Compressive Sensing in Practice: Noise, Quantization, and Real-World Signals

Mark A. Davenport

Stanford University

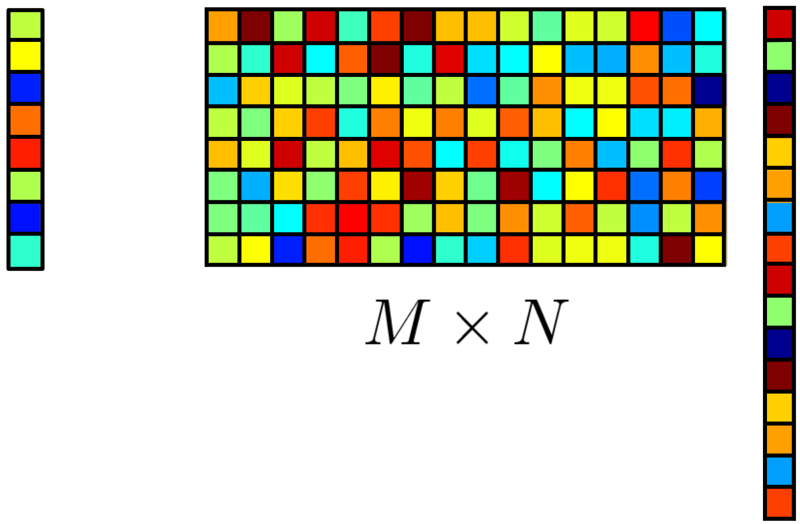
Department of Statistics

FoCM 2011 Workshop on Computational Harmonic Analysis, Image and Signal Processing

Supported by National Science Foundation

Compressive Sensing (CS)

$$y = \Phi x$$



Can we really acquire analog signals with “CS”?

An Apology for CS

Objection 1: CS is discrete, finite-dimensional

Objection 2: Impact of signal noise

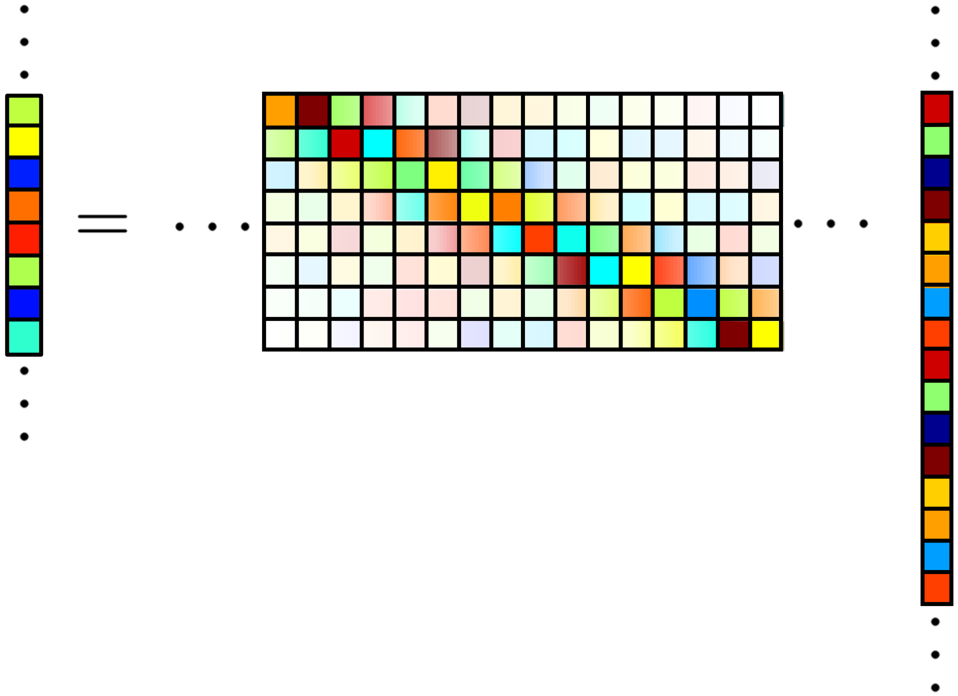
Objection 3: Impact of finite-bit quantization

Objection 4: Analog sparse representations

Objection 1

For any bandlimited signal $x(t)$,

$$
\begin{aligned}
y[m] &= \langle \phi_m(t), x(t) \rangle \\
&= \sum_{n=-\infty}^{\infty} x[n] \, \langle \phi_m(t), \mathrm{sinc}(t/T_s - n) \rangle
\end{aligned}
$$



Potential Obstacles

Objection 1: CS is discrete, finite-dimensional

Objection 2: Impact of signal noise

Objection 3: Impact of finite-bit quantization

Objection 4: Analog sparse representations

Recovery in noise

Suppose that we now observe

$$y = \Phi x + e$$

and that Φ satisfies the RIP, i.e., for all $\|x\|_0 \leq S$

$$(1-\delta)\|x\|_2^2 \leq \|\Phi x\|_2^2 \leq (1+\delta)\|x\|_2^2.$$

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

If e is white Gaussian noise with variance σ^2, then

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

White Signal Noise

What if our signal x is contaminated with noise?

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

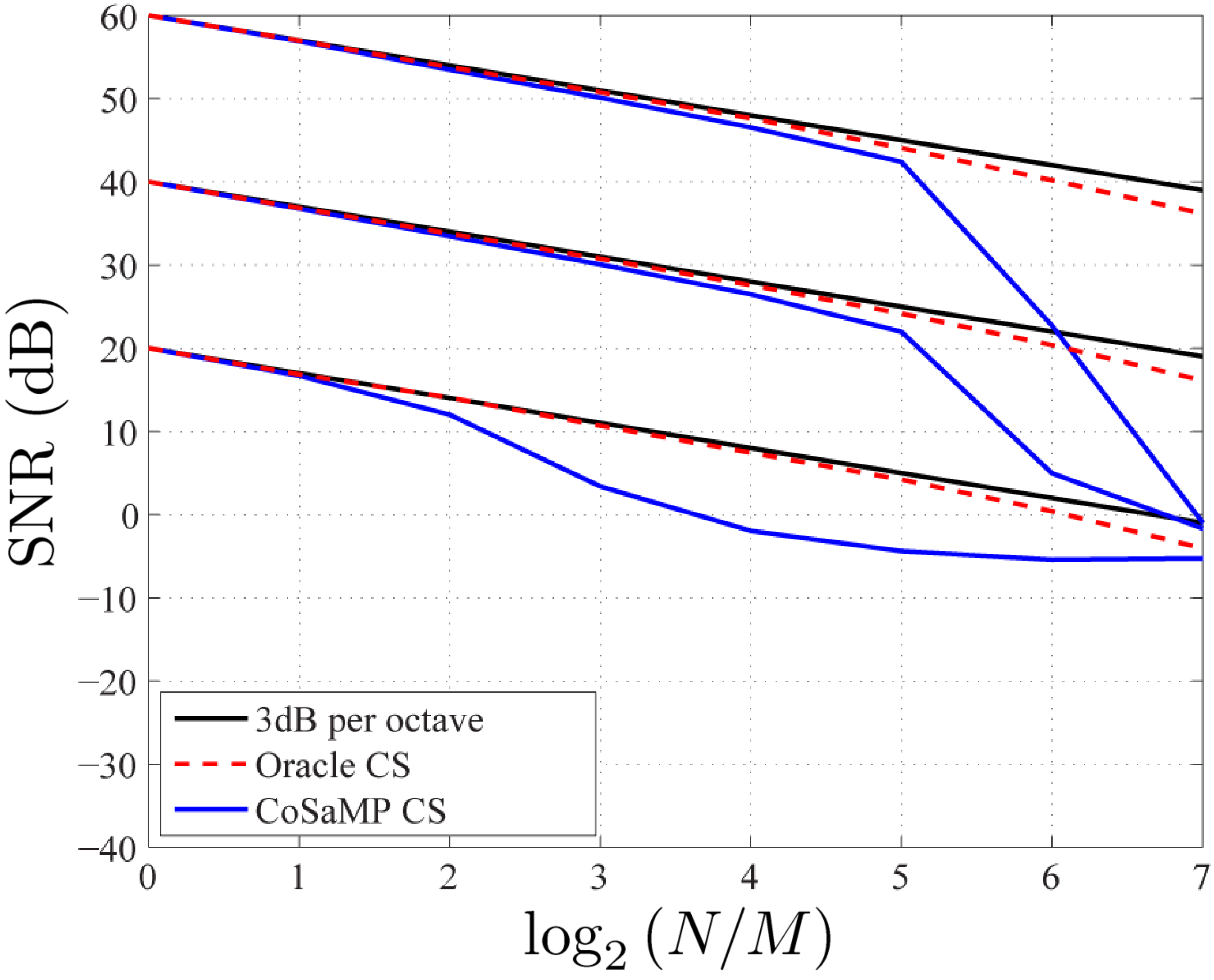
Suppose Φ satisfies the RIP and has orthogonal and equal-norm rows. If n is white noise with variance σ^2, then Φn is white noise with variance $\sigma^2 \frac{N}{M}$.

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

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

$\Rightarrow$ 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]

Potential Obstacles

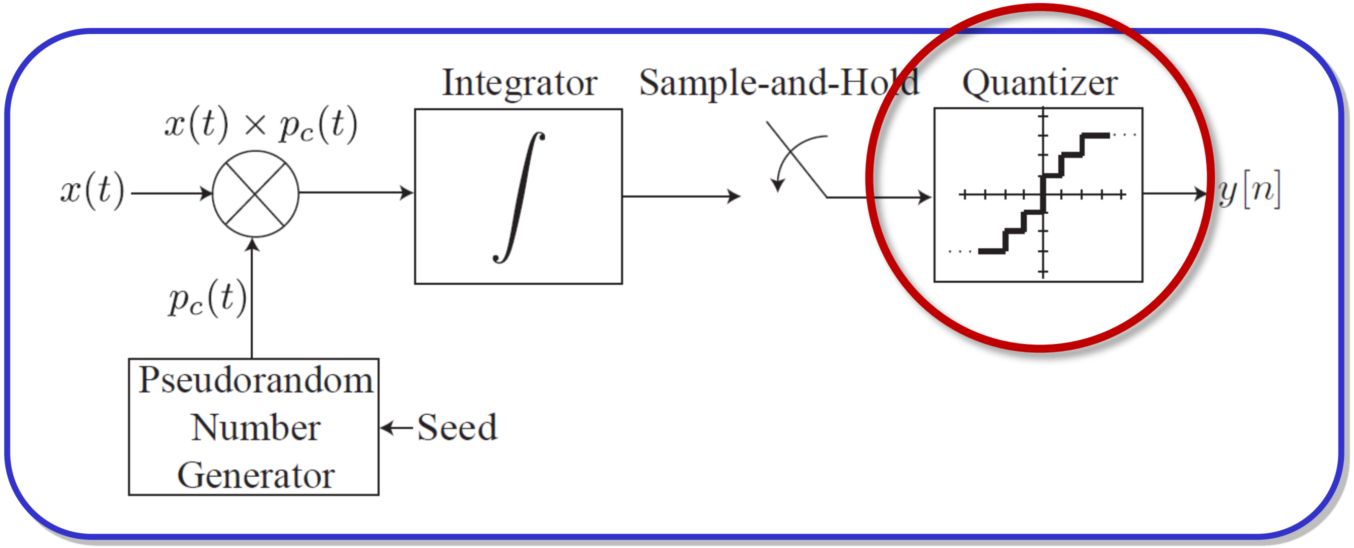
Objection 1: CS is discrete, finite-dimensional

Objection 2: Impact of signal noise

Objection 3: Impact of finite-bit quantization

Objection 4: Analog sparse representations

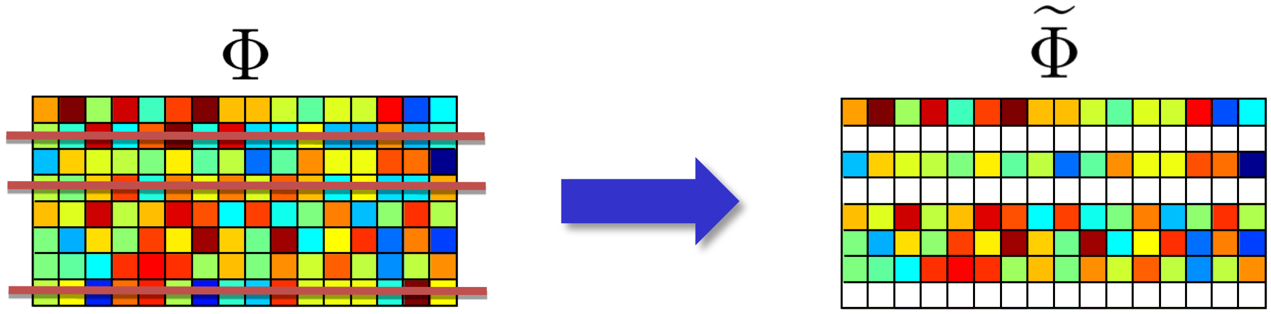
Signal Recovery with Quantization



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

Saturation Strategies

- **Rejection:** Ignore saturated measurements

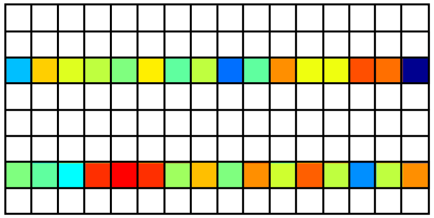


- **Consistency:** Retain saturated measurements. Use them only as inequality constraints on the recovered signal

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

Rejection and Democracy

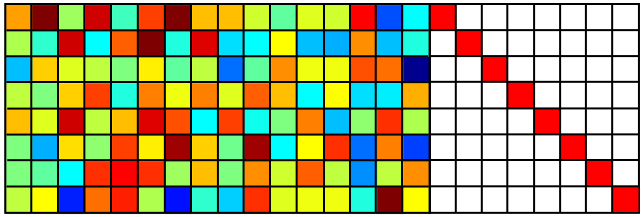
- The RIP is *not sufficient* for the rejection approach
- Example: $\Phi = I$
  - perfect isometry
  - *every* measurement must be kept
- We would like to be able to say that *any* submatrix of Φ with sufficiently many rows will still satisfy the RIP



- Strong, *adversarial* form of “democracy”

Sketch of Proof

- Step 1: Concatenate the identity to Φ



Theorem:

If Φ is a sub-Gaussian matrix with

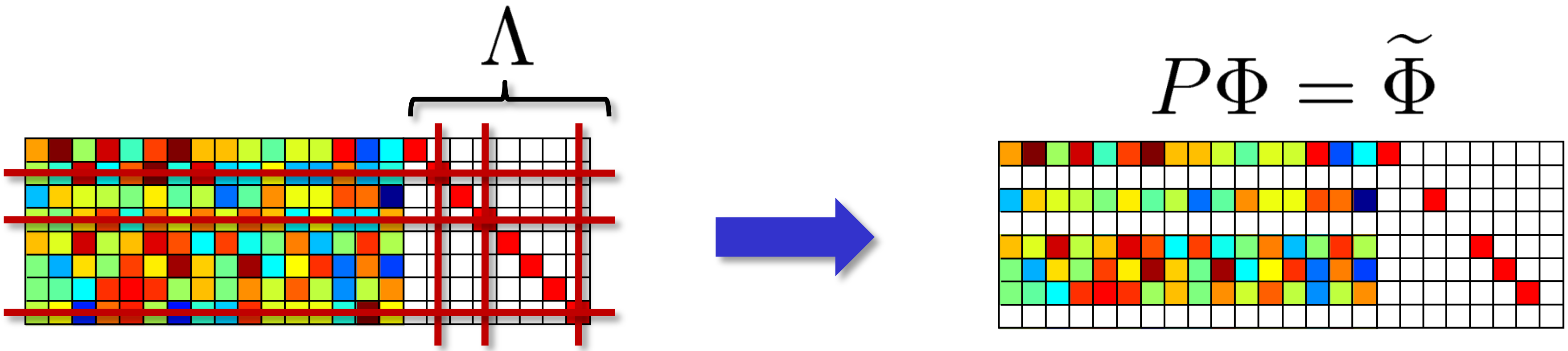
$$M = O\left(S \log\left(\frac{N}{S}\right)\right)$$

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

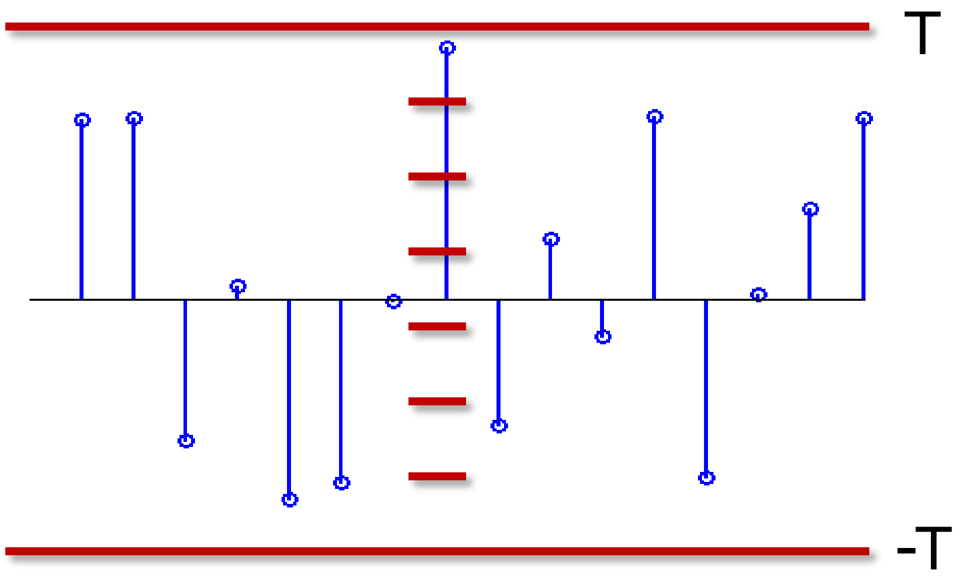
- Step 2: Combine with the "interference cancellation" lemma



- 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

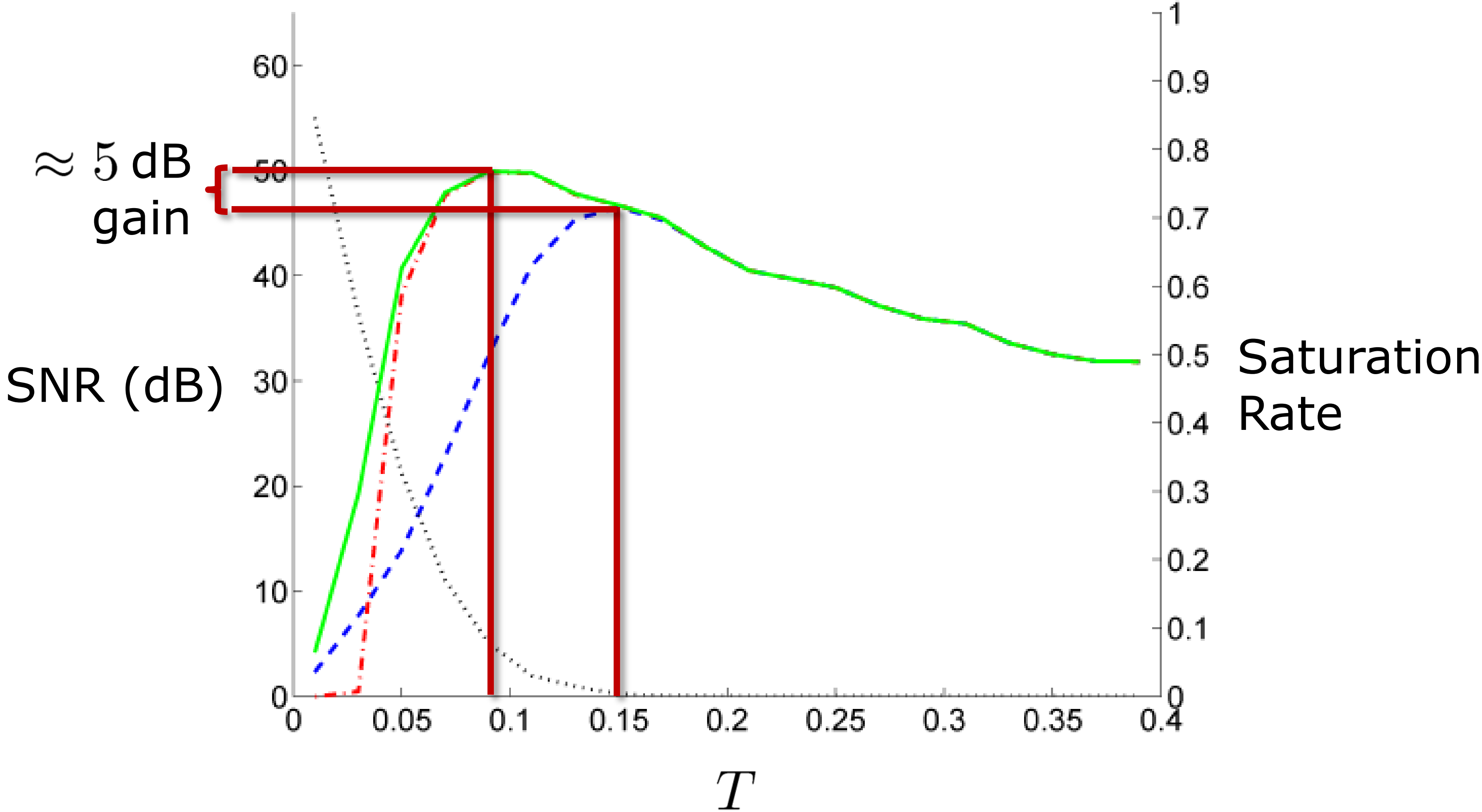
[D, Laska, Boufounos, and Baraniuk, 2009]

Rejection In Practice



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

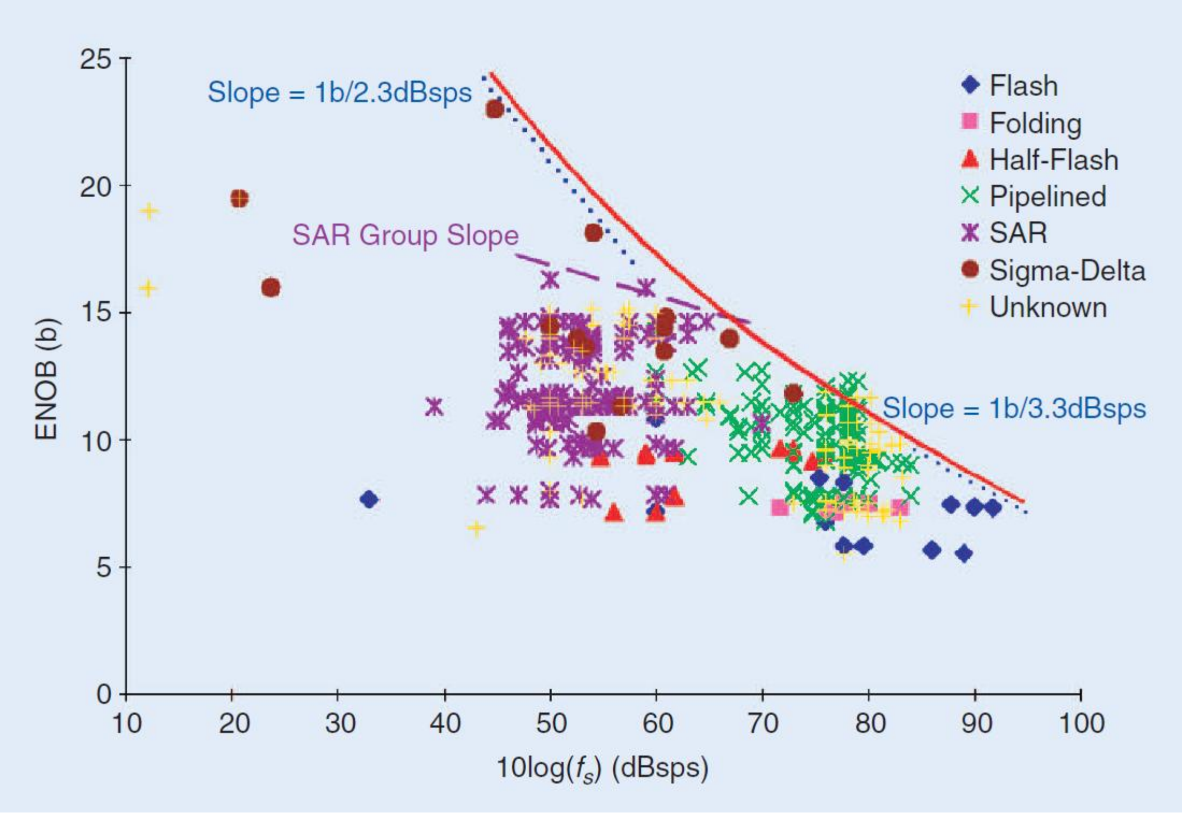
Benefits of Saturation



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

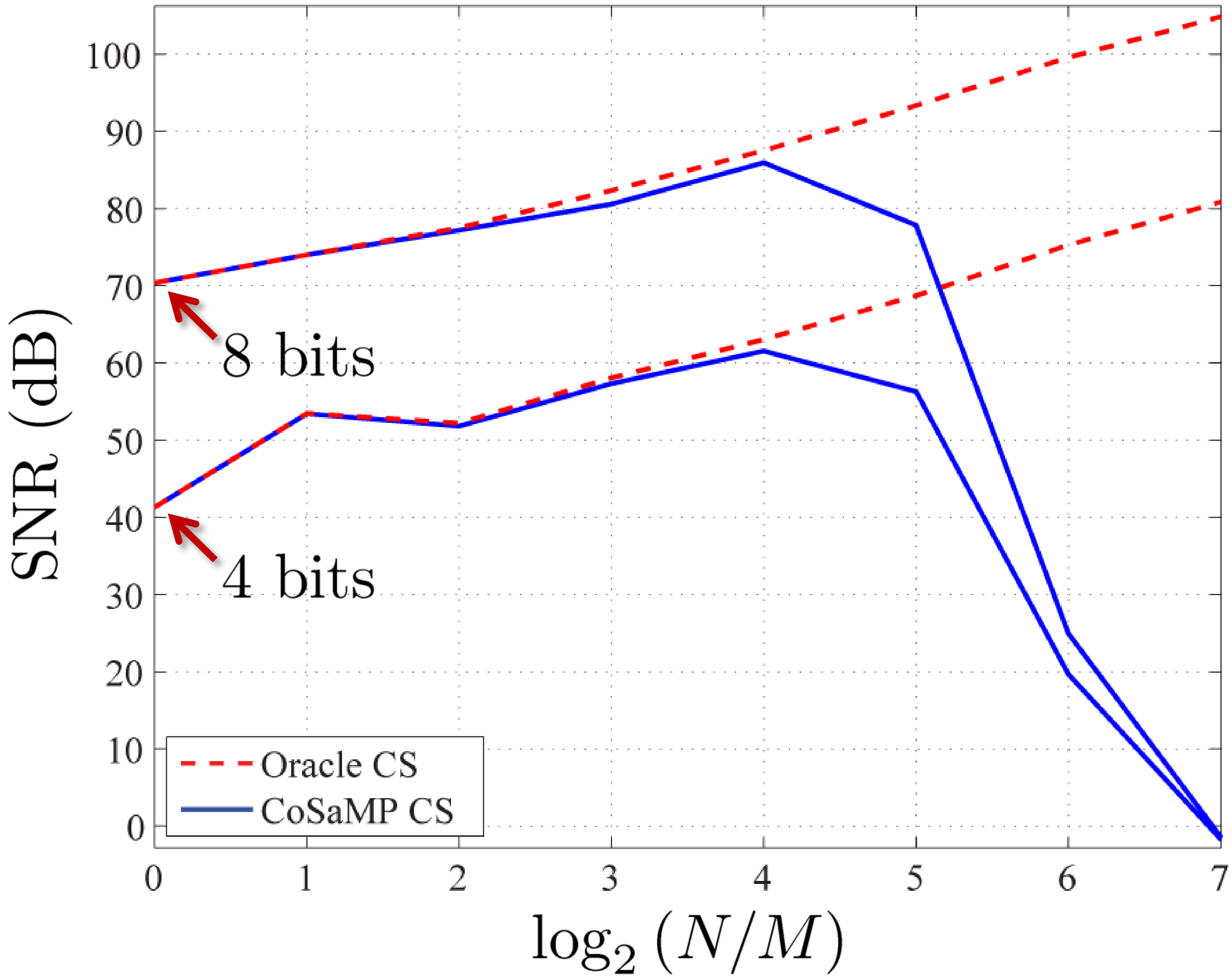
Potential for SNR Improvement?

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]

Potential Obstacles

Objection 1: CS is discrete, finite-dimensional

Objection 2: Impact of signal noise

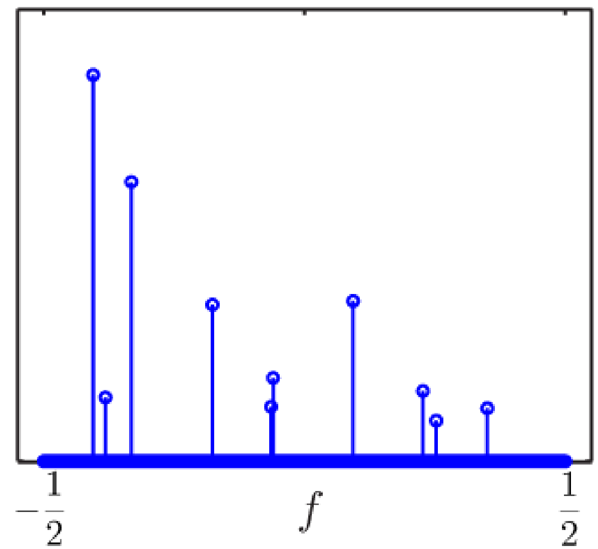
Objection 3: Impact of finite-bit quantization

Objection 4: Analog sparse representations

Candidate Analog Signal Models

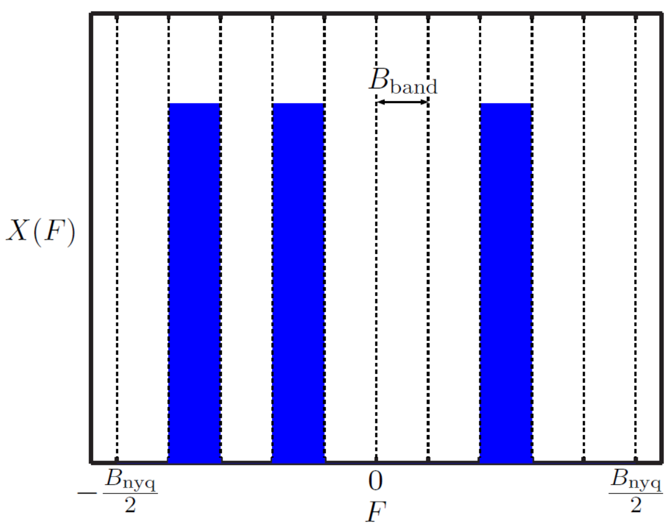
Multitone model:

- periodic signal
- DFT with S *tones*
- unknown *amplitude*



Multiband model:

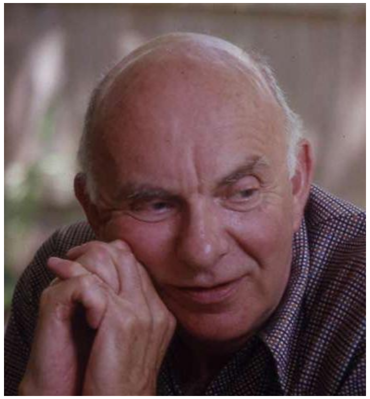
- aperiodic signal
- DTFT with K *bands* of bandwidth B_{band}
- unknown *spectra*



Discrete Prolate Spheroidal Sequences (DPSS's)

DPSS's (Slepian sequences)

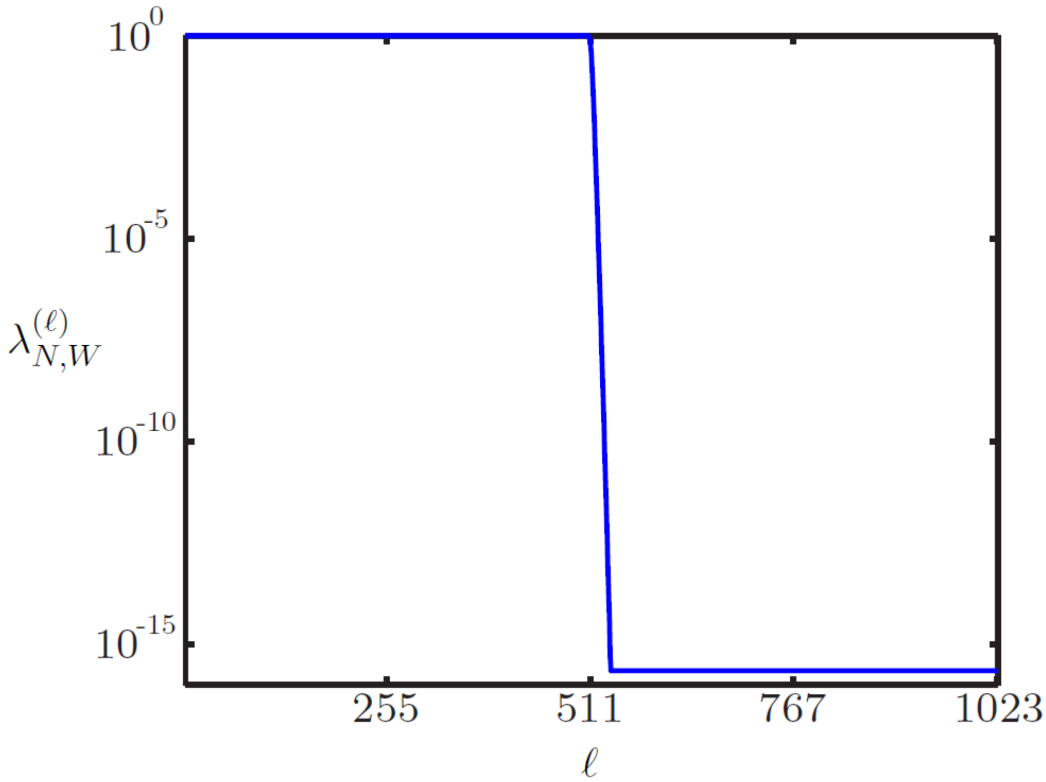
Given N and $W \leq \frac{1}{2}$, the DPSS's are a collection of N real-valued discrete-time sequences $s_{N,W}^{(0)}, s_{N,W}^{(1)}, \ldots, s_{N,W}^{(N-1)}$ such that for all ℓ



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

The DPSS's are perfectly bandlimited, but when $\lambda_{N,W}^{(\ell)} \approx 1$ they are highly concentrated in time.

DPSS Eigenvalue Concentration

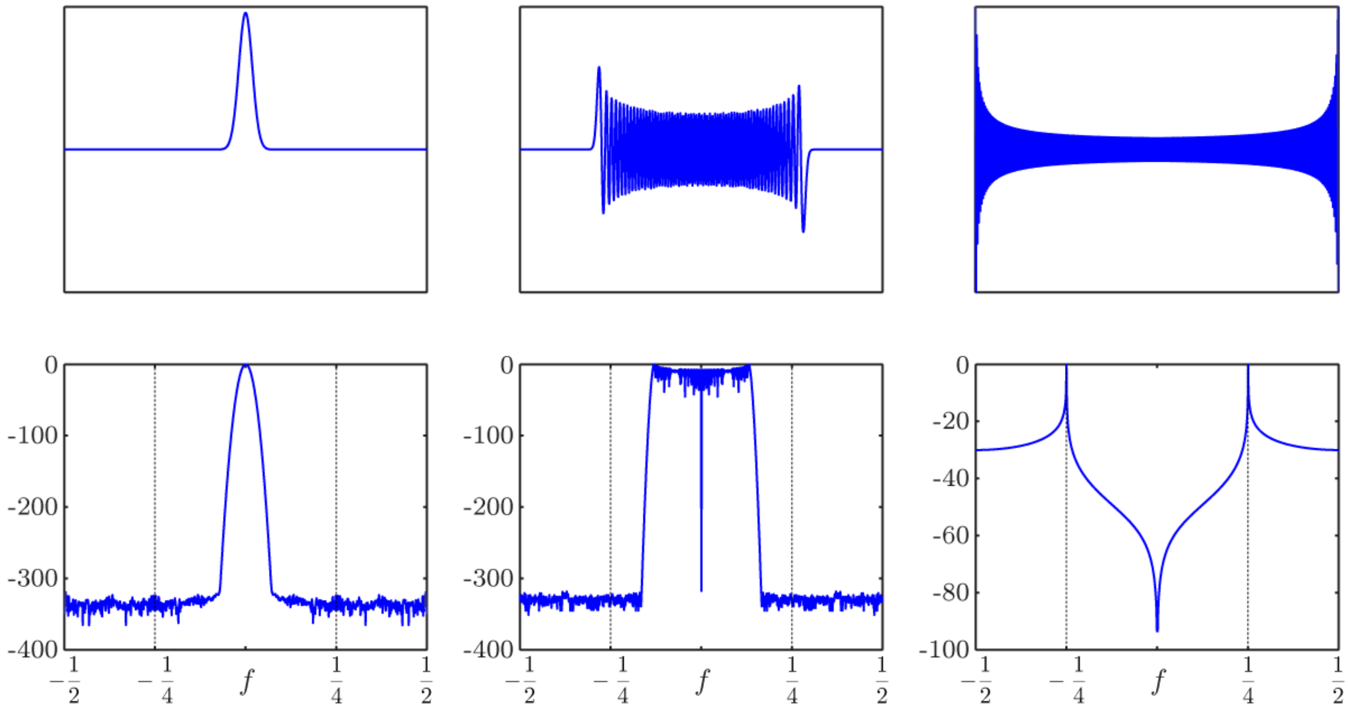


$$N = 1024$$
$$W = \frac{1}{4}$$

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

DPSS Examples

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



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

Why DPSS's?

Suppose that we wish to minimize

$$\frac{1}{2W} \cdot \int_{-W}^{W} \|e_f - P_Q e_f\|_2^2 \, df$$

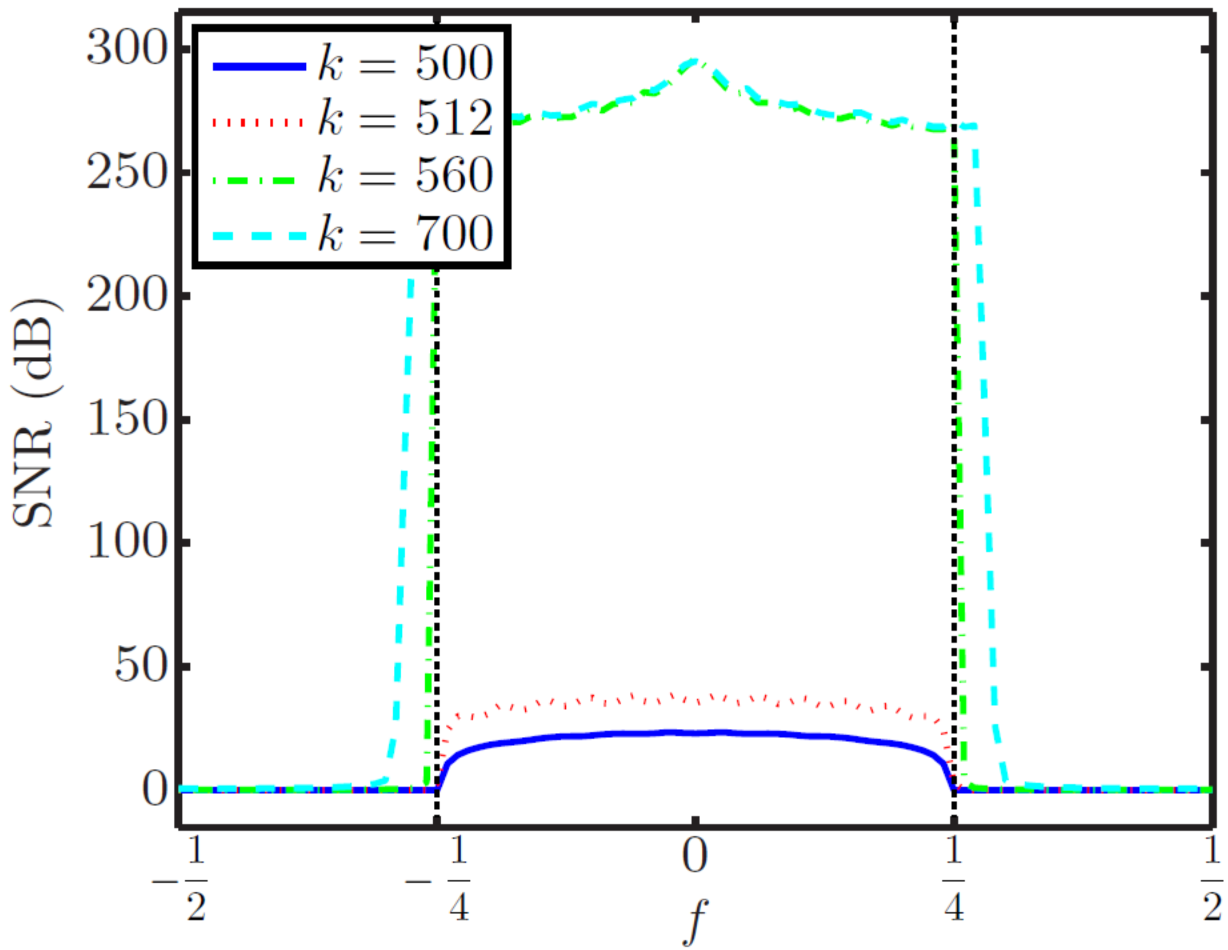
over Q where $e_f := \left[e^{j2\pi f 0}, e^{j2\pi f}, \ldots, e^{j2\pi f(N-1)}\right]^T.$

Optimal subspace of dimension k is the one spanned by the first k DPSS vectors.

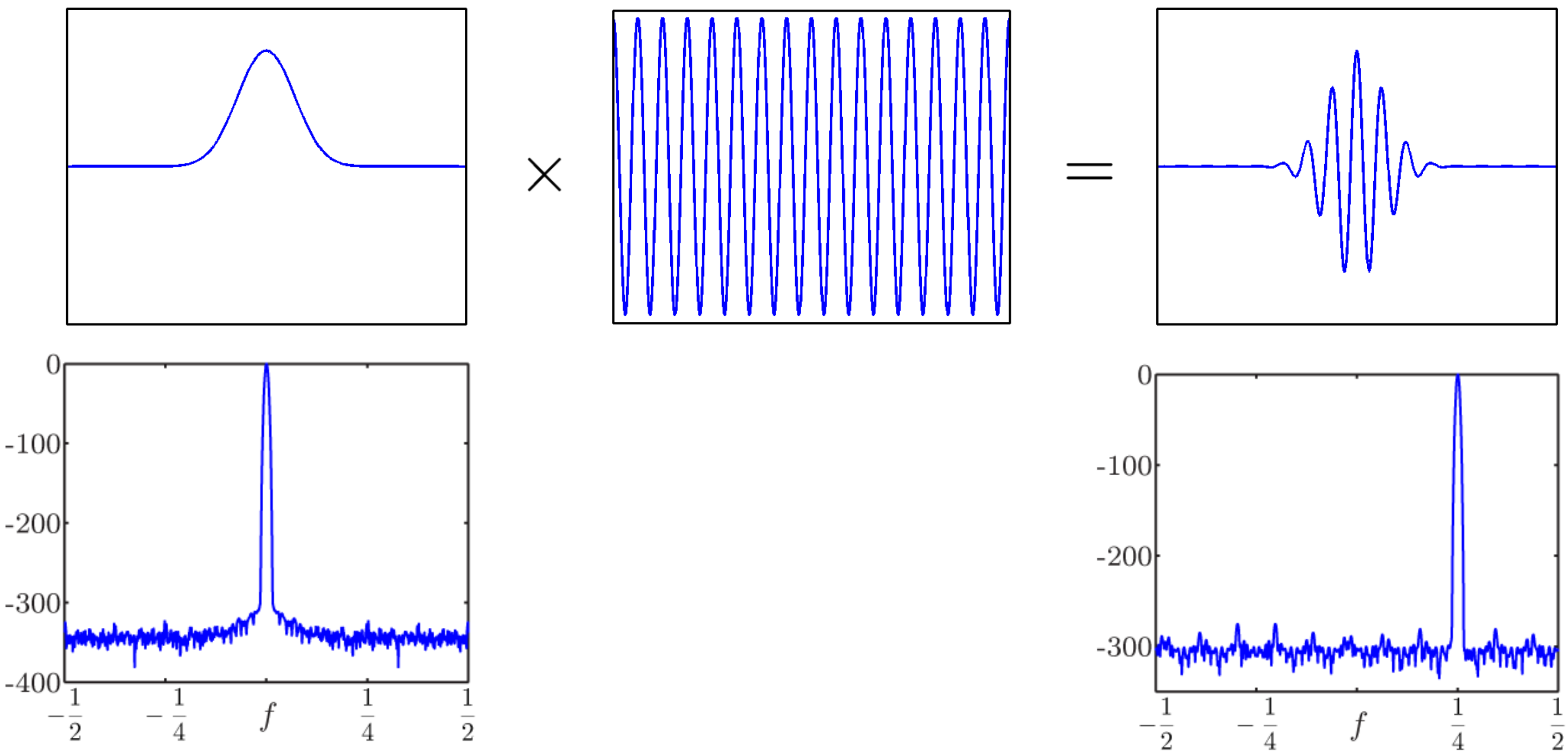
$$\frac{1}{2W} \cdot \int_{-W}^{W} \|e_f - P_Q e_f\|_2^2 \, df = \frac{1}{2W} \sum_{\ell=k}^{N-1} \lambda_{N,W}^{(\ell)}$$

Approximation Performance

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



DPSS's for Passband Signals



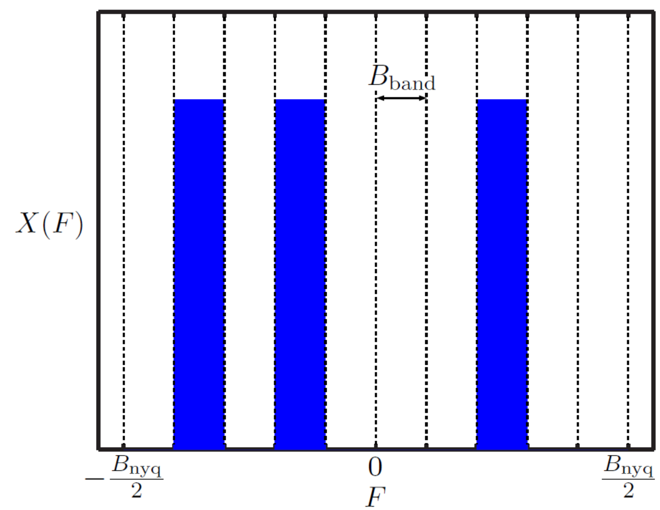
DPSS Dictionaries for CS

Construct dictionary Ψ as

$$\Psi = [\Psi_1, \Psi_2, \ldots, \Psi_J]$$

where Ψ_i is the matrix of the first k DPSS's modulated to $f_i = -\frac{1}{2} + \left(i + \frac{1}{2}\right)(B_{\text{band}}/B_{\text{nyq}})$.

Ψ sparsely and accurately represents *most* sampled multiband signals.



[D and Wakin, 2011]

DPSS Dictionaries and the RIP

Theorem:
Let $W = \frac{1}{2}(B_{\text{band}}/B_{\text{nyq}})$. 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 $\Rightarrow$ $S \approx KNB_{\text{band}}/B_{\text{nyq}}$

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

[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"

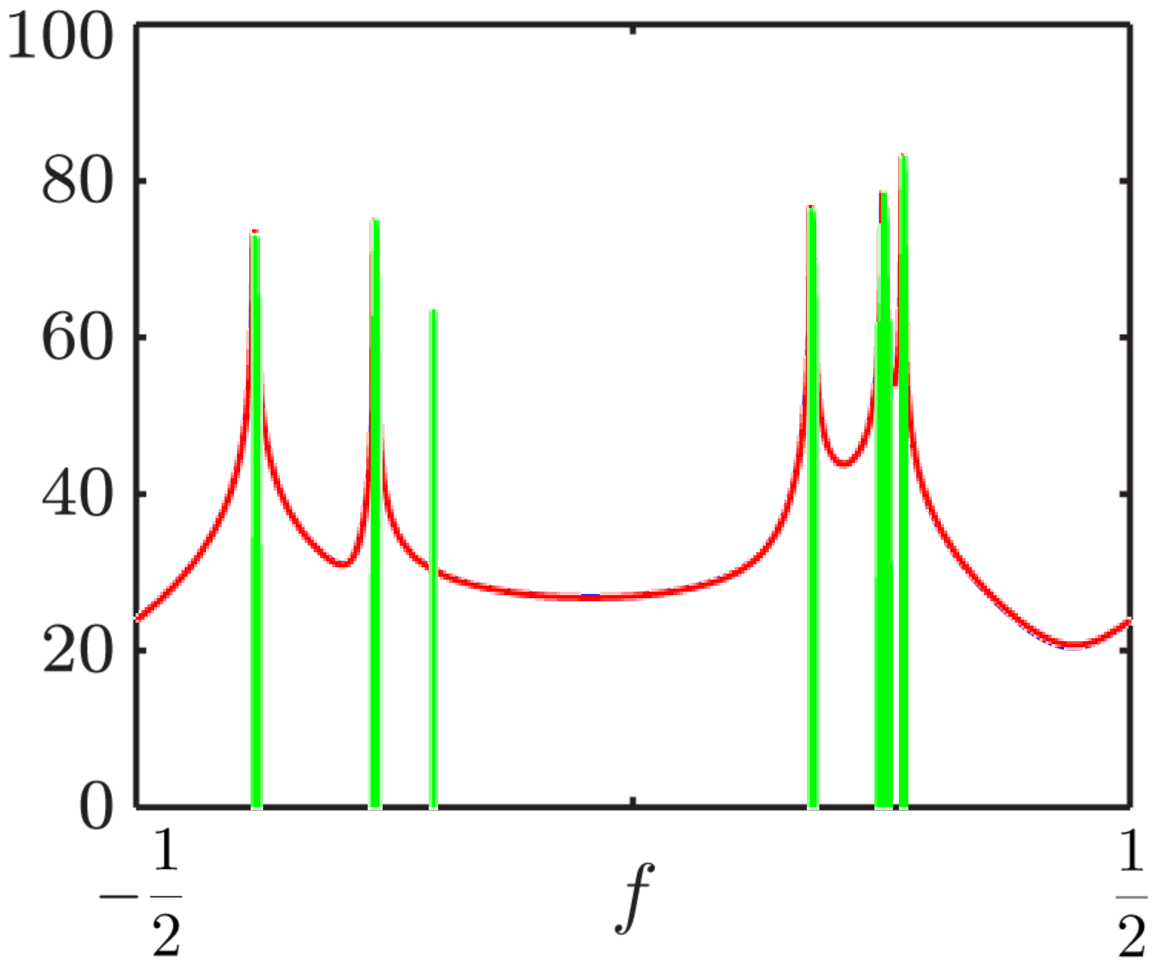
–Baraniuk et al. [2008, 2009, 2010]

–Blumensath and Davies [2009, 2011]

Recovery: DPSS vs DFT

$$N = 1024 \quad M = 128 \quad K = 5 \quad \frac{B_{\text{band}}}{B_{\text{nyq}}} = \frac{1}{512}$$

$$S \approx 45$$



$$\text{DPSS}: \quad \text{SNR} = 54\text{dB}$$

$$\text{DFT}: \quad \text{SNR} = 12\text{dB}$$

[D and Wakin, 2011]

Summary

- It is indeed possible to deal with analog signals using the traditional discrete CS formalism
- Noise can be an issue, but this is a fundamental limitation independent of the techniques used in CS
- Quantization noise can be less harmful than might be expected – CS allows for new design tradeoffs
- To give CS a fair chance we must both:
  - carefully design the sparsity basis
  - exploit any additional structure

References

- E.J. Candès and M.A. Davenport , **"How well can we estimate a sparse vector?"** *Preprint*, April 2011.
- M.A. Davenport, J.N. Laska, J.R. Treichler, and R.G. Baraniuk, **"The pros and cons of compressive sensing for wideband signal acquisition: Noise folding vs. dynamic range,"** *Preprint*, April 2011.
- J.N. Laska, P.T. Boufounos, M.A. Davenport, and R.G. Baraniuk, **"Democracy in action: Quantization, saturation, and compressive sensing,"** to appear in *Appl. Comput. Harmon. Anal.,* 2011.
- M.A. Davenport, J.N. Laska, P.T. Boufounos, and R.G. Baraniuk, **"A simple proof that random matrices are democratic,"** Rice University ECE Technical Report TREE 0906, November 2009.
- M.A. Davenport and M.B. Wakin, **"Reconstruction and cancellation of sampled multiband signals using discrete prolate spheroidal sequences,"** *SPARS 2011*.