VLSI Design of a Nonparametric Linear MMSE Equalizer (NOPE)

Gulnar Mirza (Advised by R. Ghods, C. Jeon, and Prof. C. Studer)

Massive Multi-User Multiple-Input Multiple-Output (MU-MIMO)

- Operating principle: Tens of users communicate in the same time-frequency resource with a base-station having hundreds of antennas:

$$\mathbf{y} = \mathbf{H}\mathbf{s} + \mathbf{n}$$

- $\mathbf{y} \in \mathbb{C}^B$ received vector at base station
- $\mathbf{H} \in \mathbb{C}^{B\times U}$ MIMO channel matrix
- $\mathbf{s} \in \mathcal{O}^U$ transmit vector; $\mathcal{O}$ constellation set (e.g., 16-QAM)
- $\mathbf{n} \in \mathbb{C}^B$ noise; i.i.d. zero-mean Gaussian with variance N_0

- Massive MU-MIMO offers improved spectral efficiency, coverage, and range compared to traditional small-scale MIMO technology
- Massive MU-MIMO will be a key technology in 5G wireless systems

Data Detection in Massive MU-MIMO

- Goal: Recover U-dimensional vector $\mathbf{s} \in \mathcal{O}^U$ with constellation $\mathcal{O}$ from the standard MIMO input-output relation $\mathbf{y} = \mathbf{H}\mathbf{s} + \mathbf{n}$
- Maximum-likelihood (ML) detection: $\hat{\mathbf{s}}^{\text{ML}} = \arg\min_{\mathbf{s}\in\mathcal{O}^U} \|\mathbf{y} - \mathbf{H}\mathbf{s}\|$

The ML detection problem is NP-hard → **prohibitive complexity**

Solution: Approximate Data Detection using Linear Equalization

- Most prominent equalization method is linear minimum mean-square error (L-MMSE) equalization followed by quantization:

$$\hat{\mathbf{s}} = (\mathbf{H}^{\text{H}}\mathbf{H} + N_0/E_s\mathbf{I})^{-1}\mathbf{H}^{\text{H}}\mathbf{y}$$

$$\hat{\mathbf{s}}^{\text{MMSE}} = \mathcal{Q}_{\mathcal{O}}(\hat{\mathbf{s}})$$

- Linear methods perform near-optimal for massive MU-MIMO [1]

L-MMSE requires knowledge of signal-to-noise ratio (SNR) E_s/N_0

- Knowledge of SNR requires estimation blocks or manual tuning
- Signal and noise variances may change over time (fading)

NOPE: Nonparametric Linear MMSE Equalizer [1]

We develop a **nonparametric** data detection algorithm for massive MU-MIMO that requires no knowledge of the SNR

- Algorithm builds on mismatched LAMA [2] with a Gaussian prior, which requires knowledge of the true prior (e.g., 16-QAM constellation)
- We use Stein's unbiased risk estimate (SURE) [3] to obtain an estimate for the prior using only knowledge of the received vector $\mathbf{y}$
- Advantages of NOPE:
  - Only requires knowledge of the channel matrix $\mathbf{H}$ and received vector $\mathbf{y}$
  - Low computational complexity (avoids a matrix inversion)
  - Robust to fading, system model mismatches, and channel estimation errors

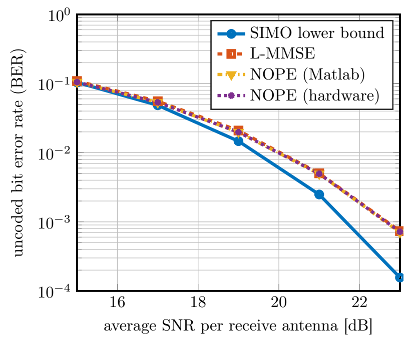
Theorem (NOPE is L-MMSE [1])

Fix $\beta = U/B \leq 1$ and let $U \to \infty$. Assume that $H_{b,u} \sim \mathcal{CN}(0, 1/B)$ and let $t \to \infty$. Then, NOPE achieves the same performance as L-MMSE.

NOPE Algorithm (Simplified)

1. **inputs:** $\mathbf{H} \in \mathbb{C}^{B\times U}$, $\mathbf{y} \in \mathbb{C}^B$, $\beta = U/B$
2. **initialize:** $t = 1$, $\hat{s}^t_\ell = 0$, $\forall \ell$, $\mathbf{r}^t = \mathbf{y} - \mathbf{H}\hat{\mathbf{s}}^t$
3. **for** $t = 1, 2, \ldots$
4. $\quad \mathbf{z}^t = \hat{\mathbf{s}}^t + \mathbf{H}^H\mathbf{r}^t$ (compute signal estimate)
5. $\quad \gamma^t = \frac{1}{\beta}\frac{\|\mathbf{z}^t\|_2^2}{\|\mathbf{r}^t\|_2^2} - 1$ (estimate SURE parameters)
6. $\quad \hat{\mathbf{s}}^{t+1} = \frac{\gamma^t}{\gamma^t+1}\mathbf{z}^t$ (compute L-MMSE estimate)
7. $\quad \mathbf{r}^{t+1} = \mathbf{y} - \mathbf{H}\hat{\mathbf{s}}^{t+1} + \beta\mathbf{r}^t\frac{\gamma^t}{\gamma^t+1}$ (compute residual error)
8. **end**
9. **outputs:** L-MMSE estimate $\hat{\mathbf{s}}^{t+1}$

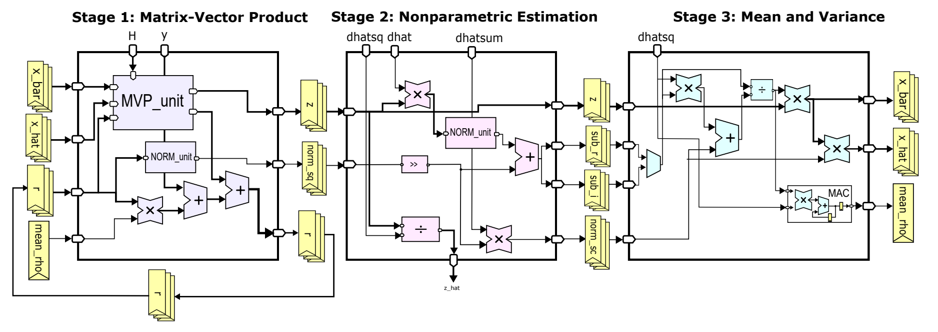
Fixed-Point Bit Error-Rate Performance



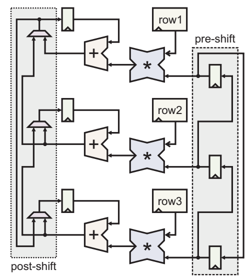
- Bit error-rate simulation in a $B = 64$ and $U = 16$ massive MU-MIMO system
- Channel matrix $\mathbf{H}$: Entries require 0 integer bits, 10 fraction bits
- Receive vector $\mathbf{y}$: Entries require 9 integer bits, 9 fraction bits
- L-MMSE estimate $\hat{\mathbf{s}}^{t+1}$: Entries require 2 integer bits, 10 fraction bits

NOPE performance matches that of L-MMSE in finite-dimensional systems, even with highly-optimized fixed-point arithmetic

VLSI Architecture of NOPE



- Architecture consists of three stages: (1) matrix-vector product, (2) nonparametric estimation, and (3) mean and variance computation
- Coarse-grained pipeline interleaving is used to increase parallelization, i.e., each stage processes an independent equalization problem



- Matrix vector product (MVP) engine performs $\mathbf{H}\mathbf{r}^t$ and $\mathbf{H}\hat{\mathbf{s}}^{t+1}$ in a column-parallel manner while avoiding memory-access contentions
- We use a custom high-throughput division unit that uses Newton-Raphson iterations
- We use pipelining within each stage to optimize the maximum clock frequency

Preliminary Synthesis Results in 40nm CMOS (TSMC)

Silicon area [mm^2]	0.40
Maximum clock frequency [MHz]	300
Maximum throughput [Mb/s]	100
Efficiency [Mb/s/mm^2]	250

- No comparable VLSI designs in the literature!
- I am still optimizing the architecture with pipelining, retiming, and fixed-point parameter tuning; extensive verification ongoing

First VLSI design of a nonparametric equalizer for massive MU-MIMO; results will appear in [4]

References

[1] R. Ghods, C. Jeon, **G. Mirza**, A. Maleki, C. Studer, "Optimally-Tuned Nonparametric Linear Equalization for Massive MU-MIMO Systems," to be presented at IEEE Intl. Symp. Inf. Theory (ISIT), 2017

[2] C. Jeon, A. Maleki, C. Studer, "On the Performance of Mismatched Data Detection in Large MIMO Systems," IEEE Intl. Symp. Inf. Theory (ISIT), 2016

[3] C. M. Stein, "Estimation of the mean of a multivariate normal distribution," The Annals of Stat., 1981

[4] **G. Mirza**, R. Ghods, C. Jeon, A. Maleki, C. Studer, "VLSI Design of a Nonparametric Equalizer for Massive MU-MIMO," to be presented at Asilomar Conf. on Signals, Systems and Computers, 2017 (invited paper)