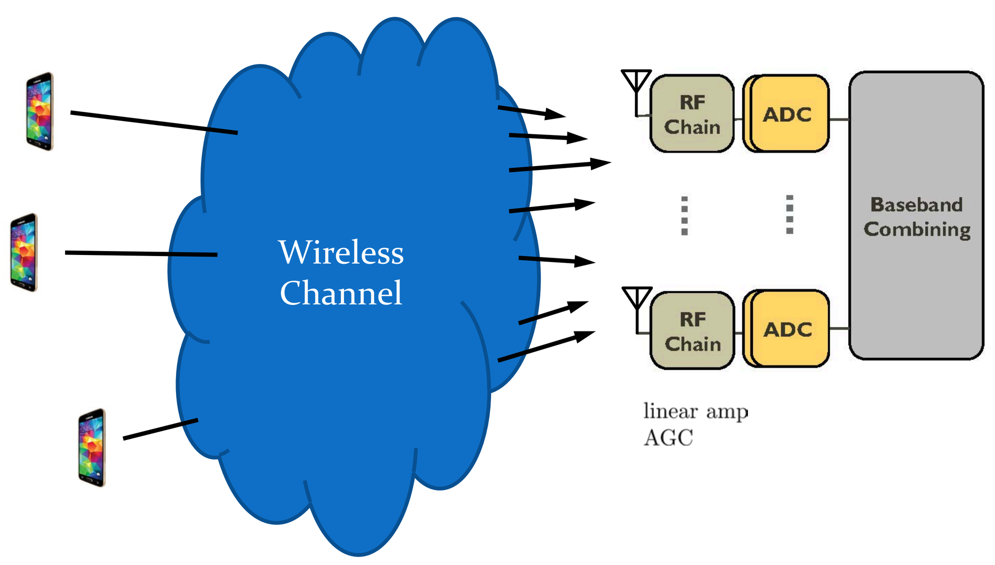
One-Bit Quantization for mmWave Massive MIMO Systems

A. Lee Swindlehurst, Amine Mezghani, Yongzhi Li, Amodh Saxena

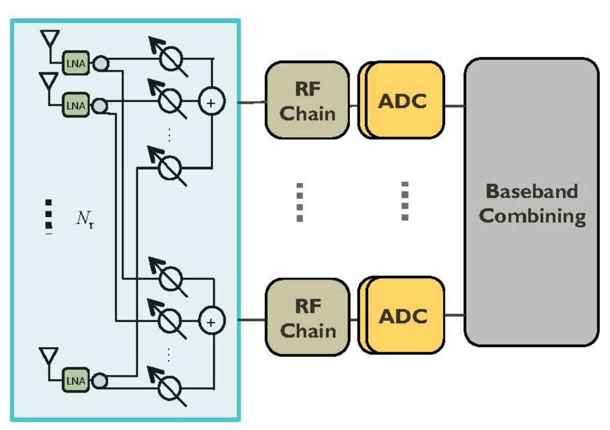
Center for Pervasive Communications & Computing

University of California Irvine

Background



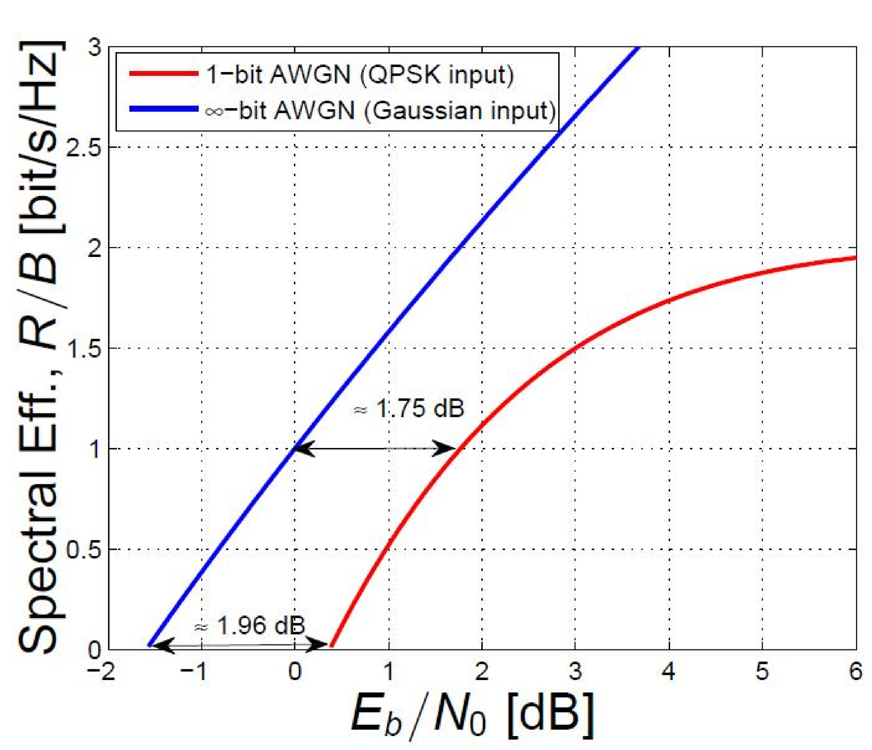
- Full precision ADC requires linear, low-noise amplifiers and AGC
- ADC power consumption grows exponentially with resolution
- A commercial TI 1 Gs/s 12-bit ADC requires 4W
- Not practical for ideal massive MIMO



- Reduce dimensionality with RF beamforming network
- Complicates receiver design, scalability issues for wideband operation
- Phase shifters are typically quantized
- Power/cost still an issue

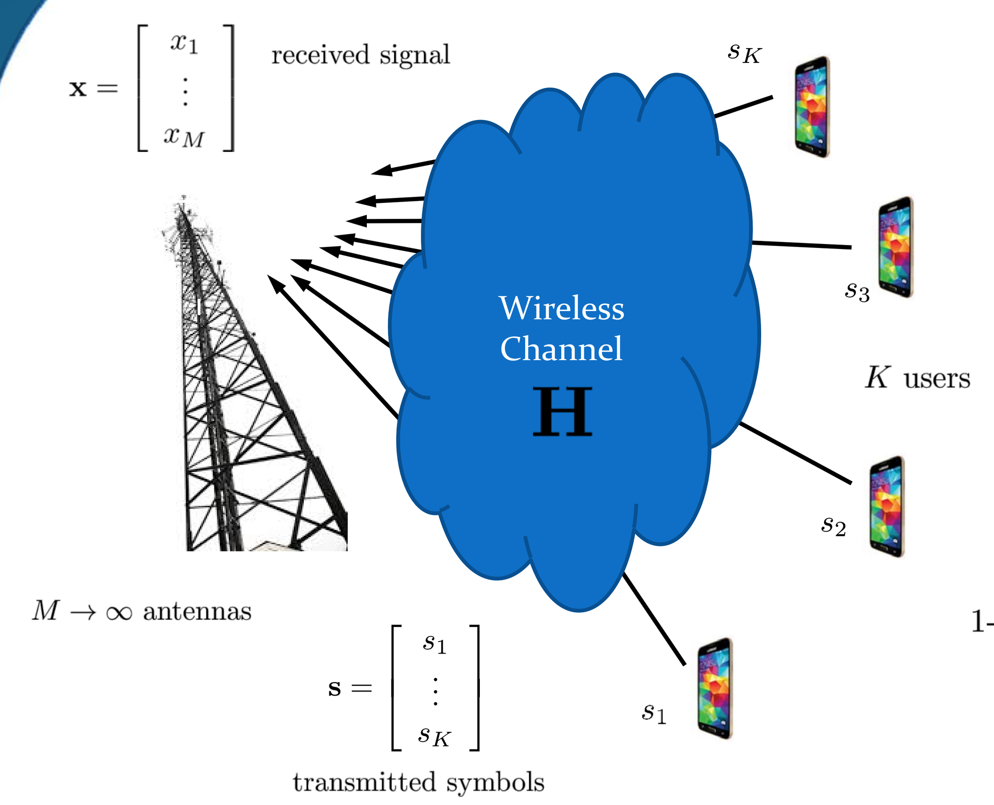
One-Bit ADCs

- One-bit ADC $\Rightarrow$ simple RF, no AGC or high cost LNA
- Operates at a fraction of the power
- Low SNR loss (typical operating point for mmWave massive MIMO) only 2dB
- Compensate for quantization error with signal processing



loss in power efficiency < 2dB when SE < 1.4 bpcu

Uplink



$$\mathbf{x} = \sqrt{\rho}\mathbf{H}\mathbf{s} + \mathbf{n}$$

Use $K \times \tau$ uplink training data $\mathbf{\Phi}$

$$\mathbf{X} = \sqrt{\rho}\mathbf{H}\mathbf{\Phi} + \mathbf{N}$$

Vectorized model

$$\mathbf{x} = \mathrm{vec}(\mathbf{X}) = \sqrt{\rho}\left(\mathbf{\Phi}^T \otimes \mathbf{I}\right)\mathrm{vec}(\mathbf{H}) + \mathrm{vec}(\mathbf{N}) = \tilde{\mathbf{\Phi}}\mathbf{h} + \mathbf{n}$$

1-bit quantization $\mathcal{Q}(\cdot)$ maps complex data to $\pm 1 \pm j$

$$\mathbf{r} = \mathcal{Q}(\mathbf{x}) = \mathcal{Q}\left(\tilde{\mathbf{\Phi}}\mathbf{h} + \mathbf{n}\right)$$

Bussgang Theorem

Let $x(t)$ be a Gaussian random process, and $r(t) = \mathcal{Q}(x(t))$ be the output of some nonlinear function. Then for a certain constant a, we have

$$a r_{xx}(\tau) = r_{xr}(\tau)$$

for the auto-correlation and cross-correlation functions $r_{xx}(\tau)$ and $r_{xr}(\tau)$, respectively.

Represent nonlinear quantization by "equivalent" linear operator:

$$\mathbf{r} = \mathcal{Q}(\mathbf{x}) = \mathcal{Q}\left(\tilde{\mathbf{\Phi}}\mathbf{h} + \mathbf{n}\right) = \mathbf{A}\mathbf{x} + \mathbf{q}$$

where

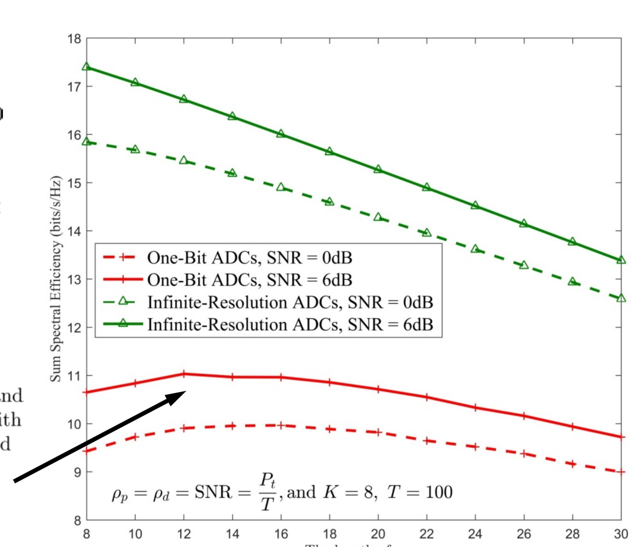
$$\mathbf{A}\mathbf{C}_{xx} = \mathbf{C}_{xr}^H$$

Under this model $\mathbf{x}$ and $\mathbf{q}$ are uncorrelated, and $\mathbf{A}$ minimizes the equivalent quantization noise:

$$\mathbf{A} = \arg\min_{\mathbf{A}} \|\mathbf{r} - \mathbf{A}\mathbf{x}\|^2$$

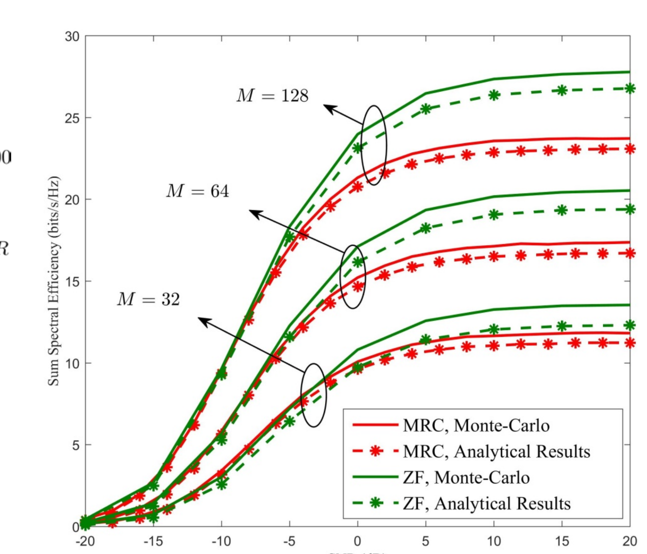
Linear model simplifies algorithm design and analysis

- $M = 32$
- coherence interval $T = 100$
- sum spec. eff. $\mathcal{S} = \frac{T-K}{T}KR$



To find optimal τ and analyze trade-off with throughput, we need analysis of spectral efficiency

- $K = 8$ users
- equal power
- coherence interval $T = 200$
- sum spec. eff. $\mathcal{S} = \frac{T-K}{T}KR$

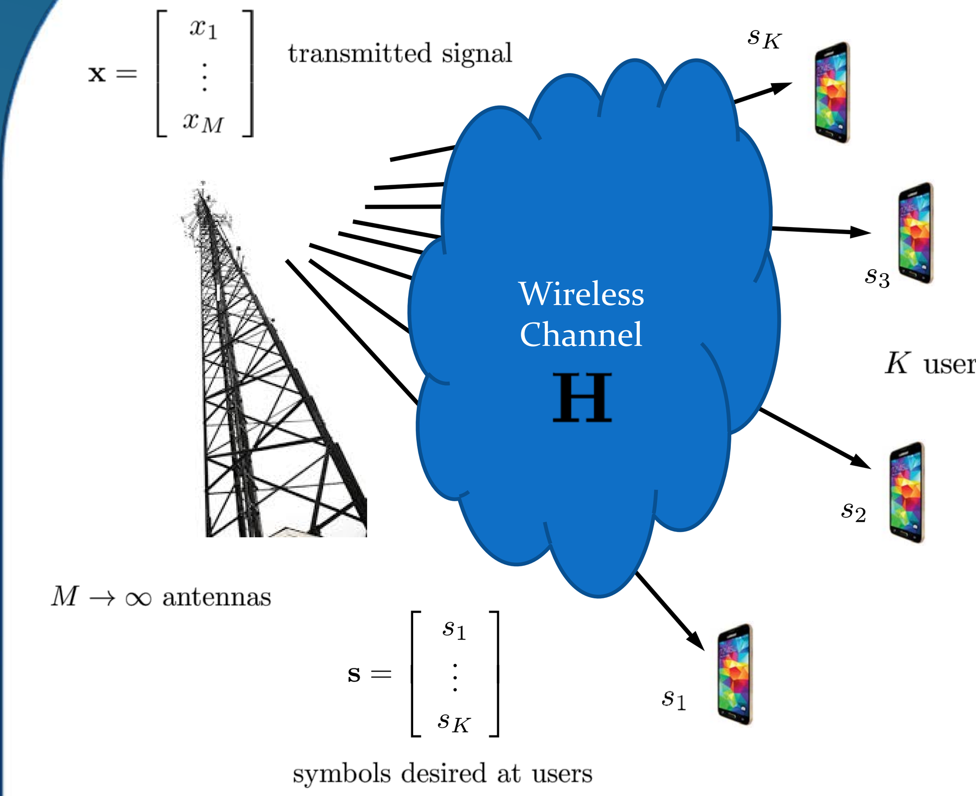


	Typical Massive MIMO	One-Bit Massive MIMO
MRC	$C\left(\frac{\rho^2 K M_{typ}}{K\rho(K\rho+1)+K\rho+1}\right)$	$C\left(\frac{\rho^2\alpha^4 K M_{one}}{\rho\alpha^2 K+\alpha^2+1-2/\pi}\right)$
ZF	$C\left(\frac{\rho^2 K(M_{typ}-K)}{2K\rho+1}\right)$	$C\left(\frac{\rho^2\alpha^4 K(M_{one}-K)}{\rho\alpha^2 K\eta+\alpha^2+1-2/\pi}\right)$

$$C(x) = \frac{T-K}{T}K\log_2(1+x) \qquad \alpha = \sqrt{\frac{2}{\pi(1+\rho K)}}$$

- $C_{typ} = C_{one}$ for MRC when $\frac{M_{one}}{M_{typ}} = \frac{\pi^2}{4} \simeq 2.5$
- $C_{typ} = C_{one}$ for ZF when $\frac{M_{one}}{M_{typ}} = \frac{\pi^2(1+\rho K)^2 - 4\rho^2K^2}{4+8\rho K} \longrightarrow \frac{\pi^2}{4} \qquad \rho \to 0$

Downlink



$$\hat{\mathbf{s}} = \mathbf{H}^T\mathbf{x} + \mathbf{n}$$

Use ML Encoding?

Since $\mathbf{x}$ is constrained to QPSK alphabet due to 1-bit quantization, suggests ML encoding:

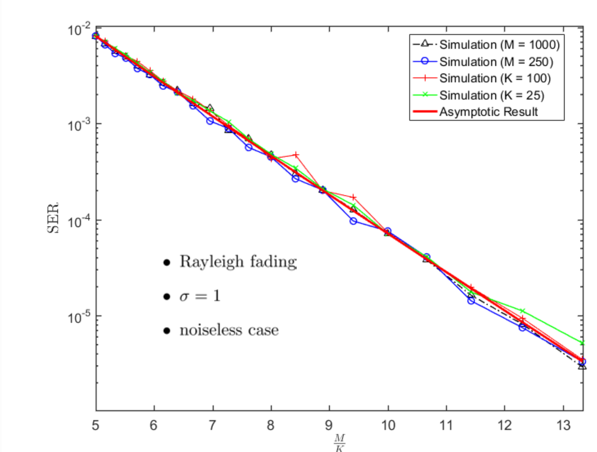
$$\mathbf{x} = \arg\min_{x_i \in \pm 1 \pm j} \|\mathbf{s} - \mathbf{H}^T\mathbf{x}\|$$

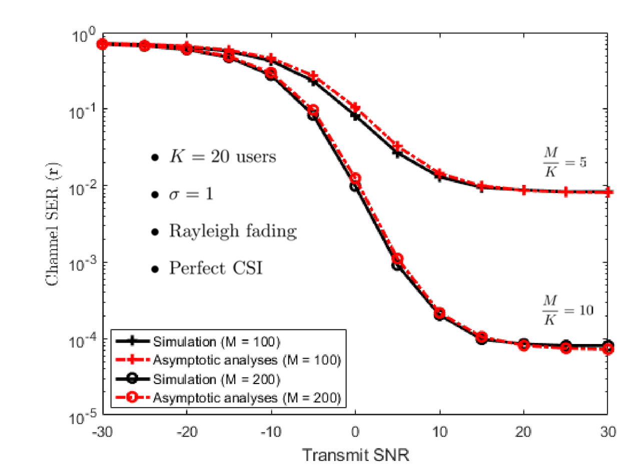
- Prohibitively complex, especially for a massive antenna array ($\mathbf{H}^T$ is $K \times M$)
- Requires special "handling" since $\mathbf{H}^T$ is a fat matrix
- Even a sphere encoding approach is too costly
- We will see the ML encoding is outperformed by something much simpler …

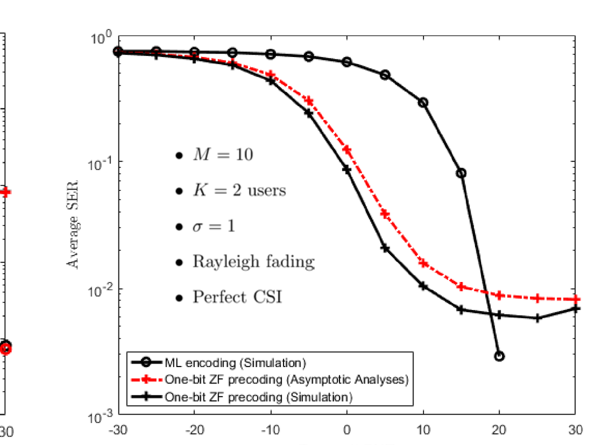
Quantized Linear Precoding

Output of linear precoder $\mathbf{x}_P = \mathbf{P}\mathbf{s}$ is 1-bit quantized prior to transmission: $\hat{\mathbf{s}} = \frac{1}{\sqrt{M}}\mathbf{H}^T\mathcal{Q}(\mathbf{P}\mathbf{s}) + \mathbf{n}$

$$P_e = 2Q\left(\sqrt{\frac{\frac{4\sigma^2(M-K)^2}{MK\pi}}{\frac{2\sigma^2}{M}\left(1-\frac{2}{\pi}\right)(M-K)+\sigma_n^2}}\right) \underset{\text{high SNR}}{\longrightarrow} 2Q\left(\sqrt{\frac{\frac{2}{\pi}}{1-\frac{2}{\pi}}\left(\frac{M}{K}-1\right)}\right)$$



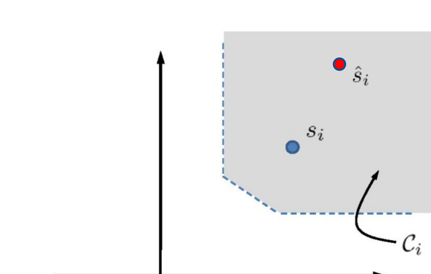




We can show that $\mathbf{R}_{\hat{s}s} \simeq \frac{2\sigma}{\sqrt{\pi}}\sqrt{\frac{M}{K}}\mathbf{I}_K$

As M/K grows, symbols move farther from decision boundaries

ML encoding overconstrains the problem if the desired signal at the receiver is digital. For example, if the elements of $\mathbf{s}$ should be QPSK (e.g., due to one-bit quantization), then all we need is that s_i lie in the right decision region $\mathcal{C}_i$.



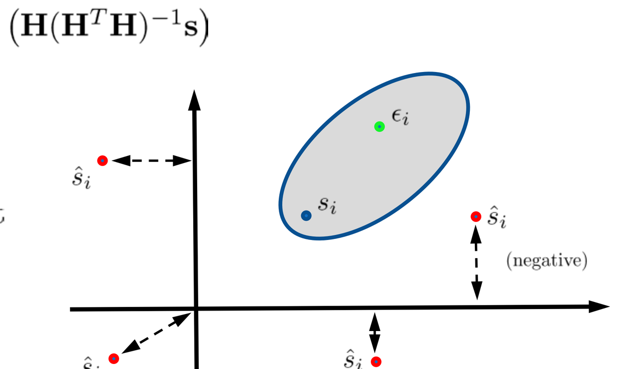
Quantization ZF precoder $\mathcal{Q}(\mathbf{P}\mathbf{s}) = \mathcal{Q}(\mathbf{H}(\mathbf{H}^T\mathbf{H})^{-1}\mathbf{s})$ eliminates scaling due to $(\mathbf{H}^T\mathbf{H})^{-1} \to \beta\mathbf{I}$ scaling which "props up" weak channels and scales down strong channels in order to enforce $\mathbf{s} \simeq \mathbf{H}^T\mathbf{x}$.

Perturbed Quantized ZF Precoding

1-bit Quantized ZF: $\mathbf{x} = \mathcal{Q}\left(\mathbf{H}(\mathbf{H}^T\mathbf{H})^{-1}\mathbf{s}\right)$

- Idea: All we need is a correct detection at the receivers … we don't necessarily need $\mathbf{s} \simeq \mathbf{H}^T\mathbf{x}$
- Instead of zero-forcing to $\mathbf{s}$, try to zero-force to a better point in the correct detection region
- "better" not defined by distance to $\mathbf{s}$, but rather by distance to avoid probability of error

$$\min_{\boldsymbol{\epsilon}} \; d\left(\mathbf{s}, \mathbf{H}^T\hat{\mathbf{x}}\right) \quad \text{s.t.} \quad \hat{\mathbf{x}} = \mathcal{Q}\left(\mathbf{H}(\mathbf{H}^T\mathbf{H})^{-1}(\mathbf{s}+\boldsymbol{\epsilon})\right), \; \boldsymbol{\epsilon} \in \mathcal{S}$$



$\hat{s}_i = \left(\mathbf{H}^T\mathbf{x}\right)_i$

- $M = 128$
- $K = 8$ users
- Rayleigh fading
- Perfect CSI