Leveraging Quantum Annealing for Large MIMO Processing in Centralized Radio Access Networks

Kyle Jamieson, Princeton University

QUBITS North America 2019

PRINCETON UNIVERSITY

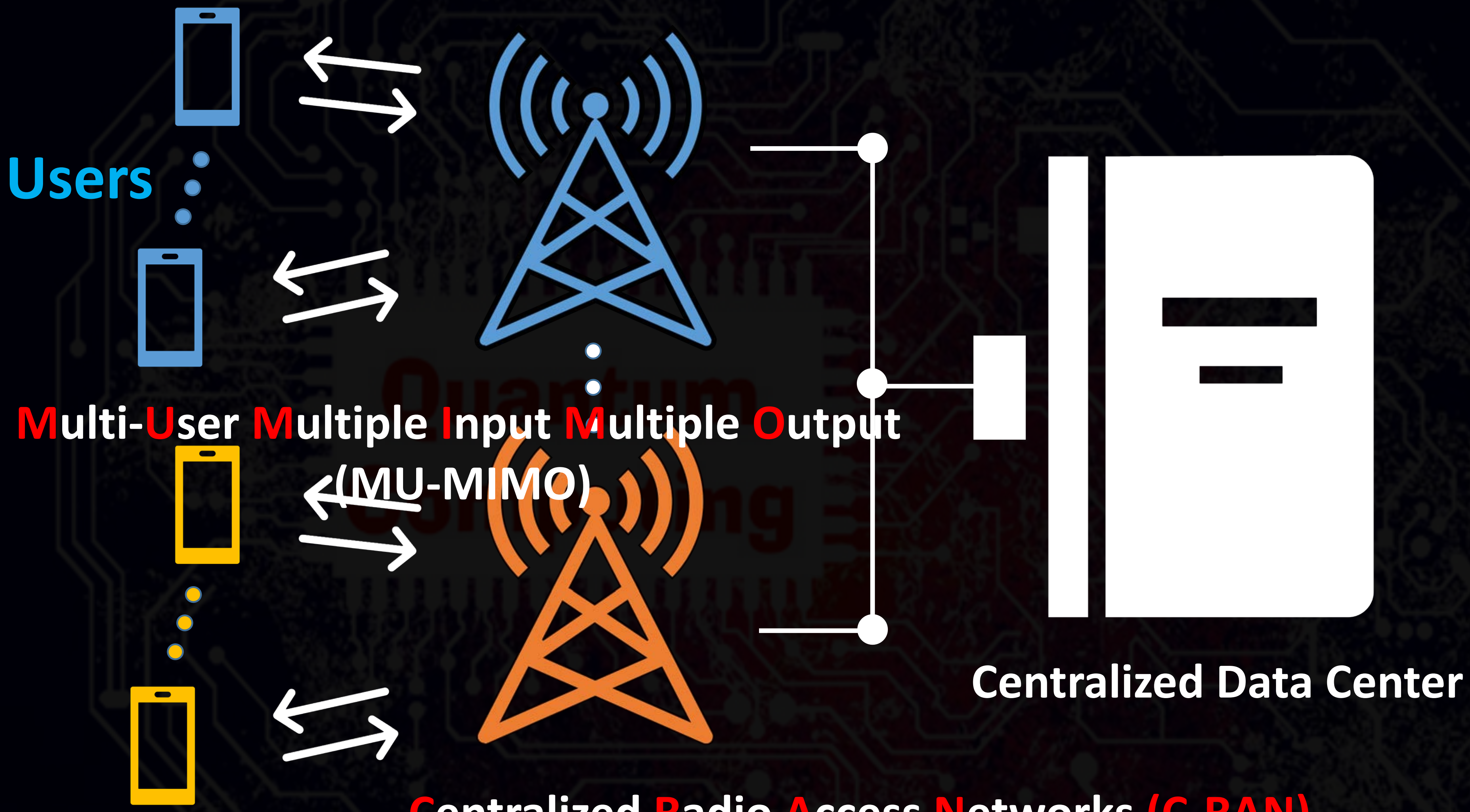
USRA

Minsung Kim, Davide Venturelli, Kyle Jamieson

New Services!

- Global mobile data traffic is increasing **exponentially**
- User demand for high data rate outpaces supply

Drives Wireless Capacity to increase!



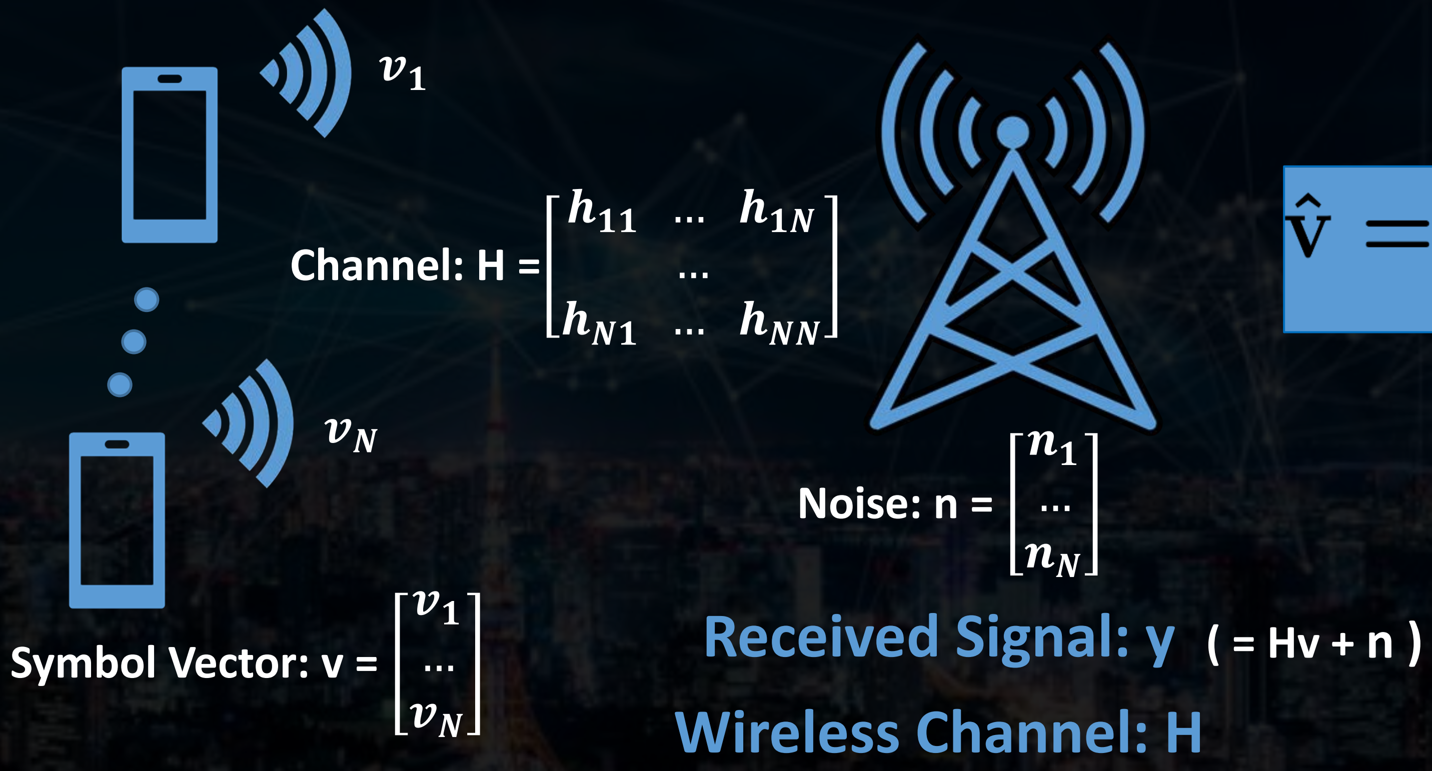
Centralized Radio Access Networks (C-RAN)

MIMO Detection

Demultiplex Mutually Interfering Streams

Maximum Likelihood (ML) MIMO Detection

: Non-Approximate but High Complexity



$$\hat{\mathrm{v}} = \arg \min_{\text{possible v}} \|\mathrm{y} - \mathrm{Hv}\|^2$$

$2^{N \log_2 M}$ possibilities for *N* x *N* MIMO with *M* modulation

Time available for processing is at most 3-10 ms.

Sphere Decoder (SD)
: Non-Approximate but High Complexity

Maximum Likelihood (ML) Detection → Tree Search with Constraints

Reduce search operations but fall short for the same reason

BPSK	QPSK	16-QAM	Complexity (Visited Nodes)
12×12	7×7	4×4	≈ 40 (♡)
21×21	11×11	6×6	≈ 270 (△)
30×30	15×15	8×8	≈ 1900 (×)

Parallelization of SD
[Flexcore, NSDI 17],
[Geosphere, SIGCOMM 14],
...
Approximate SD
[K-best SD, JSAC 06],
[Fixed Complexity SD, TWC 08],
....

Linear Detection

: Low Complexity but Approximate & Suboptimal

v_1

v_N

Channel: H = $\begin{bmatrix} H_{11} & \cdots & H_{1N} \\ & \cdots & \\ H_{N1} & \cdots & H_{NN} \end{bmatrix}$

Noise: n = $\begin{bmatrix} n_1 \\ \cdots \\ n_N \end{bmatrix}$

Symbol Vector: v = $\begin{bmatrix} v_1 \\ \cdots \\ v_N \end{bmatrix}$

Received Signal: y (= Hv + n)

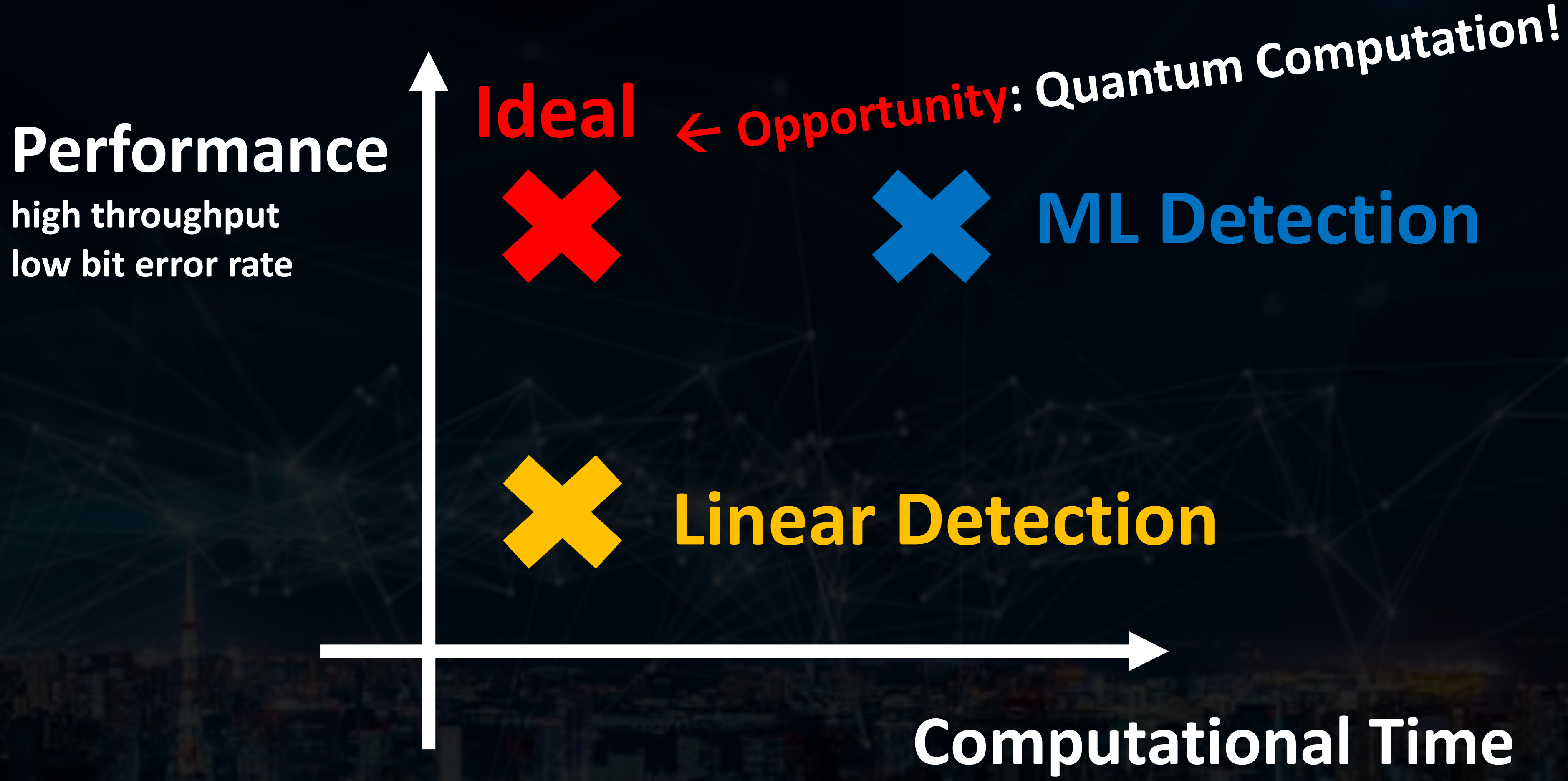
Wireless Channel: H

Zero-Forcing [BigStation, SIGCOMM 13], [Argos, MOBICOM 12], …

Nullifying Channel Effect:

$$\mathbf{H}^{-1}y = \mathbf{H}^{-1}\mathbf{H}\mathbf{v} + \mathbf{H}^{-1}\mathbf{n}$$

Performance Degradation due to Noise Amplification



Ideal: High Performance & Low Computational Time

QuAMax: Main Idea

MIMO Detection

Maximum Likelihood (ML) Detection

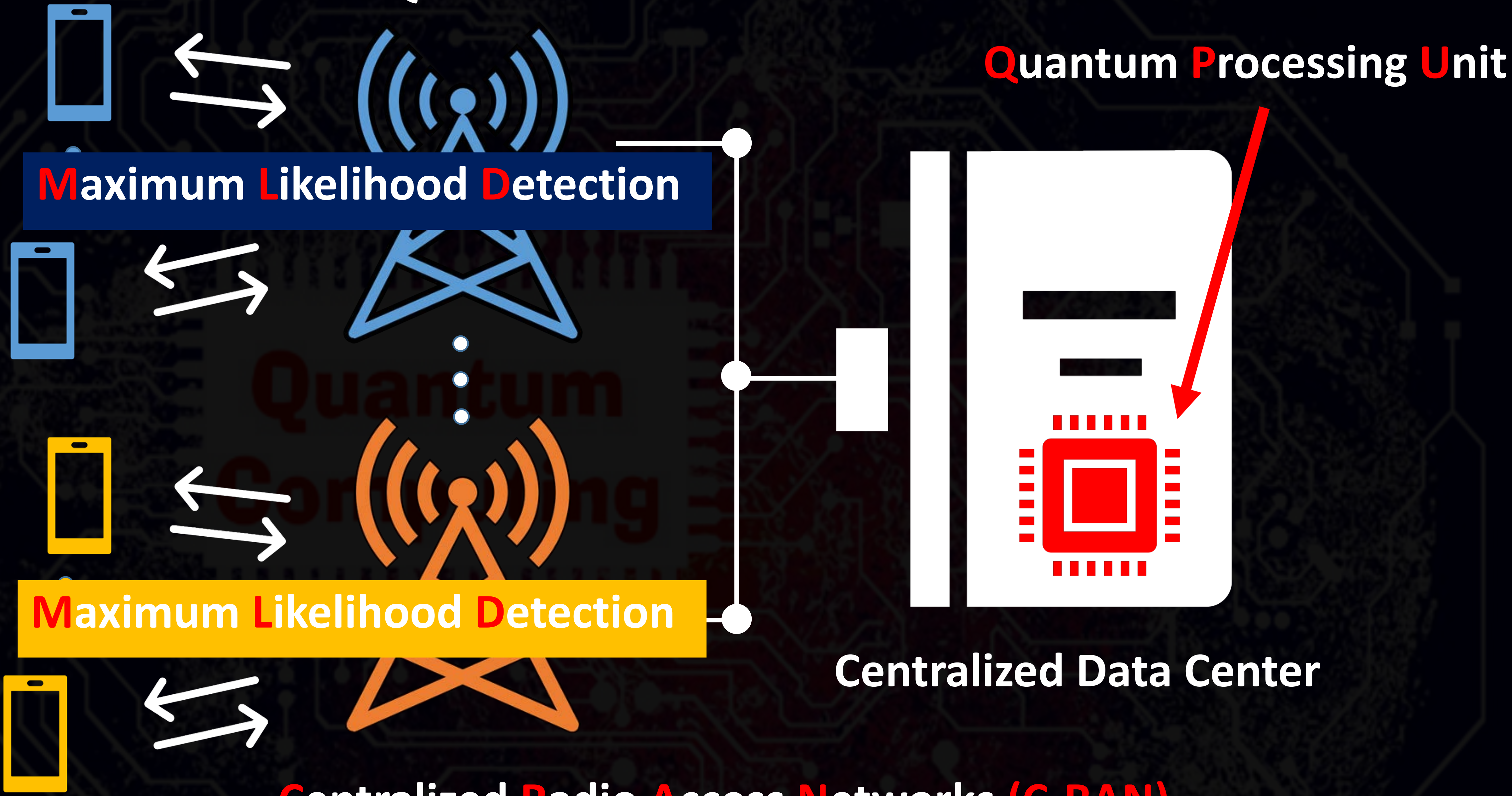
Quantum Computation

Quantum Annealing

Better Performance ?

Motivation: Optimal + Fast Detection = Higher Capacity

QuAMax Architecture



Centralized Radio Access Networks (C-RAN)

Maximum Likelihood Detection

↓

Quadratic Unconstrainted Binary Optimization

↓

Quantum Processing Unit

D-Wave 2000Q
(Quantum Annealer)

Rest of Today

1. QUAMAX: SYSTEM DESIGN

2. QUANTUM ANNEALING & EVALUATION

Key Idea of ML-to-QUBO Problem Reduction

- **Maximum Likelihood MIMO detection:**

$$\hat{\mathbf{v}} = \arg\min_{\mathbf{v}} \|\mathbf{y} - \mathbf{H}\mathbf{v}\|^2$$

- **QUBO Form:**

$$\hat{q}_1, \ldots, \hat{q}_N = \arg \min_{\{q_1, \ldots, q_N\}} \sum_{i \le j}^{N} Q_{ij} q_i q_j$$

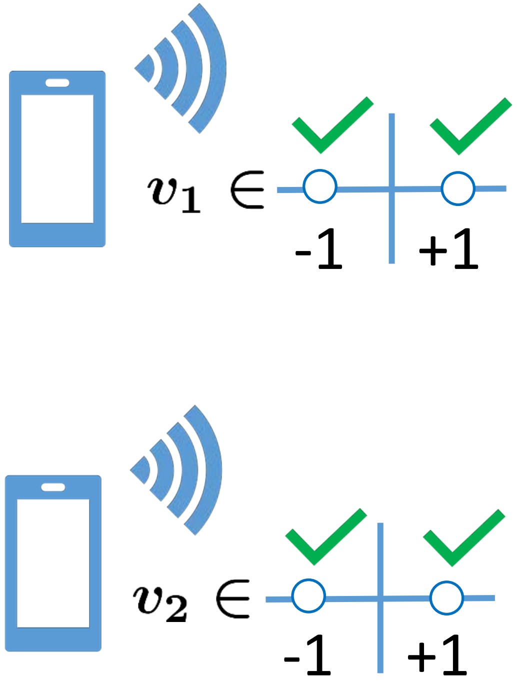
QUBO Form!

The key idea is to represent possibly-transmitted **symbol v** with **0,1 variables.**
If this is **linear**, the expansion of the norm results in **linear & quadratic** terms.

Linear variable-to-symbol transform T

Revisiting ML Detection

Example: 2x2 MIMO with Binary Modulation



Symbol Vector: $\mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}$

Received Signal: y

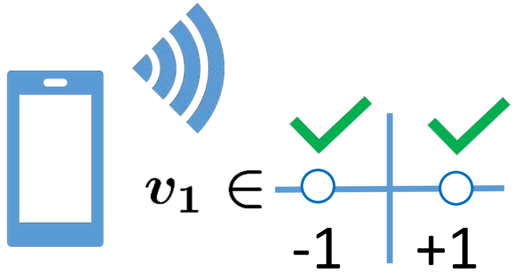
Wireless Channel: H

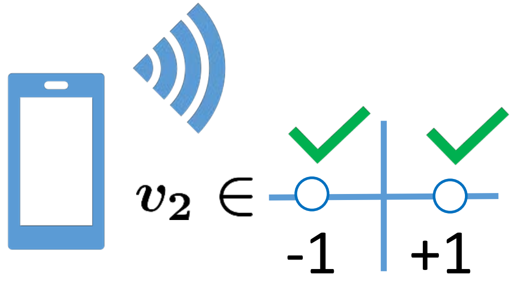
$$\hat{\mathbf{v}} = \arg \min_{\text{possible } \mathbf{v}} \|\mathbf{y} - \mathbf{H}\mathbf{v}\|^2$$

$$\text{possible} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} \in \begin{bmatrix} +1 \\ +1 \end{bmatrix}, \begin{bmatrix} +1 \\ -1 \end{bmatrix}, \begin{bmatrix} -1 \\ -1 \end{bmatrix}, \begin{bmatrix} -1 \\ +1 \end{bmatrix}$$

QuAMax's ML-to-QUBO Problem Reduction

Example: 2x2 MIMO with Binary Modulation





Symbol Vector: $\mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}$

1. Find linear variable-to-symbol transform T:

$$2q_i - 1 \leftrightarrow v_i \qquad \begin{matrix} (\text{if } q_i = 1)\ 2q_i - 1 = +1 \\ (\text{if } q_i = 0)\ 2q_i - 1 = -1 \end{matrix}$$

2. Replace symbol vector v with transform T in $\|\mathbf{y} - \mathbf{H}\mathbf{v}\|^2$:

$$\text{possible} \begin{bmatrix} q_1 \\ q_2 \end{bmatrix} \in \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix} \Leftrightarrow \text{possible} \begin{bmatrix} 2q_1 - 1 \\ 2q_2 - 1 \end{bmatrix} \in \begin{bmatrix} +1 \\ +1 \end{bmatrix}, \begin{bmatrix} +1 \\ -1 \end{bmatrix}, \begin{bmatrix} -1 \\ -1 \end{bmatrix}, \begin{bmatrix} -1 \\ +1 \end{bmatrix}$$

3. Expand the norm $(q^2 = q)$

$$\hat{q}_1, \hat{q}_2 = \underset{q_1, q_2}{\arg\min}\ f_1(\mathbf{H}, \mathbf{y}) q_1 + f_2(\mathbf{H}, \mathbf{y}) q_2 + g_{12}(\mathbf{H}) q_1 q_2$$

$$\mathbf{Q} = \begin{bmatrix} f_1(\mathbf{H}, \mathbf{y}) & g_{12}(\mathbf{H}) \\ 0 & f_2(\mathbf{H}, \mathbf{y}) \end{bmatrix}$$

QUBO Form!

QuAMax's linear variable-to-symbol Transform T

BPSK (2 symbols)	$v_i \leftrightarrow 2q_i - 1$
QPSK (4 symbols)	$v_i \leftrightarrow 2q_{2i-1} - 1 + j(2q_{2i} - 1)$
16-QAM (16 symbols)	$v_i \leftrightarrow 3q_{4i-3} - 2q_{4i-2} - 1 + j(3q_{4i-1} - 2q_{4i} - 1)$

- Coefficient functions *f(H, y)* and *g(H)* are generalized for different modulations.
- Computation required for ML-to-QUBO reduction is insignificant.

Maximum Likelihood Detection

↓

Quadratic Unconstrainted Binary Optimization

↓

Quantum Processing Unit

D-Wave 2000Q
(Quantum Annealer)

Rest of Today

1. QUAMAX: SYSTEM DESIGN

2. QUANTUM ANNEALING & EVALUATION

QuAMax's Performance Metrics

- **One run on QuAMax includes multiple QA cycles.**
  - Number of anneals (N_a) is another input.
- **Solution (state) that has the lowest energy is selected as a final answer.**

Evaluation Metric: How Many Anneals Are Required?

↑ **Target**
Bit Error Rate (BER)

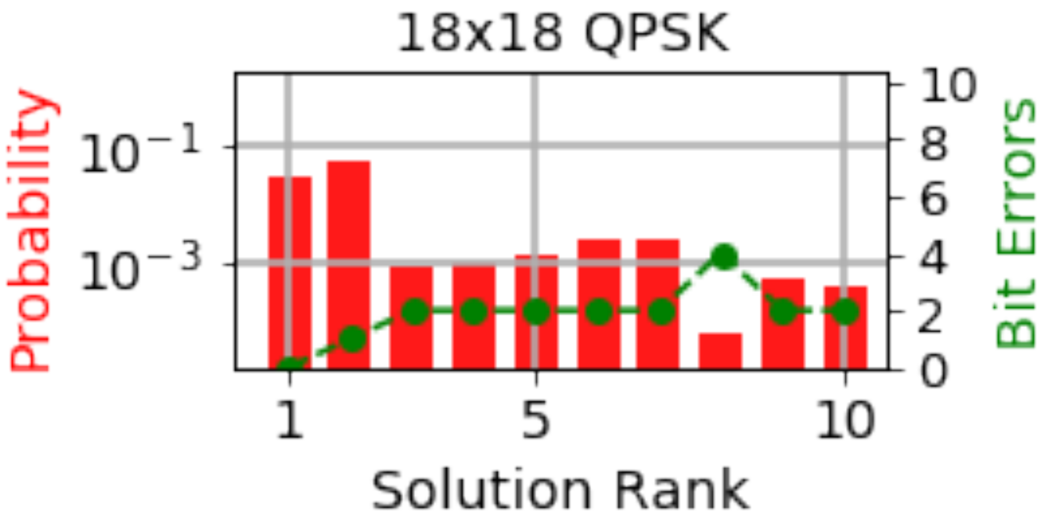
↑ **Solution's Probability**
Empirical QA Results

Experimental Methodology

1. Run enough number of anneals N_a for statistical significance.

2. Sort the L ($\leq N_a$) results in order of QUBO energy.

3. Obtain the corresponding probabilities and numbers of bit errors.



● ● ● L^{th} Solution

Wireless Performance Metric: Bit Error Rate (BER)

QuAMax's BER = BER of the lowest energy state after N_a Anneals

$$\mathbf{E}(BER(N_a)) = \sum_{k=1}^{L}$$ Probability of *k*-th solution being selected after N_a anneals ✖ Corresponding BER of *k*-th solution

=

Probability of [never finding a solution better than k-th solution; finding k-th solution at least once]

This probability depends on number of anneals N_a

Expected Bit Error Rate (BER) as a Function of Number of Anneals (N_a)

QuAMax's Comparison Schemes

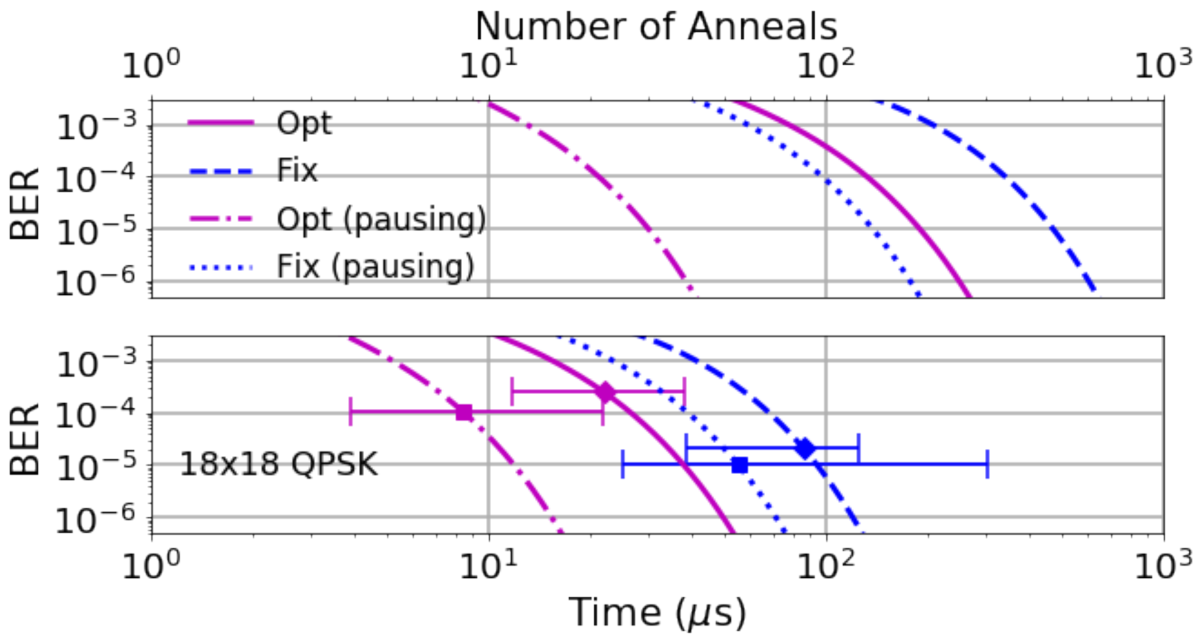
QA parameters: embedding, anneal time, pause duration, pause location, …

- **Opt:** run with optimized QA parameters per instance (Oracle)
- **Fix:** run with fixed QA parameters per classification (QuAMax)

Quantum Compute-Wireless Performance Metric: *Time-to-BER*

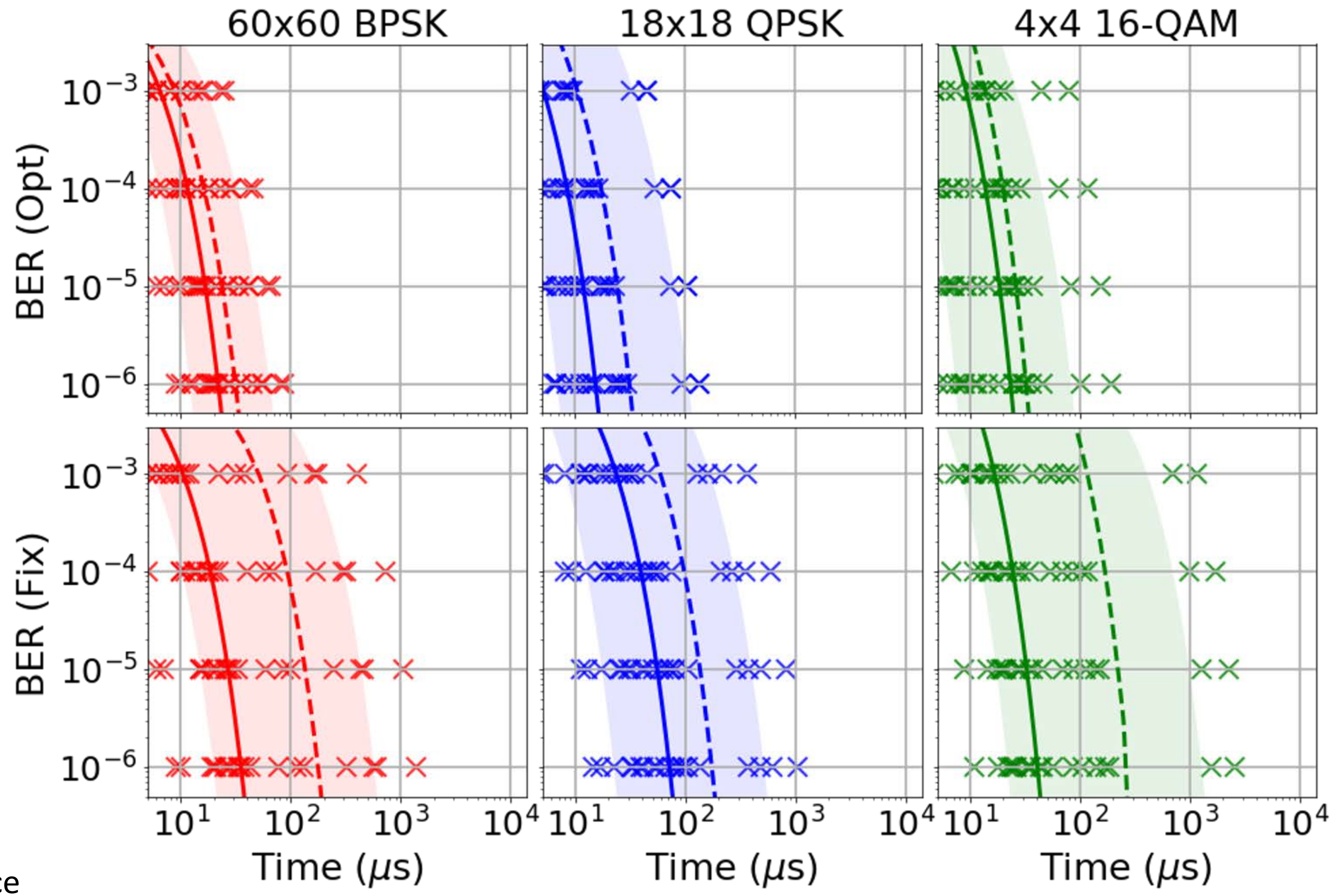
- Opt: run with optimized QA parameters per instance (oracle)
- Fix: run with fixed QA parameters per classification (QuAMax)

Expected Bit Error Rate (BER) as a Function of Number of Anneals (N_a)



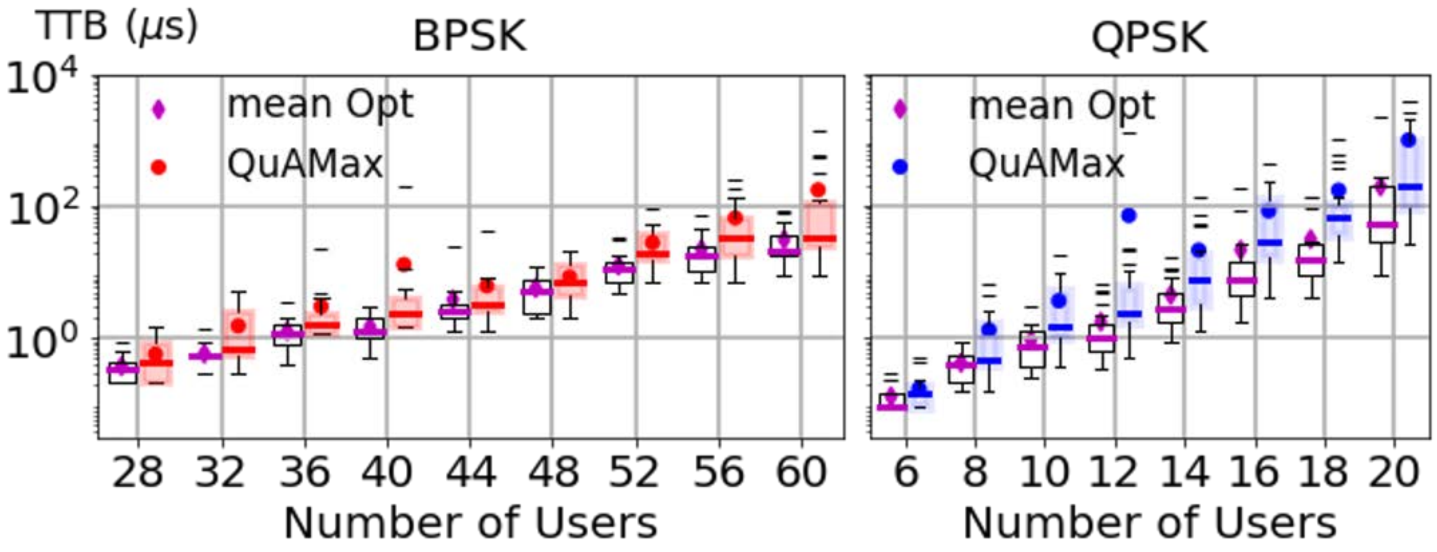
***Time-to-BER* (TTB)**

Time-to-BER for Various Modulations



Lines: Median
Dash Lines: Average
x symbols: Each Instance

QuAMax's Time-to-BER (10^{-6}) Performance

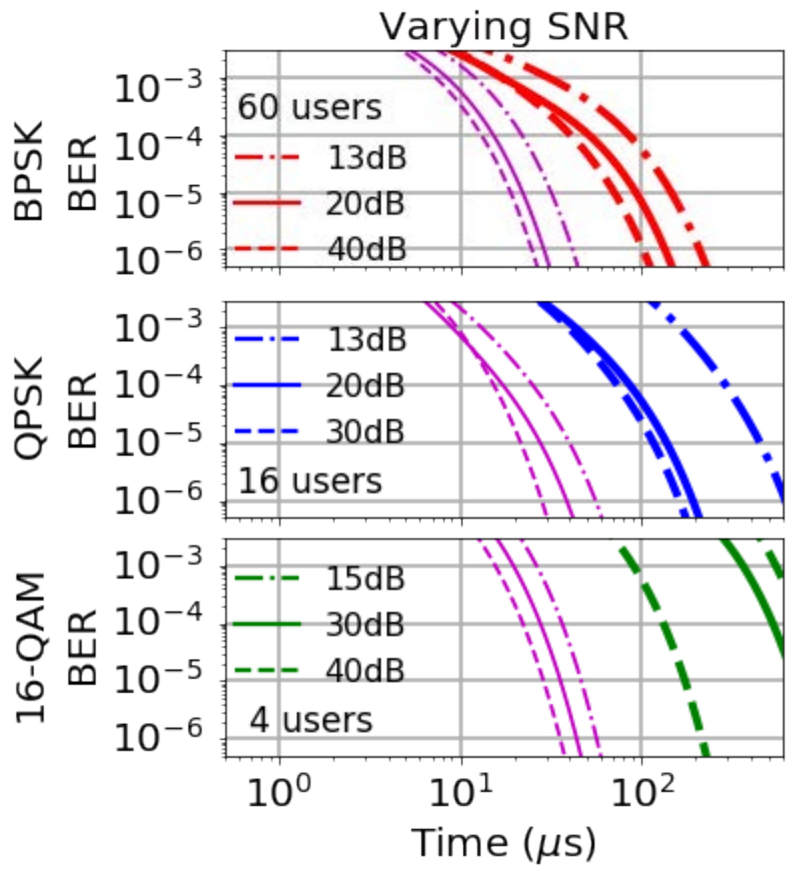


Practicality of Sphere Decoding

BPSK	QPSK	16-QAM	Complexity (Visited Nodes)
12×12	7×7	4×4	≈ 40 (♡)
21×21	11×11	6×6	≈ 270 (△)
30×30	15×15	8×8	≈ 1900 (×)

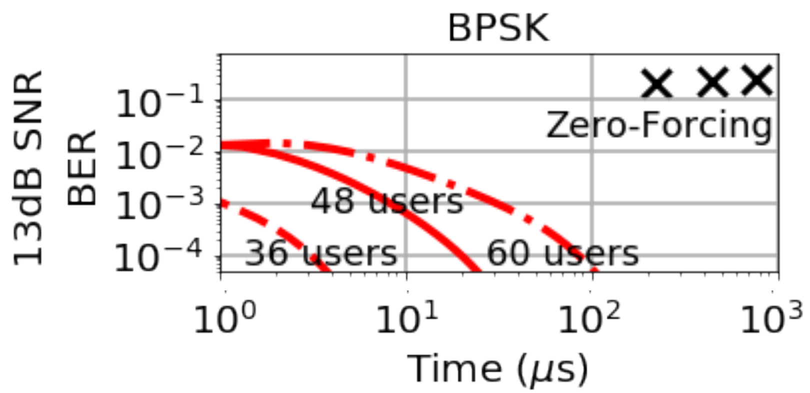
Well Beyond the Borderline of Conventional Computer

QuAMax's Time-to-BER Performance with Noise



Same User Number
Different SNR

- When user number is fixed, higher TTB is required for lower SNRs.



Comparison against Zero-Forcing

- Better BER performance than zero-forcing can be achieved.

Practical Considerations

- Significant Operation Cost:
  About USD $17,000 per year

- Processing Overheads (as of 2019):
  Preprocessing, Read-out Time,
  Programming Time = hundreds of *ms*

D-Wave 2000Q (hosted at NASA Ames)

Future Trends in QA Technology

More Qubits (x2), More Flexibility (x2), Low Noise (x25), Advanced Annealing Schedule, …

Contributions

- First application of QA to MIMO detection
- New metrics: BER across anneals & Time-to-BER (TTB)
- New techniques of QA: Anneal Pause & Improved Range
- Comprehensive baseline performance for various scenarios

Conclusion

- QA may hold the potential to overcome computational limits in wireless networks, but technology & integration yet to mature
- Our work paves the way for quantum hardware and software to contribute to improved performance envelope of Massive MIMO

Supported by

NSF

PRINCETON UNIVERSITY

USRA

NASA