Beyond Nyquist

Joel A. Tropp

Applied and Computational Mathematics
California Institute of Technology
jtropp@acm.caltech.edu

With M. Duarte, J. Laska, R. Baraniuk (Rice DSP), D. Needell (UC-Davis), and J. Romberg (Georgia Tech)

Research supported in part by NSF, DARPA, and ONR

The Sampling Theorem

Theorem 1. *Suppose f is a continuous-time signal whose highest frequency is at most $W/2$ Hz. Then*

$$f(t) = \sum_{n \in \mathbb{Z}} f\left(\frac{n}{W}\right) \operatorname{sinc}(Wt - n).$$

where $\operatorname{sinc}(x) = \sin(\pi x)/\pi x$*.*

- The *Nyquist rate* W is twice the highest frequency
- The *cardinal series* represents a bandlimited signal by uniform samples taken at the Nyquist rate

Reference: [Oppenheim et al. 2000]

Analog-to-Digital Converters (ADCs)

- An ADC consists of a *low-pass filter*, a *sampler* and a *quantizer*
- For sampling rate R, low-pass filter has cutoff $R/2$ to prevent aliasing
- Ideal sampler produces a sequence of amplitude values:

$$f \longmapsto \{f(nT) : n \in \mathbb{Z}\}$$

  where the sampling interval $T = R^{-1}$

- The quantizer maps the real sample values to a discrete set of levels
- Commonly, analog signals are acquired by sampling at the Nyquist rate and processing information with digital technology

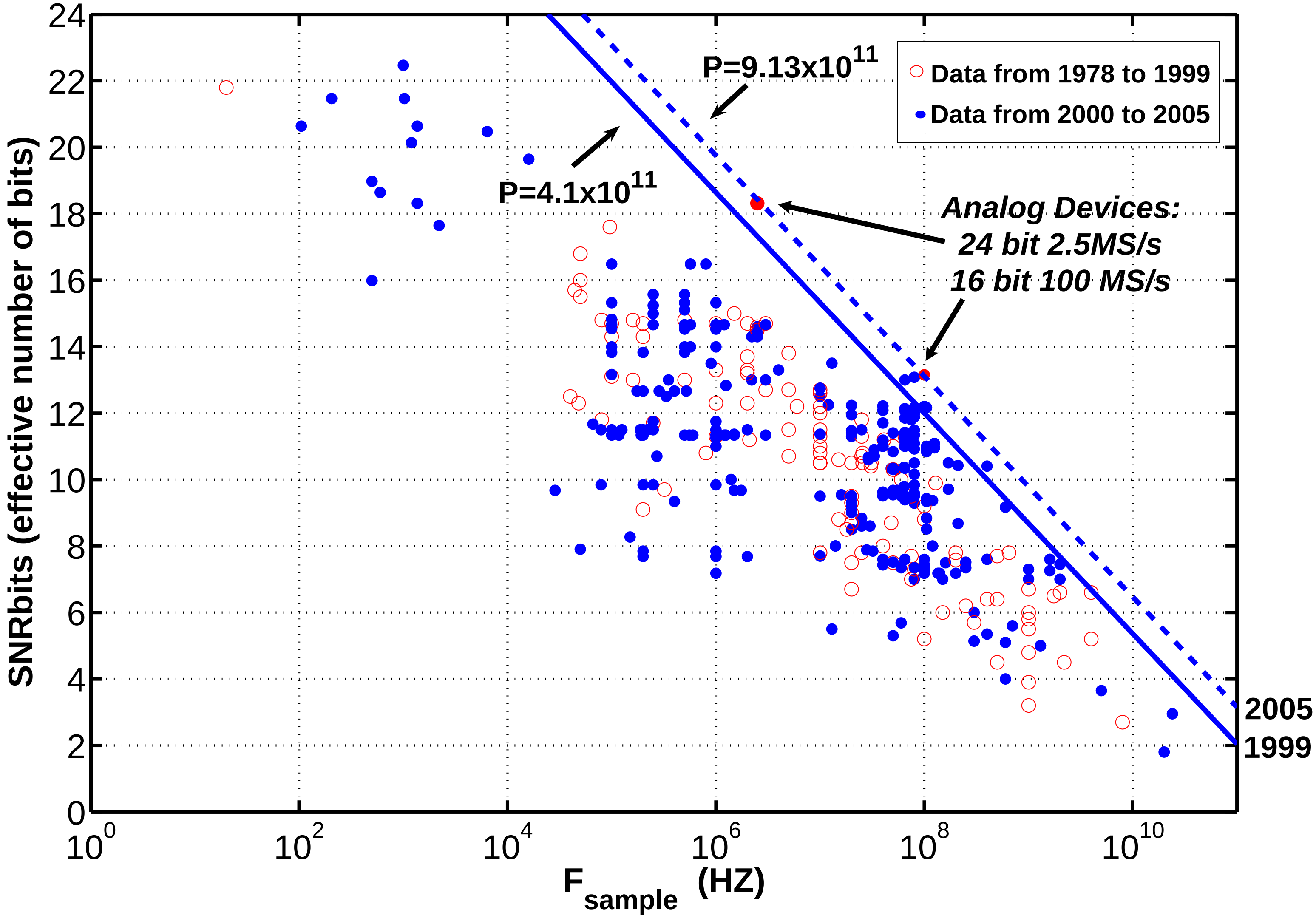
ADCs: State of the Art

- The best current technology (2005) gives
  - 18 effective bits at 2.5 MS/s (MegaSamples/sec)
  - 13 effective bits at 100 MS/s
- Performance degradation about 1 effective bit per frequency octave
- The standard performance metric is

$$P = 2^{\#\text{ effective bits}} \cdot \text{sampling frequency}$$

- At all sampling rates, one effective bit improvement every 6 years

References: [Walden 1999, 2006]


P=9.13x10^11
P=4.1x10^11
Data from 1978 to 1999
Data from 2000 to 2005
Analog Devices:
24 bit 2.5MS/s
16 bit 100 MS/s
2005
1999
SNRbits (effective number of bits)
F sample (HZ)
0
2
4
6
8
10
12
14
16
18
20
22
24
10^0
10^2
10^4
10^6
10^8
10^10

Train Wreck

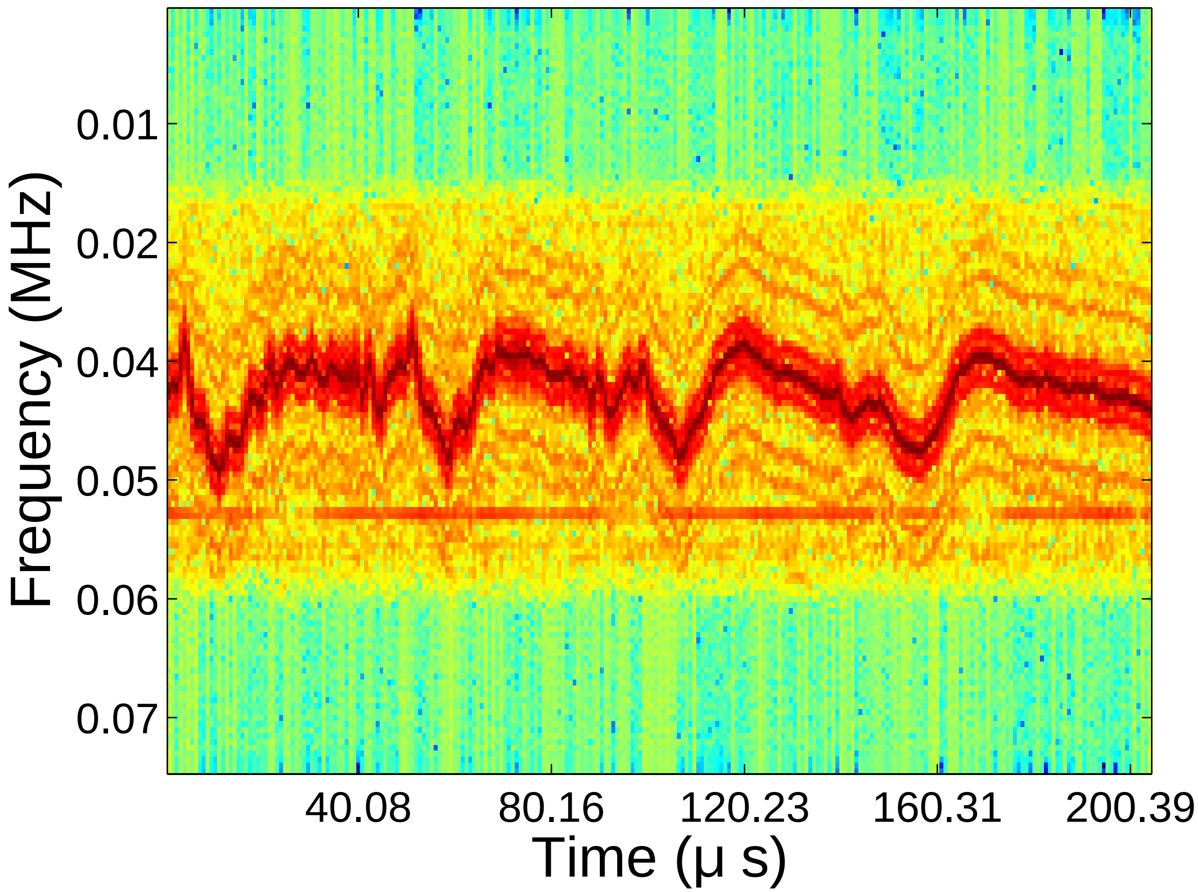
- Modern applications already exceed ADC capabilities
- The Moore's Law for ADCs is too shallow to help

Conclusion:
We need fundamentally new approaches

Idea: Exploit Structure

- Absent additional structure, Nyquist-rate sampling is optimal
- Need to identify and exploit other properties of signals
- Signals of interest do not contain much information relative to their bandwidth
- In communications applications, signals often contain few significant frequencies

Example: An FM Signal



Data provided by L3 Communications

Sparse, Bandlimited Signals

A *normalized* model for signals sparse in time–frequency:

- Let W exceed the signal bandwidth (in Hz)
- Let $\Omega \subset \{-W/2+1, \dots, -1, 0, 1, \dots, W/2\}$ be *integer* frequencies
- For each one-second time interval, signal has the form

$$f(t) = \sum_{\omega \in \Omega} a(\omega)\, \mathrm{e}^{2\pi \mathrm{i} \omega t} \qquad \text{for } t \in [0, 1)$$

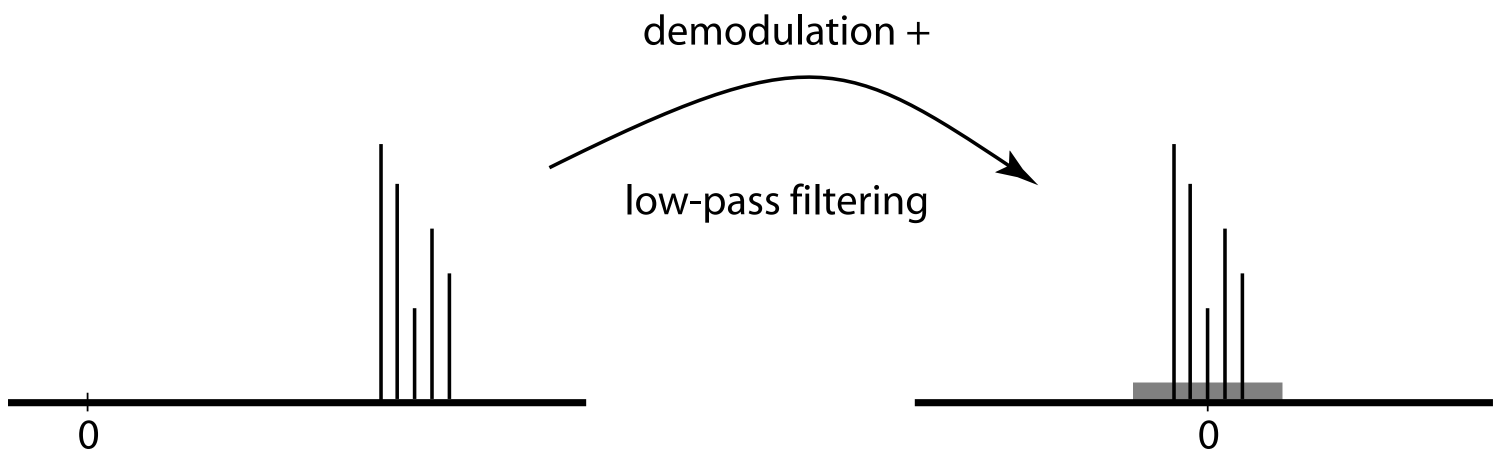
- The set Ω of frequencies can change every second
- In each time interval, number of frequencies $|\Omega| = K \ll W$

Information and Signal Acquisition

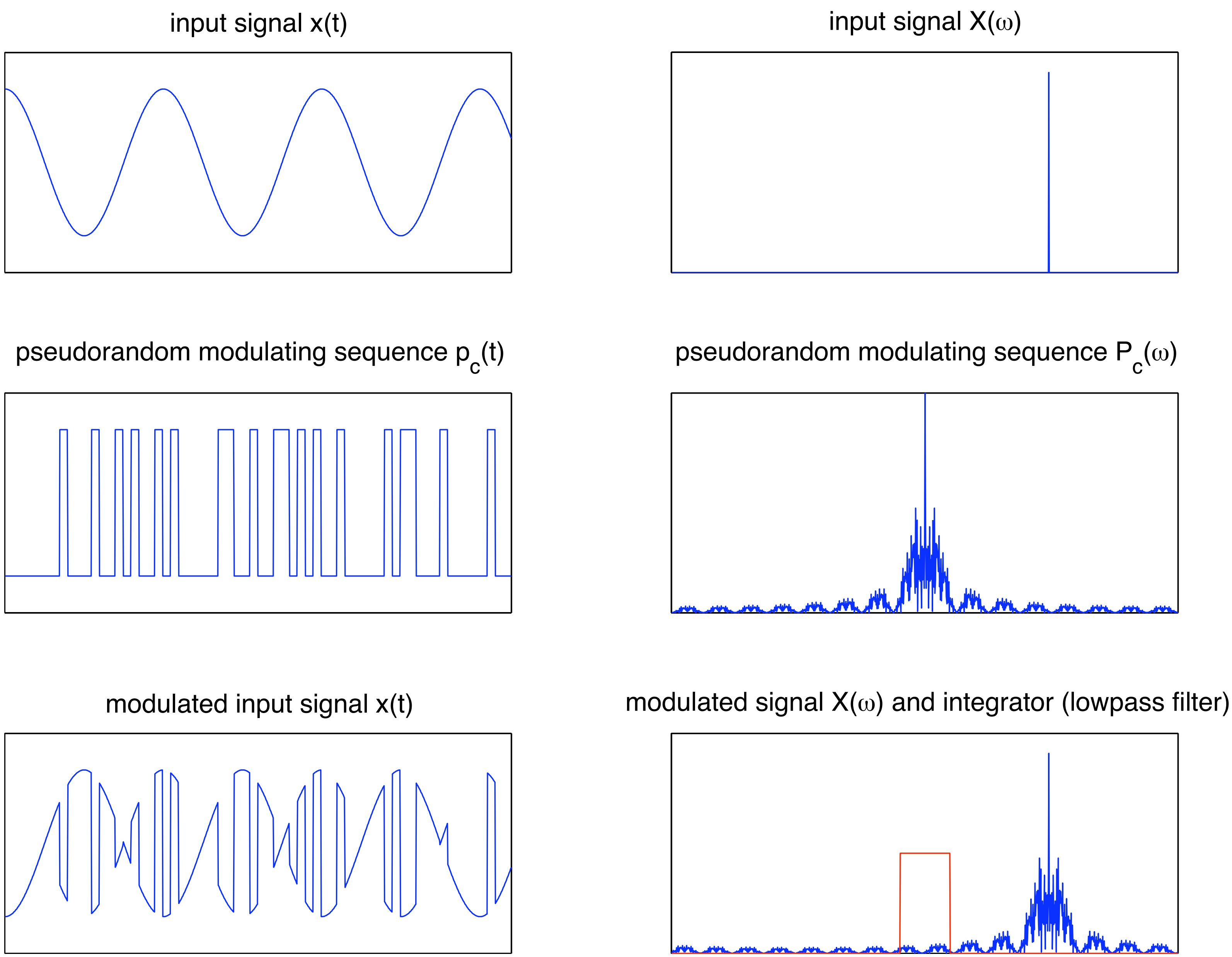
- Signals in our model contain little information
  - In each time interval, have K frequencies and K coefficients
  - Total: About $K \log W$ bits of information
- **Idea:** We should be able to acquire signals with about $K \log W$ nonadaptive measurements
- **Challenge:** Achieve goal with current ADC hardware
- **Approach:** Use randomness!

Random Demodulator: Intuition

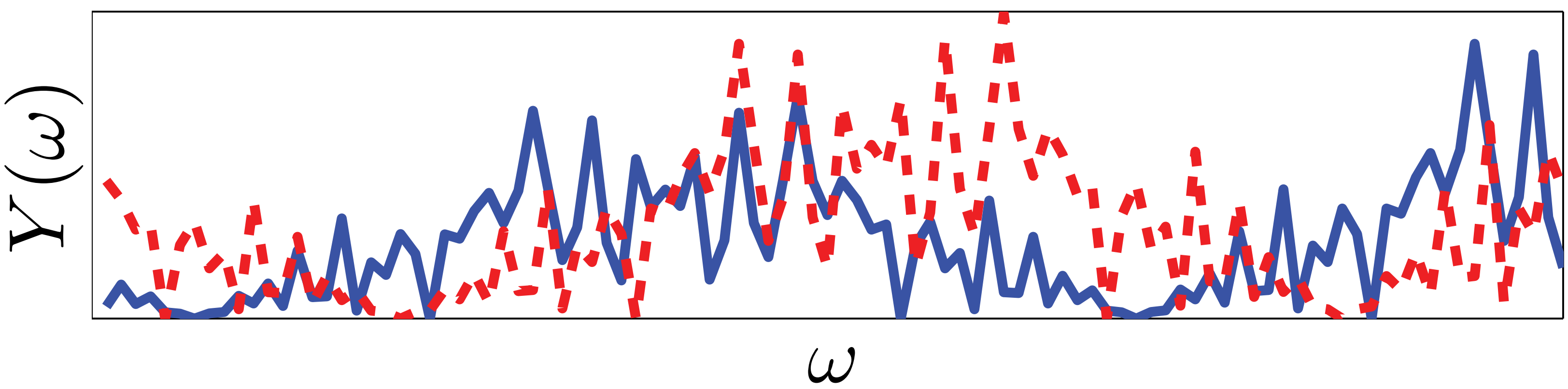
- With clustered frequencies, demodulate to baseband and low-pass filter



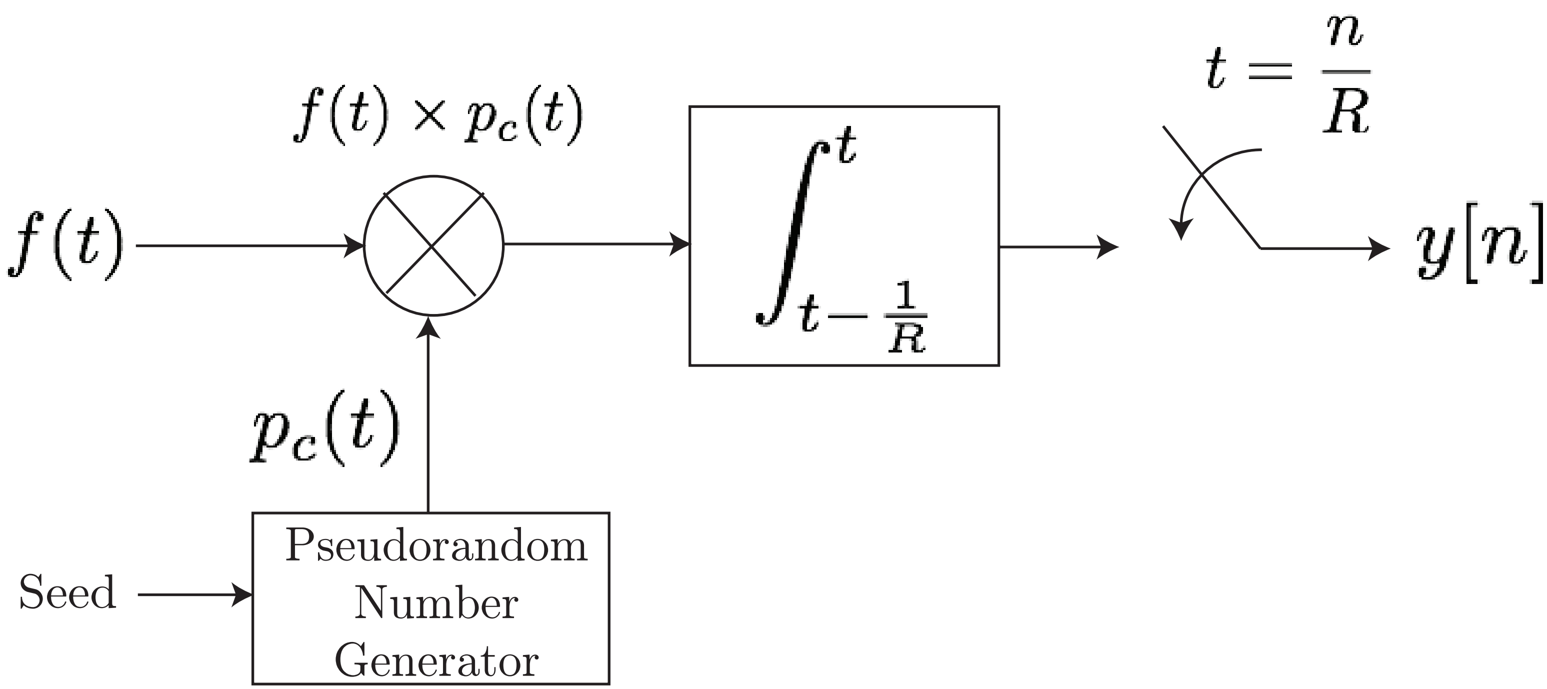
- Don't know locations, so demodulate *randomly* and low-pass filter


input signal x(t)
input signal X(ω)
pseudorandom modulating sequence $p_c(t)$
pseudorandom modulating sequence $P_c(\omega)$
modulated input signal x(t)
modulated signal X(ω) and integrator (lowpass filter)

Exploded View of Passband



Random Demodulator: System Model



- $p_c(t)$ alternates randomly between levels ± 1 at Nyquist rate W
- Sampler runs at rate $R \ll W$

Matrix Formulation I

- The continuous signal has the form

$$f(t) = \sum\nolimits_{\omega \in \Omega} a(\omega)\, \mathrm{e}^{2\pi \mathrm{i} \omega t} \qquad \text{for } t \in [0, 1)$$

- Time-averaging for $1/W$ seconds at $t_n = n/W$ yields

$$\begin{aligned} \int_{t_n}^{t_n + 1/W} f(t)\, \mathrm{d}t &= \sum\nolimits_{\omega \in \Omega} a(\omega) \left[\frac{\mathrm{e}^{2\pi \mathrm{i} \omega / W} - 1}{2\pi \mathrm{i} \omega} \right] \mathrm{e}^{2\pi \mathrm{i} \omega t_n} \\ &= \sum\nolimits_{\omega \in \Omega} s(\omega)\, \mathrm{e}^{2\pi \mathrm{i} \omega t_n} \end{aligned}$$

- Can express time-averaged signal as a vector $\boldsymbol{x} = \boldsymbol{F}\boldsymbol{s} \in \mathbb{C}^W$
  - $\boldsymbol{s}$ is sparse and supported on Ω
  - $\boldsymbol{F}$ is essentially a DFT matrix
  - $\boldsymbol{x}$ contains the same (discrete) frequencies as f

Matrix Formulation II

- The (ideal) action of the multiplier is given by

$$\boldsymbol{D} = \begin{bmatrix} \pm 1 & & & & \\ & \pm 1 & & & \\ & & \pm 1 & & \\ & & & \ddots & \\ & & & & \pm 1 \end{bmatrix}$$

- The (ideal) action of the accumulate-and-dump sampler is given by

$$\boldsymbol{H} = \begin{bmatrix} 1 & 1 & \dots & 1 & & & & & & & & \\ & & & & 1 & 1 & \dots & 1 & & & & \\ & & & & & & & \ddots & \ddots & & & \\ & & & & & & & & 1 & 1 & \dots & 1 \end{bmatrix}_{R \times W}.$$

Reconstruction from Samples

- The matrix $\mathbf{\Phi}$ summarizes the action of the random demodulator

$$\mathbf{\Phi} = \boldsymbol{HDF} : \mathbb{C}^W \longrightarrow \mathbb{C}^R$$

- Maps a (sparse) amplitude vector $\boldsymbol{s}$ to a vector of samples $\boldsymbol{y}$
- Given samples $\boldsymbol{y} = \mathbf{\Phi}\boldsymbol{s}$, signal reconstruction can be formulated as

$$\widehat{\boldsymbol{s}} = \arg\min \|\boldsymbol{c}\|_0 \quad \text{subject to} \quad \mathbf{\Phi}\boldsymbol{c} = \boldsymbol{y}$$

- The ℓ_0 function counts nonzero entries of a vector

Signal Reconstruction Algorithms

Approach 1: Convex Relaxation

- Can often find sparsest amplitude vector by solving

$$\widehat{\boldsymbol{s}} = \arg\min \|\boldsymbol{c}\|_1 \quad \text{subject to} \quad \boldsymbol{\Phi c} = \boldsymbol{y} \tag{P1}$$

Approach 2: Greedy Pursuit

- Identify a small set of significant frequencies and iteratively refine
- Examples: OMP and CoSaMP

References: [Candès et al. 2006, Donoho 2006, Tropp–Gilbert 2007, Tropp–Needell 2008]

Shifting the Burden

- These algorithms are much more computationally intensive than linear reconstruction via cardinal series

- Move the work from the analog front end to the digital back end

Moore's Law for ICs
saves us from
Moore's Law for ADCs!

Theoretical Analysis

Theorem 2. [T 2007] *Suppose the sampling rate satisfies*

$$R \geq \mathrm{C} \cdot K \cdot \log^6 W$$

Then the matrix $\mathbf{\Phi}$ *has the restricted isometry property*

$$(1 - \mathrm{c}) \|\boldsymbol{x}\|_2^2 \leq \|\mathbf{\Phi}\boldsymbol{x}\|_2^2 \leq (1 + \mathrm{c}) \|\boldsymbol{x}\|_2^2 \quad \textit{when} \quad \|\boldsymbol{x}\|_0 \leq 2K$$

except with probability W^{-1}.

- Abstract property supports efficient sampling and reconstruction
- Intuition: Sampling operator preserves geometry of sparse vectors

Recovery via Convex Optimization

Theorem 3. [Candès–Romberg–Tao 2006] *Suppose that*

- *the sampling matrix* $\mathbf{\Phi}$ *has the RIP,*
- *the sample vector* $\boldsymbol{y} = \mathbf{\Phi}\boldsymbol{s} + \boldsymbol{e}$, *and*
- *the error* $\|\boldsymbol{e}\|_2 \leq \eta$.

Then the solution $\widehat{\boldsymbol{s}}$ *to the program*

$$\min \|\boldsymbol{c}\|_1 \qquad \textit{subject to} \qquad \|\boldsymbol{y} - \mathbf{\Phi}\boldsymbol{c}\|_2 \leq \eta$$

satisfies

$$\|\boldsymbol{s} - \widehat{\boldsymbol{s}}\|_2 \leq \mathrm{C}\left[\frac{1}{\sqrt{K}}\|\boldsymbol{s} - \boldsymbol{s}_K\|_1 + \eta\right].$$

Recovery via Greedy Pursuit

Theorem 4. [Needell–T 2008] *Suppose that*

- *the sampling matrix* $\mathbf{\Phi}$ *has the RIP,*
- *the sample vector* $\boldsymbol{y} = \mathbf{\Phi}\boldsymbol{s} + \boldsymbol{e}$*,*
- η *is a precision parameter,*
- $\mathscr{L}$ *bounds the cost of a matrix–vector multiply with* $\mathbf{\Phi}$ *or* $\mathbf{\Phi}^*$*.*

Then CoSaMP produces a $2K$*-sparse approximation* $\widehat{\boldsymbol{s}}$ *such that*

$$\|\boldsymbol{s} - \widehat{\boldsymbol{s}}\|_2 \leq \mathrm{C} \max\left\{\eta, \frac{1}{\sqrt{K}} \|\boldsymbol{s} - \boldsymbol{s}_K\|_1 + \|\boldsymbol{e}\|_2\right\}$$

with execution time

$$\mathrm{O}(\mathscr{L} \cdot \log(\|\boldsymbol{s}\|_2 / \eta)).$$

Simulations

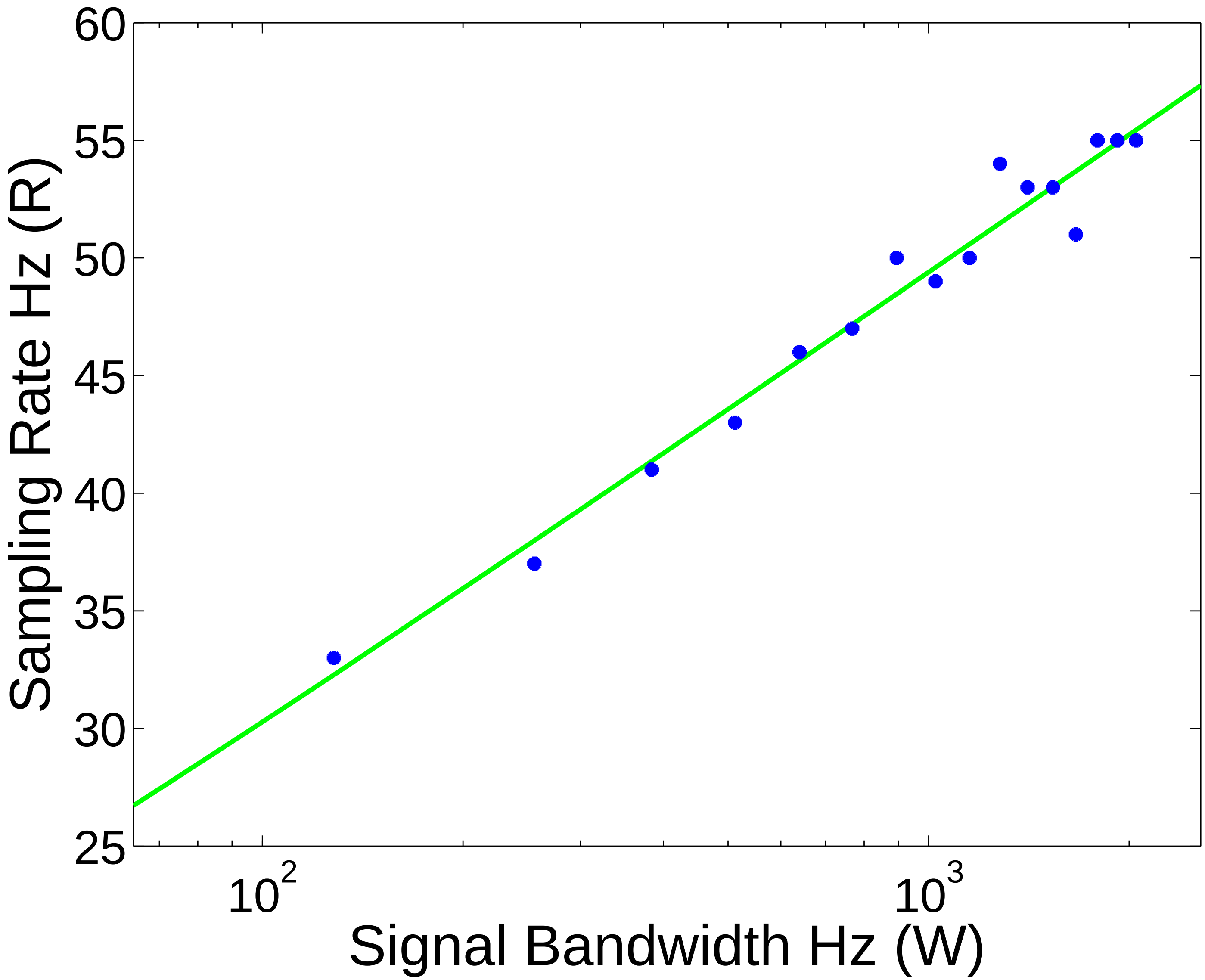
Goal: Estimate sampling rate R to achieve success probability 99%

For each of 500 trials,

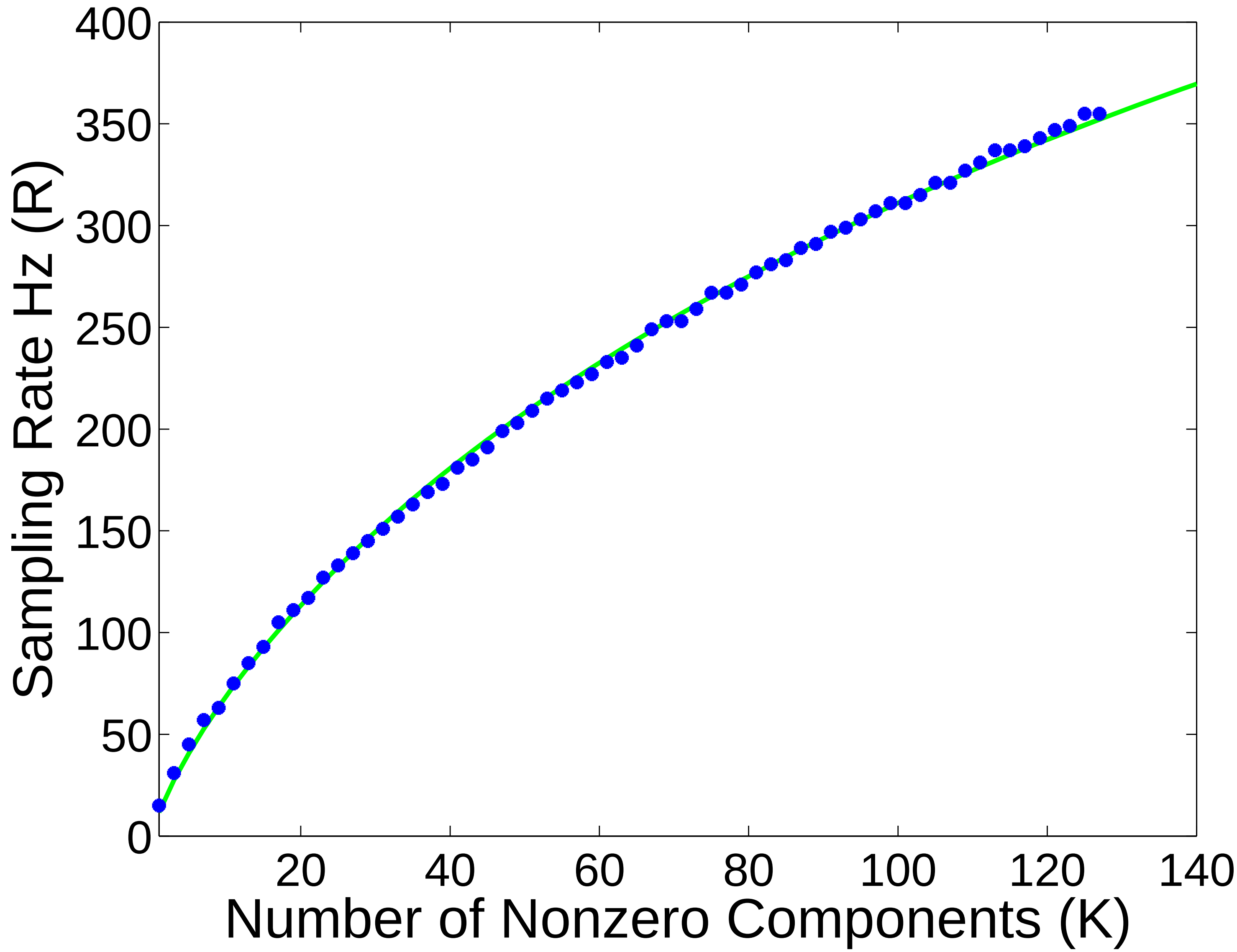
- Draw a random demodulator with dimensions $R \times W$
- Choose a random set of K frequencies
- Set their amplitudes equal to one
- Take measurements of the signal
- Recover with ℓ_1 minimization (via IRLS)

Define *success* at rate R when 99% of trials result in

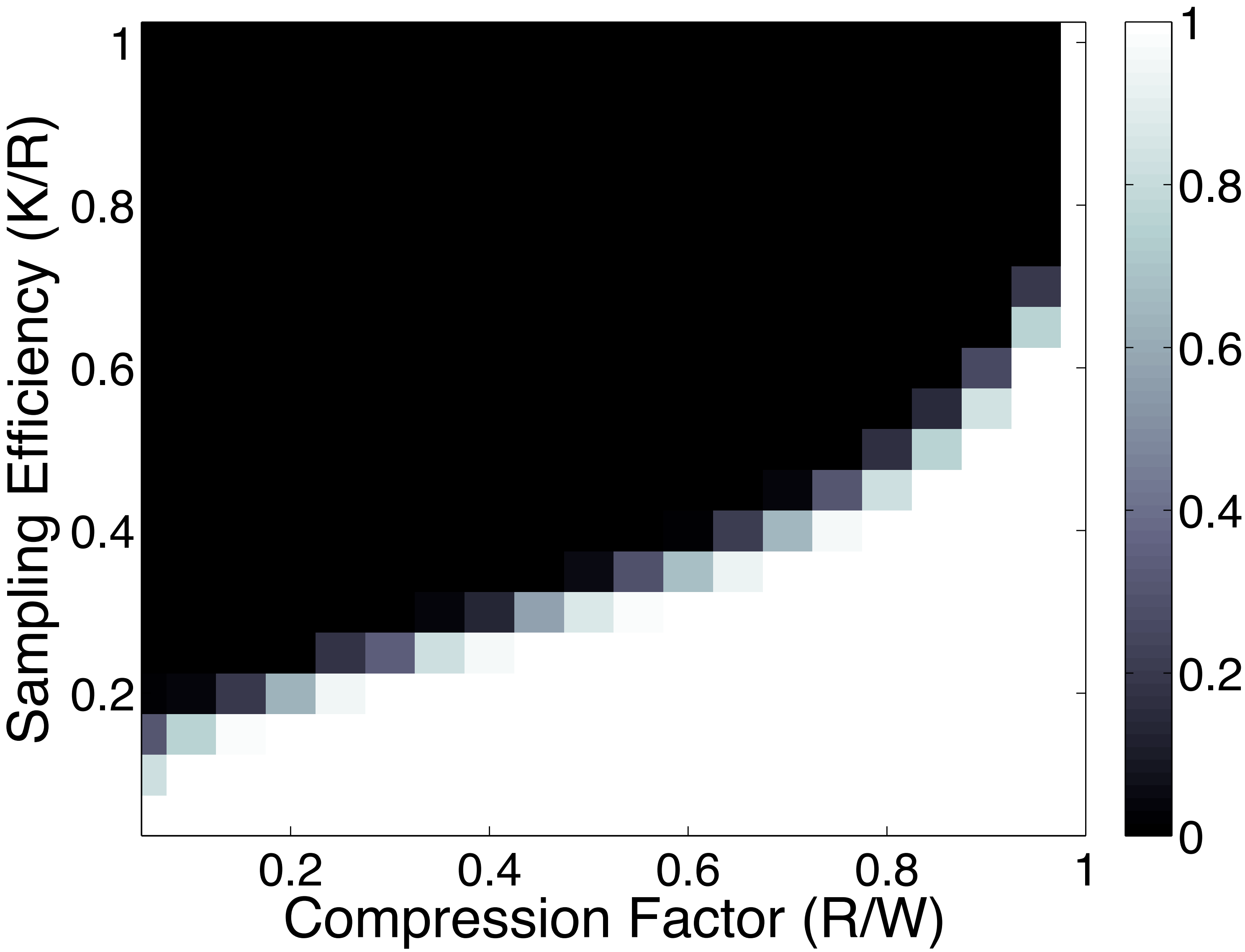
$$\|\boldsymbol{s} - \widehat{\boldsymbol{s}}\| < \varepsilon_{\text{mach}}$$



$K = 5$, regression line $R = 1.69K\log(W/K + 1) + 4.51$

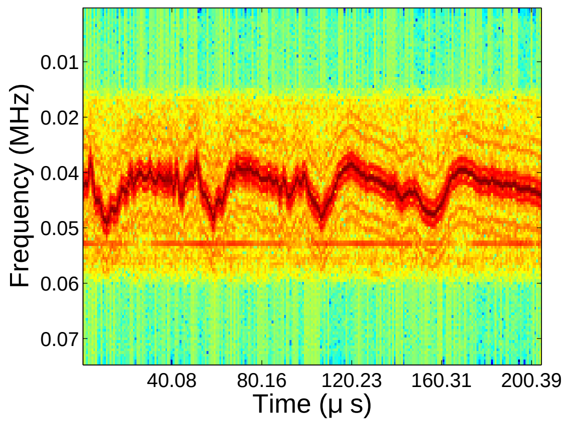


$W = 512$, regression line $R = 1.71K \log(W/K + 1) + 1.00$

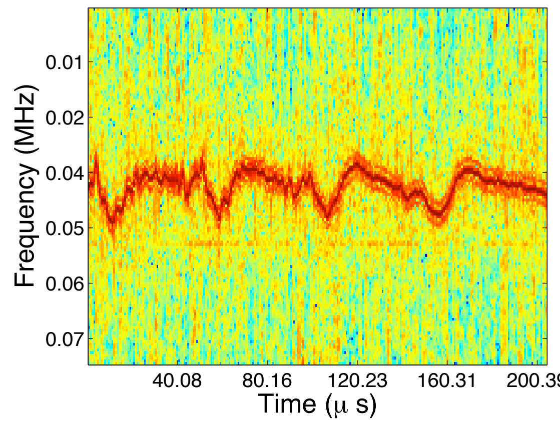


Sampling Efficiency (K/R)
1
0.8
0.6
0.4
0.2
0.2
0.4
0.6
0.8
1
Compression Factor (R/W)
1
0.8
0.6
0.4
0.2
0

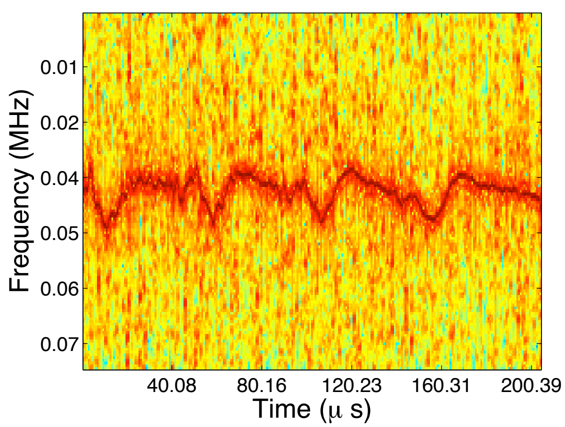
Reconstruction of FM Signal



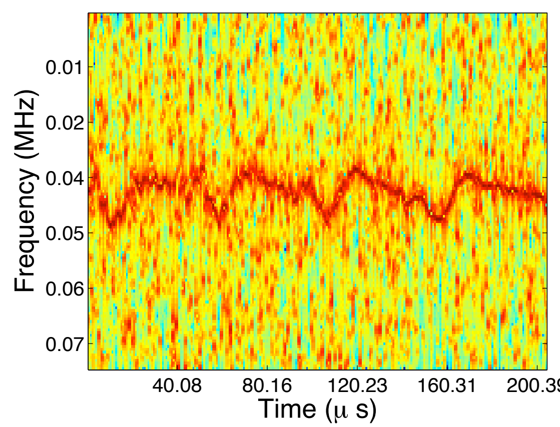
(a) Original Signal (1.25 MHz)



(b) Rand Demod (0.63 MHz)

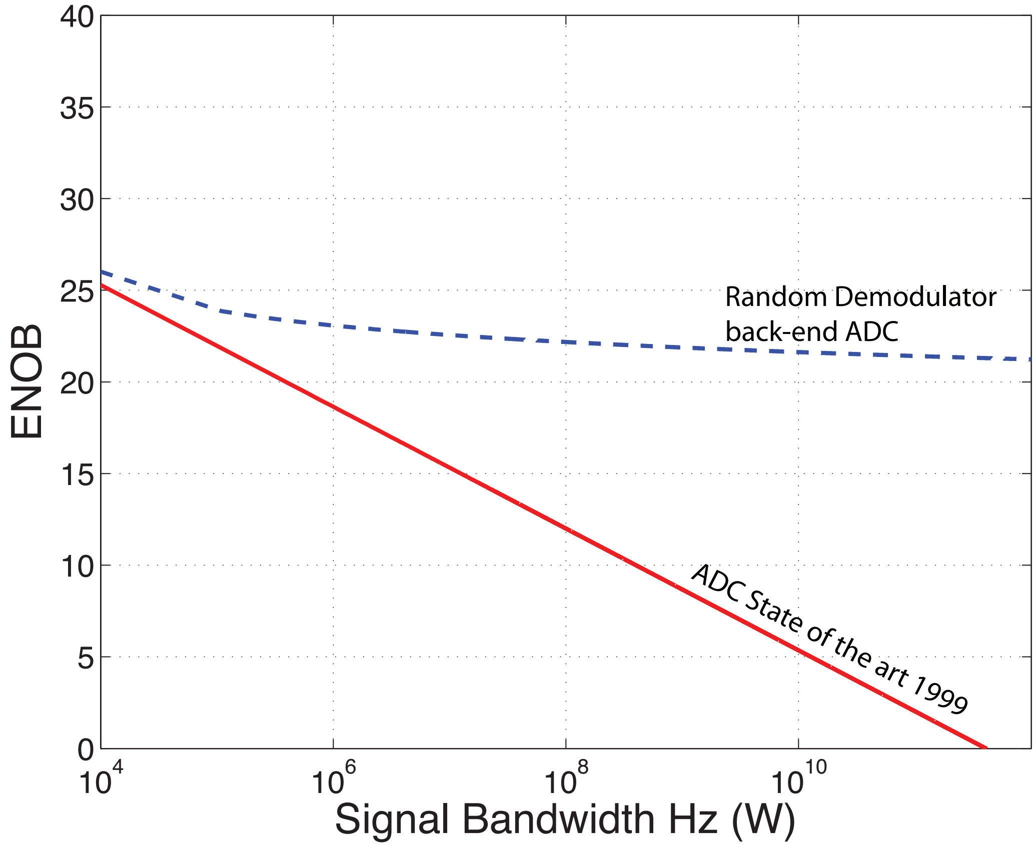


(c) Rand Demod (0.31 MHz)



(d) Rand Demod (0.16 MHz)

On Walden Pond



Fixed sparsity $K = 5000$

To learn more...

E-mail: `jtropp@acm.caltech.edu`

Web: `http://acm.caltech.edu/~jtropp`

`http://www.dsp.rice.edu/cs/`

`http://www.dsp.rice.edu/a2i/`

Papers

- Needell and T, "CoSaMP: Iterative Signal Recovery from Incomplete and Inaccurate Measurements," ACHA 2008
- T, Romberg, Rice CSP, "Beyond Nyquist: Efficient Sampling of Sparse, Bandlimited Signals." In preparation.