in optics

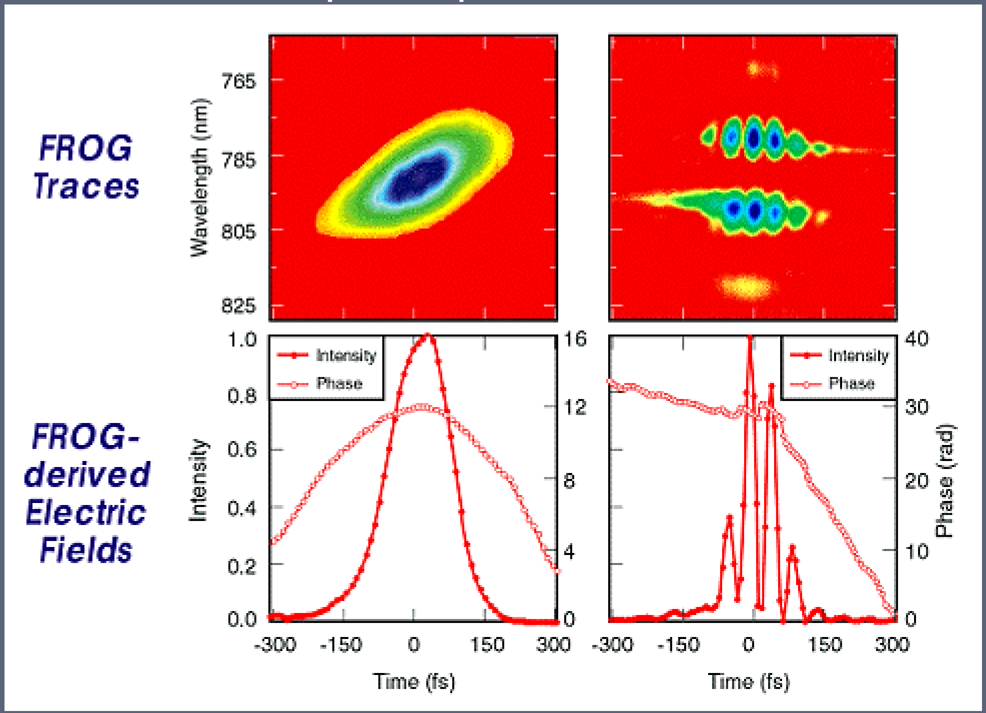
Phase Is Important!



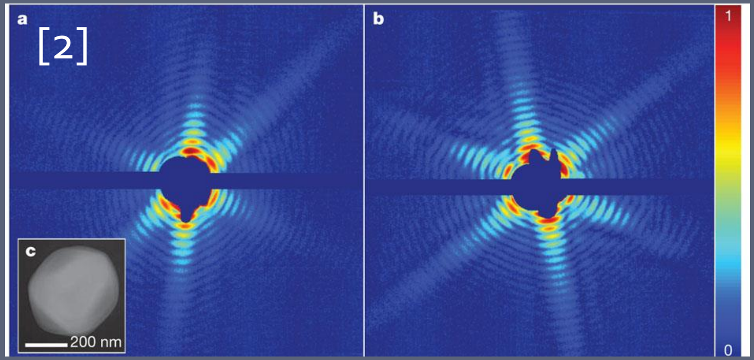
Phase Retrieval Applications

Ultra-short optical pulse measurement

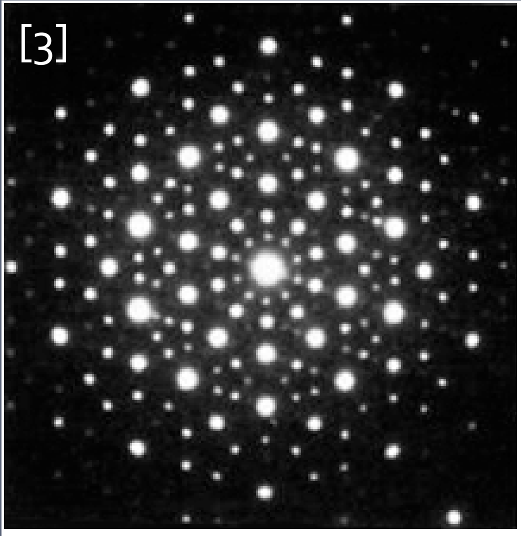
[1]



Coherent Diffractive Imaging



Crystallography



[1] R. Trebino et al., JOSA A **10**, 5 1101-1111 (1993)
[2] MM Seibert *et al. Nature* **470**, 78-81 (2011)
[3] D Shechtman et al. PRL **53**, 20, 1951-1952 (1984)

The Goal: Subwavelength CDI

- Several challenges in optical imaging:
  - Bandwidth extrapolation
  - Nonlinear measurements: Phase retrieval
- Exploit sparsity

Requires extension of compressed sensing methods

- Extend sparse recovery to lowpass data
- Nonlinear compressed sensing

Outline

- Bandwidth extrapolation
- Phase retrieval methods
- Optimality conditions for nonlinear sparse recovery
- Algorithms for nonlinear sparse recovery
  - **Extended IHT:** Nonlinear Iterative hard thresholding
  - **Extended OMP:** Greedy sparse simplex
- Phase retrieval: GESPAR
- Some application results

References for nonlinear compressed sensing:

- A. Beck and Y. C. Eldar, "Sparsity Constrained Nonlinear Optimization: Optimality Conditions and Algorithms"
- Y. Shechtman, A. Beck and Y. C. Eldar "GESPAR: Efficient Phase Retrieval of Sparse Signals"
- Y. C. Eldar and S. Mendelson "Phase Retrieval: Stability and Recovery Guarantees"

BW Extrapolation: Challenges

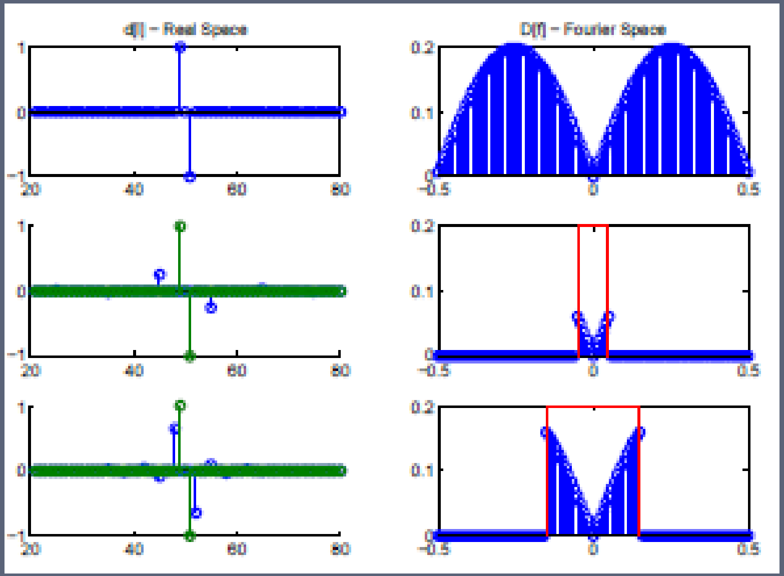
- In subwavelength imaging the measurement matrix corresponds to the low pass Fourier coefficients
- Does not satisfy conditions of compressed sensing (RIP or coherence)
- BP can be shown to recover spikes only if separation is *2/f* (Candes et. al.' 12)
- Nonlocal hard thresholding (NLHT): A specialized algorithm targeting sparse recovery from lowpass coefficients
- An iterative recovery that searches for the zeros rather than the signal peaks

Can recover spikes closer than *2/f*

S. Gazit, A. Szameit, Y. C. Eldar and M. Segev, Optics Express, (2009)

Recovery From Lowpass Measurements

- Close spikes cannot be recovered by basis pursuit



- Erroneous spikes tend to occur but typically in the vicinity of the true spikes with width proportional to *1/f*
- Stretches of zero values tend to be correct

NLHT: Nonlocal Hard Thresholding

NLHT algorithm

- S – off-support
- μ - nearest-neighbor window size
- ζ_0 - threshold
- $\Delta\zeta$ - increment in threshold
- ε – noise parameter
- k – max cardinality of sparse solution

Initialize: $S = \varnothing,\ \ \mu = \mu_0,\ \ \zeta = \zeta_0$

Repeat

Solve:

$$\min_d \|d\|_1 \ \ \textit{subject to} \ \ \|b - Fd\|_2 \le \varepsilon,$$

$$d[l] = 0 \ \forall l \in S$$

Allocate off-support:

1. Find all $\tilde{l}$ such that $\hat{d}[l] < \zeta \cdot \max|\hat{d}|$ for all l's that are distanced from $\tilde{l}$ to the right or to the left by μ, or that are distanced from both sides by μ/2.
2. Add $\tilde{l}$ to $\tilde{S}$.

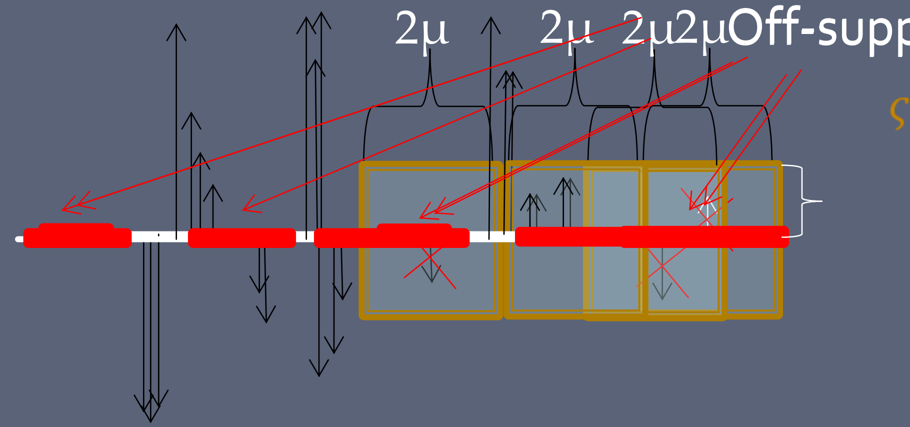
Update $S = S \cup \tilde{S}$

If support was not updated, increase ζ by $\Delta\zeta$ and decrease μ by 1.

Until $|S| \le k$

Output $\hat{d}$

Iteration #1 (BPDN):
Iteration #N (with given support):

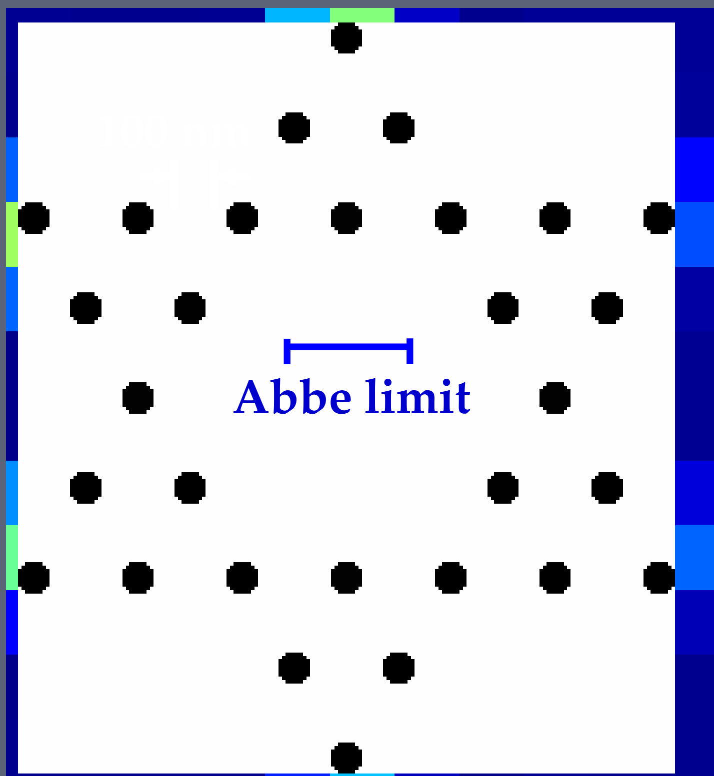


Window width decreases and height increases with the iterations…

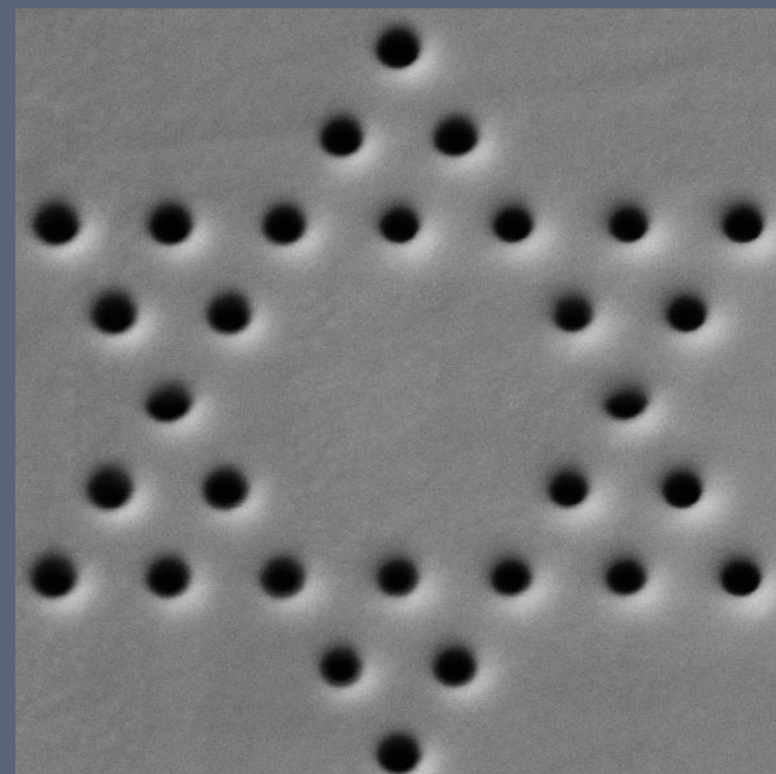
Experimental Results

microscope image

reconstructed image



SEM image



A. Szameit et al., Nature Materials 2012

Can be extended to incoherent light and partially incoherent light
Y. Shechtman et al. (2010, 2011)

Subwavelength CDI: Lensless Imaging

- CDI: Recovering near field from freely diffracting intensity pattern

  (i.e. from intensity of the Fourier transform)
- Phase information is lost

  Can we combine bandwidth extrapolation and phase retrieval?
- Sub-wavelength image recovery from highly truncated Fourier spectrum

  Mathematical formulation:

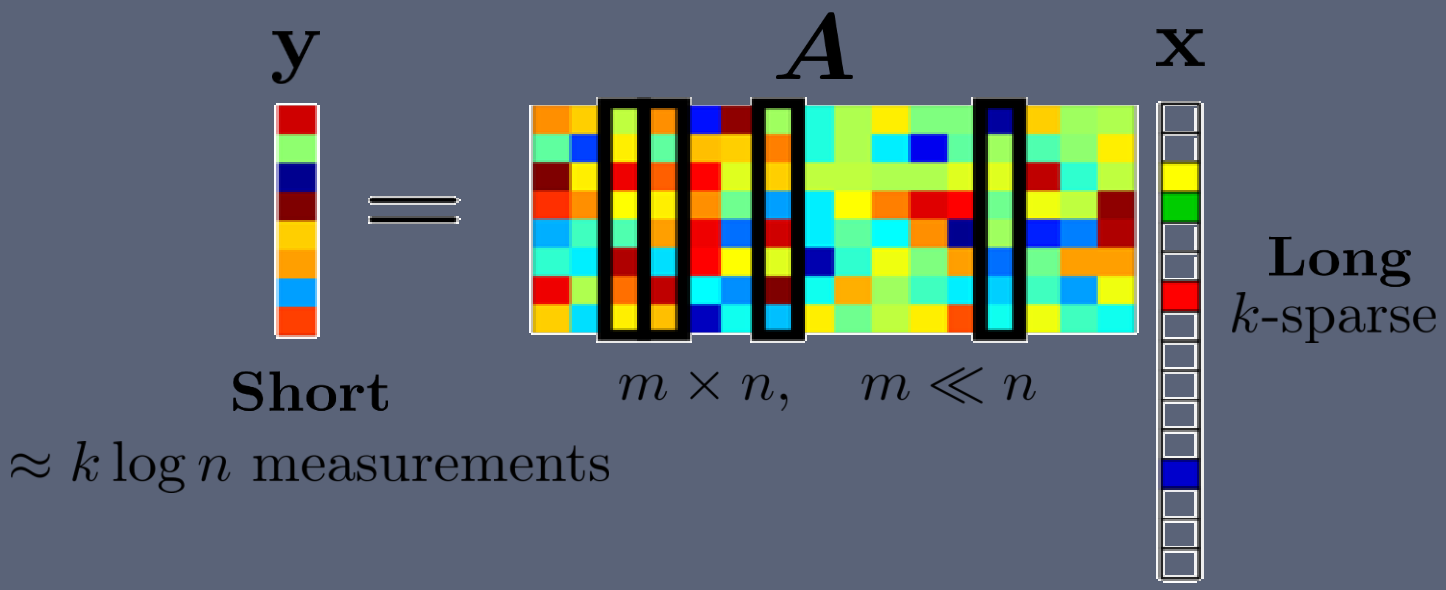
$$\text{Find image } a \text{ such that } y \approx |Fa|^2$$

- y are the measurements
- F represents the low pass Fourier coefficients
- Highly ill-posed problem!

Impose sparsity on input image

Nonlinear Compressed Sensing

- There has been an explosion of work on recovery of sparse vectors from linear measurements
- Basis of compressed sensing



- Many applications in which the measurements are nonlinear $y = g(x)$
- Can we extend compressed sensing results to the nonlinear case?
- Minimize $||y - g(x)||^2$ subject to a sparsity constraint on x

Nonlinear Sparse Recovery

- We consider the problem

$$\min f(x) \quad s.t. \ \|x\|_0 \leq k$$

  - $f(x)$ is an arbitrary continuously differentiable function
  - Not necessarily convex!

- Special case: **Phase retrieval**

$$f(x) = \sum_{i=1}^{n} \left| |F_i x|^2 - y_i \right|^2$$

  - F_i - ith row of Fourier matrix

Beck and Eldar, (2013)

Phase Retrieval

- No uniqueness in 1D problems (Hofstetter 64)
- Uniqueness in 2D if oversampled by factor 2 (Hayes 82)
- No guarantee on stability
- No known algorithms to achieve unique solution

Need for regularization:

- Support knowledge
- Input magnitude
- Positivity

Not sufficient to guarantee uniqueness, stability and algorithms

Phase Retrieval: Random Measurements

- Since analyzing Fourier measurements is hard we can look at random measurements

$$y_i = |\langle a_i, x\rangle|^2 + w_i \quad \longleftarrow \text{noise} \quad x \in R^N$$

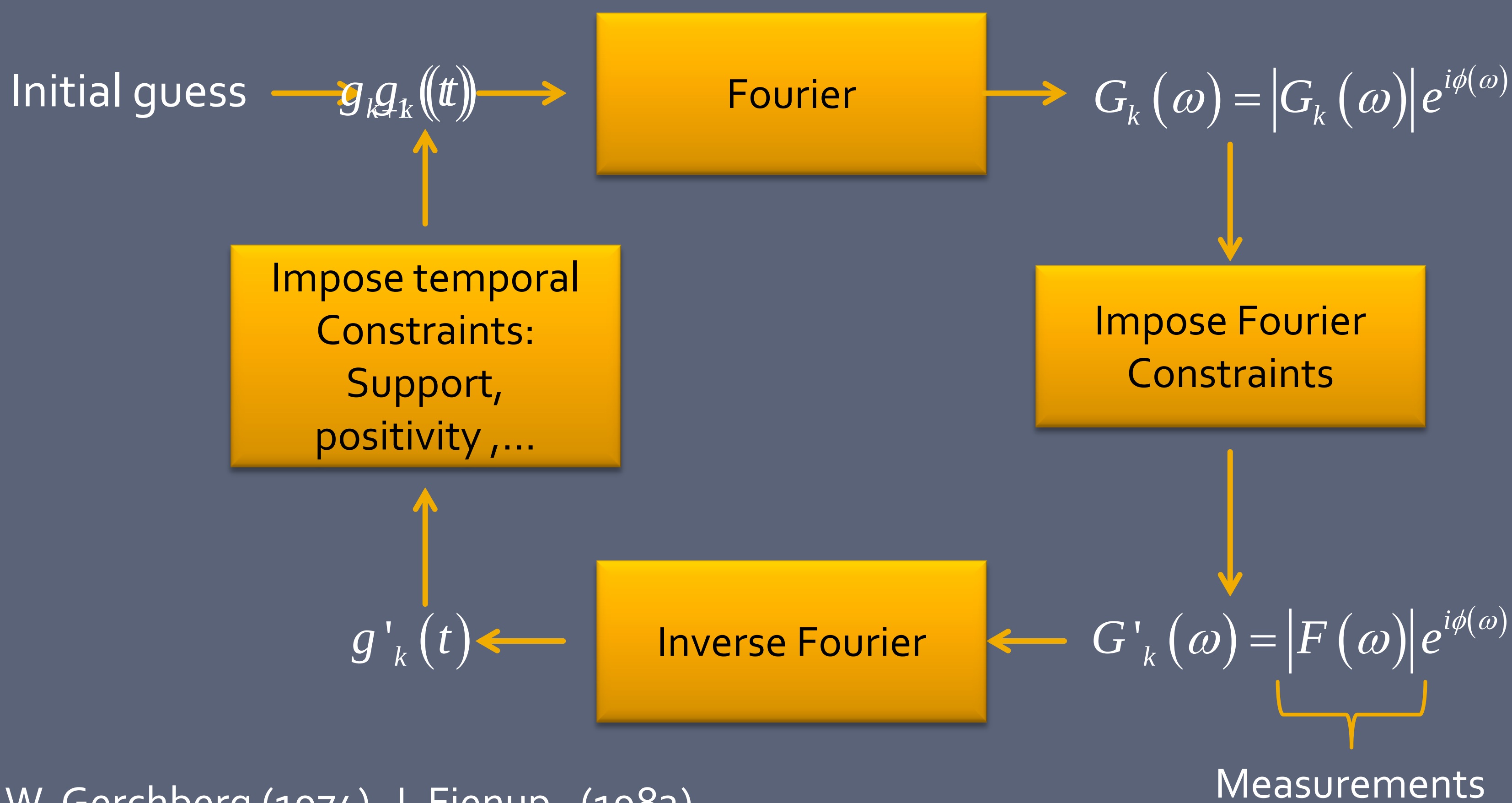
↑ random vector

- $M = 4N - 2$ measurements needed for uniqueness (Balaw, Casazza, Edidin 06, Bandira et.al 13)
- $M = N\log(N)$ measurements needed for stability (Eldar and Mendelson 12)
- Solving $\sum_{i=1}^{M} \left| y_i - |\langle a_i, x\rangle|^2 \right|^p \quad 1 < p \leq 2$ provides stable solution (Eldar and Mendelson 12)

How to solve objective function?

Existing Techniques – Iterative ("Fienup") Algorithms

Basic Scheme (Variants exist):



R. W. Gerchberg (1974), J. Fienup, (1982)
S. Mukherjee, ICASSP, (2012)

Semidefinite Relaxation

- $|\langle a_k, x\rangle|^2 = \mathrm{Tr}(A_k X)$ with $A_k = a_k {a_k}^T$, $X = xx^T$
- Phase retrieval can be written as

$$\begin{aligned} &\text{minimize} && \mathrm{rank}(X) \\ &\text{subject to} && A(X) = b \\ & && X \geq 0 \end{aligned}$$

- SDP relaxation: replace $\mathrm{rank}(X)$ by $\mathrm{Tr}(X)$ or by $\mathrm{logdet}(X + \varepsilon I)$ and apply reweighting

Advantages/Disadvantages:

- Generally improves performance over Fienup
- Computationally demanding
- No optimality conditions for general measurements

Shechtman, Eldar, Szameit, Segev, (2011)
Candes, Eldar, Strohmer , Voroninski, (2012)

Phase Retrieval: Sparsity

- PR methods often perform poorly, no general optimality conditions
- Sparsity can be added to improve performance (Moravec et. al 07, Shechtman et. al 11, Bahman et. al 11, Ohlsson et. al 12, Janganathan 12)
- Both Fienup methods and SDP-based techniques can be modified to account for sparsity (Shechtman et. al 11, Mukherjee et. al 12)
- SDP methods require $k^2 \log(N)$ measurements (Candes et. al 12, Li 12)
- $k \log(N/k)$ measurements needed for stability (Eldar and Mendelson 12)
- Solving $\sum_{i=1}^{M} |y_i - |\langle a_i, x\rangle|^2|^p \quad 1 < p \leq 2$ subject to a sparsity constraint provides stable solution (Eldar and Mendelson 12)

Can we efficiently solve the least-squares nonlinear sparse recovery problem?

Our Approach

$$\min f(x) \quad s.t. \ \|x\|_0 \leq k$$

- General theory and algorithms for nonlinear sparse recovery
- Derive conditions for optimal solution
- Use them to generate algorithms
- Since our problem in non-convex no simple necessary and sufficient conditions
- We develop three necessary conditions
  - Basic feasibility
  - L-stationarity ⇒ Iterative Hard Thresholding
  - CW-minima ⇒ Greedy Sparse Simplex (OMP)

Beck and Eldar, (2013)

Basic Feasibility

- For unconstrained differentiable problems a necessary condition is that $\nabla f(x^*) = 0$
- We expect a similar condition over the support S

Theorem: Any optimal solution is a basic feasible vector:
1. when $\|x^*\|_0 < k$, $\nabla f(x^*) = 0$
2. when $\|x^*\|_0 = k$, $\nabla f(x^*) = 0$ for all $i \in S$

- Condition is quite weak – there can be many BF points

L - Stationarity

- For constrained problems $\min\{f(x) : x \in C\}$ where C is convex, a necessary condition is stationarity

$$< \nabla f(x^*), x - x^* > \geq 0 \; for \; all \; x \in C$$

- For any $L > 0$ a vector x^* is stationary if and only if

$$x^* = P_c(x^* - \frac{1}{L}\nabla f(x^*))$$

  - Does not depend on L!

- For our nonconvex setting we define an L-stationary point:

$$x^* \in P_c(x^* - \frac{1}{L}\nabla f(x^*))$$

  - The projection is no longer unique

Theorem: Let $\nabla f(x)$ be Lipschitz continuous.
Then L-stationarity with *L>L(f)* is necessary for optimality

L – Stationarity: Discussion

Advantages:

- L – stationarity is necessary for optimality
- Implies basic feasibility
- Simple algorithm that finds L – stationary points

Iterative Hard Thresholding:

$$x^{k+1} \in H\left(x^k - \frac{1}{L}\nabla f\left(x^k\right)\right)$$

where $H(x)$=hard thresholding of x

- Algorithm popular for compressed sensing when $y = Ax$ (Blumensath and Davies, 08)

Disadvantages:

- Requires knowledge of Lipschitz constant
- Often converges to wrong solution – not a strong condition

Coordinate–Wise Minima

- For an unconstrained problem x^* is a coordinate-wise minima (CW) if x_i^* is a minimum with respect to the ith component

$$x_i^* \in argmin\, f(x_1^*, .., x_{i-1}^*, x_i, x_{i+1}^*, \dots x_n^*)$$

- For our constrained problem we define CW as:

1. $\|x^*\|_0 < k \text{ and } f(x^*) = \min_{t \in R} f(x^* + te_i)$
2. $\|x^*\|_0 = k \text{ and for every } i \in S$

$$f(x^*) \le \min_{t \in R} f(x^* - x_i^* e_i + te_j)$$

Theorem: Any optimal solution is a CW minima

CW–Minima: Discussion

- Does not require Lipschitz continuity
- Any optimal solution is a CW minima
- Implies basic feasibility
- If gradient is Lipschitz continuous:

Optimal solution $\Rightarrow$ CW-minima $\Rightarrow$ L-stationarity $\Rightarrow$ BF

- In fact, CW minima implies L' stationarity with $L' < L$

Optimal solution $\Rightarrow$ CW-minima $\Rightarrow$ L'-stationarity $\Rightarrow$ L-stationarity

Stronger than L- stationarity

Greedy Sparse Simplex Method

- Generalization of matching pursuit to nonlinear objectives
- Additional correction step which improves OMP

General step $\|x^*\|_0 \leq k$: find coordinate that minimizes $f(x)$

$$t_i \in arg\min_t f(x^k + te_i) \qquad f_i = \min_t f(x^k + te_i)$$

$$x^{k+1} = x^k + t_{i^*} e_{i^*}$$

Swap $\|x^*\|_0 = k$: Swap index i with best j if lower objective

- Converges to CW minima (stronger than IHT guarantee)
- Does not require Lipschitz continuity

Summary: Greedy Sparse Simplex

- Converges to CW-minima: strongest necessary condition
- Applicable to general functions
- No need to know Lipschitz constant
- Generalizes OMP to nonlinear setting
- Improves on OMP in linear case
  - Arbitrary initial points
  - Correction stage

Example

- Generate a matrix A of size 4x5
- $y = Ax$ where $x = [-1\ \ -1\ 0\ 0\ 0\ ]^T$
- In this problem there are 10 BF points
- We implement IHT and the greedy method with 1000 randomly generated values x_0

Number of times method converged to each BF

BF vector	1	2	3	4	5	6	7	8	9	10
IHT	329	50	63	92	229	0	130	0	61	46
Greedy	813	0	0	112	0	0	75	0	0	0

L - stationarity

CW minima

Example – Cont'd

- Iterations of greedy method with $x_0 = [0\ 1\ 5\ 0\ 0\ 0]^T$

iteration number	x_1	x_2	x_3	x_4	x_5
0	0	1	5	0	0
1	0	1.0000	1.5608	0	0
2	0	0	1.5608	0	-0.6674
3	1.6431	0	0	0	-0.6674
4	1.6431	-0.8634	0	0	0
5	1.0290	-0.8634	0	0	0
6	1.0290	-0.9938	0	0	0
7	1.0013	-0.9938	0	0	0
8	1.0013	-0.9997	0	0	0
9	1.0001	-0.9997	0	0	0
10	1.0001	-1.0000	0	0	0
11	1.0000	-1.0000	0	0	0

$$x = [-1\ \ -1\ 0\ 0\ 0\ ]^T$$

GESPAR: GrEedy Sparse PhAse Retrieval

- Specializing our general algorithm to phase retrieval
- Local search method with update of support
- For given support solution found via Damped Gauss Newton
- Efficient and more accurate than current techniques

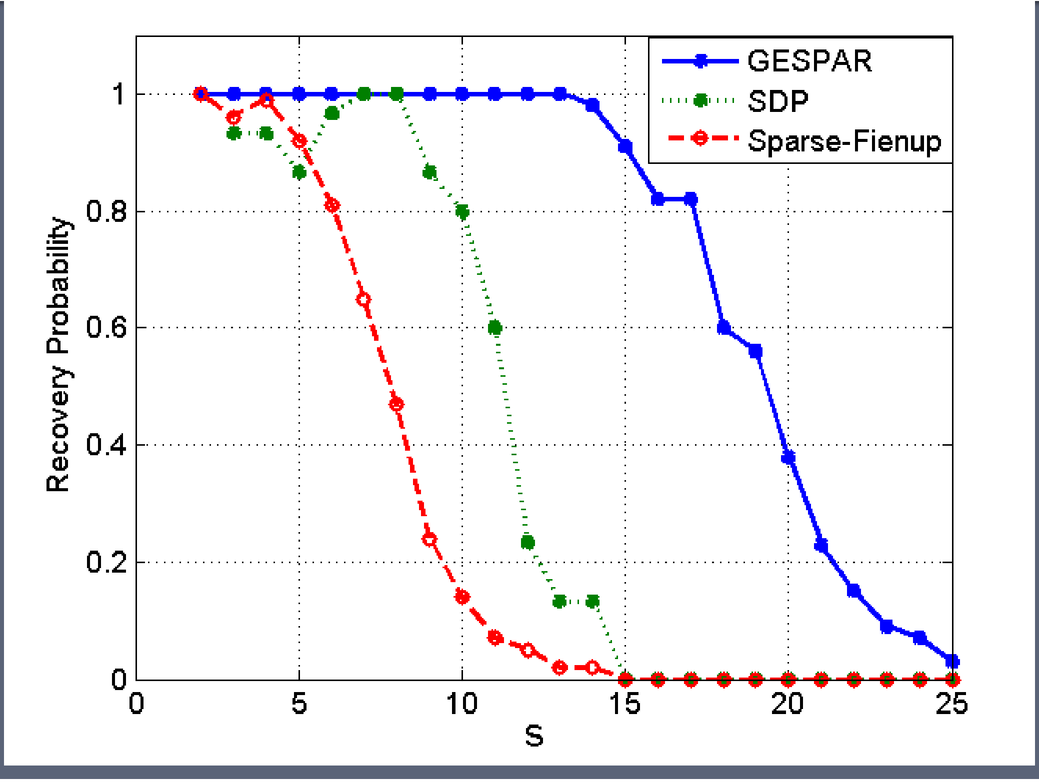
1. For a given support: minimizing objective over support by linearizing the function around current support and solve for y_k

$$z_k = z_{k-1} + \underbrace{t_k}_{\text{determined by backtracking}}(y_k - z_{k-1})$$

2. Find support by finding best swap: swap index with small value $|x_i|$ with index with large value $|\nabla f(x_j)|$

Shechtman, Beck and Eldar, (2013)

Performance Comparison



n=64
m=128

Each point = 100 signals

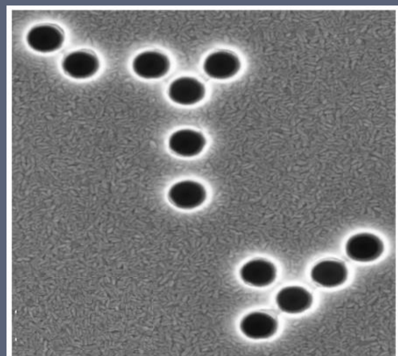
	SDP		Sparse-Fienup		GESPAR	
	recovery	runtime	recovery	runtime	recovery	runtime
$s=3$	0.93	1.32 sec	0.96	0.09 sec	1	0.12 sec
$s=5$	0.86	1.78 sec	0.92	0.12 sec	1	0.12 sec
$s=8$	1	3.85 sec	0.47	0.16 sec	1	0.23 sec

Sparsity Based Subwavelength CDI

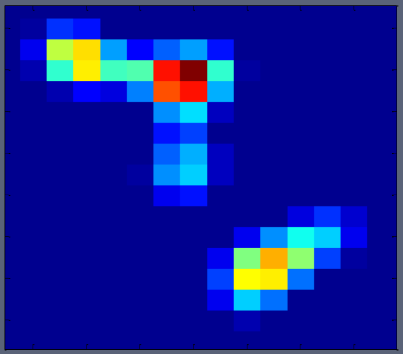
Circles are 100 nm diameter

Wavelength 532 nm

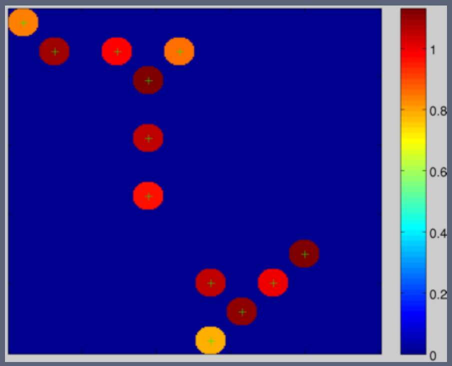
SEM image



Blurred image

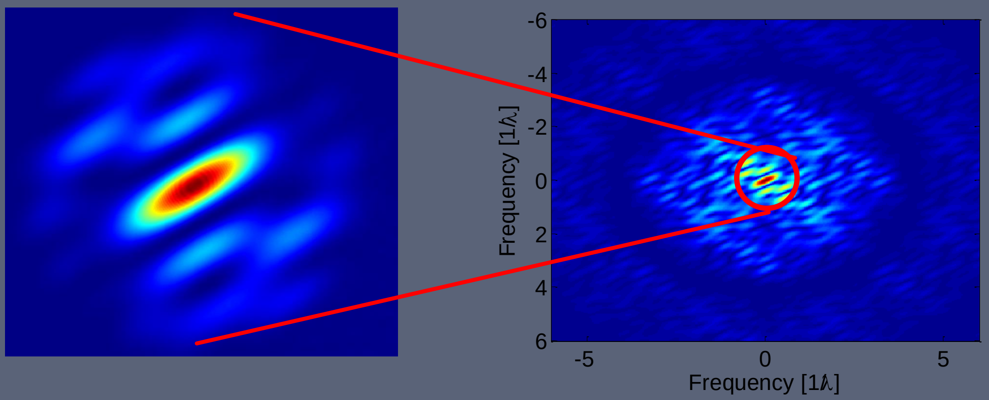


Sparse recovery



Diffraction-limited (low frequency) intensity measurements

Model Fourier transform



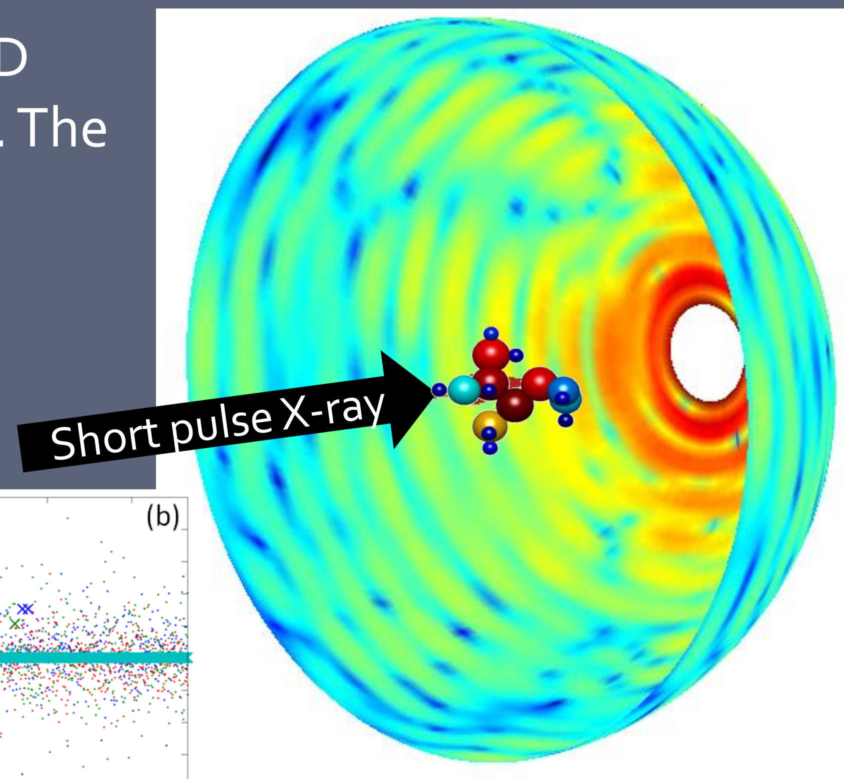
Szameit et al., Nature Materials, (2012)

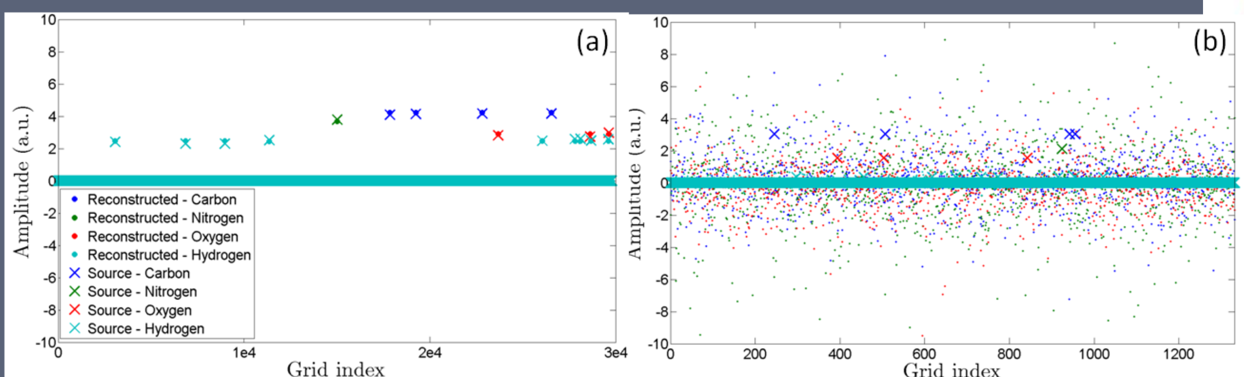
Sparsity Based Ankylography

Concept :

- A short x-ray pulse is scattered from a 3D molecule combined of known elements. The 3D scattered diffraction pattern is then sampled in a single shot

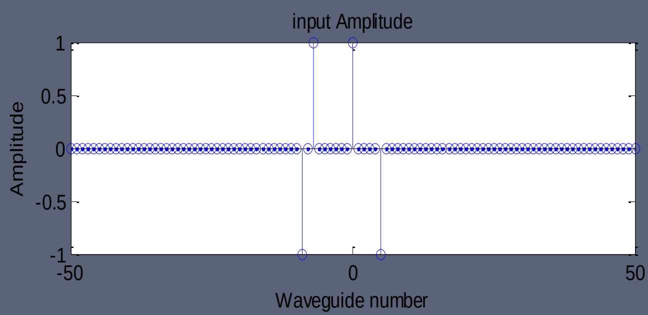
- Recover a 3D molecule using 2D sample

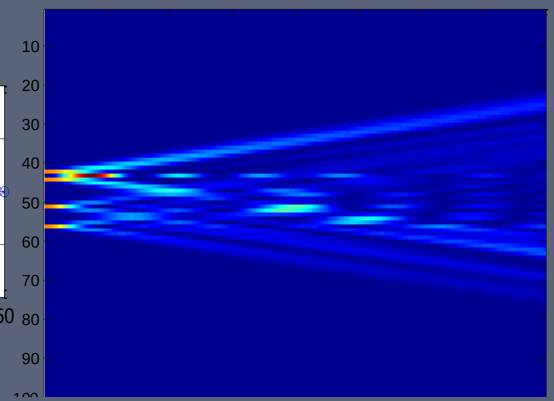


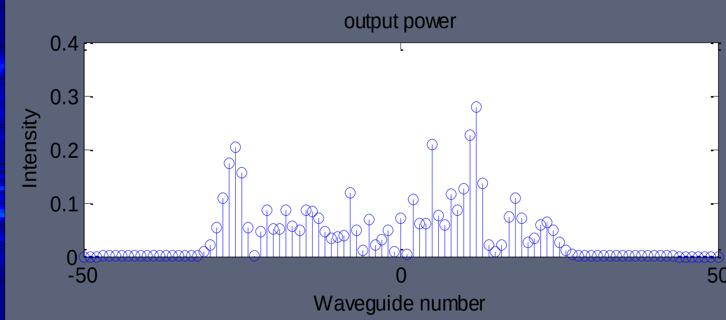


K.S. Raines et al. Nature 463, 214 ,(2010).
Mutzafi et. al., (2013).

Coupled Waveguide Arrays





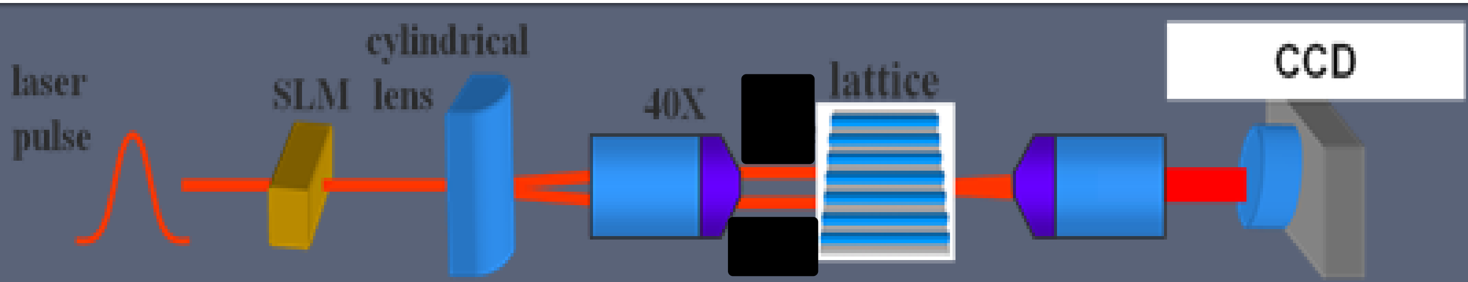


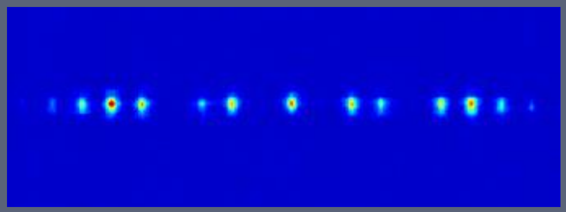
$$E_{in} = E\big|_{z=0} = \sum_k a_k \phi\left(x - kd\right) \qquad E_{out} = \left(a * h\right)\left[k\right]$$

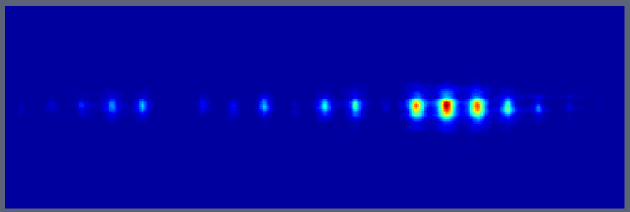
Only output intensity is measured: $I_{out} = \left|E_{out}\right|^2$

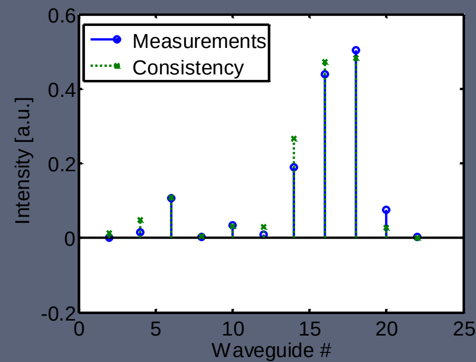
Leading to the quadratic problem: $\boxed{y_i = \left|\left(H \cdot x\right)_i\right|^2 = x^T H_i x}$

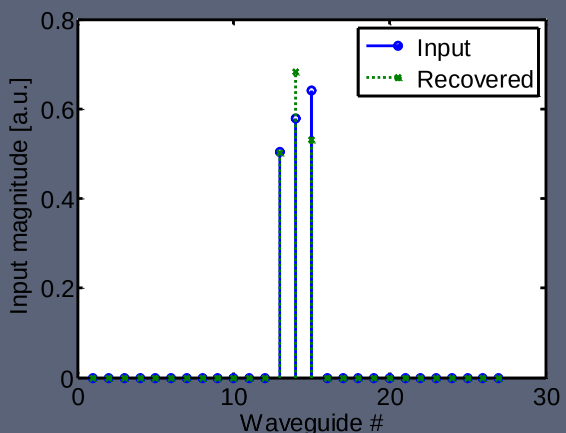
Experimental Results Using GESPAR

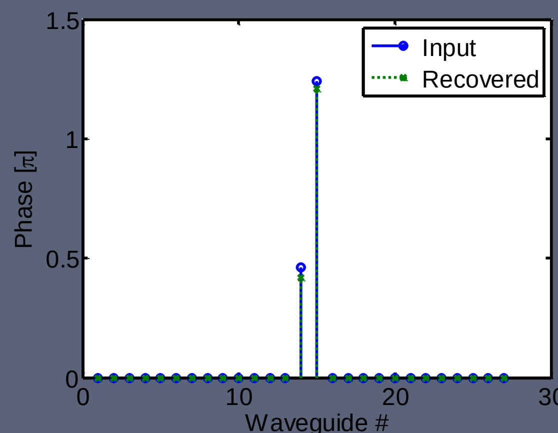


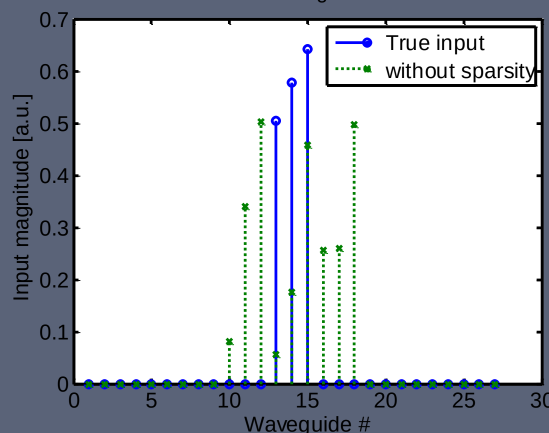






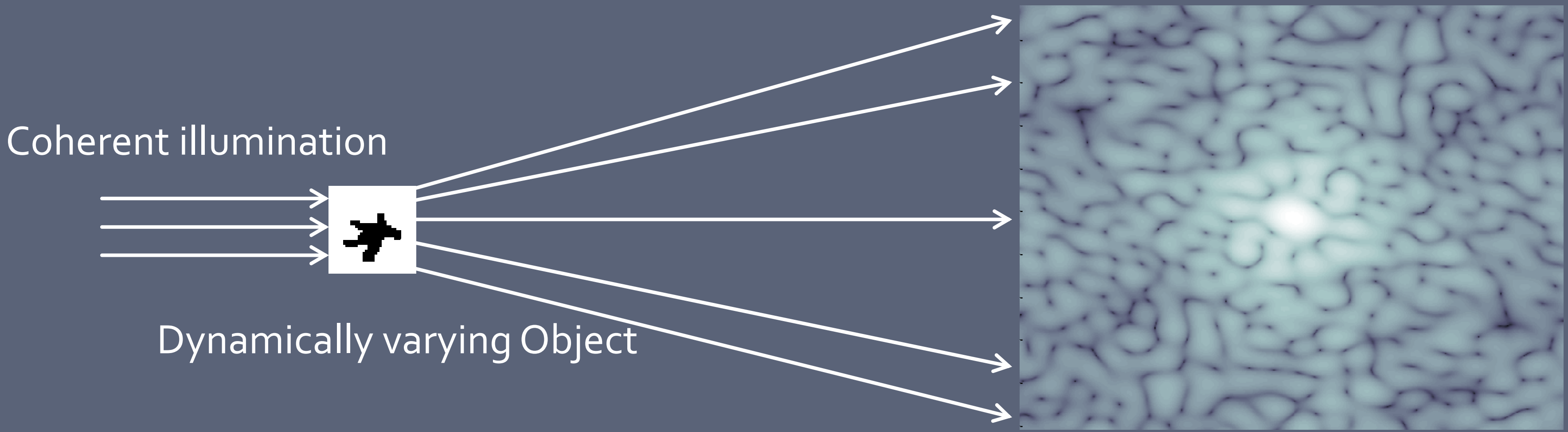






Y. Shechtman et. al. (2013)

Application – Sparse Dynamic CDI

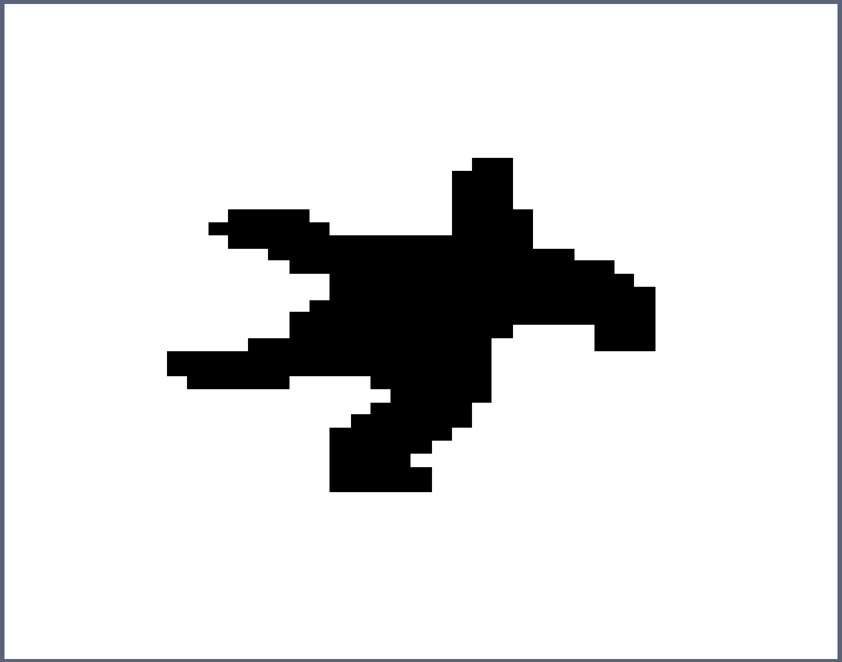


Far Field Diffraction Measurement:
Fourier Magnitude

- Assumption: Sparse difference between consecutive frames
- Can be used to decrease acquisition time → improve temporal resolution

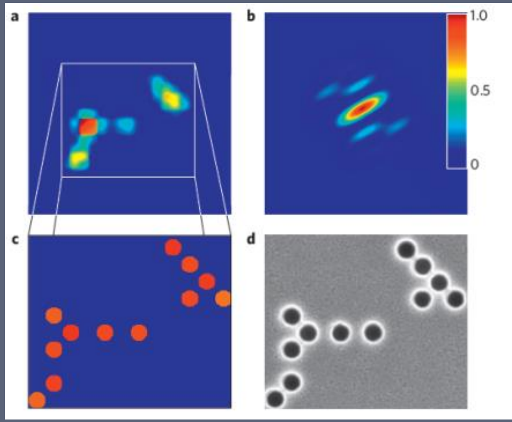
Y. Shechtman et al. 2013

Small Difference – Few Samples



Our Problem Is Quadratic

Sparse CDI (**quadratic CS**): $y_i = \left| F_i x \right|^2, \; x \text{ is sparse}$

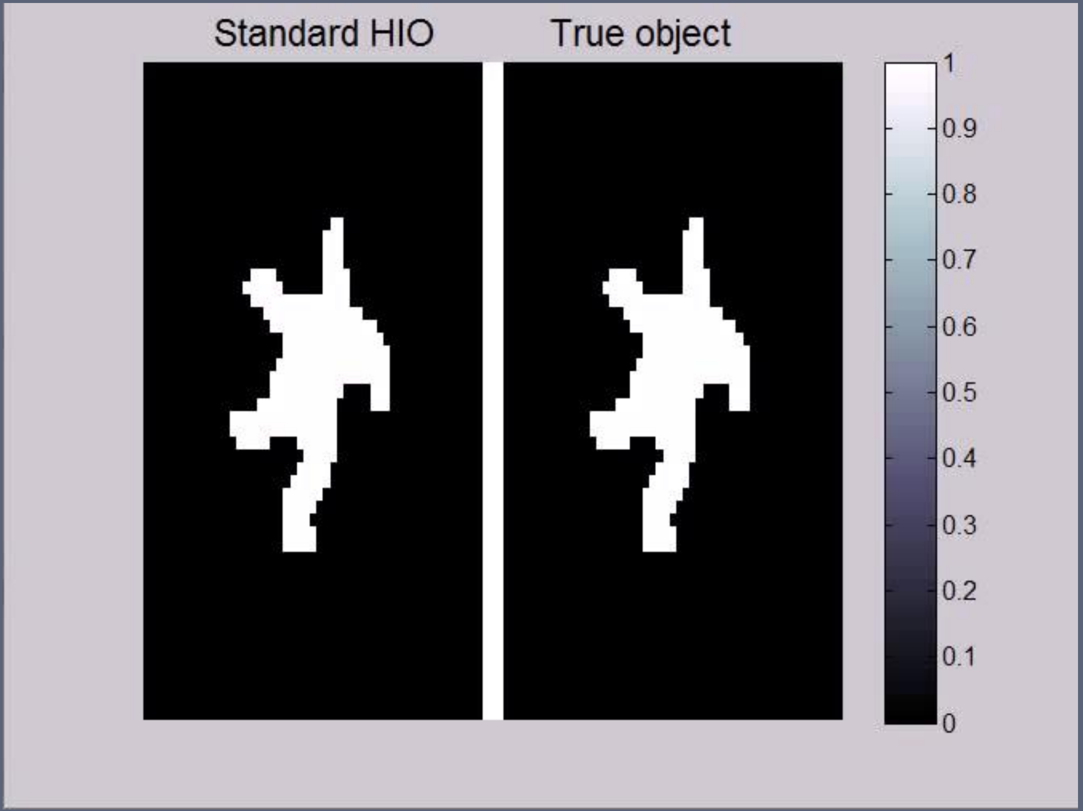


Sparse difference CDI (**quadratic CS**):

$$\boxed{y_i = \left| F_i \left(x + \Delta x \right) \right|^2, \; \Delta x \text{ is sparse} \qquad x \text{ is known}}$$

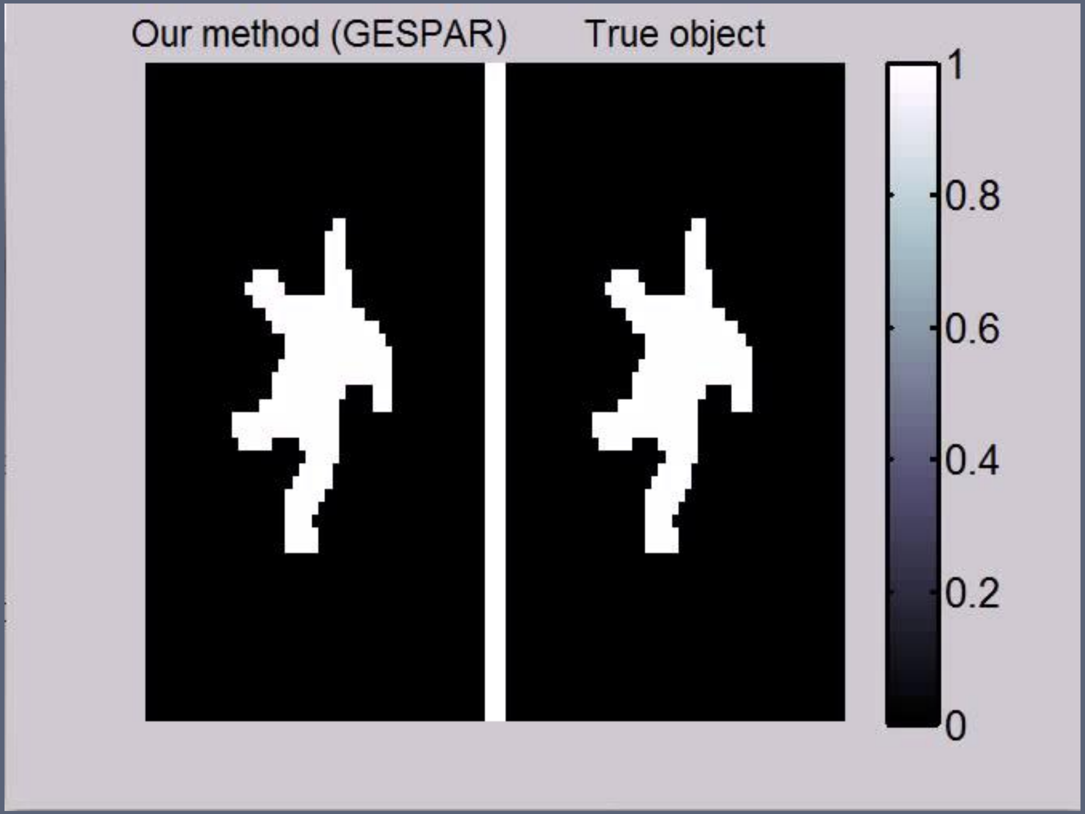
“Clean” first frame

Simulation: Fienup (HIO)



1225 unknowns, 676 samples, practically noiseless

Simulation: Our Method (GESPAR)

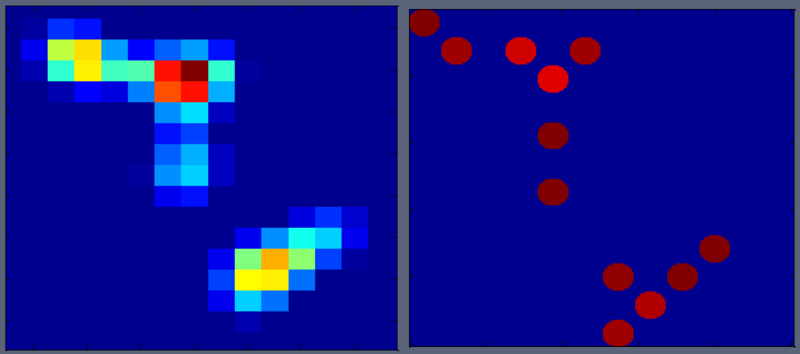


1225 unknowns, 676 samples, practically noiseless

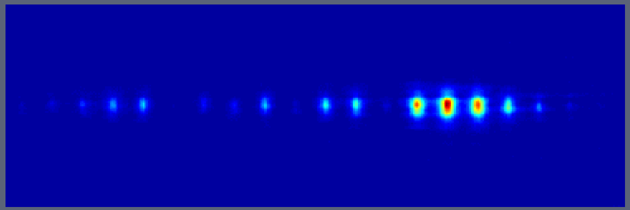
Conclusion

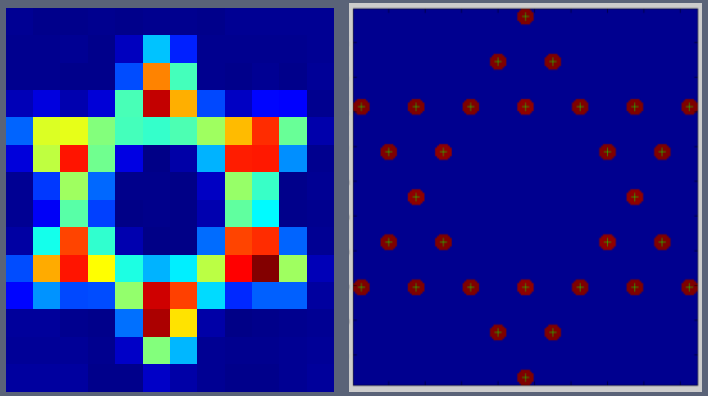
- Prior knowledge of sparsity compensates for loss of optical information
- Subwavelength imaging and CDI
- Necessary optimality conditions for nonlinear sparse recovery
- Efficient greedy algorithm that extends OMP to general nonlinear sparse recovery
- GESPAR: Efficient phase retrieval method
- Applications to optics

Exploring connections between sparse recovery and optical imaging can be fruitful in both directions



Thank you!





If you found this interesting …
Looking for a post-doc in information processing!