The Pros and Cons of Compressive Sensing

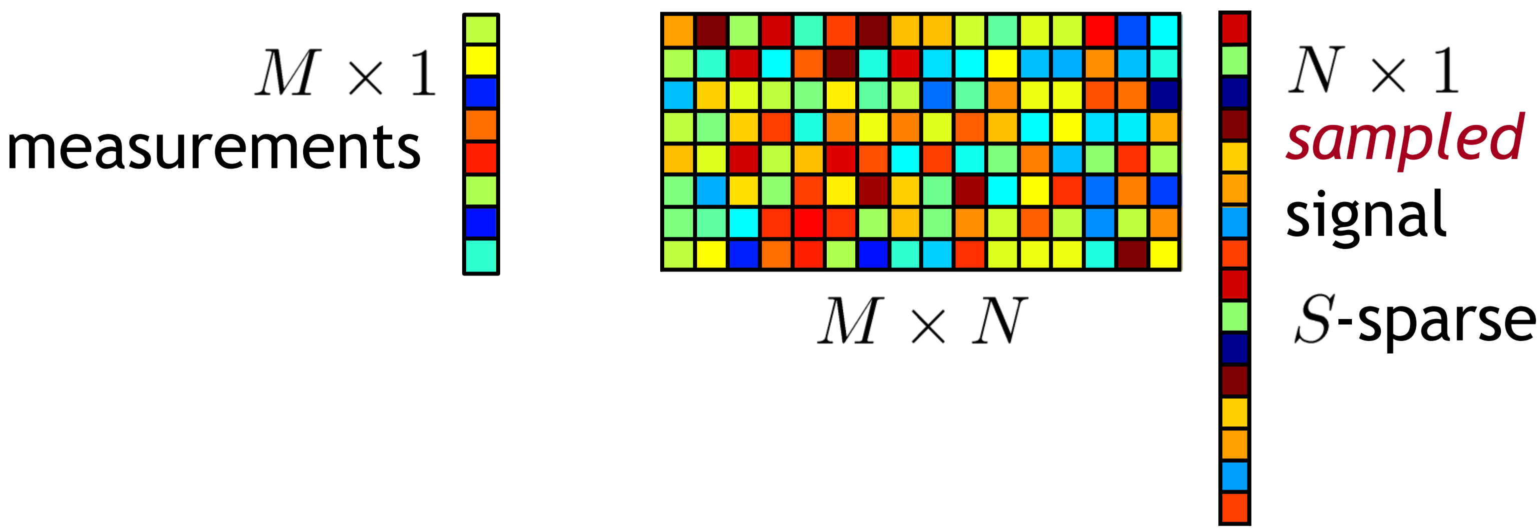
Mark A. Davenport

Stanford University
Department of Statistics

Compressive Sensing

Replace samples with general *linear measurements*

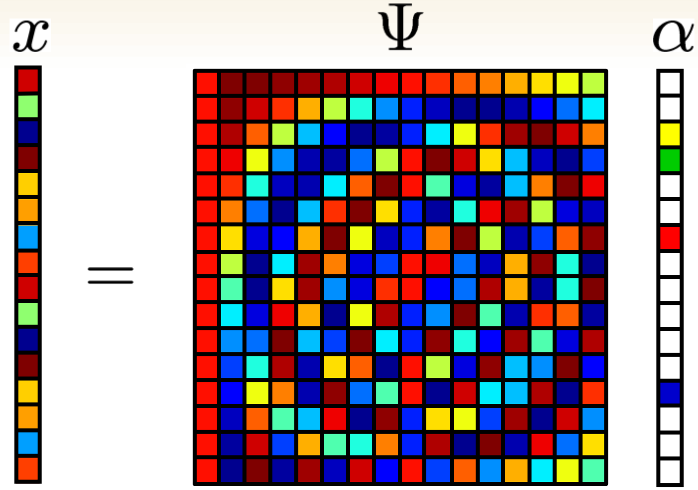
$$y = \Phi x$$



What are the pros and cons of “CS” in practice?

Review: Sparsity

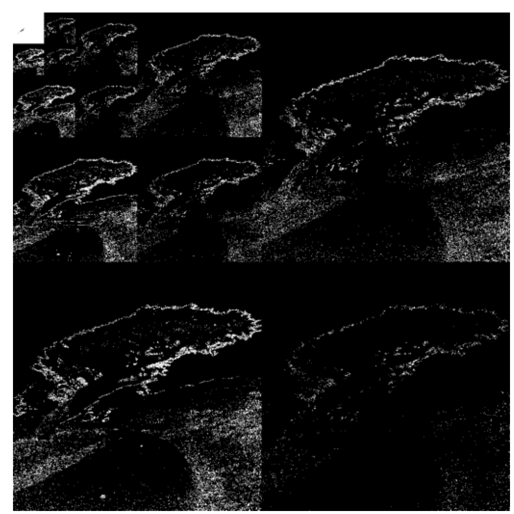
$$x = \sum_{j=1}^{N} \alpha_j \psi_j$$
$$= \Psi \alpha$$



S nonzero entries

$\|\alpha\|_0 = S$

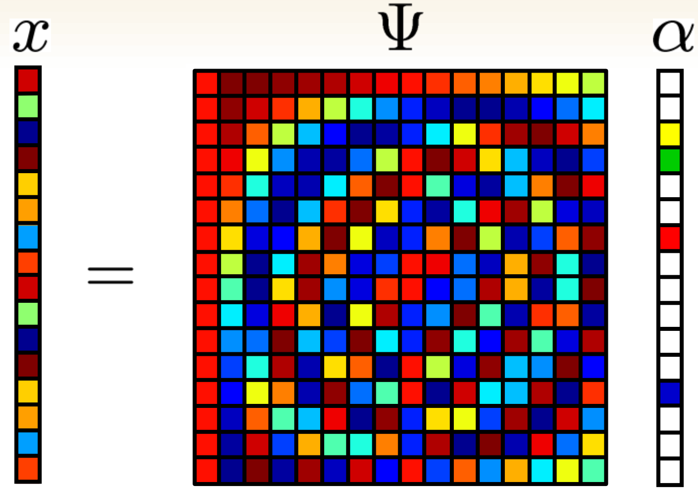
N pixels



$S \ll N$ large wavelet coefficients

Review: Sparsity

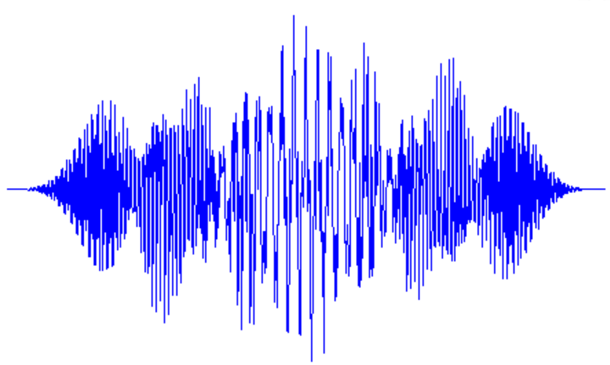
$$x = \sum_{j=1}^{N} \alpha_j \psi_j$$
$$= \Psi \alpha$$

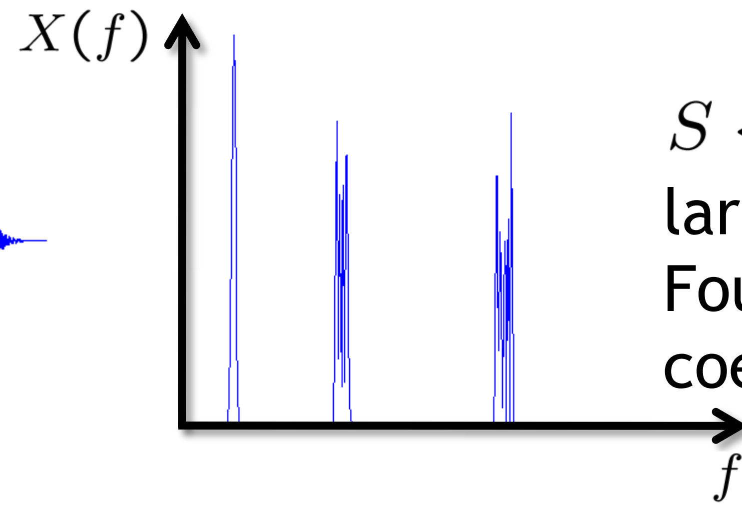


S nonzero entries

$\|\alpha\|_0 = S$

N samples

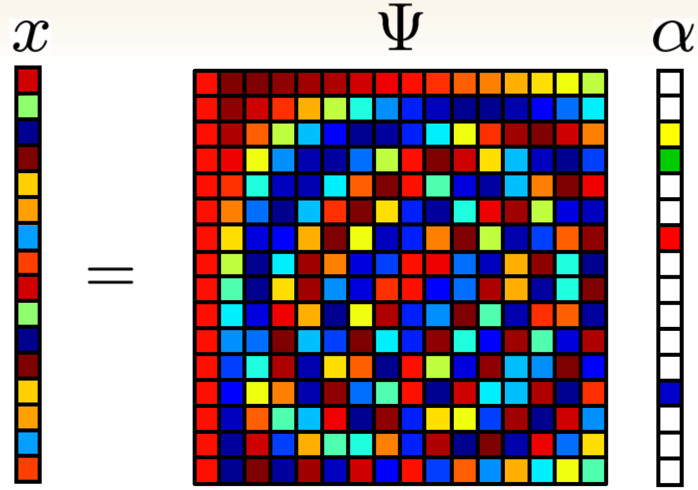




$S \ll N$ large Fourier coefficients

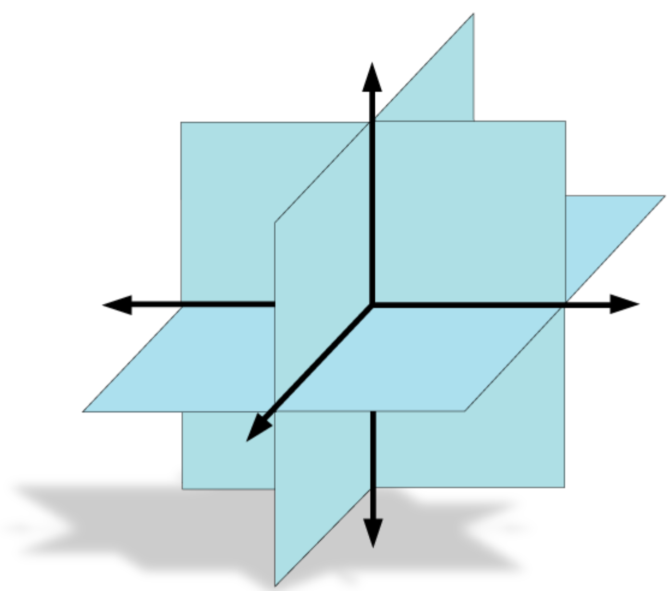
Review: Sparsity

$$x = \sum_{j=1}^{N} \alpha_j \psi_j$$
$$= \Psi \alpha$$



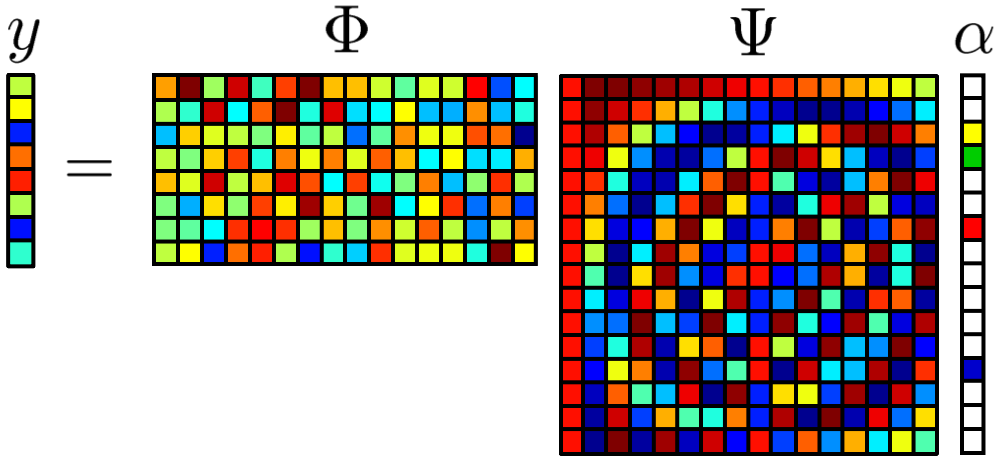
S nonzero entries

$$\|\alpha\|_0 = S$$



Core Theoretical Challenges

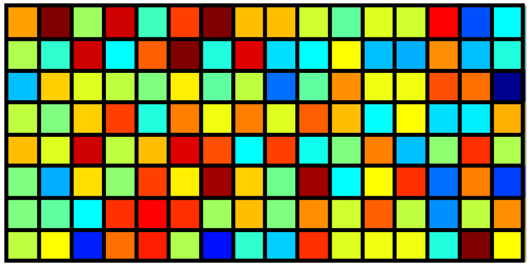
- How should we design the matrix Φ so that M is as small as possible?



- How can we recover $x = \Psi\alpha$ from the measurements y ?

Answers

- Choose a ***random matrix***
  - fill out the entries of Φ with i.i.d. samples from a sub-Gaussian distribution
  - project onto a “random subspace”



$$M = O(S \log(N/S)) \ll N$$

- Lots and lots of algorithms

Compressive Sensing: An Apology

Objection 1: CS is discrete, finite-dimensional

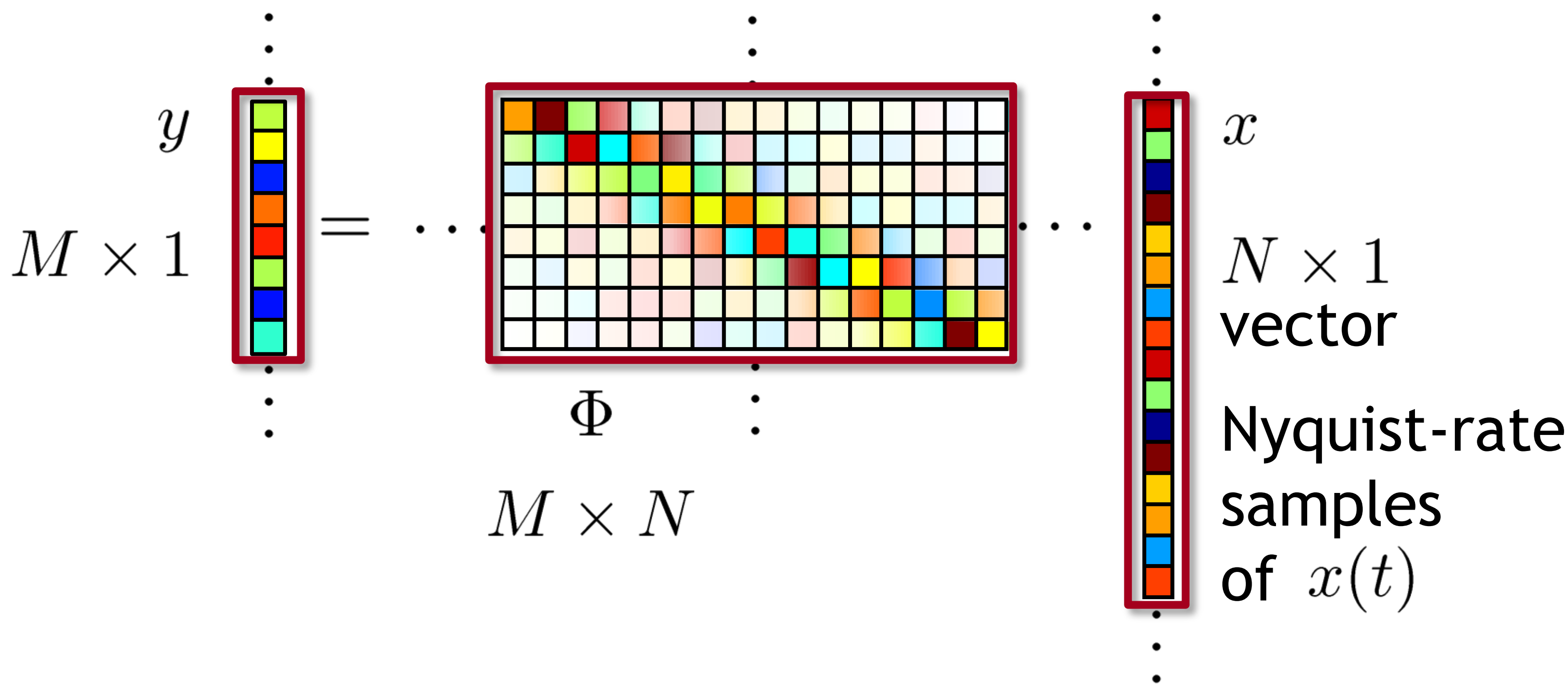
Objection 2: Impact of noise

Objection 3: Impact of quantization

Analog Sensing *is* Matrix Multiplication

If $x(t)$ is bandlimited,

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



Compressive Sensing: An Apology

Objection 1: CS is discrete, finite-dimensional

Objection 2: Impact of noise

Objection 3: Impact of quantization

Recovery from Noisy Measurements

Given $y = \Phi x + e$ or $y = \Phi(x + n)$,
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 $2S$.

$$\boxed{(1-\delta)\|x\|_2^2 \le \|\Phi x\|_2^2 \le (1+\delta)\|x\|_2^2 \qquad \|x\|_0 \le 2S}$$

Typical (worst-case) guarantee

$$\boxed{\|\widehat{x} - x\|_2^2 \le C\|e\|_2^2}$$

Even if $\Lambda = \mathrm{supp}(x)$ is provided by an oracle, the error can still be as large as $\|\widehat{x} - x\|_2^2 = \|e\|_2^2/(1-\delta)$.

Stable Signal Recovery: Part II

Suppose now that Φ satisfies

$$\boxed{A(1-\delta)\|x\|_2^2 \leq \|\Phi x\|_2^2 \leq A(1+\delta)\|x\|_2^2 \qquad \|x\|_0 \leq 2S}$$

In this case our guarantee becomes

$$\boxed{\|\widehat{x} - x\|_2^2 \leq \frac{C}{A}\|e\|_2^2}$$

Unit-norm rows $\Rightarrow$ $\|\widehat{x} - x\|_2^2 \leq C\frac{N}{M}\|e\|_2^2$

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

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

  - if e is Gaussian, then for ℓ_1 -minimization

$$\mathbb{E}\left(\|\widehat{x} - x\|_2^2\right) \leq \frac{C'}{A} S\sigma^2 \log N$$

White Signal Noise

What if our signal x is contaminated with noise?

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

Suppose Φ has orthogonal rows with norm equal to $\sqrt{B}$.
If n is white noise with variance σ^2, then Φn is white noise with variance $B\sigma^2$.

$$\boxed{\mathbb{E}\left[\|\widehat{x} - x\|_2^2\right] \leq C' \frac{B}{A} S \sigma^2 \log N}$$

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

3dB loss per octave of subsampling

White Signal Noise

What if our signal x is contaminated with noise?

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

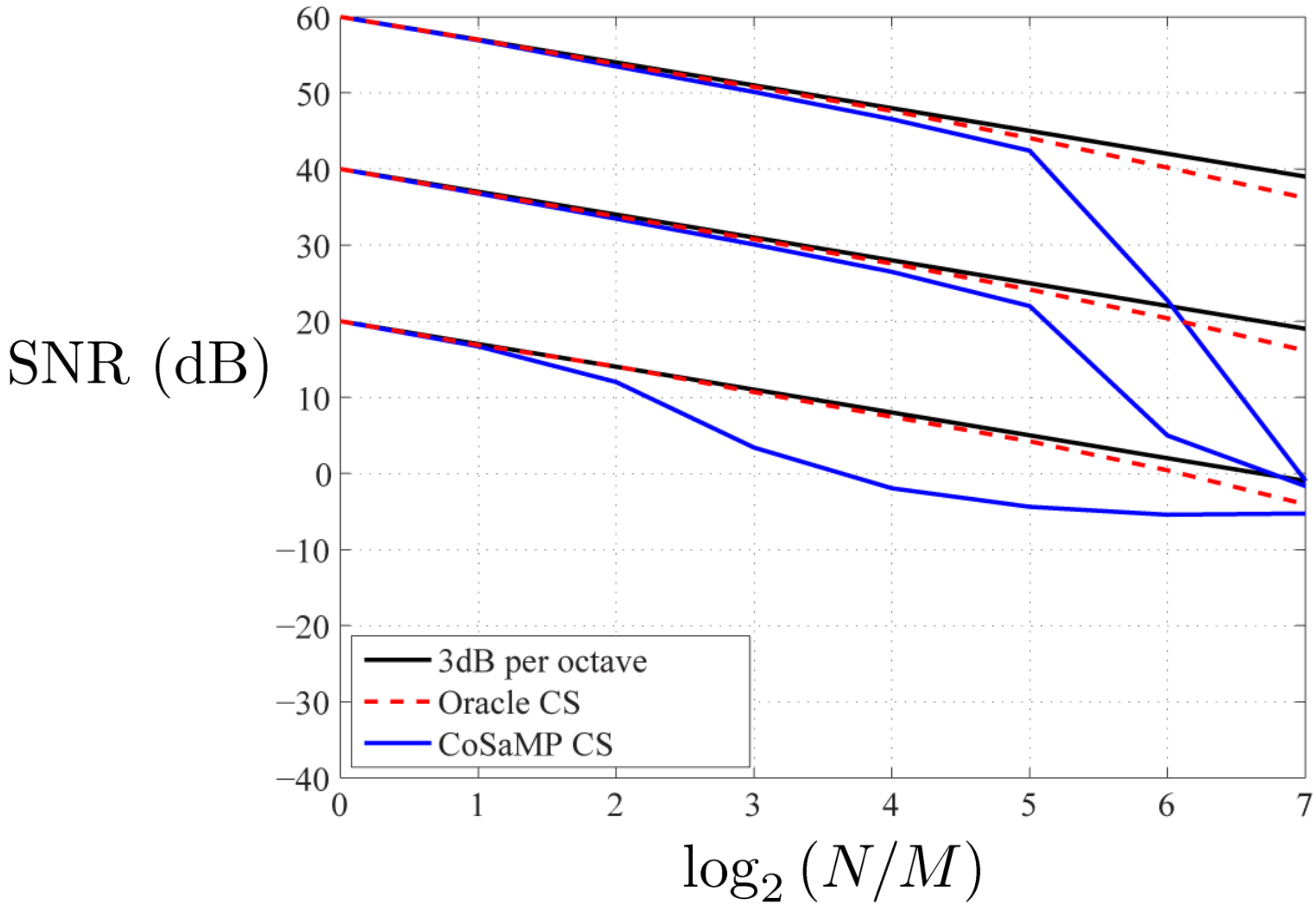
Suppose Φ has orthogonal rows with norm equal to $\sqrt{B}$.
If n is white noise with variance σ^2, then Φn is white noise with variance $B\sigma^2$.

$$\boxed{\mathbb{E}\left[\|\widehat{x} - x\|_2^2\right] \leq C' \frac{N}{M} S \sigma^2 \log N}$$

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

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

$$\mathbb{E}\left[\|\widehat{x}-x\|_2^2\right] \approx \frac{S}{M} S\sigma^2 \ll C' \frac{N}{M} S\sigma^2 \log N$$

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

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

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

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

$$\Phi = U\Sigma V^* \quad y' = \Sigma^{-1}U^* y = V^* x + V^* n \quad \|V^*\|_F^2 = M$$

See also: Raskutti, Wainwright, and Yu (2009)
Ye and Zhang (2010)

[Candès and D - 2011]

Proof Recipe

Ingredients [Makes $\sigma^2 = 1$ servings]

- Lemma 1: Suppose $\mathcal{X}$ is a set of S-sparse points such that
  $\|x_i - x_j\|_2^2 \geq 8R^*_{\mathrm{mm}}(\Phi)$ for all $x_i, x_j \in \mathcal{X}$.
  Then $\frac{1}{2}\log|\mathcal{X}| - 1 \leq \frac{1}{2|\mathcal{X}|^2}\sum_{i,j}\|\Phi x_i - \Phi x_j\|_2^2$.

- Lemma 2: There exists a set $\mathcal{X}$ of S-sparse points such that
  - $|\mathcal{X}| = (N/S)^{S/4}$
  - $\|x_i - x_j\|_2 \geq \frac{1}{2}$ for all $x_i, x_j \in \mathcal{X}$
  - $\|\frac{1}{|\mathcal{X}|}\sum_i x_i x_i^* - \frac{1}{N}I\| \leq \frac{\beta}{N}$ for some $\beta > 0$

Instructions

Combine ingredients and add a dash of linear algebra.

Proof Outline

$$\mu = \frac{1}{|\mathcal{X}|}\sum_i x_i \qquad Q = \frac{1}{|\mathcal{X}|}\sum_i x_i x_i^*$$

$$\begin{aligned}
\frac{S}{4}\log(N/S) - 2 &\leq \frac{1}{|\mathcal{X}|^2}\sum_{i,j} \|\Phi x_i - \Phi x_j\|_2^2 \\
&= \mathrm{Tr}\left(\Phi^*\Phi\left(\frac{1}{|\mathcal{X}|^2}\sum_{i,j}(x_i - x_j)(x_i - x_j)^*\right)\right) \\
&= \mathrm{Tr}\left(\Phi^*\Phi\left(2(Q - \mu\mu^*)\right)\right) \\
&\leq 2\mathrm{Tr}\left(\Phi^*\Phi Q\right) \\
&\leq 2\mathrm{Tr}\left(\Phi^*\Phi\right)\|Q\| \\
&\leq 2\|\Phi\|_F^2 \cdot 16 R_{\mathrm{mm}}^*(\Phi)(1+\beta)
\end{aligned}$$

$$\Rightarrow \quad R_{\mathrm{mm}}^*(\Phi) \geq \frac{S\log(N/S)}{128(1+\beta)\|\Phi\|_F^2}$$

Recall: Lemma 2

Lemma 2: There exists a set $\mathcal{X}$ of S-sparse points such that

- $|\mathcal{X}| = (N/S)^{S/4}$
- $\|x_i - x_j\|_2 \geq \frac{1}{2}$ for all $x_i, x_j \in \mathcal{X}$
- $\|\frac{1}{|\mathcal{X}|}\sum_i x_i x_i^* - \frac{1}{N} I\| \leq \frac{\beta}{N}$ for some $\beta > 0$

Strategy

Construct $\mathcal{X}$ by sampling (with replacement) from

$$\mathcal{U} = \left\{ x \in \{0, \sqrt{1/S}, -\sqrt{1/S}\}^N : \|x\|_0 \leq S \right\}$$

Repeat for $|\mathcal{X}| = (N/S)^{S/4}$ iterations.

With probability > 0, the remaining properties are satisfied.

Key: *Matrix Bernstein Inequality* [Ahlswede and Winter, 2002]

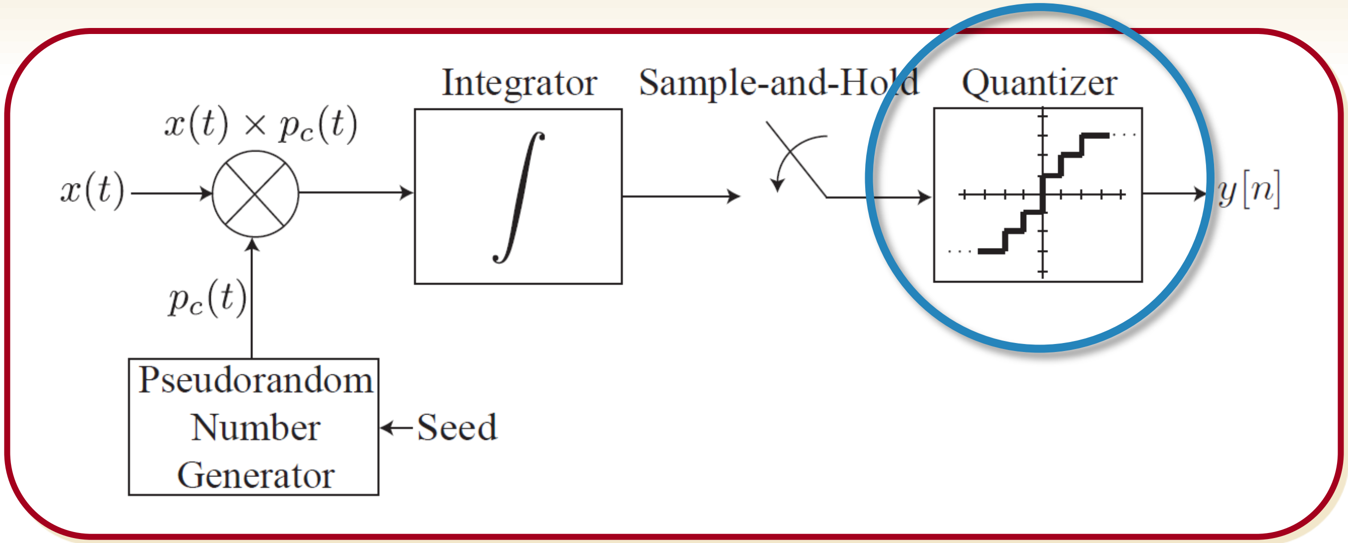
Compressive Sensing: An Apology

Objection 1: CS is discrete, finite-dimensional

Objection 2: Impact of signal noise

Objection 3: Impact of quantization

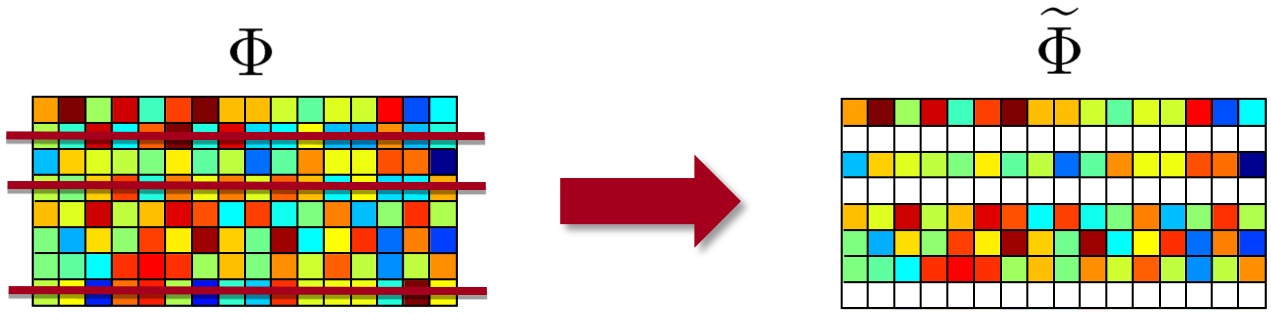
Signal Recovery with Quantization



- Finite-range quantization leads to *saturation*, i.e., *unbounded errors* on the largest measurements
- Quantization noise changes as we change the sampling rate

Saturation Strategies

- **Rejection:** Ignore saturated measurements

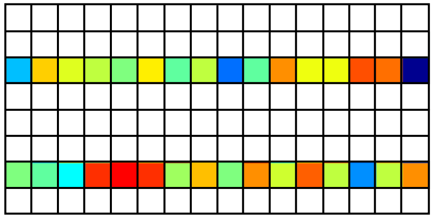


- **Consistency:** Retain saturated measurements.
  Use them only as inequality constraints on the recovered signal

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

Rejection and Democracy

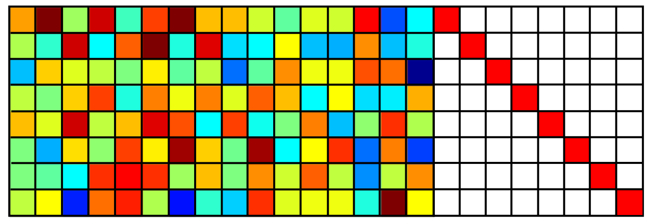
- The RIP is *not sufficient* for the rejection approach
- Example: $\Phi = I$
  - perfect isometry
  - *every* measurement must be kept
- We would like to be able to say that *any* submatrix of Φ with sufficiently many rows will still satisfy the RIP



- Strong, *adversarial* form of "democracy"

Sketch of Proof

- Step 1: Concatenate the identity to Φ



Theorem:
If Φ is a sub-Gaussian matrix with

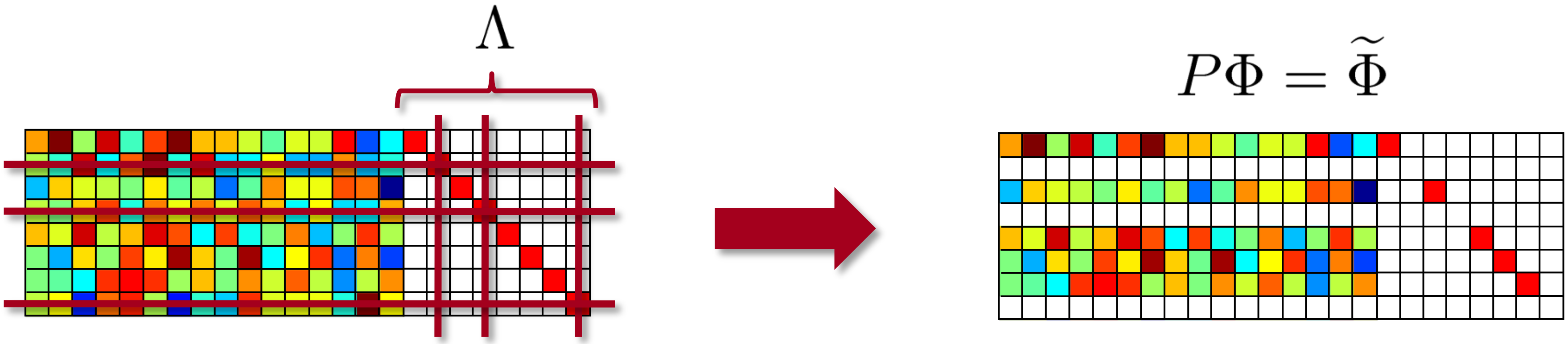
$$M = O\left(S \log\left(\frac{N+M}{S}\right)\right)$$

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

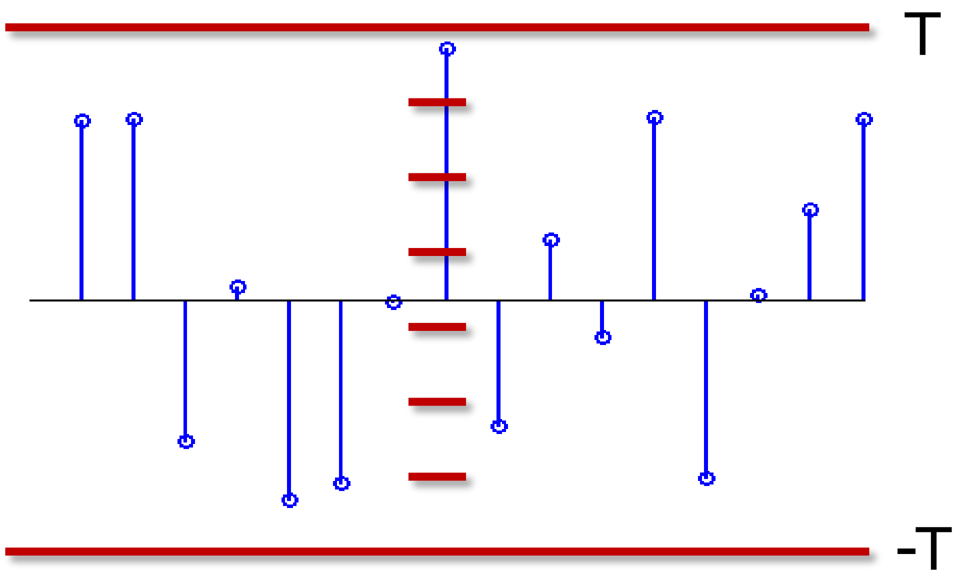


- The fact that $[\Phi \;\, I]$ satisfies the RIP implies that if we take D extra measurements, then we can delete $O(D)$ arbitrary rows of Φ and retain the RIP

- This is a strong *adversarial* notion of democracy

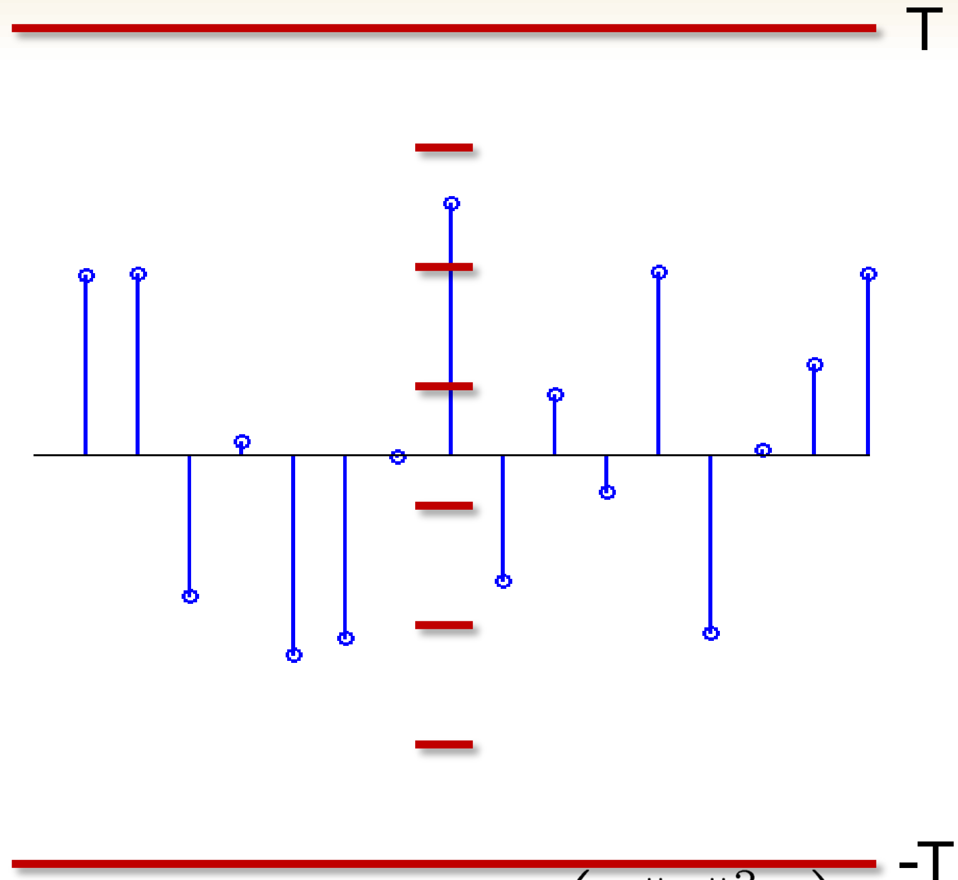
[D, Laska, Boufounos, and Baraniuk - 2009]

Rejection In Practice



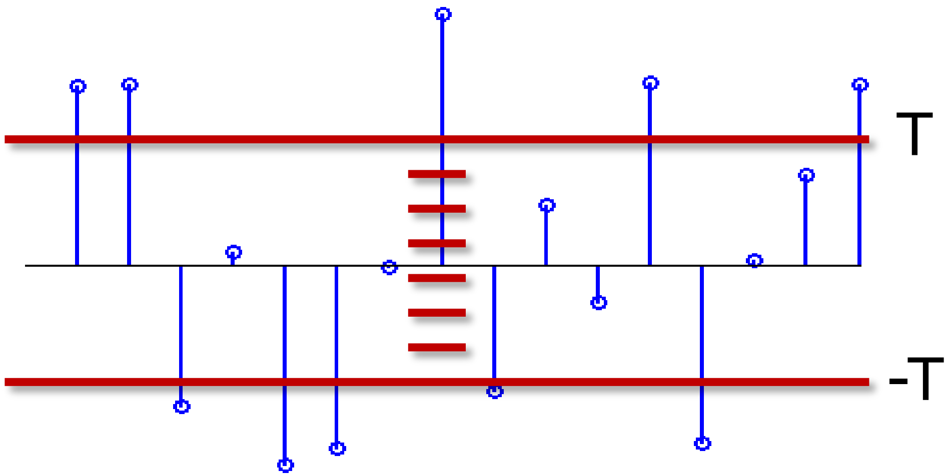
$$\mathrm{SNR} = 10\log_{10}\left(\frac{\|x\|_2^2}{\|\widehat{x}-x\|_2^2}\right)$$

Rejection In Practice



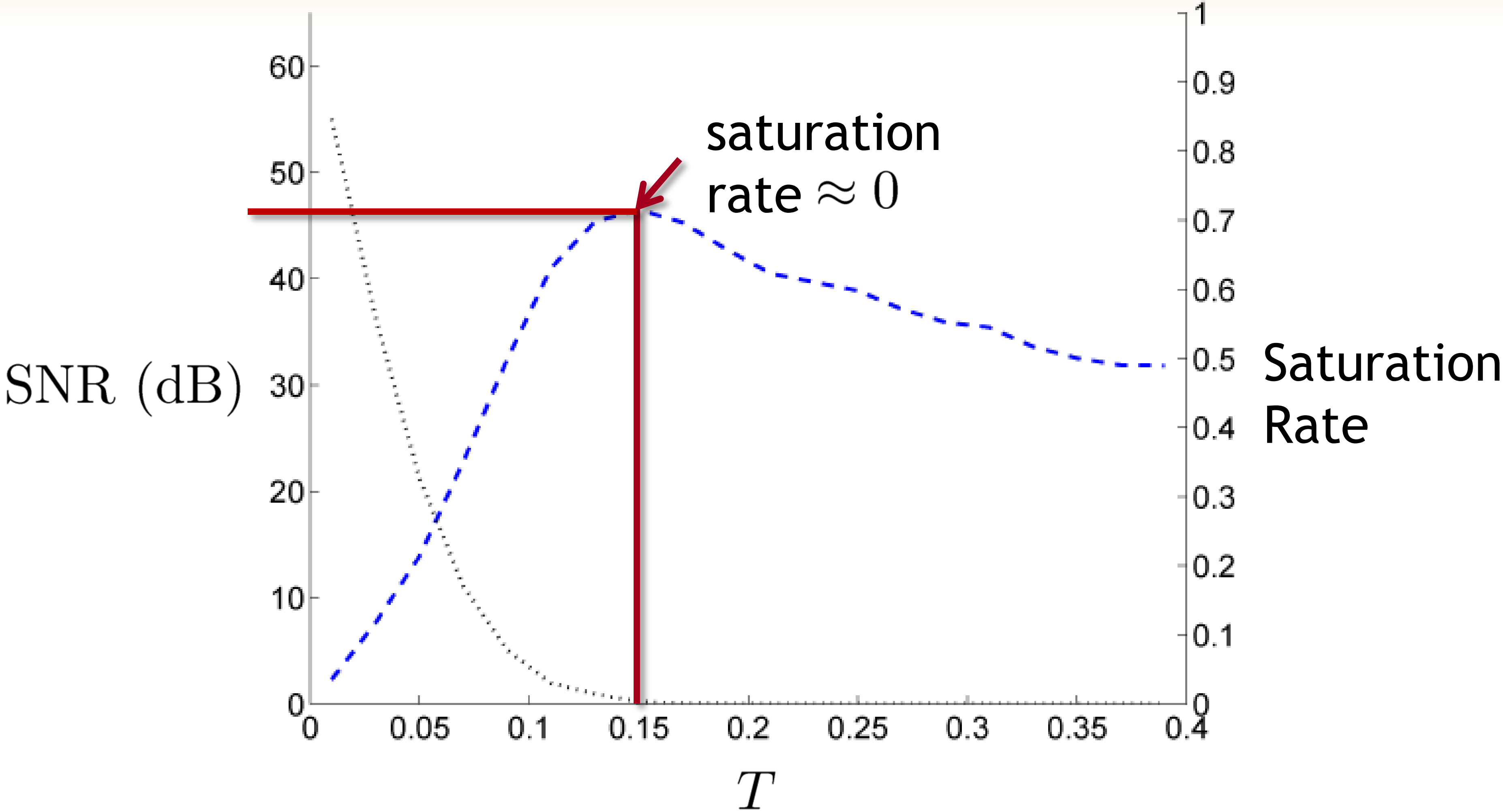
$$\mathrm{SNR} = 10\log_{10}\left(\frac{\|x\|_2^2}{\|\widehat{x}-x\|_2^2}\right)$$

Rejection In Practice



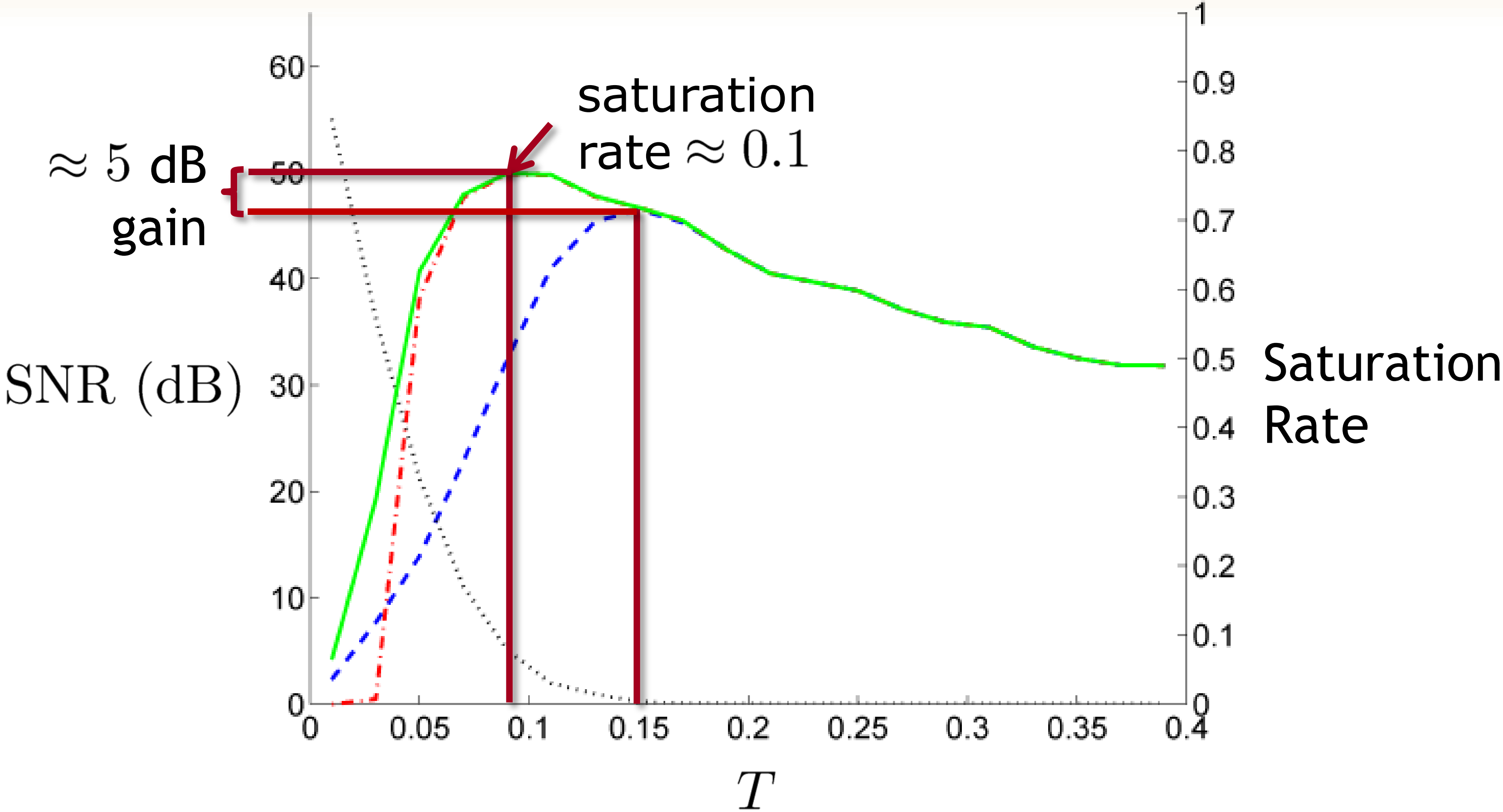
$$\mathrm{SNR} = 10\log_{10}\left(\frac{\|x\|_2^2}{\|\widehat{x}-x\|_2^2}\right)$$

Benefits of Saturation



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

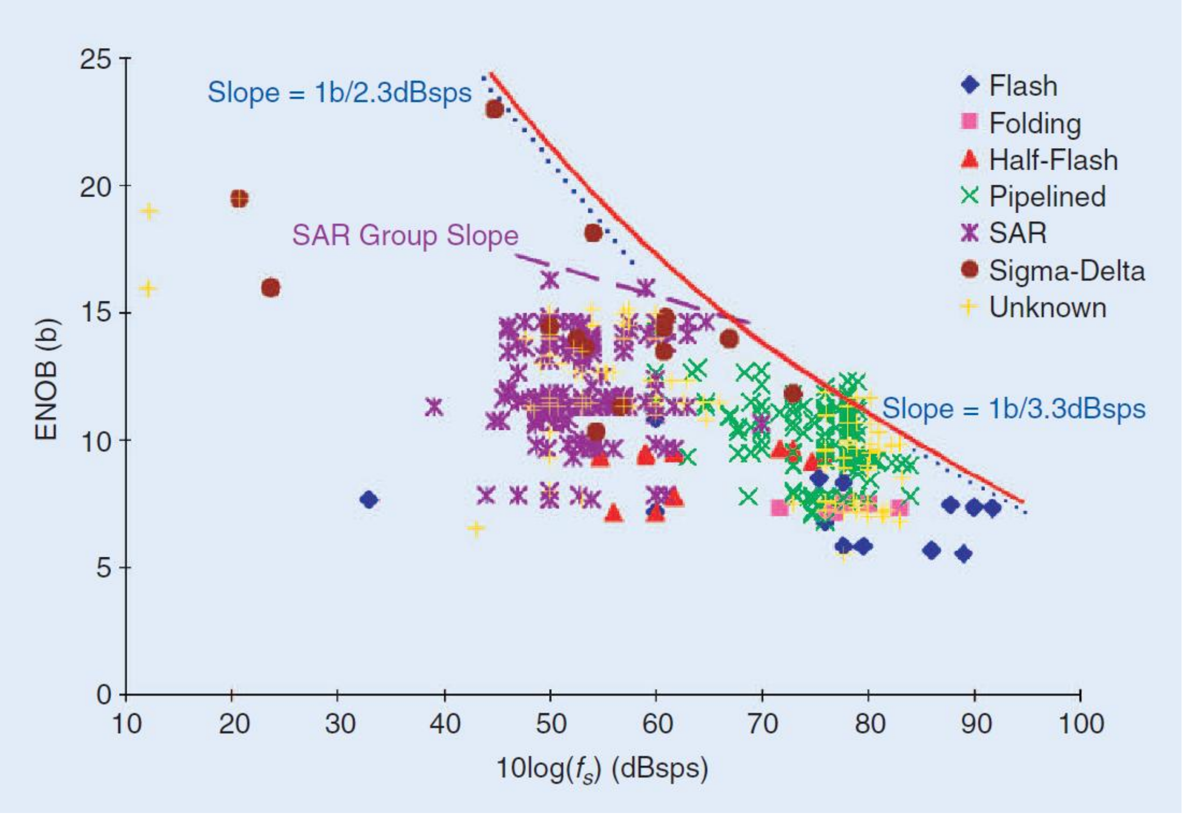
Benefits of Saturation



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

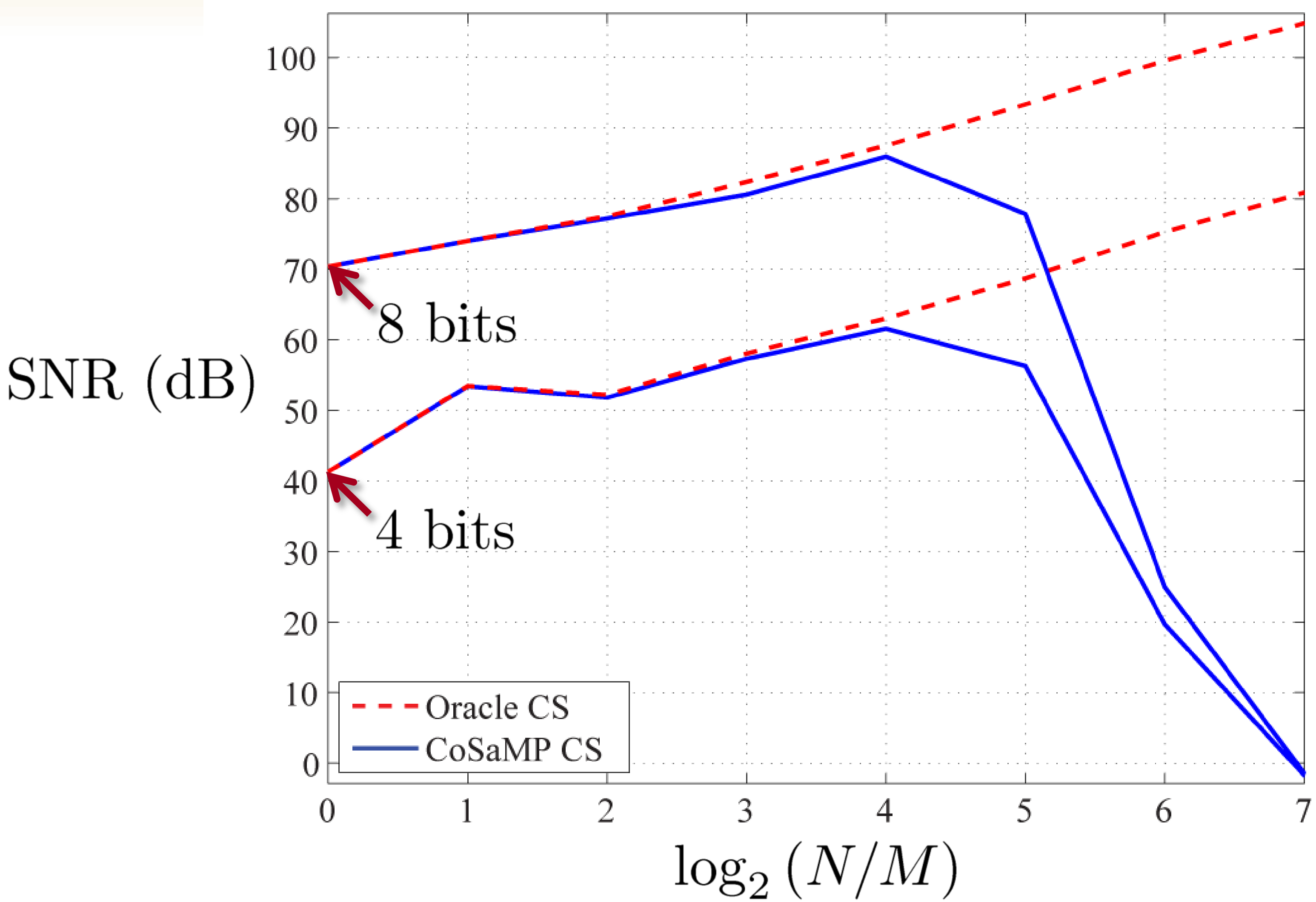
Potential for SNR Improvement?

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

Cons

- signal noise can potentially be a problem
- nonadaptivity entails a tremendous SNR loss
- if you have signal noise or can get benefits from averaging, taking fewer measurements might be a really bad idea!

Pros

- if quantization noise dominates the error, CS can potentially lead to big improvements
- novel strategies for handling saturation errors
- low-bit "CS" might be useful even when M is relatively large