Compressive Sensing

Part III: Compressive Sensing in Practice

Mark A. Davenport

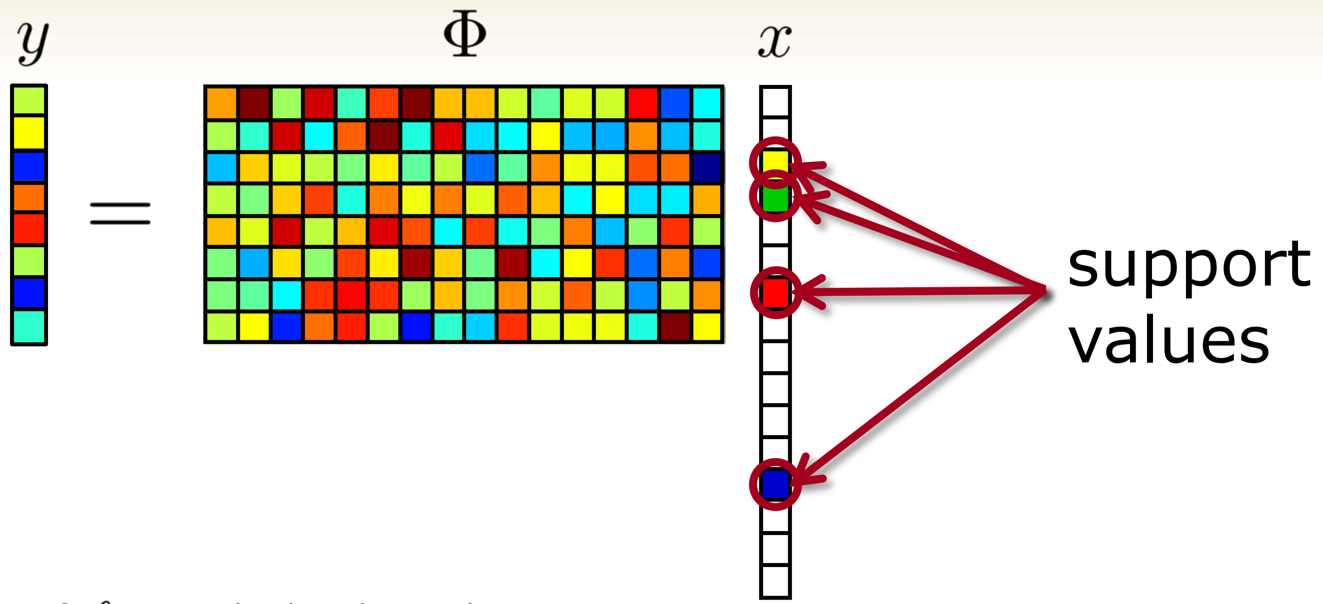
Stanford University
Department of Statistics

Important Practical Challenges

- Noise!
  - noisy measurements
  - noisy signals
  - interferene
- Quantization
  - quantization error
  - saturation effects
- Good signal models
  - is sparsity sometimes not enough?
  - what dictionaries should we use in practice?

Measurement and Signal Noise

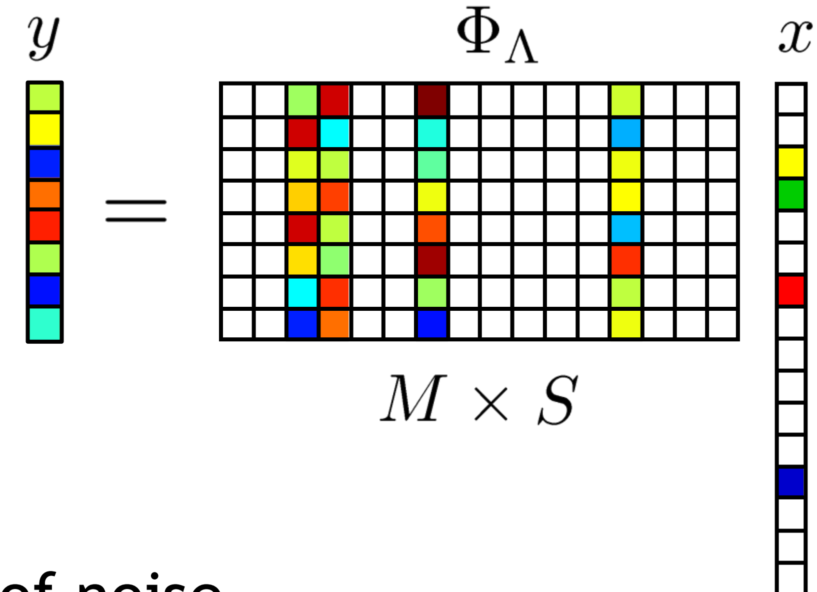
Sparse Signal Recovery



- Optimization / ℓ_1 -minimization
- Greedy algorithms
  - matching pursuit
  - orthogonal matching pursuit (OMP)
  - regularized OMP
  - CoSaMP, Subspace Pursuit, IHT, …

Exact Recovery

If we can determine $\Lambda = \mathrm{supp}(x)$, then the problem becomes *over*-determined.



In the absence of noise,

$$\begin{aligned} \Phi_\Lambda^\dagger y &= (\Phi_\Lambda^T \Phi_\Lambda)^{-1} \Phi_\Lambda^T y \\ &= (\Phi_\Lambda^T \Phi_\Lambda)^{-1} \Phi_\Lambda^T \Phi_\Lambda x \\ &= x \end{aligned}$$

Signal Recovery in Noise

Given $y = \Phi x + e$
find x

- Optimization-based methods
  - basis pursuit, basis pursuit de-noising, Dantzig selector

$$\widehat{x} = \arg\min_{x \in \mathbb{R}^N} \|x\|_1$$

$$\text{s.t.} \quad \|y - \Phi x\|_2 \leq \epsilon$$

- Greedy/Iterative algorithms
  - OMP, StOMP, ROMP, CoSaMP, Thresh, SP, IHT, …

Stable Signal Recovery

Suppose that we observe $y = \Phi x + e$ and that Φ satisfies the RIP of order S.

Typical (worst-case) guarantee

$$\boxed{\|\widehat{x} - x\|_2 \leq C\|e\|_2}$$

Even if $\Lambda = \mathrm{supp}(x)$ is provided by an oracle, the error can still be as large as

$$\boxed{\|\widehat{x} - x\|_2 = \frac{\|e\|_2}{1 - \delta}}$$

Expected Performance

- Worst-case bounds can be pessimistic
- What about the *average* error?
  - assume e is white noise with variance σ^2

$$\mathbb{E}\left(\|e\|_2^2\right) = M\sigma^2$$

  - for oracle-assisted estimator

$$\mathbb{E}\left(\|\widehat{x} - x\|_2\right) \leq \frac{S\sigma^2}{1-\delta}$$

  - if e is Gaussian, then for ℓ_1 -minimization

$$\mathbb{E}\left(\|\widehat{x} - x\|_2^2\right) \leq CS\sigma^2 \log N$$

White Signal Noise

What if our signal x is contaminated with noise?

$$y = \Phi(x + n)$$

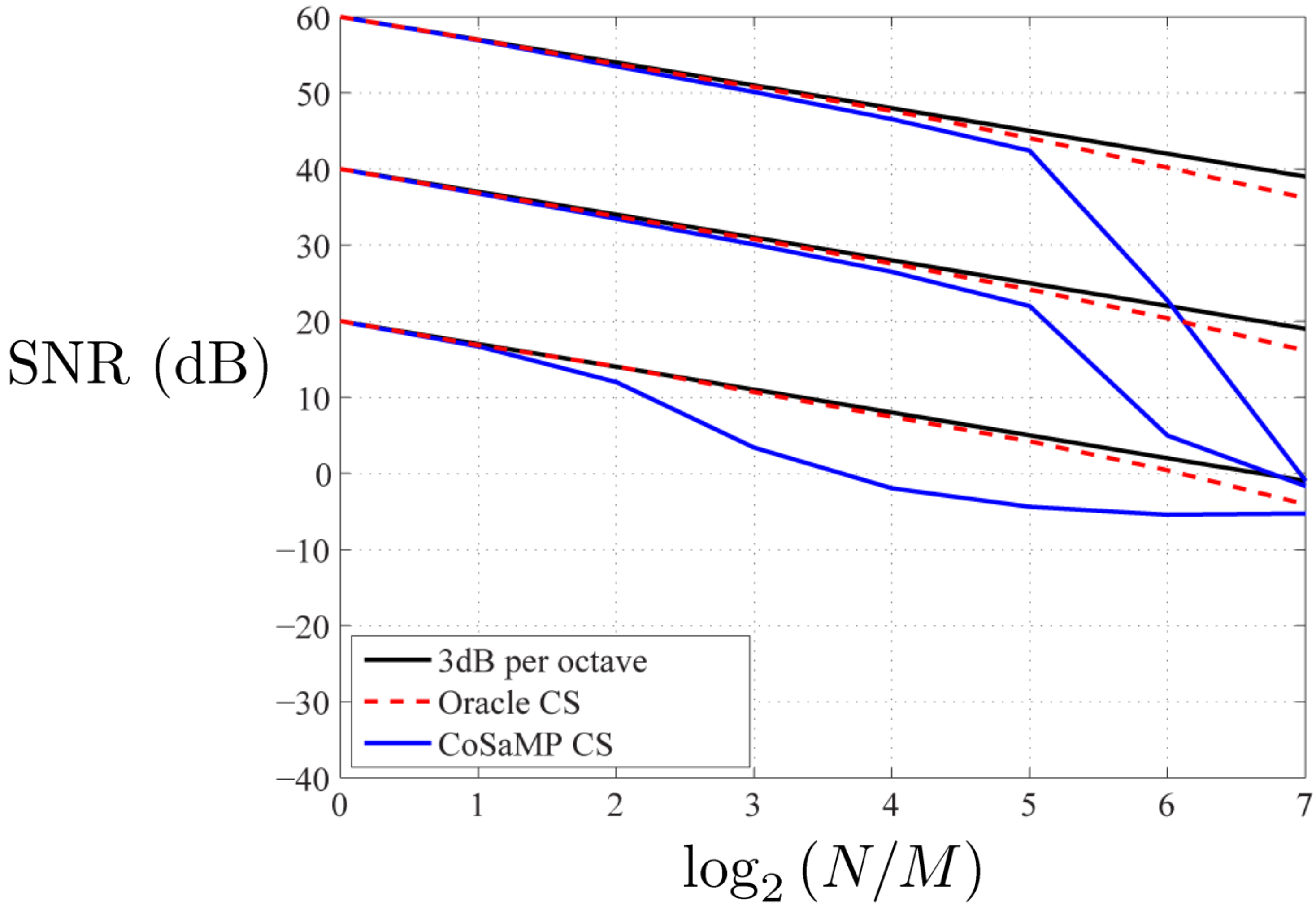
Suppose Φ satisfies the RIP and has orthogonal and equal-norm rows. If n is white noise with variance σ^2, then Φn is white noise with variance $\sigma^2 \frac{N}{M}$.

$$\boxed{\|\widehat{x} - x\|_2^2 \leq C' \frac{N}{M} S \sigma^2 \log N}$$

$$\mathrm{SNR} = 10 \log_{10} \left(\frac{\|x\|_2^2}{\|\widehat{x} - x\|_2^2} \right)$$

3dB loss per octave of subsampling

Noise Folding



[D, Laska, Treichler, and Baraniuk - 2011]

Can We Do Better?

- Better choice of Φ ?
- Better recovery algorithm?

If we knew the support of x *a priori*, then we could achieve

$$\|\widehat{x} - x\|_2^2 \approx \frac{S}{M} S\sigma^2 \ll C' \frac{N}{M} S\sigma^2 \log N$$

Is there any way to match this performance without knowing the support of x in advance?

$$R^*_{\mathrm{mm}}(\Phi) = \inf_{\widehat{x}} \sup_{\|x\|_0 \leq S} \mathbb{E}\left[\|\widehat{x}(y) - x\|_2^2\right]$$

No!

Theorem:

If $y = \Phi x + e$ with $e \sim \mathcal{N}(0, \sigma^2 I)$, then

$$R^*_{\mathrm{mm}}(\Phi) \geq C \frac{N}{\|\Phi\|_F^2} S\sigma^2 \log(N/S).$$

If $y = \Phi(x + n)$ with $n \sim \mathcal{N}(0, \sigma^2 I)$, then

$$R^*_{\mathrm{mm}}(\Phi) \geq C \frac{N}{M} S\sigma^2 \log(N/S).$$

Ingredients in proof:

- Fano's inequality
- Random construction of packing set of sparse points
- Matrix Bernstein inequality to bound empirical covariance matrix of packing set

[Candès and D - 2011]

Interference

$$y = \Phi x + e$$

- What if e represents corruption or *structured noise*, rather than Gaussian noise or arbitrary perturbations?

- Structured signal noise:

$$y = \Phi x_S + \Phi x_I$$

- Structured measurement noise:

$$y = \Phi x + \Omega e$$

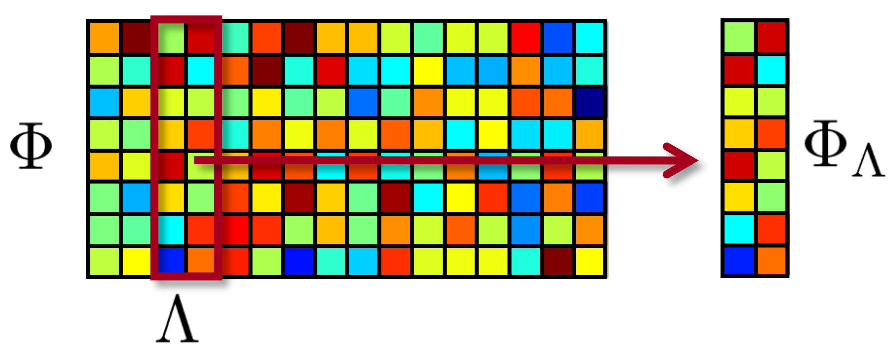
Interference Cancellation

Suppose $x = x_S + x_I$ where x_S is sparse with *unknown* support and x_I is sparse with *known* support Λ

Goal: Design an $M \times M$ matrix P such that

$$\|P(\Phi x_I)\|_2 \approx 0$$

$$\|P(\Phi x_S)\|_2 \approx \|\Phi x_S\|_2$$



$$P = I - \underbrace{\Phi_\Lambda \Phi_\Lambda^\dagger}_{\text{Projection onto } \mathcal{R}(\Phi_\Lambda)}$$

$$P\Phi_\Lambda = 0$$

Interference Cancellation

Lemma:

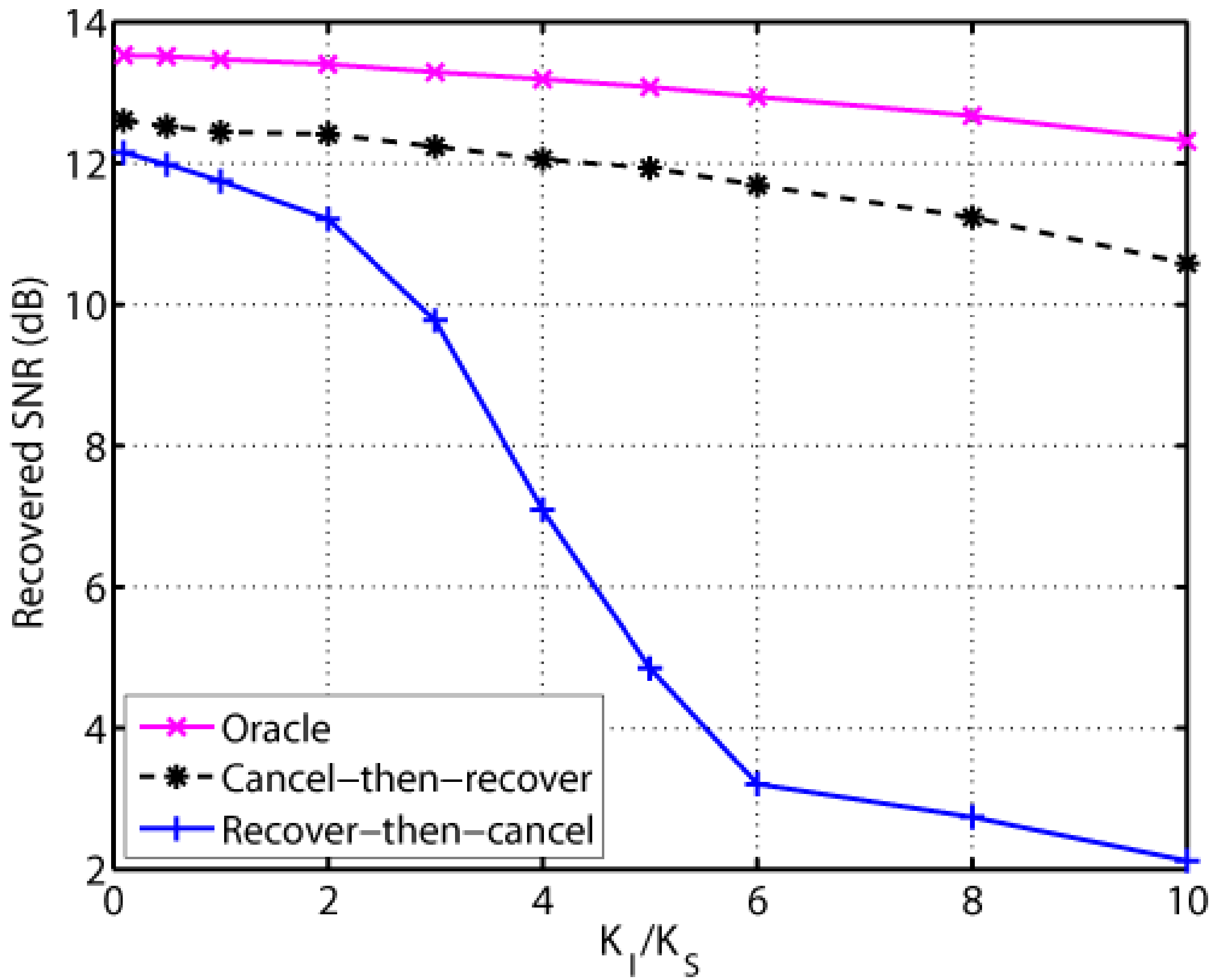
If Φ satisfies the RIP of order S, then

$$\left(1 - \frac{\delta}{1-\delta}\right)\|x\|_2^2 \leq \|P\Phi x\|_2^2 \leq (1+\delta)\|x\|_2^2$$

provided that $\|x\|_0 \leq S - |\Lambda|$ and $\mathrm{supp}(x) \cap \Lambda = \emptyset.$

[D, Boufounos, Wakin, and Baraniuk – 2010]

Interference Cancellation in Action



[D, Boufounos, Wakin, and Baraniuk – 2010]

Interference Cancellation

Lemma:

If Φ satisfies the RIP of order S, then

$$\left(1 - \frac{\delta}{1-\delta}\right)\|x\|_2^2 \le \|P\Phi x\|_2^2 \le (1+\delta)\|x\|_2^2$$

provided that $\|x\|_0 \le S - |\Lambda|$ and $\mathrm{supp}(x) \cap \Lambda = \emptyset.$

$$\Rightarrow \quad |\langle Py, P\Phi_j\rangle - x_j| \le \frac{\delta}{1-\delta}\|x_{\Lambda^c}\|_2$$

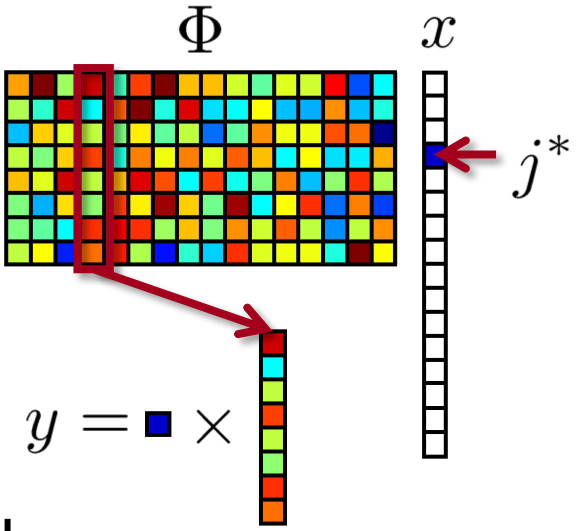
[D, Boufounos, Wakin, and Baraniuk – 2010]

Aside: Orthogonal Matching Pursuit

OMP selects one index at a time

Iteration 1:

$$\boxed{j^* = \arg\max_j |\langle y, \Phi_j \rangle|}$$



If Φ satisfies the RIP of order $\|u \pm v\|_0$, then

$$|\langle \Phi u, \Phi v \rangle - \langle u, v \rangle| \leq \delta \|u\|_2 \|v\|_2$$

Set $u = x$ and $v = e_j$

$$|\langle y, \Phi_j \rangle - x_j| \leq \delta \|x\|_2$$

Aside: Orthogonal Matching Pursuit

Subsequent Iterations:

$$\boxed{j^* = \arg\max_j |\langle Py, P\Phi_j \rangle|}$$

$$P = I - \Phi_\Lambda \Phi_\Lambda^\dagger$$

$$P\Phi_\Lambda = 0 \quad \Rightarrow \quad P\Phi x = P\Phi x_{\Lambda^c}$$

$$\Rightarrow \quad |\langle Py, P\Phi_j \rangle - x_j| \le \frac{\delta}{1-\delta} \|x_{\Lambda^c}\|_2$$

Aside: Orthogonal Matching Pursuit

Theorem:
Suppose x is S-sparse and $y = \Phi x$.
If Φ satisfies the RIP of order $S+1$ with constant $\delta < \frac{1}{3\sqrt{S}}$, then the j^* identified at each iteration will be a nonzero entry of x.

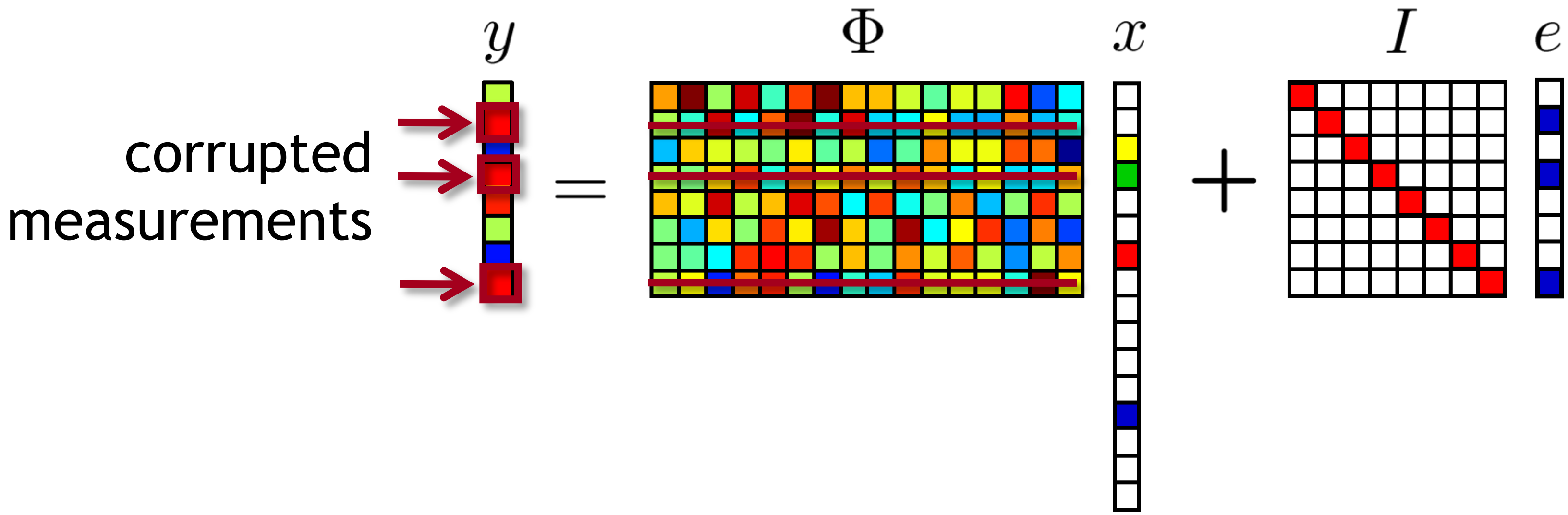
Exact recovery after S iterations.

Argument provides simplified proofs for other orthogonal greedy algorithms (e.g. ROMP) that are robust to noise

[D and Wakin – 2010]

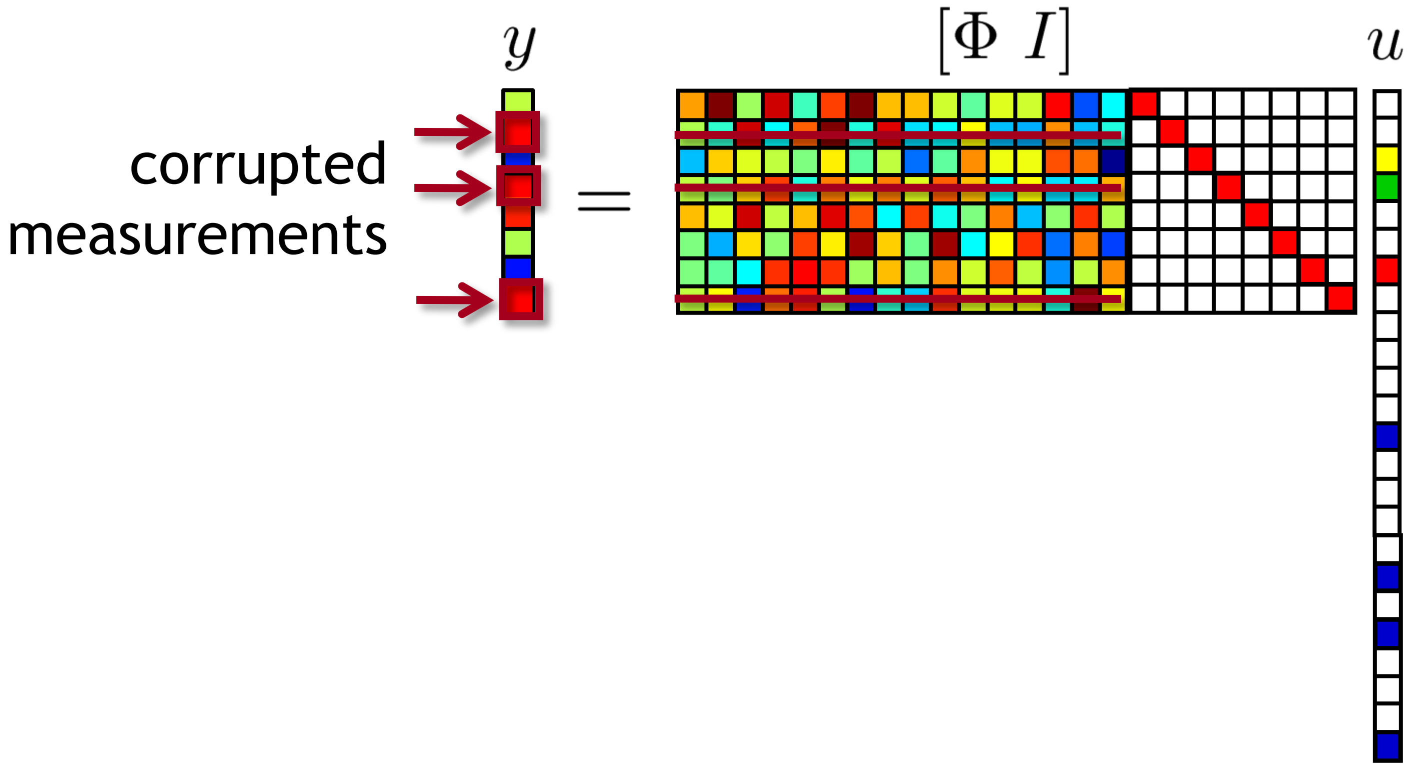
Measurement Interference Cancellation

What about structured measurement noise?



Measurement Interference Cancellation

What about structured measurement noise?



Justice Pursuit

$$\widehat{u} = \arg\min_{u} \|u\|_1$$
$$\text{s.t.} \quad y = [\Phi \;\; I]\, u$$

Does this matrix satisfy the RIP?

$$\|[\Phi \;\; I]u\|_2^2 = \|\Phi x\|_2^2 + 2e^T\Phi x + \|e\|_2^2 \approx \|x\|_2^2 + \|e\|_2^2$$

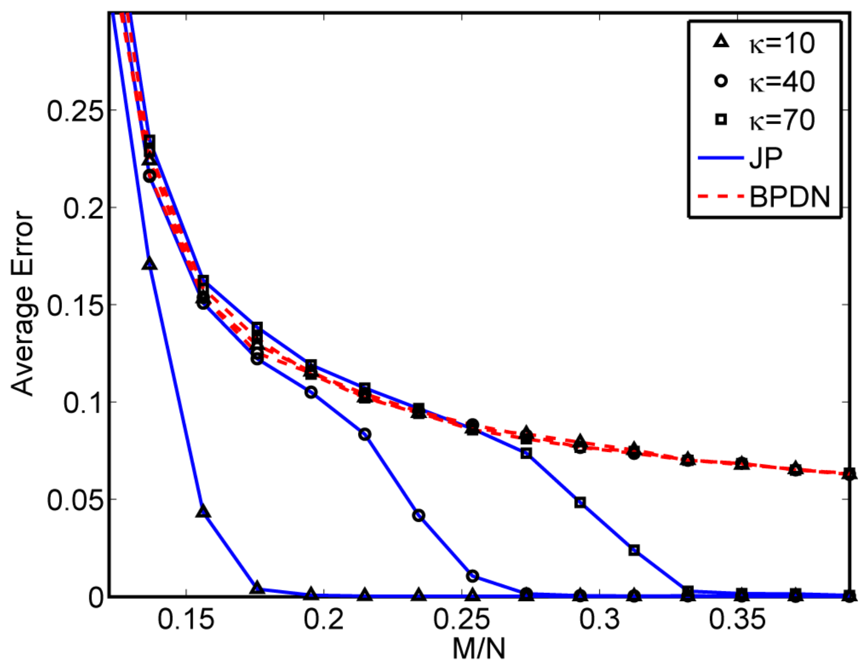
Theorem:
If Φ is a sub-Gaussian matrix with
$$M = O\left((S+\kappa)\log\left(\frac{N+M}{S+\kappa}\right)\right)$$
then $[\Phi \;\; I]$ satisfies the RIP of order $(S+\kappa)$ with probability at least $1 - 3e^{-CM}$.

[Laska, D, and Baraniuk - 2009]

Justice Pursuit

We can recover sparse signals *exactly* in the presence of *unbounded* sparse noise

Fixed $\|e\|_2 = 0.1$



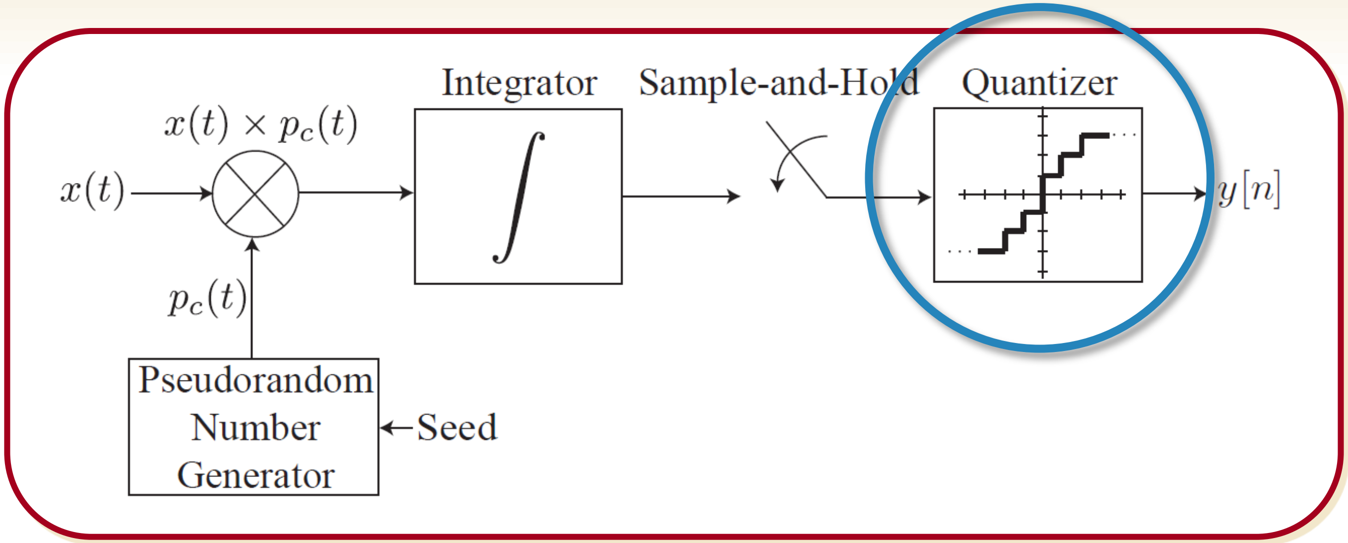
[Laska, D, and Baraniuk - 2009]

Conclusions

- CS systems are sensitive to noisy signals
  - if our input signal is very noisy, it isn't really very sparse
  - when noise is large, *measurements matter*
  - exploit sparsity in a different manner - e.g., adaptivity
- CS can be highly robust to *interference*
  - structured signal noise
  - structured measurement noise
- What about quantization noise?

Quantization Noise

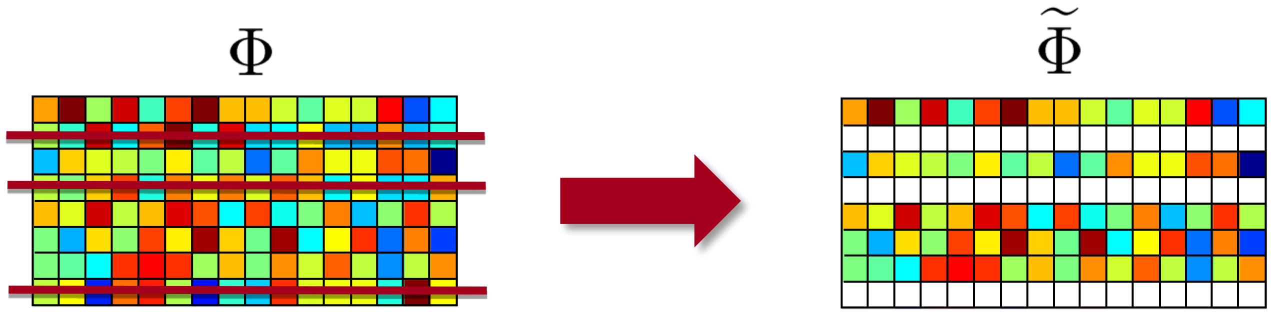
Signal Recovery with Quantization



- Finite-range quantization leads to *saturation* and *unbounded errors*
- Quantization noise changes as we change the sampling rate

Saturation Strategies

- **Rejection:** Ignore saturated measurements

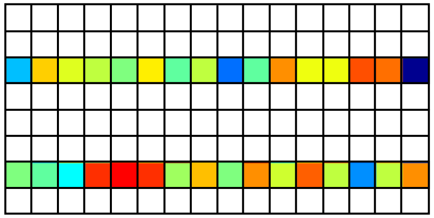


- **Consistency:** Retain saturated measurements.
  Use them only as inequality constraints on the recovered signal

- If the rejection approach works, the consistency approach should automatically do better

Rejection and Democracy

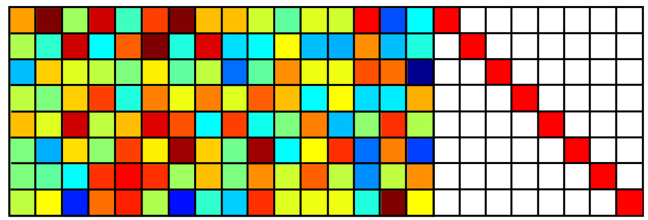
- The RIP is *not sufficient* for the rejection approach
- Example: $\Phi = I$
  - perfect isometry
  - *every* measurement must be kept
- We would like to be able to say that *any* submatrix of Φ with sufficiently many rows will still satisfy the RIP



- Strong, *adversarial* form of "democracy"

Sketch of Proof

- Step 1: Concatenate the identity to Φ



Theorem:
If Φ is a sub-Gaussian matrix with

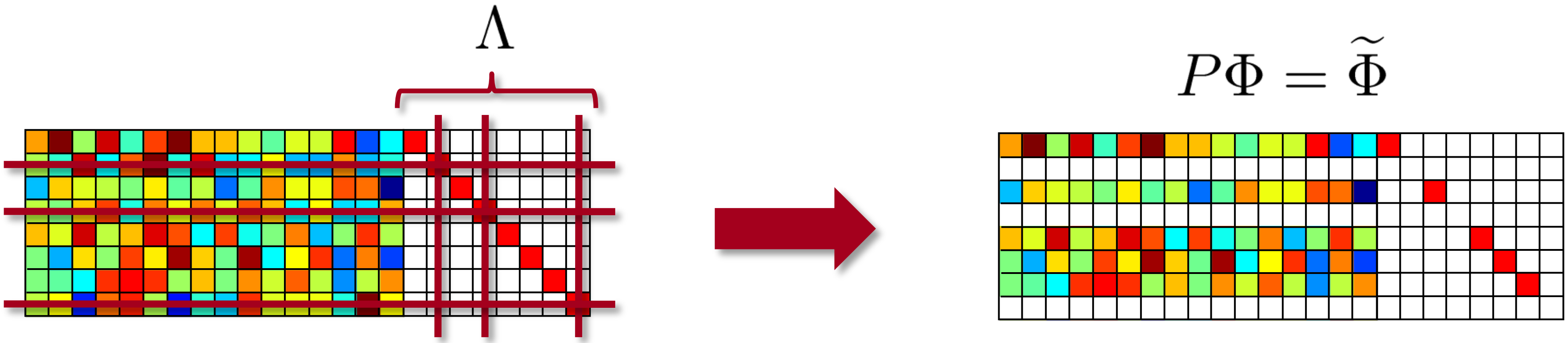
$$M = O\left(S \log\left(\frac{N+M}{S}\right)\right)$$

then $[\Phi \;\; I]$ satisfies the RIP of order S with probability at least $1 - 3e^{-CM}$.

[D, Laska, Boufounos, and Baraniuk - 2009]

Sketch of Proof

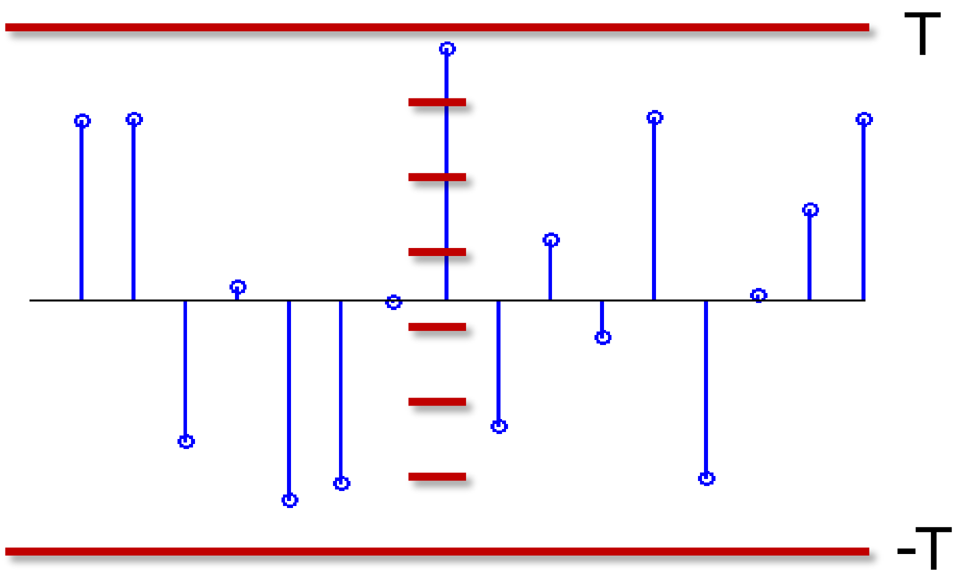
- Step 2: Combine with the "interference cancellation" lemma



- The fact that $[\Phi \ I]$ satisfies the RIP implies that if we take D extra measurements, then we can delete $O(D)$ arbitrary rows of Φ and retain the RIP
- This is a strong *adversarial* notion of democracy

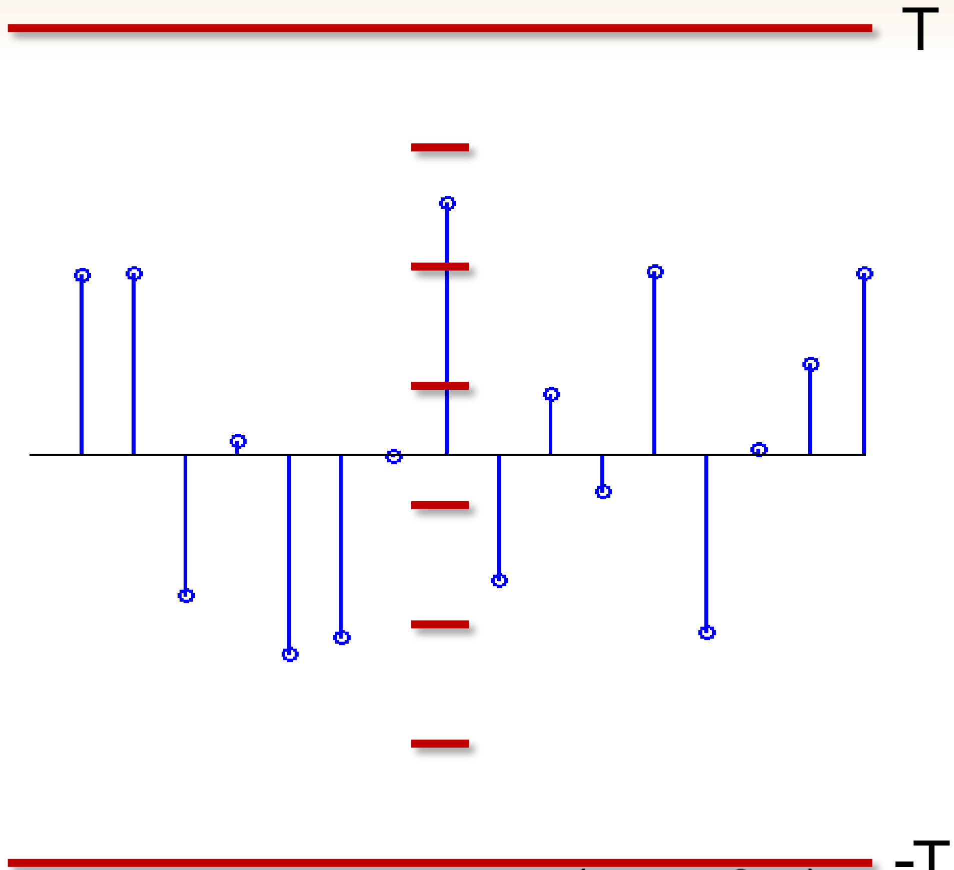
[D, Laska, Boufounos, and Baraniuk - 2009]

Rejection In Practice



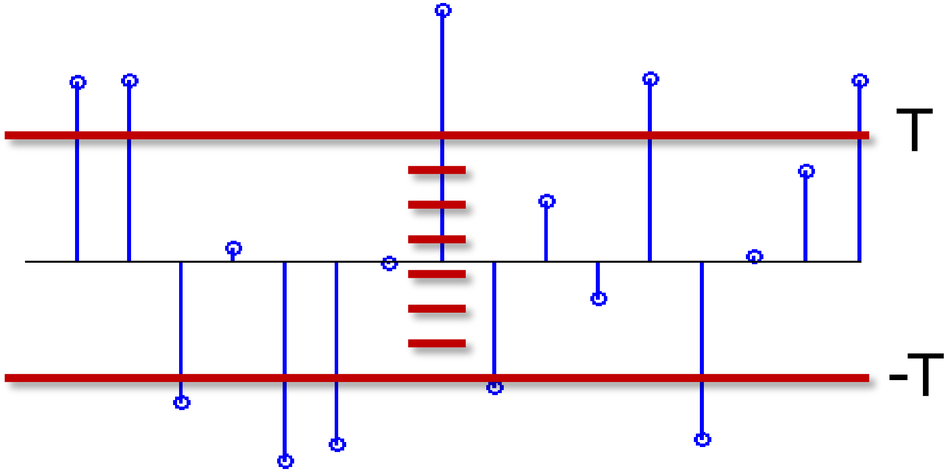
$$\mathrm{SNR} = 10\log_{10}\left(\frac{\|x\|_2^2}{\|\widehat{x}-x\|_2^2}\right)$$

Rejection In Practice



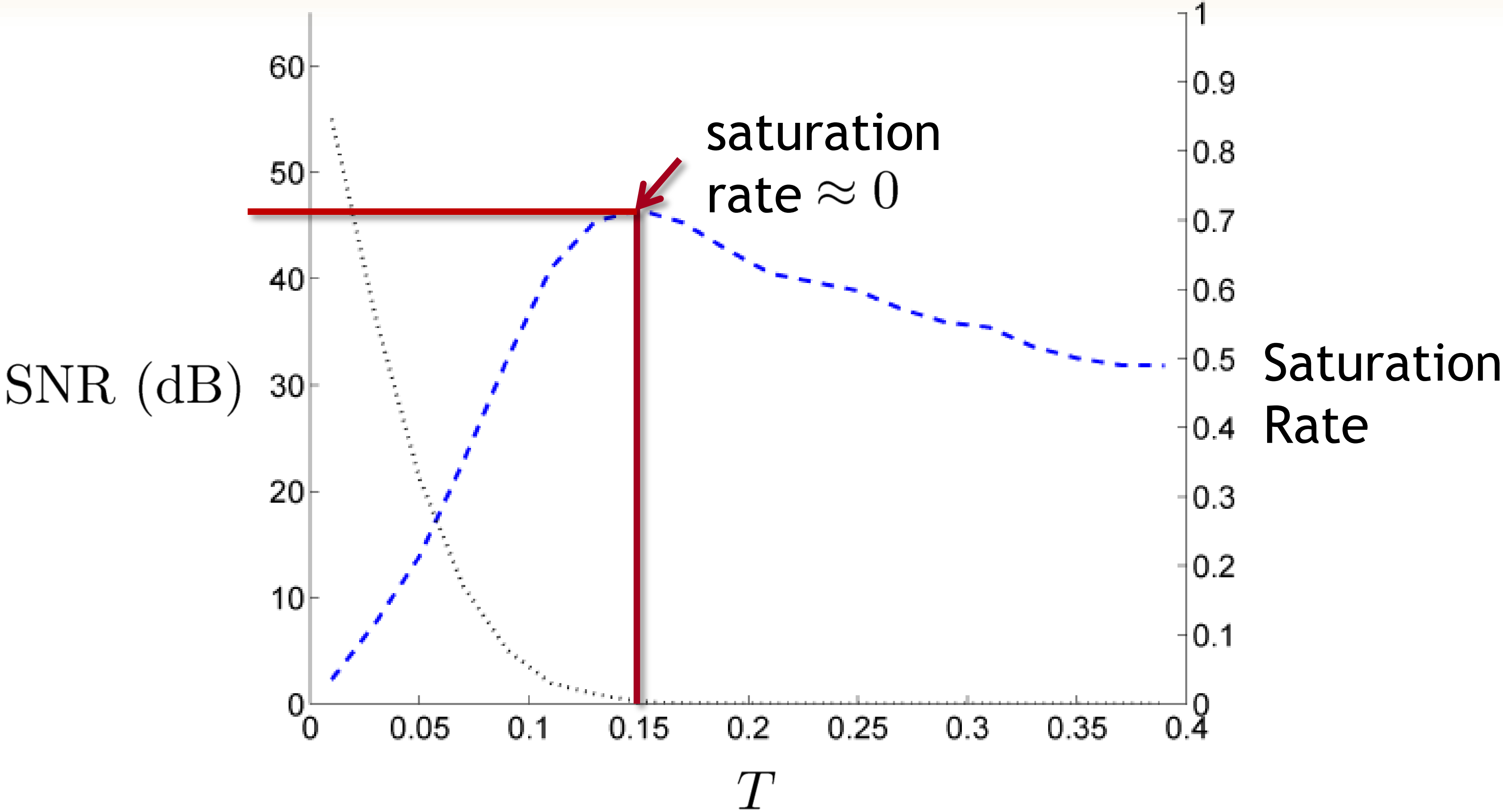
$$\text{SNR} = 10\log_{10}\left(\frac{\|x\|_2^2}{\|\widehat{x}-x\|_2^2}\right)$$

Rejection In Practice



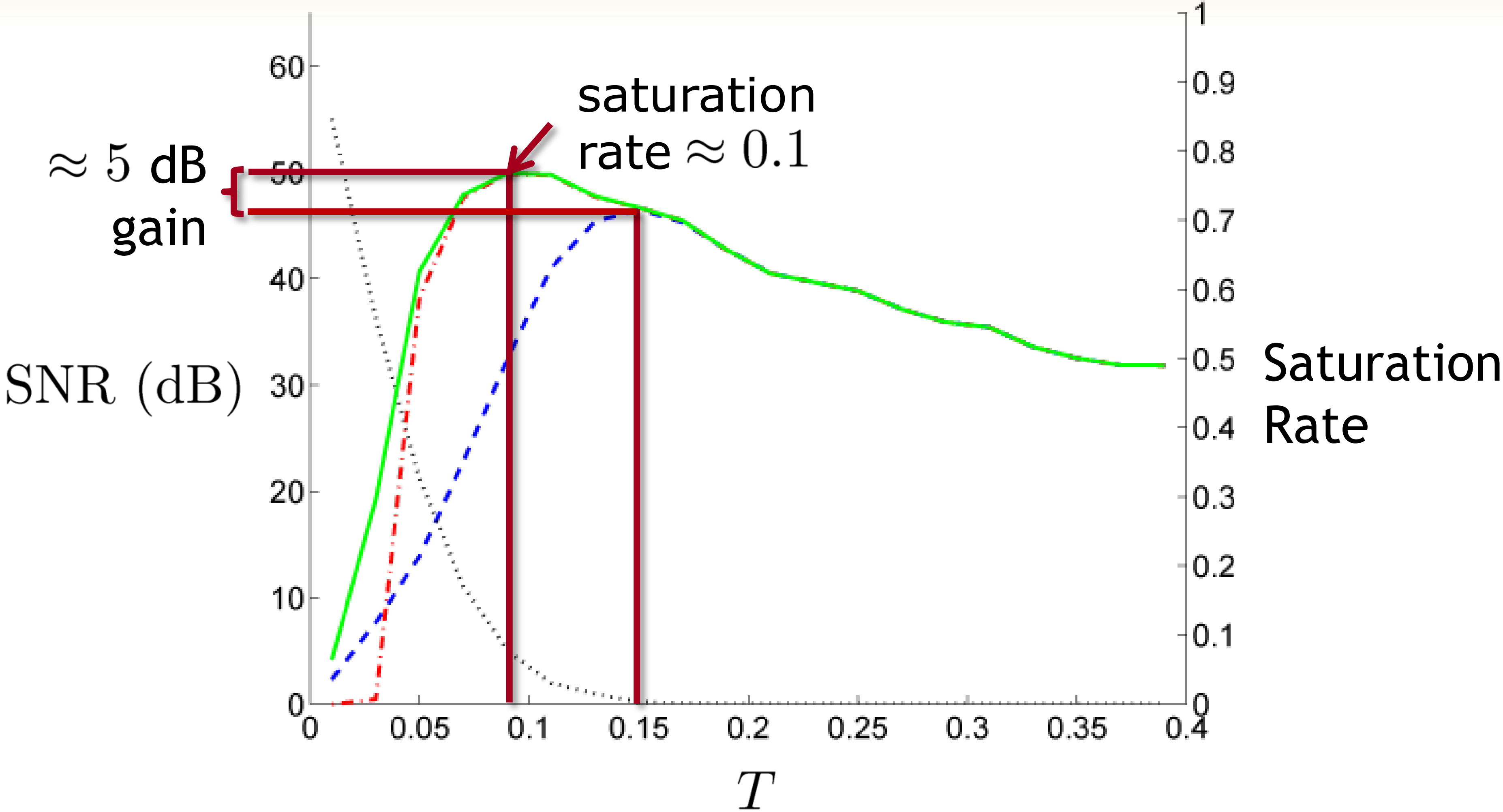
$$\mathrm{SNR} = 10\log_{10}\left(\frac{\|x\|_2^2}{\|\widehat{x}-x\|_2^2}\right)$$

Benefits of Saturation



[Laska, Boufounos, D, and Baraniuk - 2011]

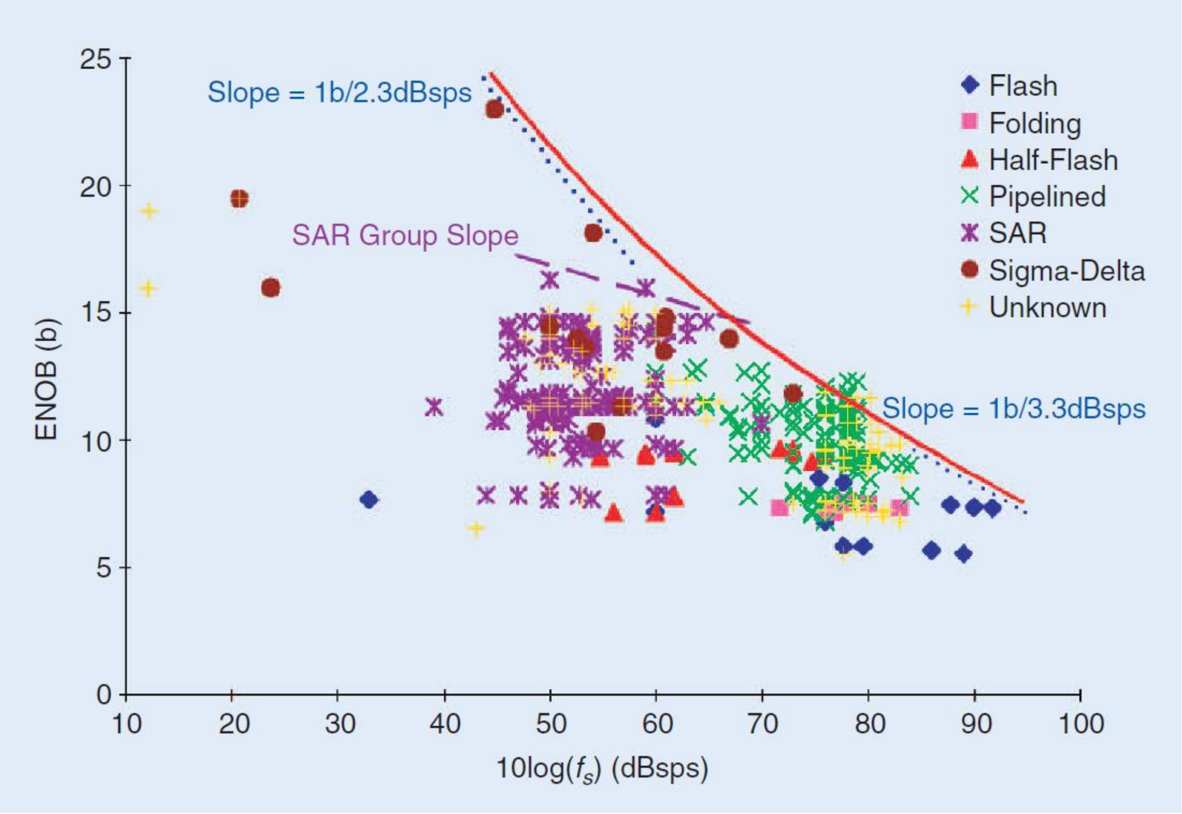
Benefits of Saturation



[Laska, Boufounos, D, and Baraniuk - 2011]

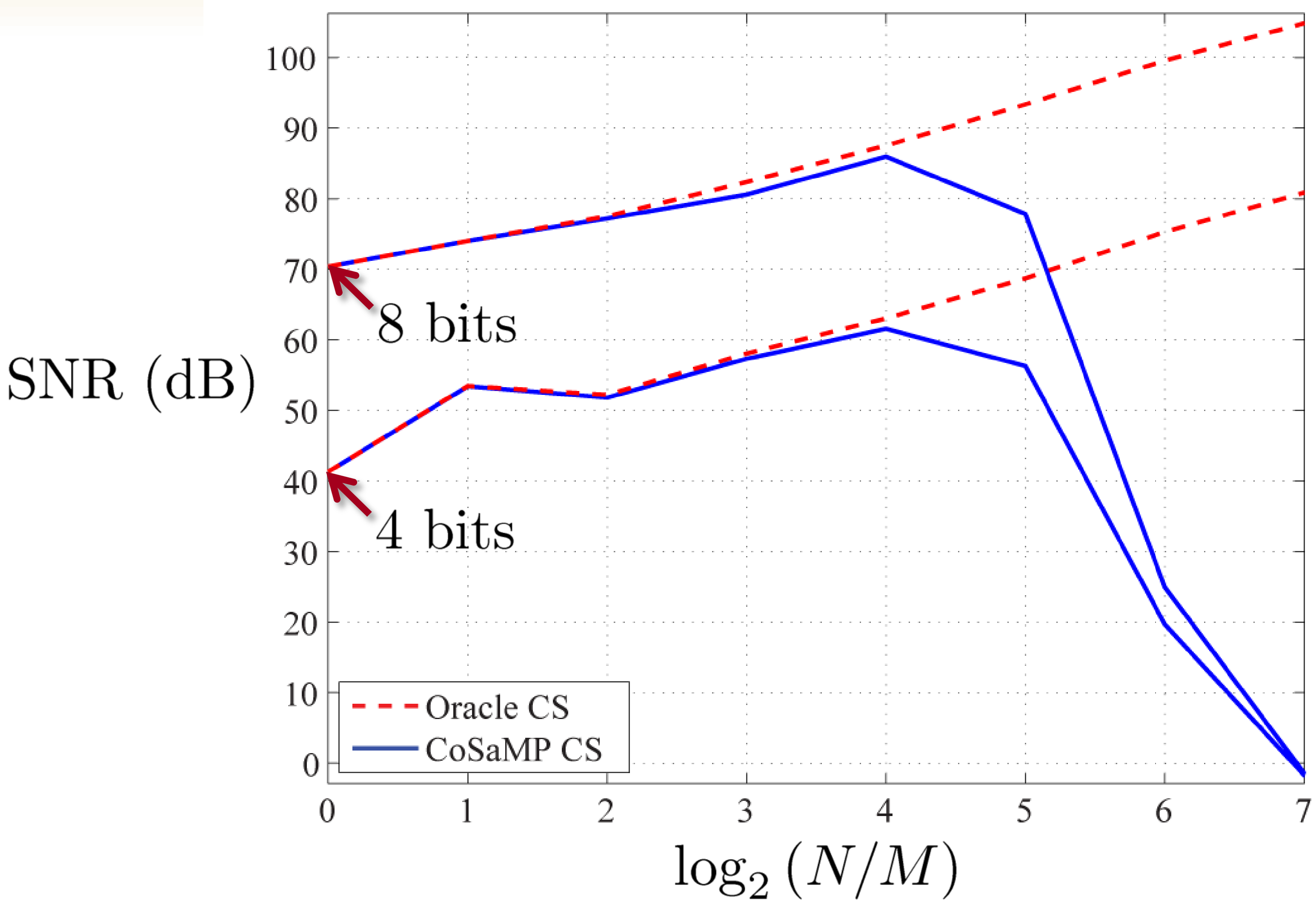
Potential for SNR Improvement?

By sampling at a lower rate, we can quantize to a higher bit-depth, allowing for potential gains



[Le et al. - 2005]

Empirical SNR Improvement



[D, Laska, Treichler, and Baraniuk - 2011]

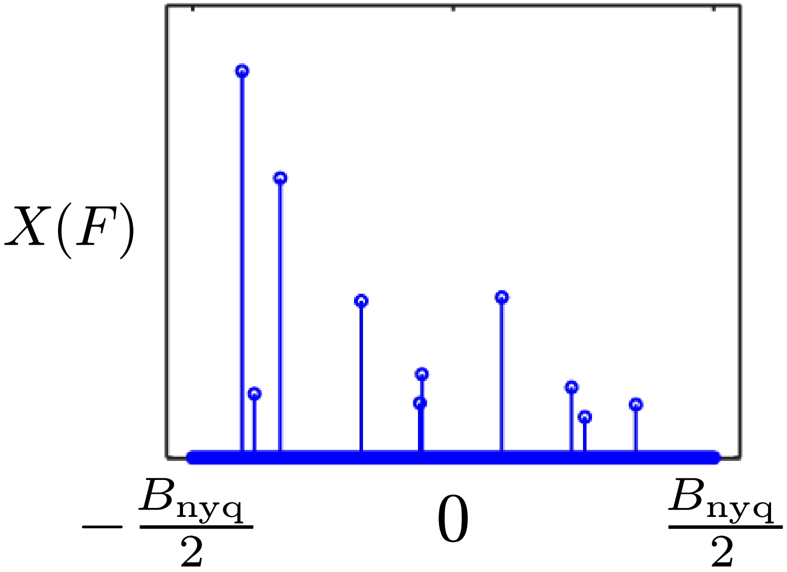
Conclusions

- CS is robust to quantization noise in a non-traditional sense
- Democracy is a major advantage of CS measurements
- CS offers the potential to significantly boost dynamic range
  - can offset drawbacks associated with noise
- When is CS most useful?
  - performance is limited by quantization (high bandwidth apps)
  - when your signal is sparse (not too noisy)

Real-World
Signal Models

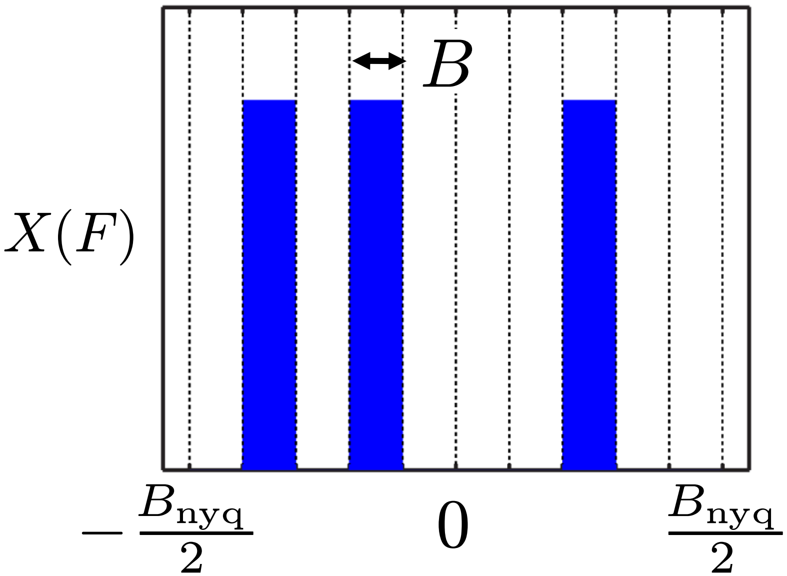
Candidate Analog Signal Models

	Model for $x(t)$	Basis for x	Sparsity level for x
multitone	sum of S "on-grid" tones	Ψ = DFT	S -sparse



Candidate Analog Signal Models

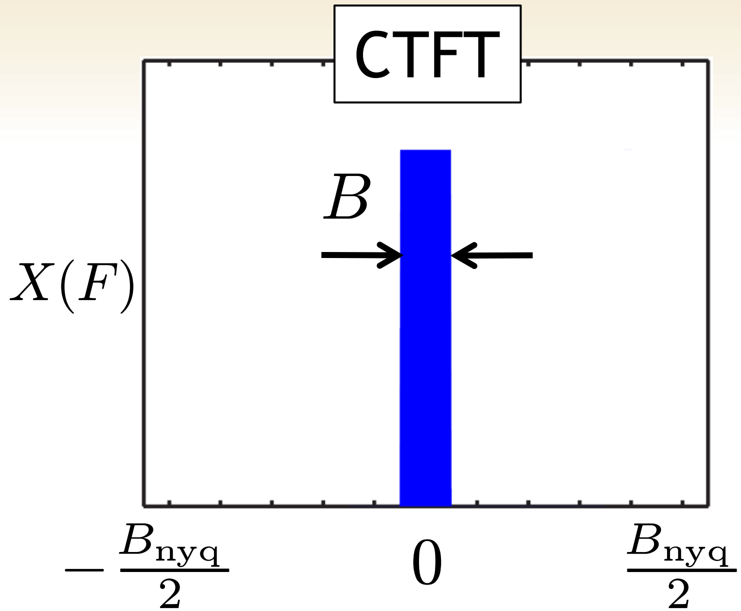
	Model for $x(t)$	Basis for x	Sparsity level for x
multitone	sum of S "on-grid" tones	Ψ = DFT	S -sparse
multiband	K occupied bands of bandwidth B	Ψ = ?	?



- Landau
- Bresler, Feng, Venkataramani
- Eldar, Mishali

The Problem with the DFT

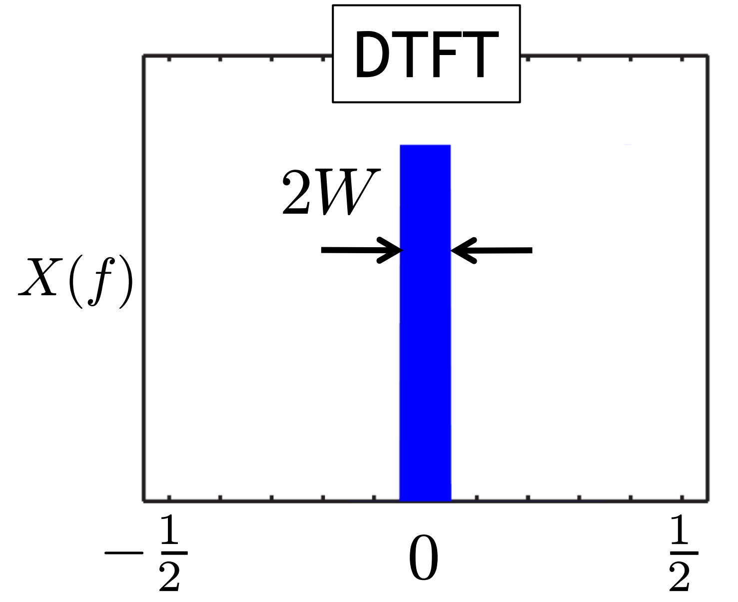
$$x(t) = \int_{-\frac{B}{2}}^{\frac{B}{2}} X(F) e^{j2\pi Ft}\, dF$$



sampling

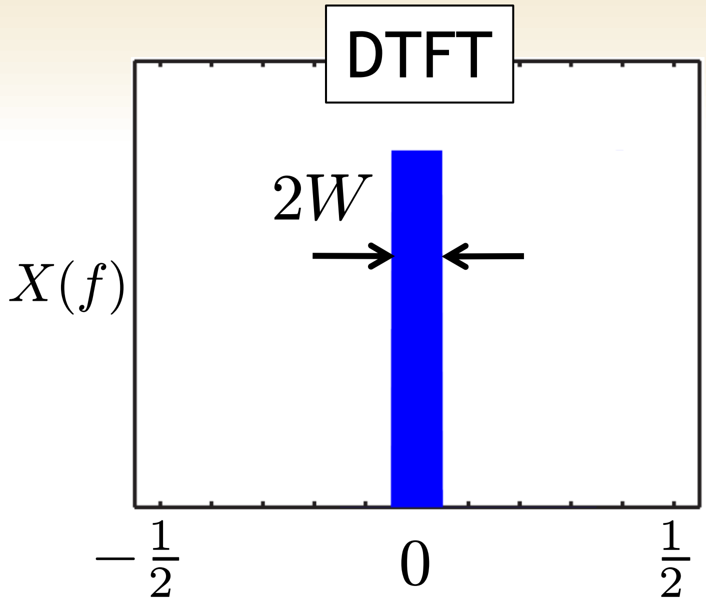
$$x[n] = \int_{-W}^{W} X(f) e^{j2\pi fn}\, df, \ \forall n$$

$$W = \frac{B}{2B_{\text{nyq}}}$$



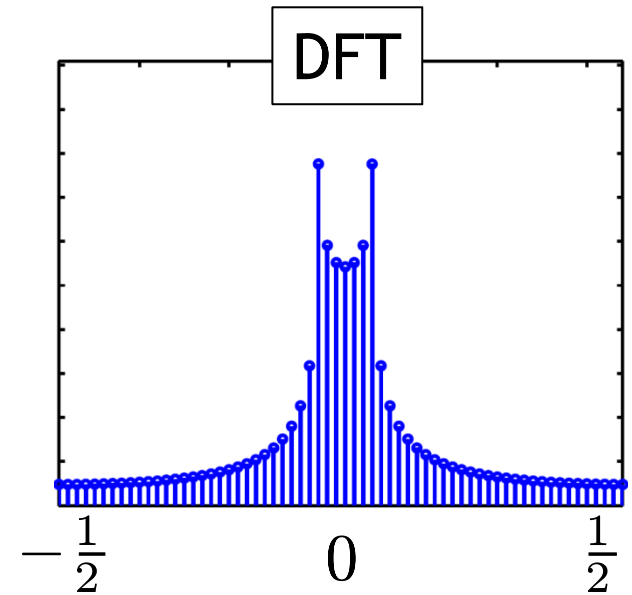
The Problem with the DFT

$$x[n] = \int_{-W}^{W} X(f) e^{j2\pi f n} \, df, \ \forall n$$



time-limiting

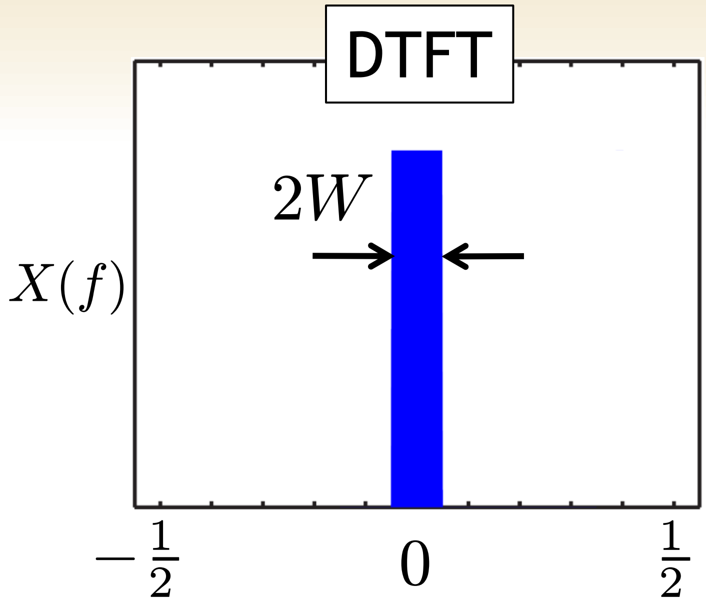
$$x = \sum_{k=0}^{N-1} X_k e_{\frac{k}{N}}, \quad e_f := \begin{bmatrix} e^{j2\pi f 0} \\ e^{j2\pi f} \\ \vdots \\ e^{j2\pi f (N-1)} \end{bmatrix}$$



NOT SPARSE

Alternative Perspective

$$x[n] = \int_{-W}^{W} X(f) e^{j2\pi f n}\, df, \ \forall n$$



time-limiting

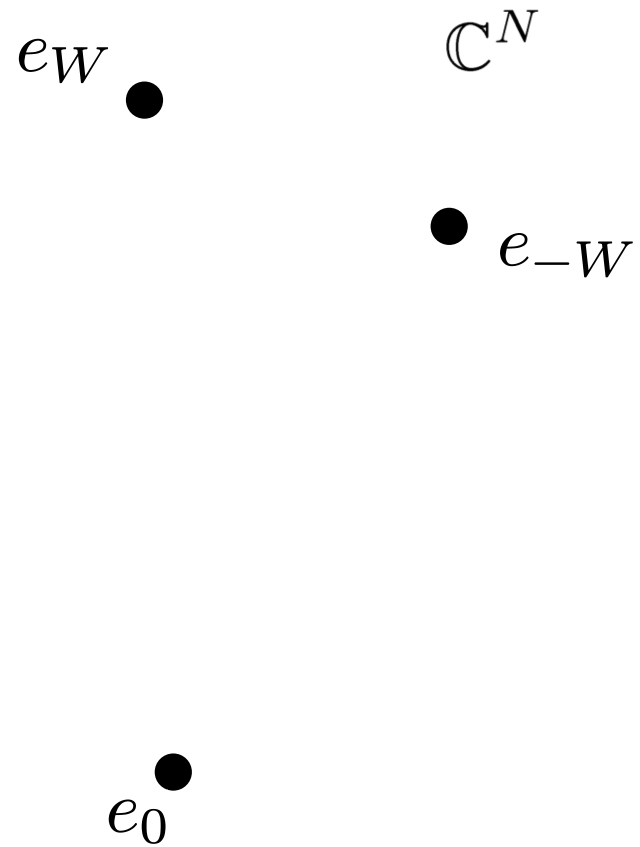
$$\mathcal{T}_N(x[n]) = \int_{-W}^{W} X(f) \mathcal{T}_N(e^{j2\pi f n})\, df, \ \forall n$$

Building Blocks for Lowpass Signals

Time-limited complex exponentials form a “basis” for bandlimited signals

$$x = \int_{-W}^{W} X(f) e_f \, df$$

$$e_f := \begin{bmatrix} e^{j2\pi f 0} \\ e^{j2\pi f} \\ \vdots \\ e^{j2\pi f (N-1)} \end{bmatrix}$$



The problem: we need infinitely many of them.

Best Subspace Fit

Suppose that we wish to minimize

$$\int_{-W}^{W} \|e_f - P_Q e_f\|_2^2 \, df$$

over all subspaces Q of dimension k.

Optimal subspace is spanned by the first k “DPSS vectors”.

Discrete Prolate Spheroidal Sequences (DPSS's)

Slepian [1978]: Given an integer N and $W \leq \frac{1}{2}$,
the *DPSS's* are a collection of N vectors

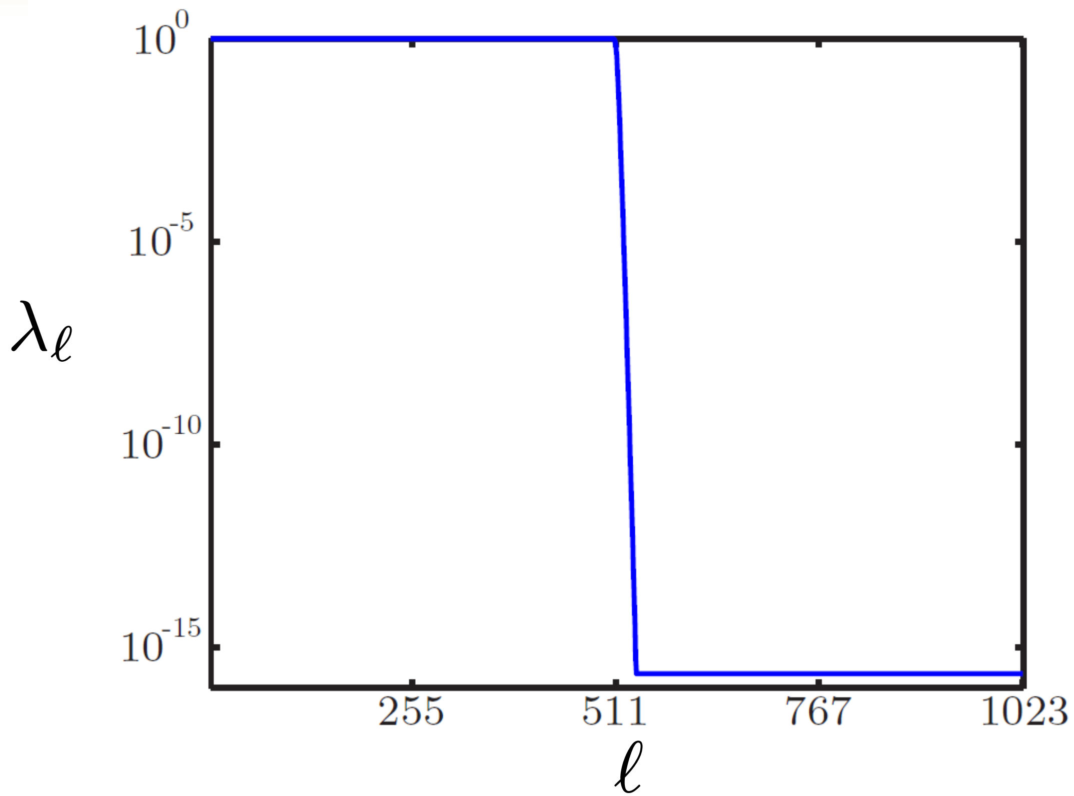
$$s_0, s_1, \ldots, s_{N-1} \in \mathbb{R}^N$$

that satisfy

$$\mathcal{T}_N(\mathcal{B}_W(s_\ell))) = \lambda_\ell s_\ell.$$

The DPSS's are perfectly time-limited, but when $\lambda_\ell \approx 1$ they are highly concentrated in frequency.

DPSS Eigenvalue Concentration



$N = 1024$

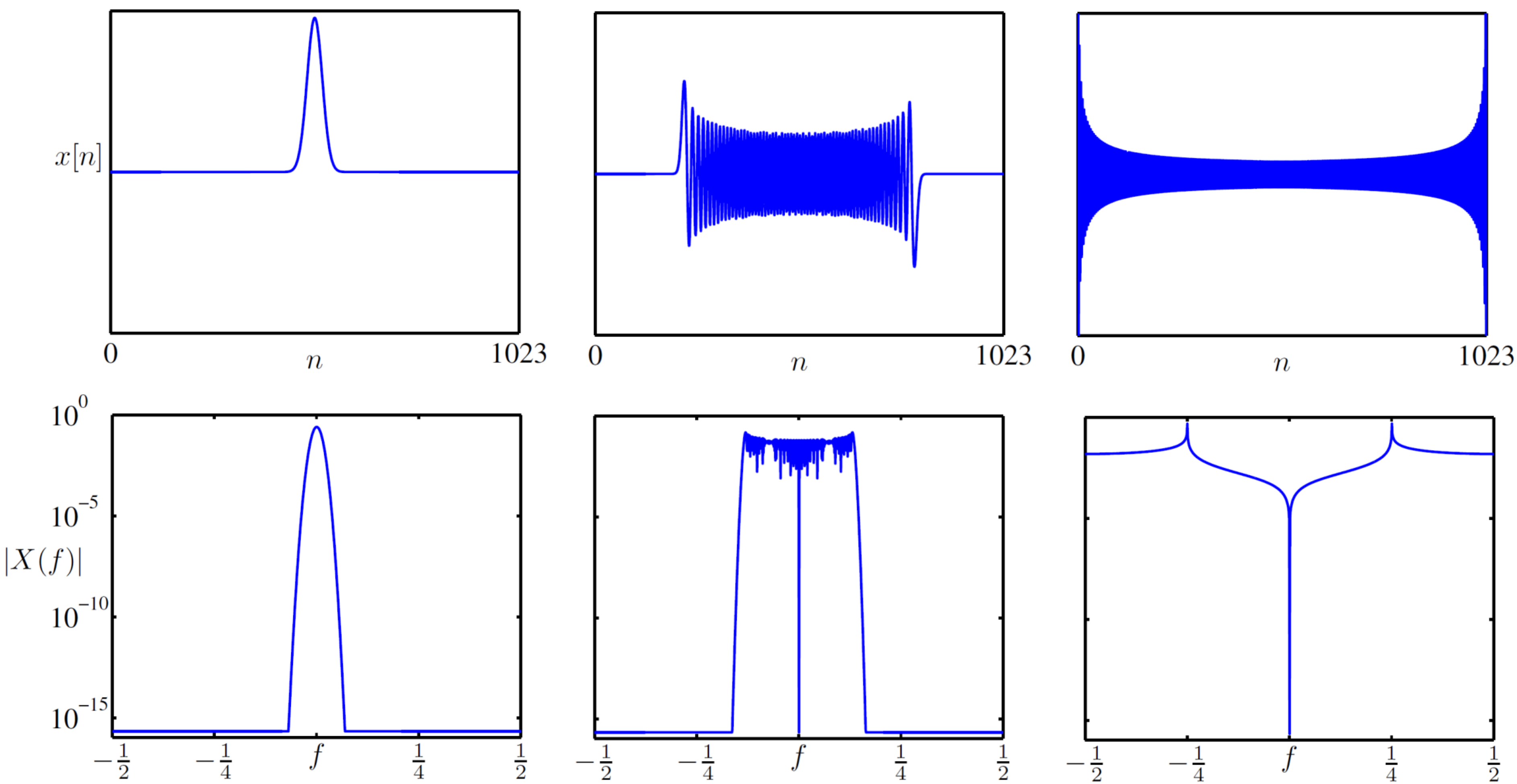
$W = \frac{1}{4}$

$2NW = 512$

The first $\approx 2NW$ eigenvalues ≈ 1.
The remaining eigenvalues ≈ 0.

DPSS Examples

$$N = 1024 \qquad W = \frac{1}{4}$$



$\ell = 0$ $\qquad$ $\ell = 127$ $\qquad$ $\ell = 511$

Recall: Best Subspace Fit

Suppose that we wish to minimize

$$\int_{-W}^{W} \|e_f - P_Q e_f\|_2^2 \, df$$

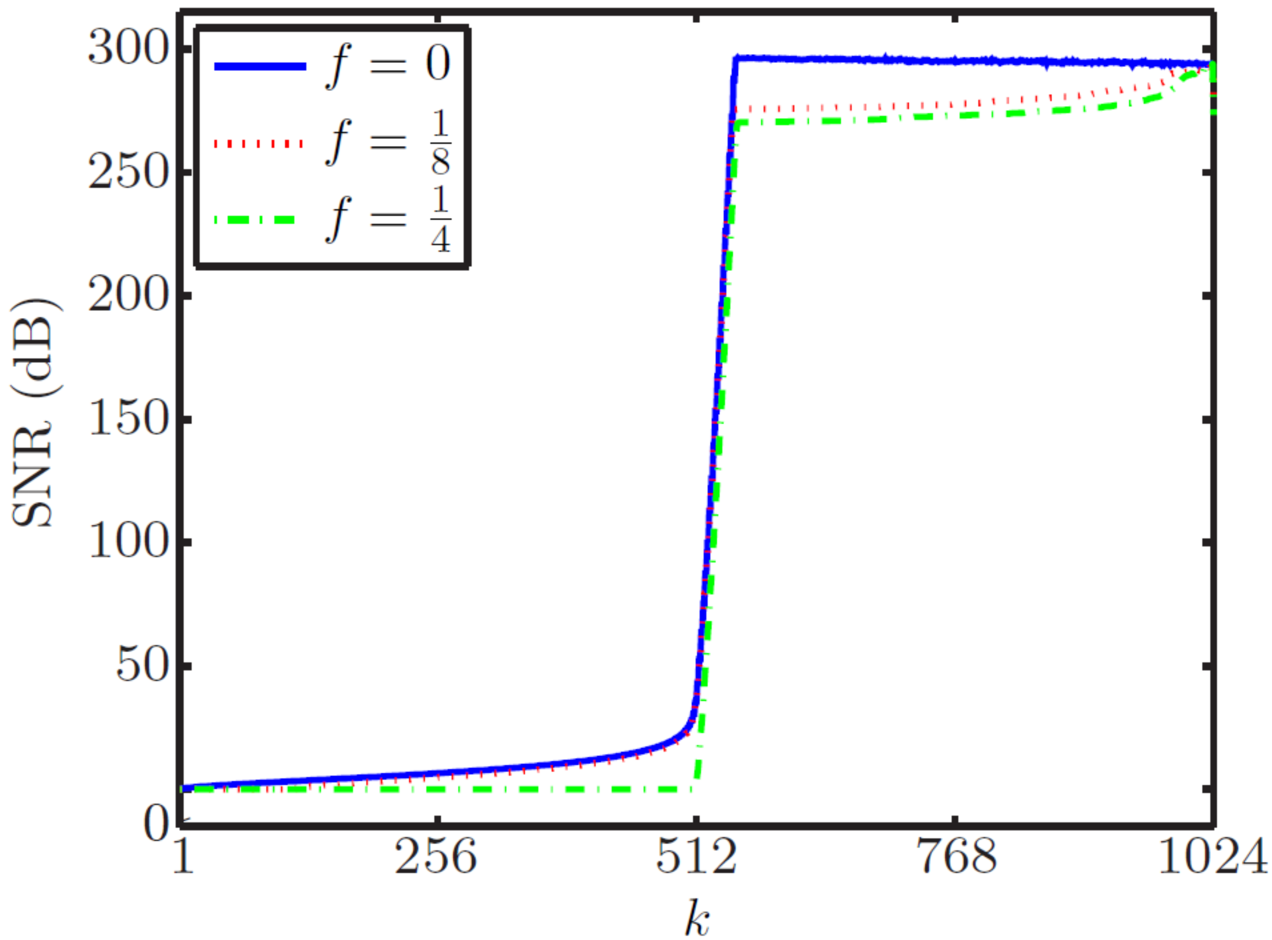
over all subspaces Q of dimension k.

Optimal subspace is spanned by the first k "DPSS vectors".

$$\int_{-W}^{W} \|e_f - P_Q e_f\|_2^2 \, df = \sum_{\ell=k}^{N-1} \lambda_\ell$$

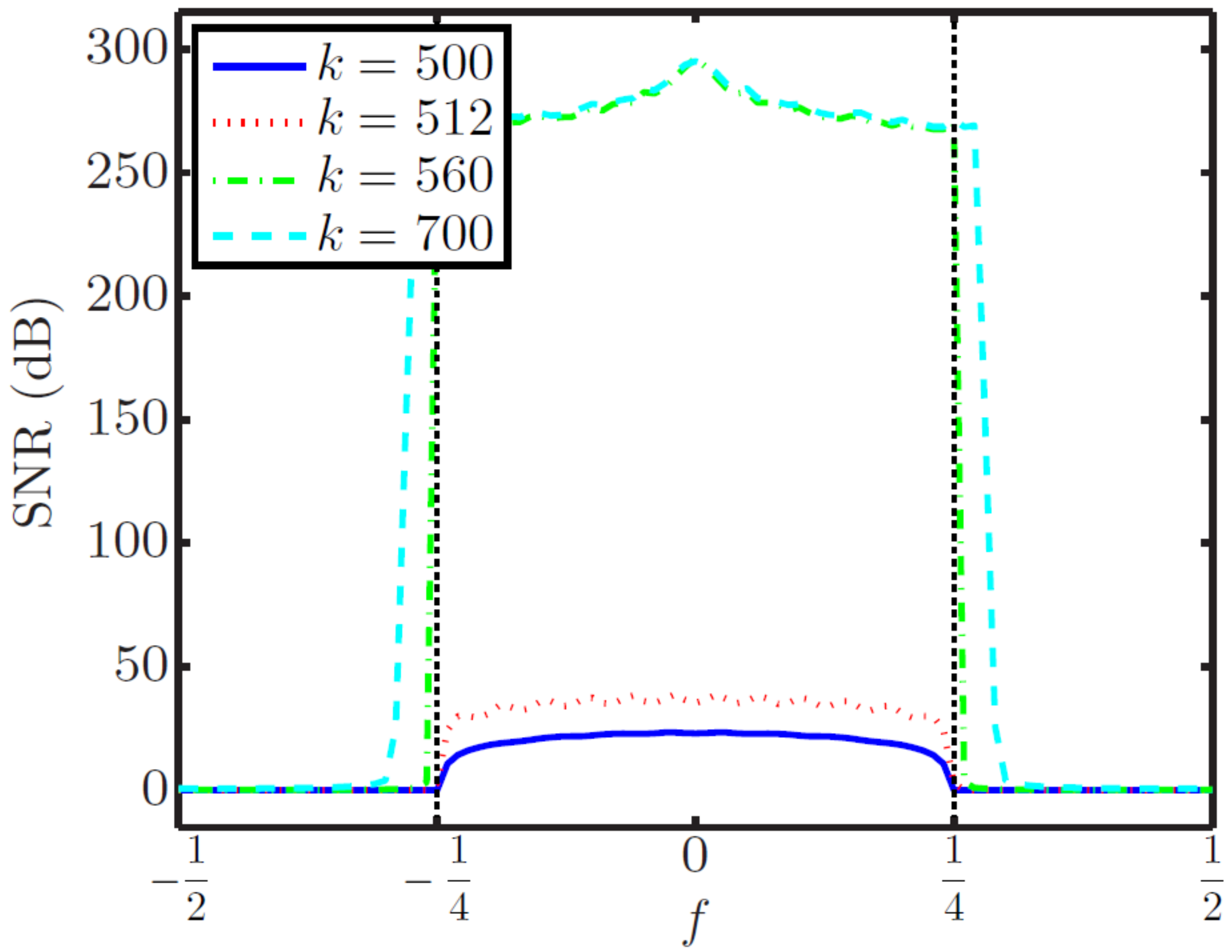
Approximation of Bandlimited Signals

$$\mathrm{SNR} = 20\log_{10}\left(\frac{\|e_f\|}{\|e_f - P_Q e_f\|}\right)\mathrm{dB}$$

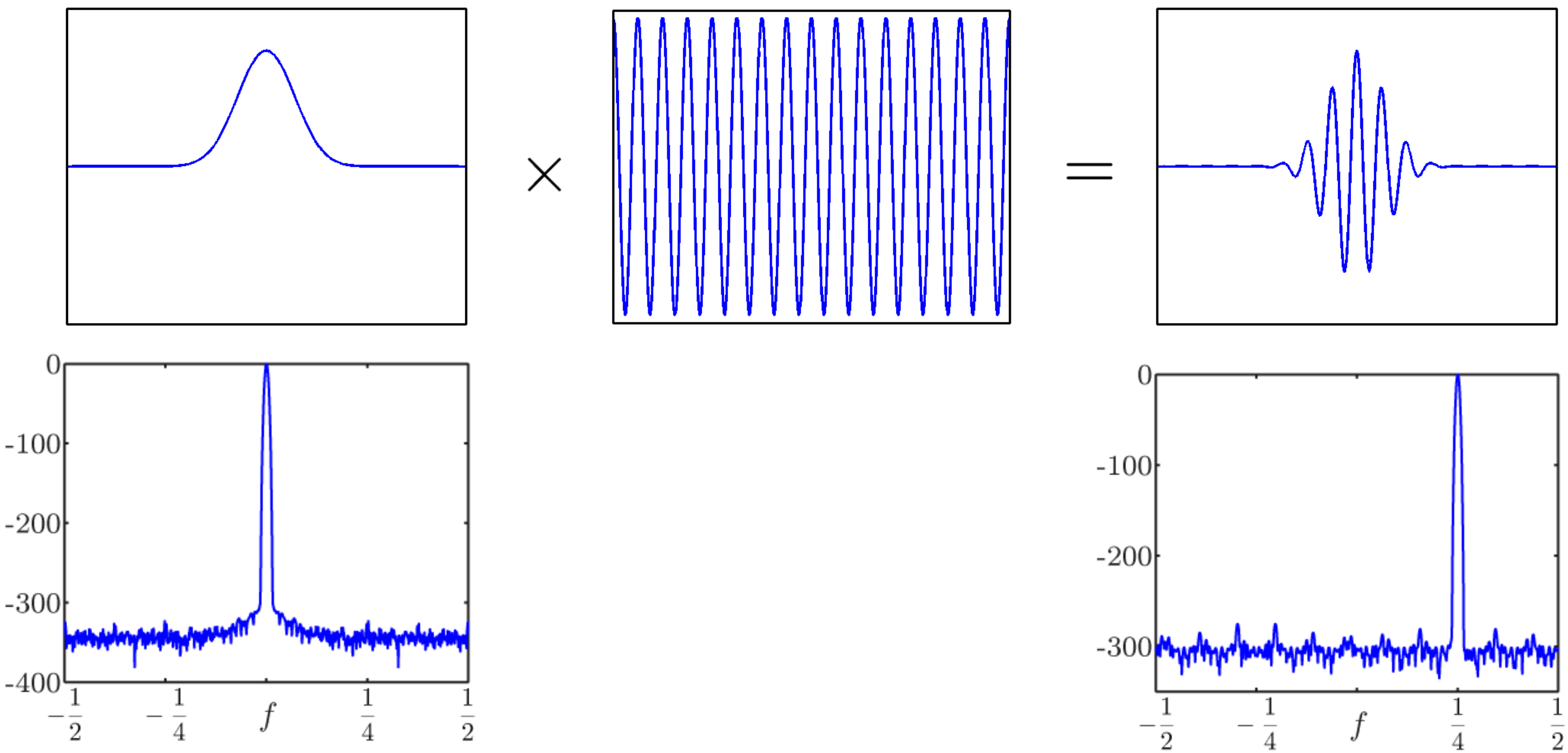


Approximation of Bandlimited Signals

$$\mathrm{SNR} = 20\log_{10}\left(\frac{\|e_f\|}{\|e_f - P_Q e_f\|}\right)\,\mathrm{dB}$$



DPSS's for Bandpass Signals

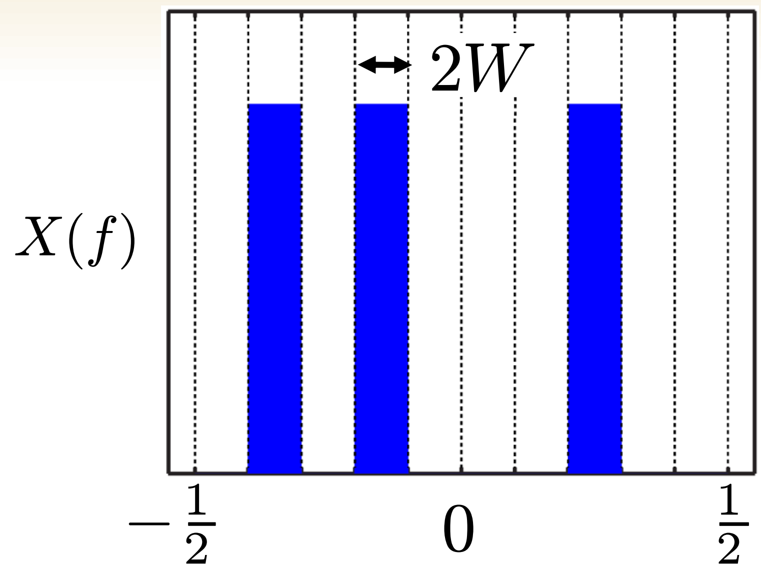


DPSS Dictionaries for CS

Modulate k DPSS vectors
to center of each band:

$$\underbrace{\Psi = [\Psi_1, \Psi_2, \ldots, \Psi_J]}$$

approximately square
if $k \approx 2NW$



J possible bands

Most multiband signals, when sampled and time-limited, are well-approximated by a sparse representation in Ψ .

DPSS Dictionaries and the RIP

Theorem:
Suppose that Φ is sub-Gaussian and that the Ψ_i are constructed with $k = (1-\epsilon)2NW$. If

$$M \geq CS \log(N/S)$$

then with high probability $\Phi\Psi$ will satisfy the RIP of order S.

K occupied bands → $S \approx KNB/B_{\text{nyq}}$

$$\boxed{\frac{M}{N} \geq C' \frac{KB}{B_{\text{nyq}}} \log\left(\frac{B_{\text{nyq}}}{KB}\right)}$$

[D and Wakin - 2011]

Block-Sparse Recovery

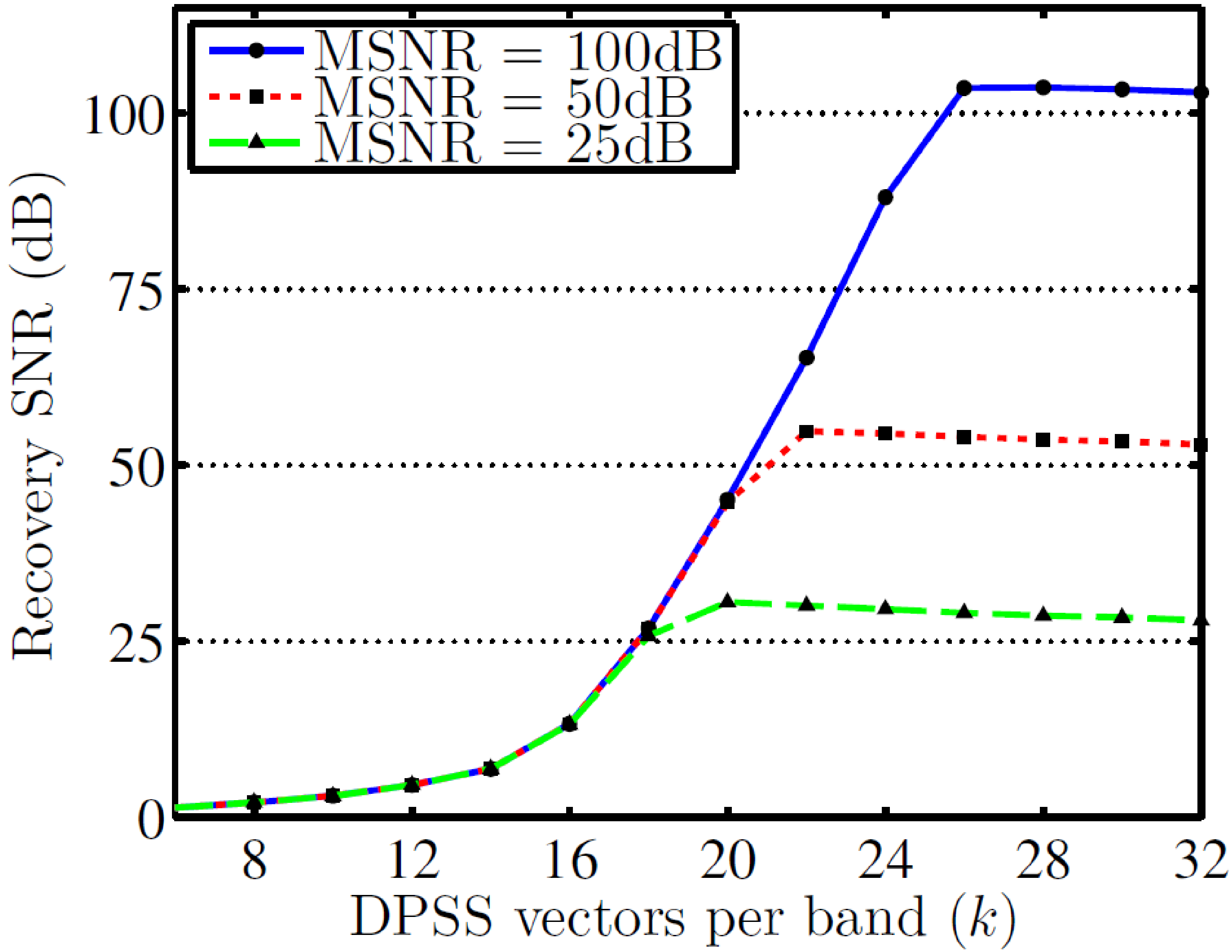
Nonzero coefficients of α should be clustered in blocks according to the occupied frequency bands

$$x = [\Psi_1, \Psi_2, \ldots, \Psi_J] \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_J \end{bmatrix}$$

This can be leveraged to reduce the required number of measurements and improve performance through "model-based CS"

- Baraniuk et al. [2008, 2009, 2010]
- Blumensath and Davies [2009, 2011]

Empirical Results: Noise



$$N = 4096$$

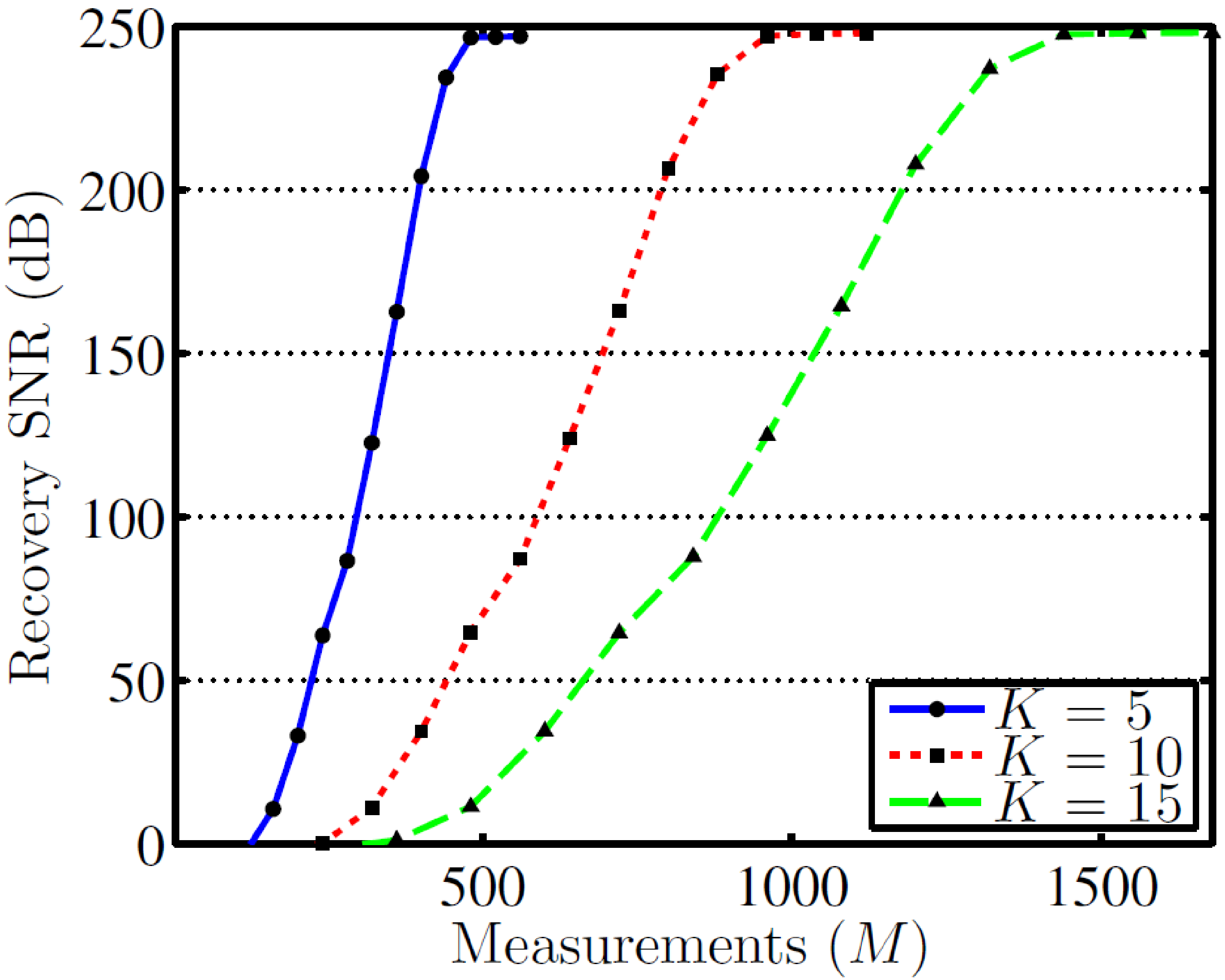
$$M = 512$$

$$K = 5$$

$$\frac{B}{B_{\mathrm{nyq}}} = \frac{1}{256}$$

[D and Wakin - 2011]

Empirical Results: Measurements

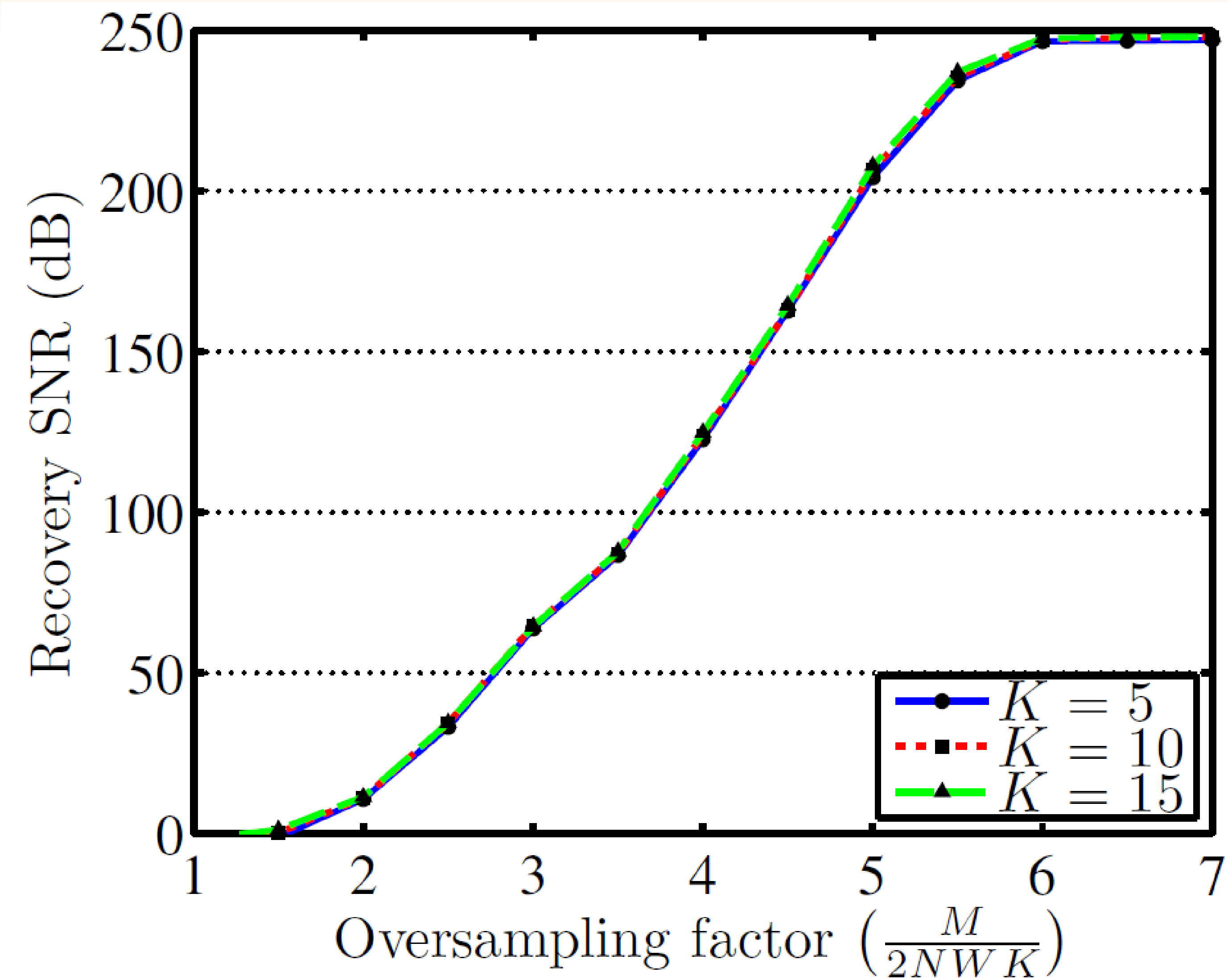


$$N = 4096$$

$$\frac{B}{B_{\text{nyq}}} = \frac{1}{256}$$

[D and Wakin - 2011]

Empirical Results: Measurements

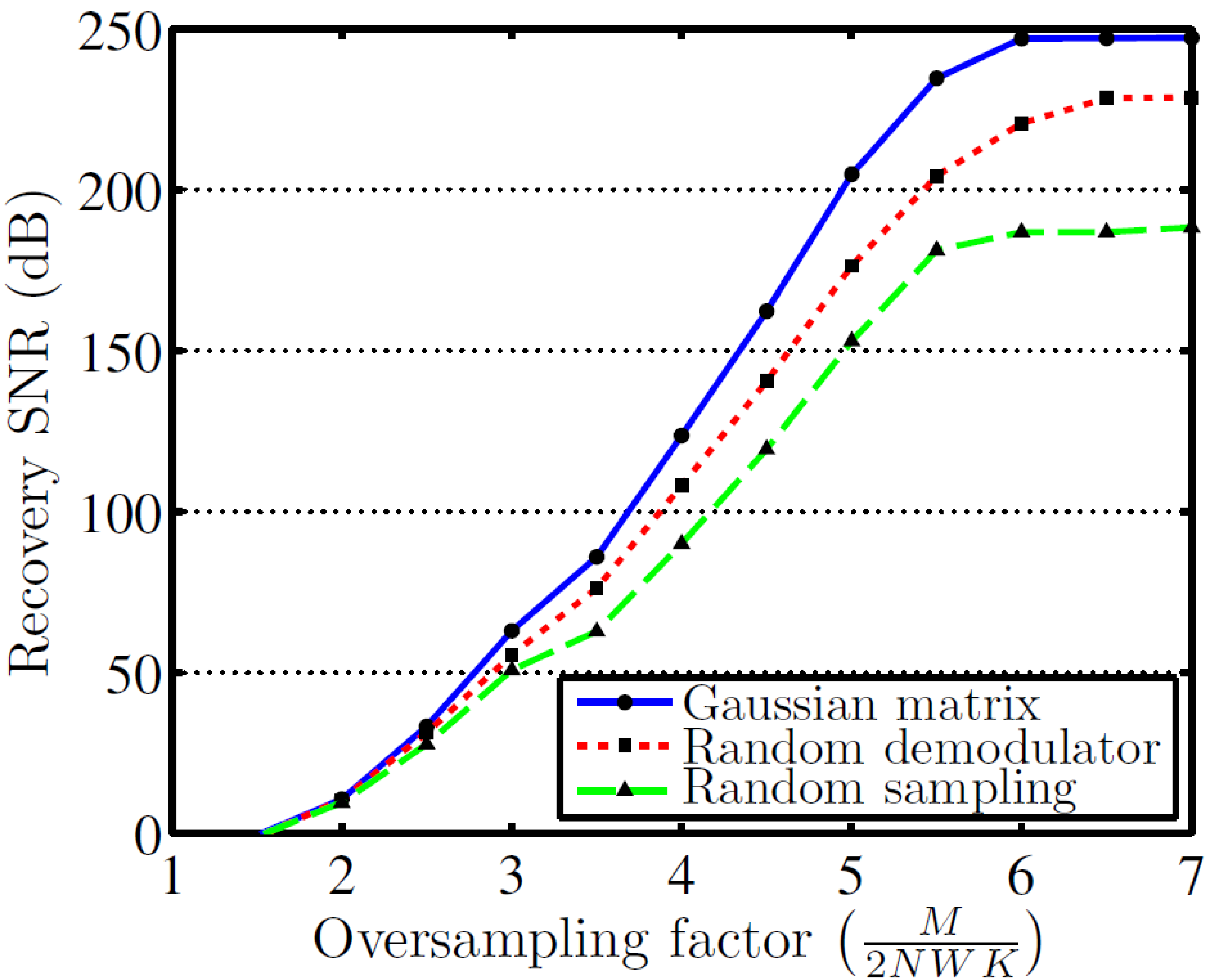


$N = 4096$

$\frac{B}{B_{\text{nyq}}} = \frac{1}{256}$

[D and Wakin - 2011]

Empirical Results: Real-World Sensors



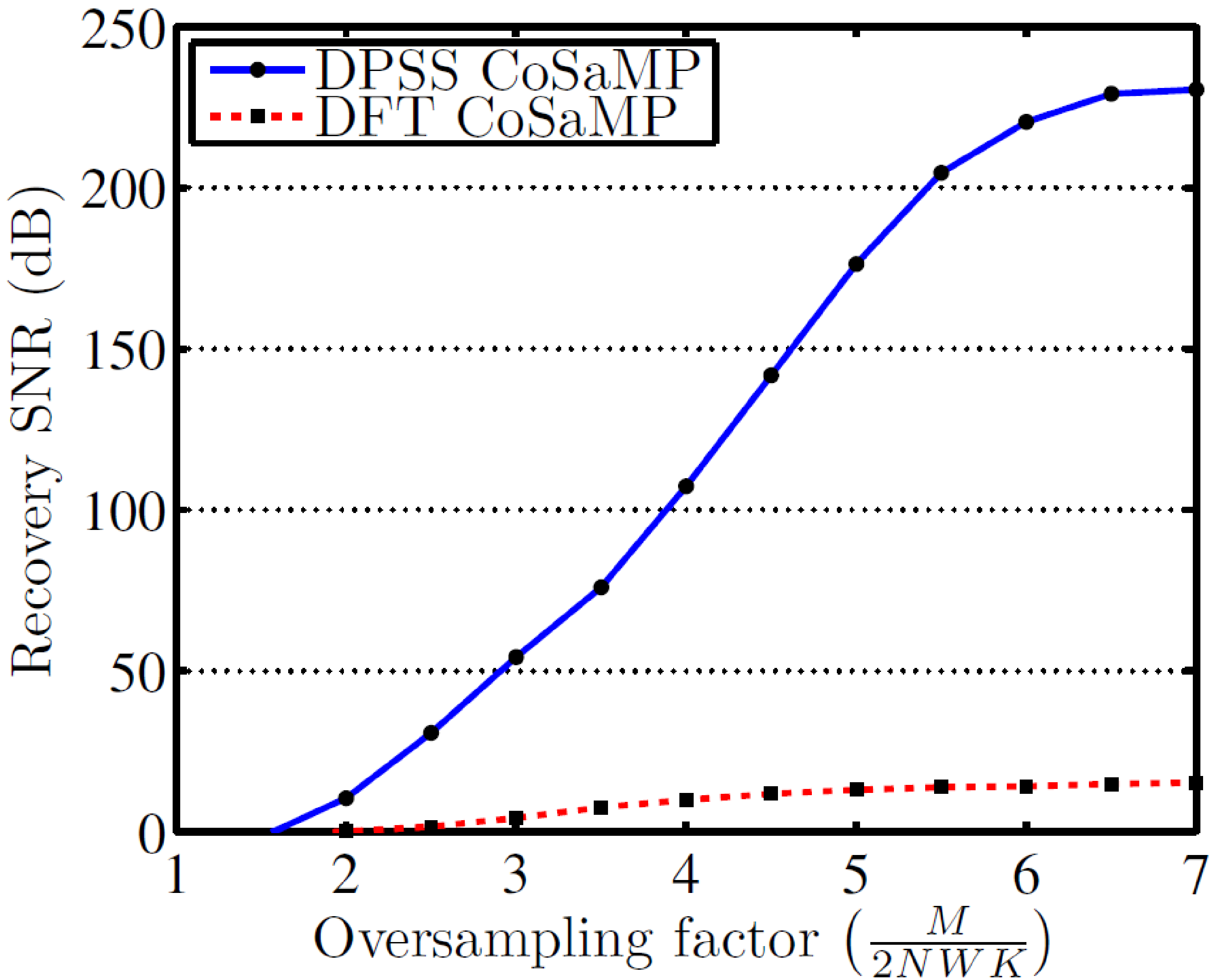
$$N = 4096$$

$$\frac{B}{B_{\text{nyq}}} = \frac{1}{256}$$

$$K = 5$$

[D and Wakin - 2011]

Empirical Results: DFT Comparison



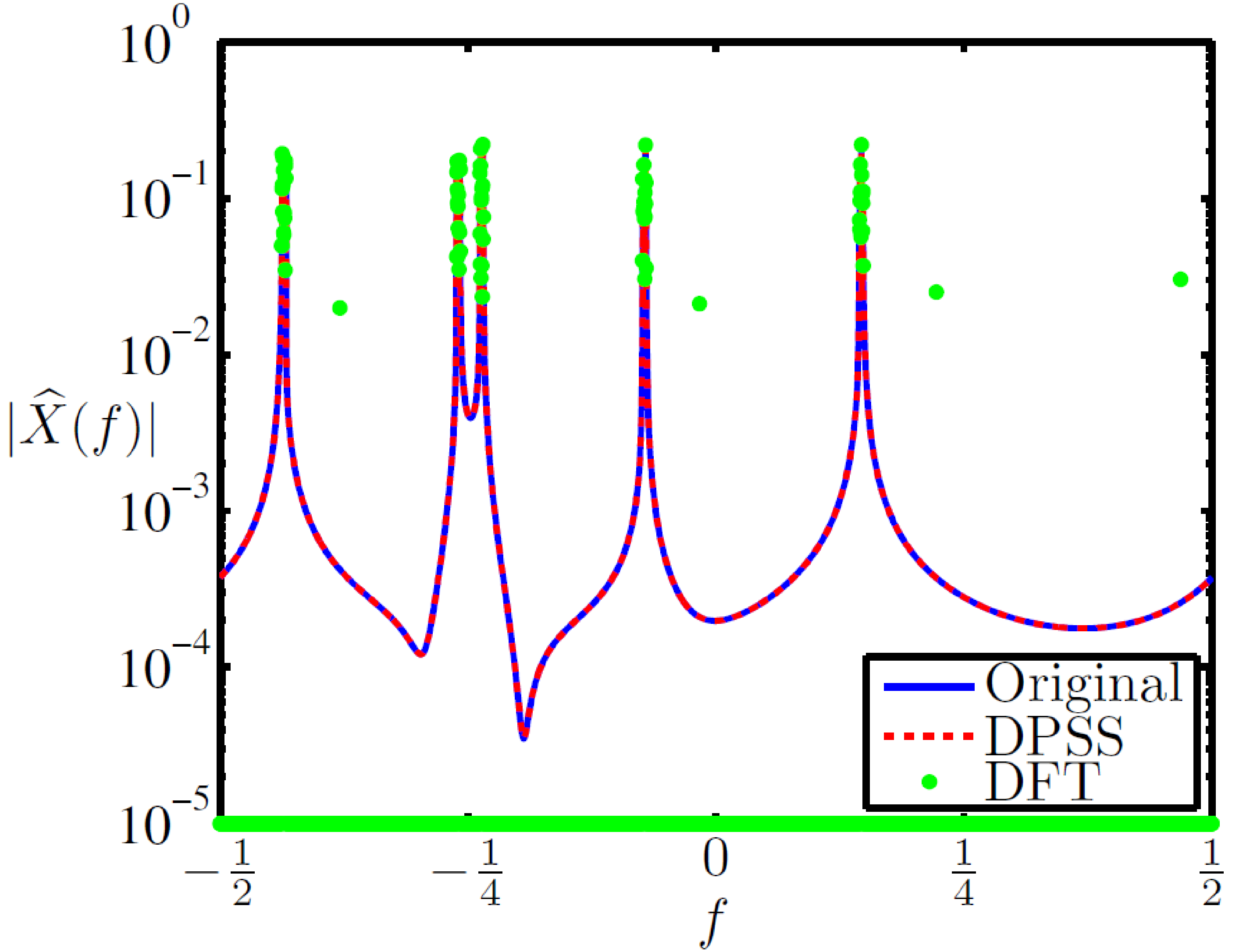
$$N = 4096$$

$$\frac{B}{B_{\mathrm{nyq}}} = \frac{1}{256}$$

$$K = 5$$

[D and Wakin - 2011]

Empirical Results: DFT Comparison



$$N = 4096$$

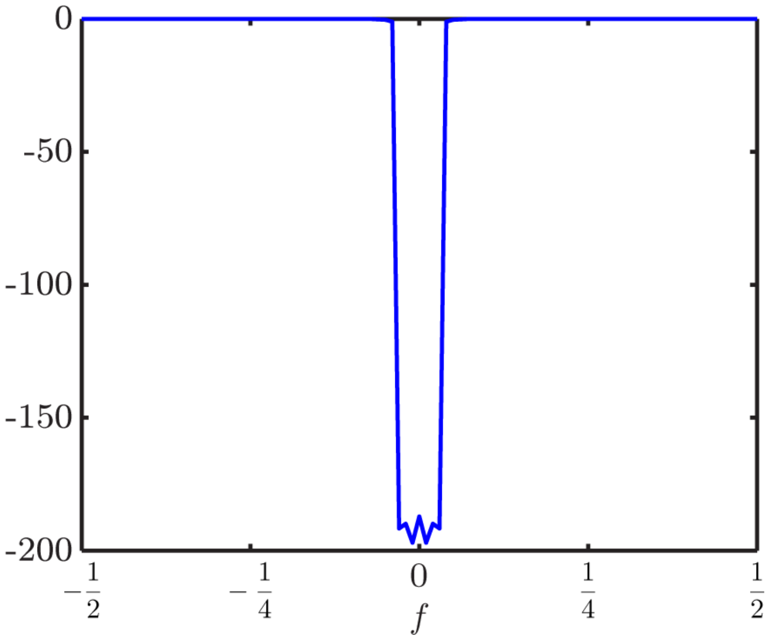
$$\frac{B}{B_{\text{nyq}}} = \frac{1}{256}$$

$$K = 5$$

[D and Wakin - 2011]

Interference Cancellation

DPSS's can be used to cancel bandlimited interferers *without reconstruction*.



$$P = I - \Phi\Psi_i(\Phi\Psi_i)^\dagger$$

Extremely useful in *compressive signal processing* applications.

Conclusions

- DPSS's can be used to efficiently represent *most* sampled multiband signals
  - far superior to DFT
- Two types of error: *approximation* + *reconstruction*
  - approximation: small for most signals
  - reconstruction: zero for DPSS-sparse vectors
  - delicate balance in practice, but there is a sweet spot
- This approach combines careful design of Ψ with more sophisticated sparse models
  - relevant in many contexts beyond ADCs