IEEE ComSoc Technical Committee on Green Communications and Computing-
IEEE Transactions on Green Communications and Networking 2nd Joint Seminar

Quantum Computation for MIMO Detection and LDPC Decoding in Wireless Networks

Kyle Jamieson

PRINCETON UNIVERSITY

With collaborators: John Kaewell (Interdigital), Srikar Kasi (Princeton), Abhishek Kumar (Princeton), Minsung Kim (Princeton), Aaron Lott (USRA), Davide Venturelli (USRA)

NSF *Quantum-Enabled Networks* (QENeTs) Project (CNS-1824357, CNS-1824470)

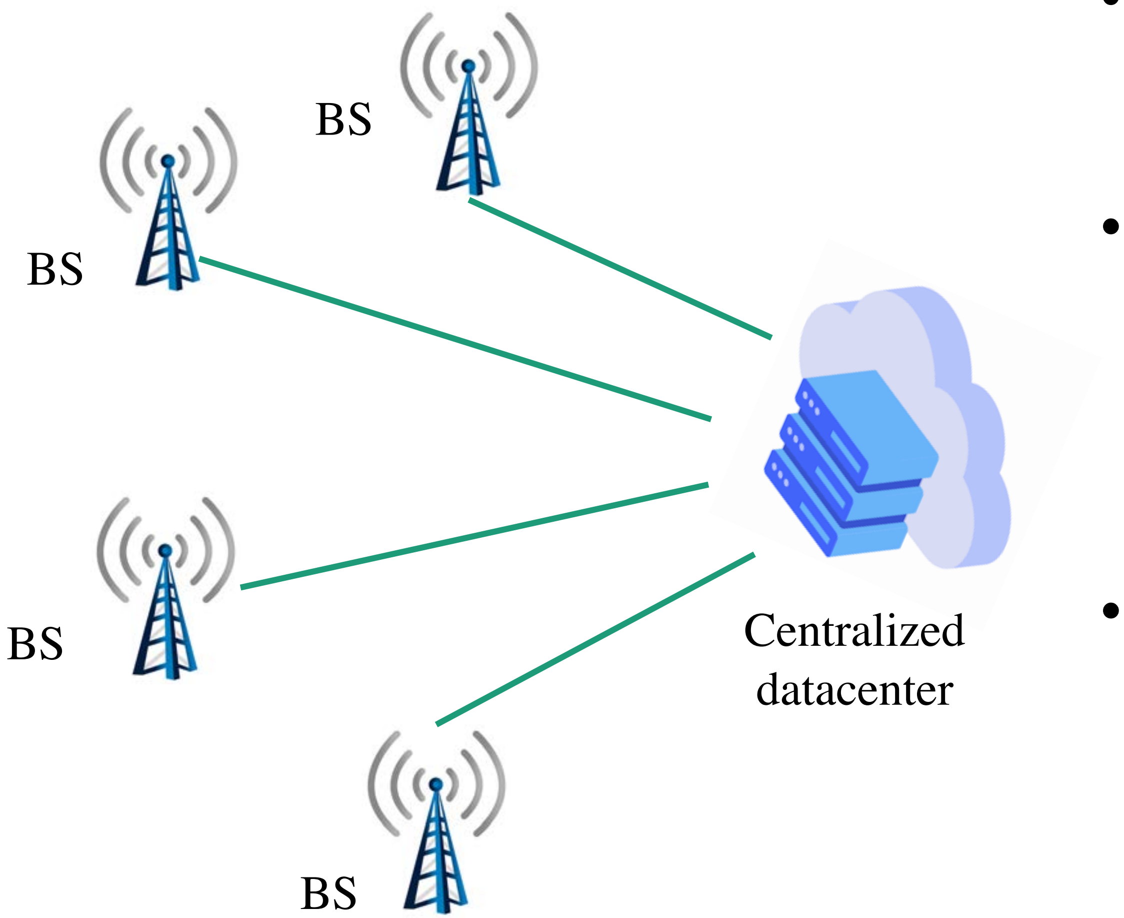
Outline

1. **LDPC decoding**
   - **Quantum LDPC decoder (MobiCom'20) [1]**

2. Large MIMO detection
   - Quantum detection algorithm (SIGCOMM '19) [2]

1. **Srikar Kasi** and Kyle Jamieson**.** Towards Quantum Belief Propagation for LDPC Decoding in Wireless Networks. MobiCom'20.
2. **Minsung Kim, Davide Venturell, Kyle Jamieson**. Leveraging quantum annealing for large MIMO processing in centralized radio access networks. ACM SIGCOMM '19.

Research Goals



- BS' computational processing is being aggregated (*e.g.*, C-RAN)
- ***Centralized Radio Access Network*** locations:
  - Process heavy computation
  - Maintain latency requirements
  - Energy efficiency
- Silicon hardware tradeoffs:
  - Accuracy – Throughput
    - (*e.g.*, bit-precision vs parallelism)
  - Potential of algorithms are limited by hardware

Quantum-Enabled Wireless Networks: Research Goals

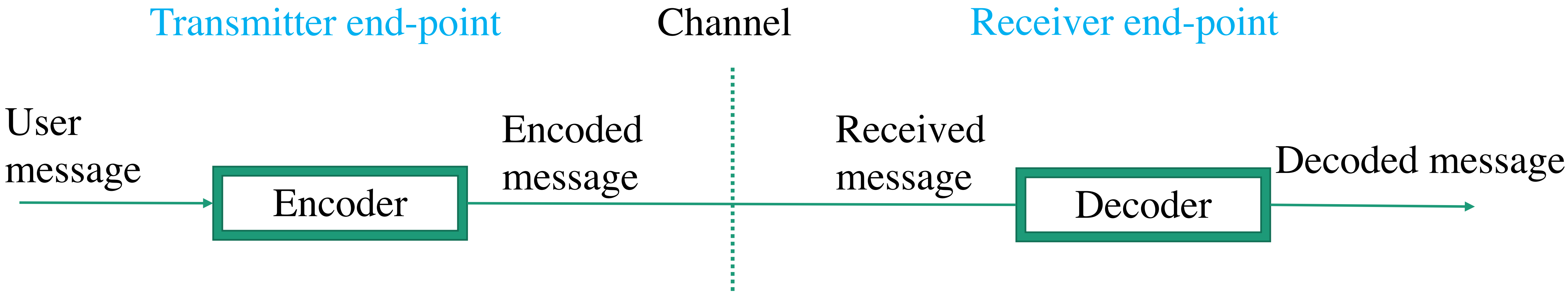
- Explore bottlenecks in Classical computation
  - Algorithms
  - Hardware

- Investigate Quantum computation (Pros and Cons)
  - Quantum Annealing
  - Quantum-Classical Hybrid (future work)
  - Quantum Gate-model (future work)

- Demonstrate head-to-head comparisons
  - Performance, throughput, energy efficiency

Channel coding

- One key component of baseband processing is the error correction code



- Bit Flips (data corruption)
- Error correction codes seek to correct these bit flips (e.g., LDPC, Polar codes etc.)

LDPC codes

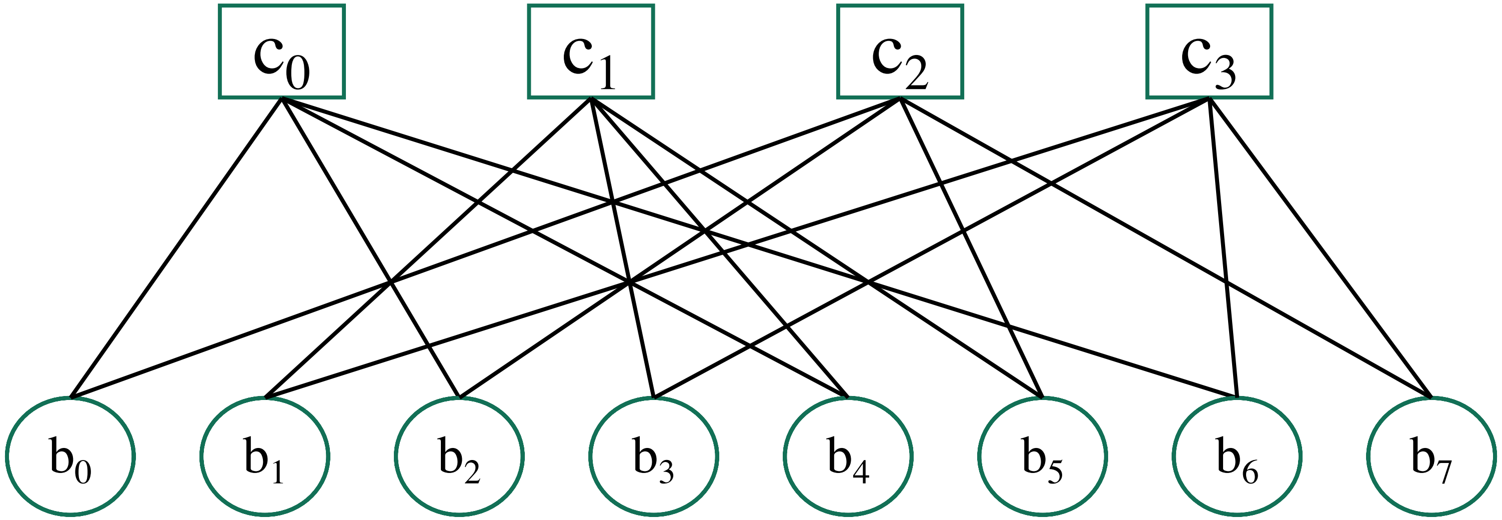
- Low Density Parity Check (LDPC) codes:
  - Capacity-achieving
    - Capacity is max transmission rate for reliable communication
  - Use in protocols:
    - 5G-NR, DVB-S/S2, 802.11, Near-Earth (< 200,000 km), Deep space
  - Fairly simple encoding
  - Computationally complex decoding: *belief propagation (BP)* algorithm

LDPC: Encoding

- Characterized by a parity check matrix $\mathbf{H}_{MxN}$

$\mathbf{H}_{4x8} =$

1	0	1	0	1	0	1	0
0	1	0	1	1	1	0	0
1	0	1	0	0	1	0	1
0	1	0	1	0	0	1	1



- Bit sum at every check node is zero
- Code is (4,2)-regular in example

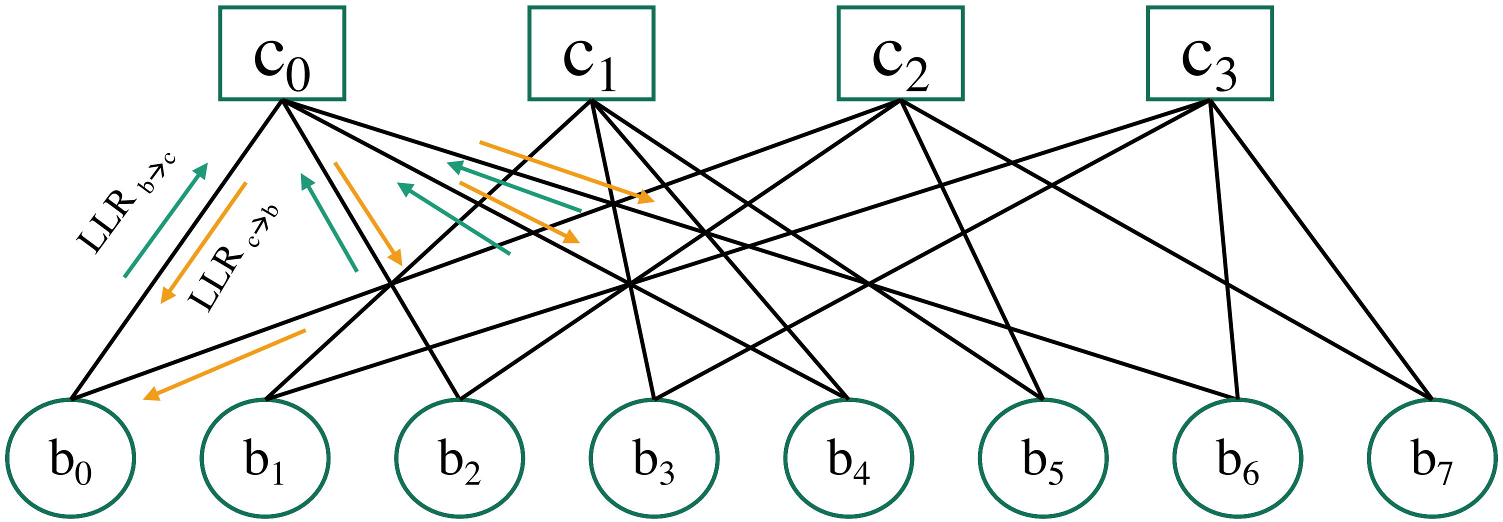
- Encoding:
- Message = **m** (1xK)
- Generator = **G** (KxN)
  - Gauss Jordan elimination **H**

- Encoded = **u** (1xN) = **mG**

- N is block length

LDPC: Decoding

- Belief Propagation (BP) decoding
- LLR = Log-Likelihood ratio

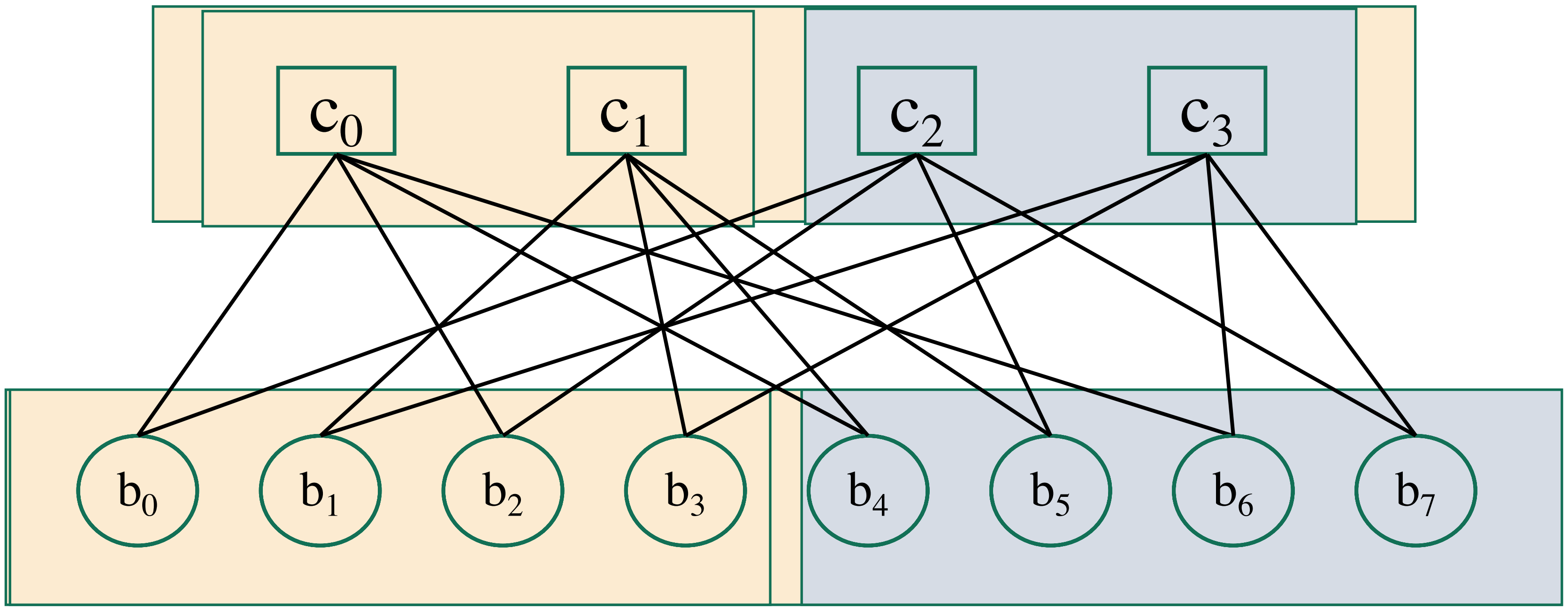


- Iteration:
  1. Initialize (bit to check messages)
  2. Check node computation
  3. Bit node computation

- Iterate sequentially Steps 1-3
- Final iteration LLR gives bit value

LDPC: Decoding

- Hardware (FPGAs/ASICs): Decoding Parallelism



- Fully parallel decoder
- Partially-parallel decoder
- Fully sequential decoder

Problems with classical decoding

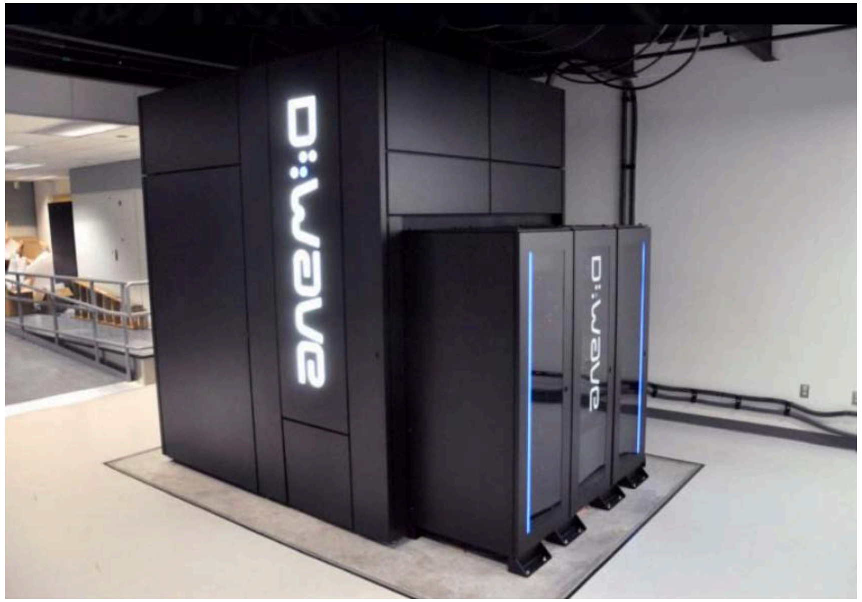
- Decoded via the *belief propagation (BP)* algorithm on FPGA/ASIC hardware
  - Accurate decoding = **high likelihood bit precision** (more resources)
  - Greater throughput = **high decoding parallelism** (more resources)
  - BP algorithm requires several **serial iterations** (impedes throughput)
- Network designers compromise between decoder accuracy and throughput
  - Fully parallel decoders with 8-bit precision (xcvu440 FPGA)
  - A (2,3)-regular code, block length 1944 bits, covers 72% of resources
  - A (4,8)-regular code, block length 2048 bits, exhausts resources

Problems with classical decoding

- But practical protocol block lengths are higher
    - Wi-Fi : up to 1944 bits
    - WiMax: up to 2304 bits
    - DVB-S/S2 : up to 64800 bits
- BP decoders today = *partially-parallel* decoding architectures
- Full potential of LDPC codes is not being realized.

Outline

- Intro: Quantum Annealing
- LDPC decoding: Quantum Belief Propagation
- Experimental Results



D-Wave Quantum Annealer

Host : NASA Ames Research Center

Quantum Annealing

- Analog computation
- Quantum bits
- Heuristic algorithm

- Input : QUBO/Ising model problem
- Output : Lowest energy configuration of the Input

- QUBO = **Q**uadratic **U**nconstrained **B**inary **O**ptimization

Design a QUBO → Map the QUBO onto QA hardware (Embedding) → Solve the problem

QUBO Form

QUBO form: $E = \sum_i h_i q_i + \sum_{i<j} J_{ij} q_i q_j$

Energy

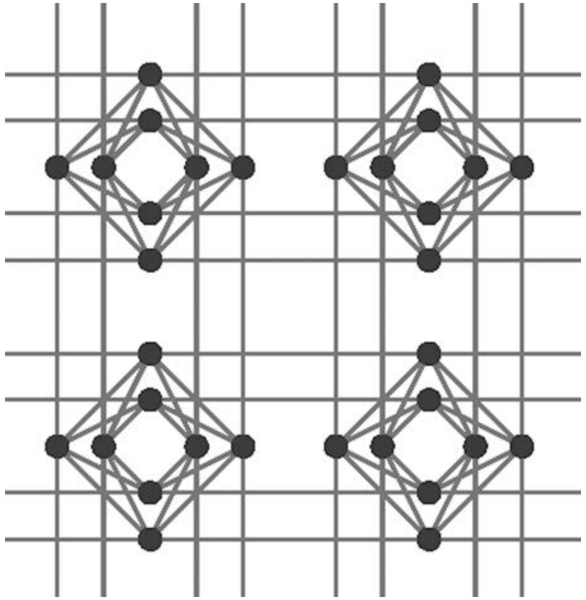
Parameters

Variables (binary)

Programmed onto the QA hardware
using on-chip control circuitry

QA Hardware

QUBO form: $E = \sum_i h_i q_i + \sum_{i<j} J_{ij} q_i q_j$

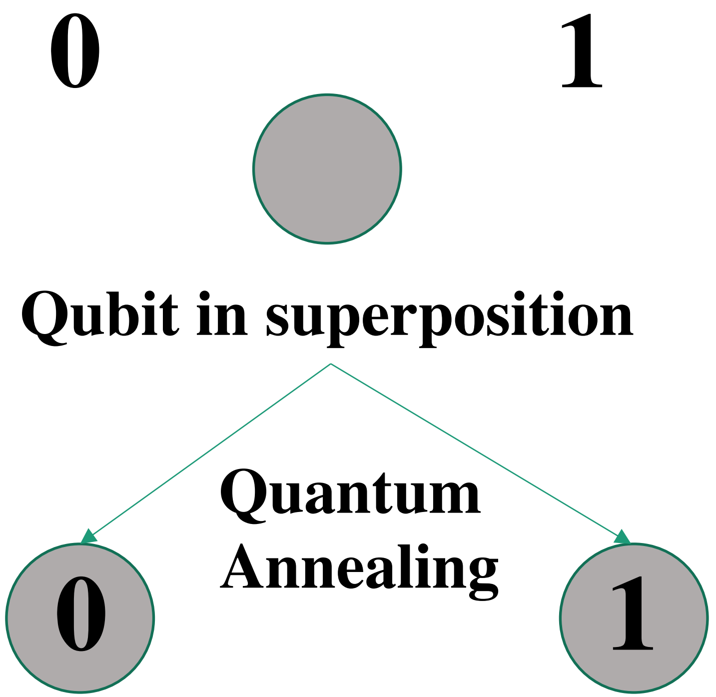


QA Hardware

- Nodes are *qubits*
- Edges are *couplers*
- h_i is programmed onto qubits (external magnetic field)
- J_{ij} is programmed onto couplers (magnetic coupling)

Quantum Annealing (QA)

Superconducting Qubits:



Magnetic couplings:

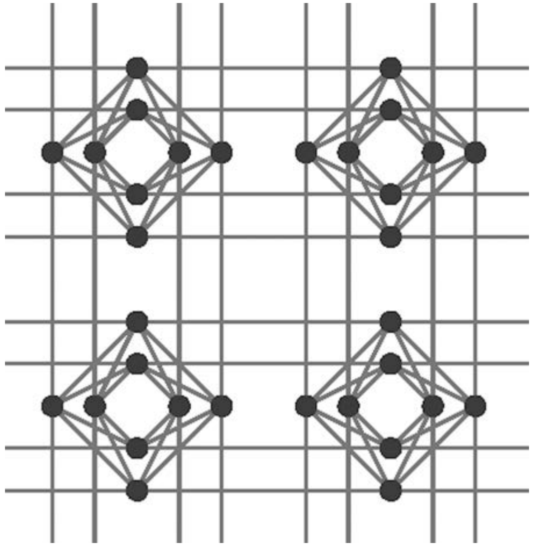
$$b = \{0, 1\}$$

Strong negative coupling: Qubits agree

Strong positive coupling: Qubits disagree

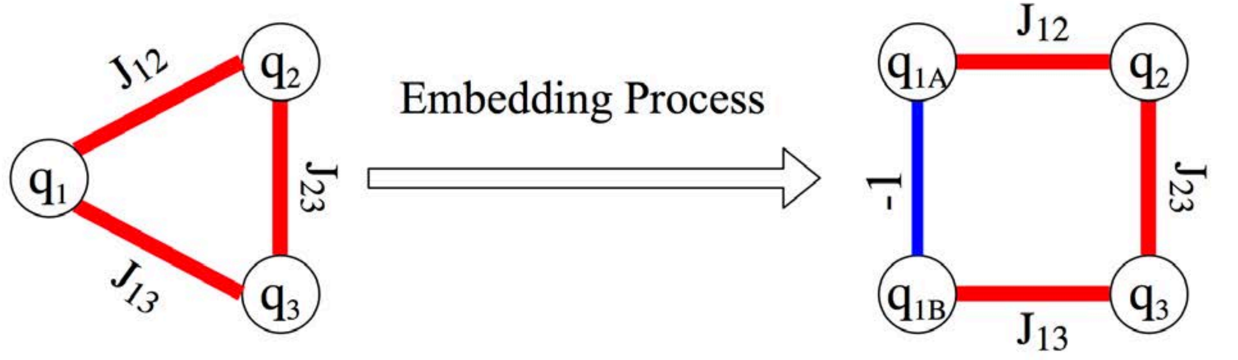
Embedding

QA hardware: Chimera Graph



Mapping a 3-variable fully connected problem

$$E = J_{12}\, q_1 q_2 + J_{13} q_1 q_3 + J_{23} q_2 q_3$$



(a) Before Embedding

(b) After Embedding

Design Contributions

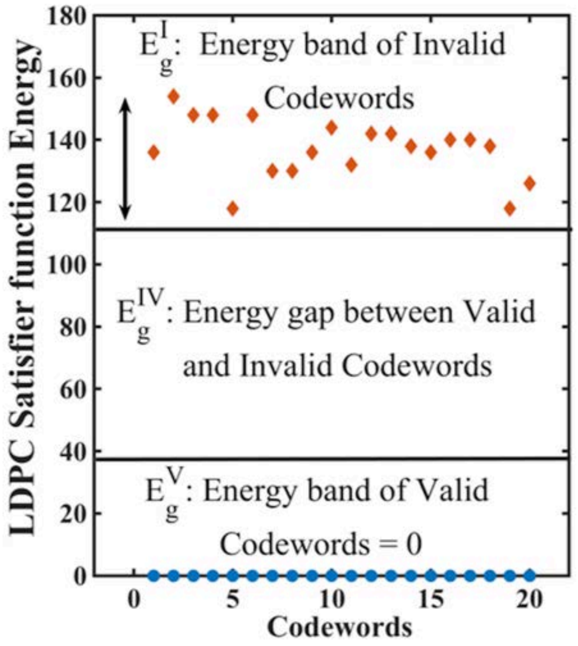
- QUBO Formulation (LDPC codes)
- QA hardware custom Embedding (LDPC codes)

Quantum Belief Propagation (QBP)

QUBO: $\min_{\mathbf{q}} \{ W_1 \sum_i (L_{sat}(c_i)) + W_2 \sum_j (\Delta_j) \}$

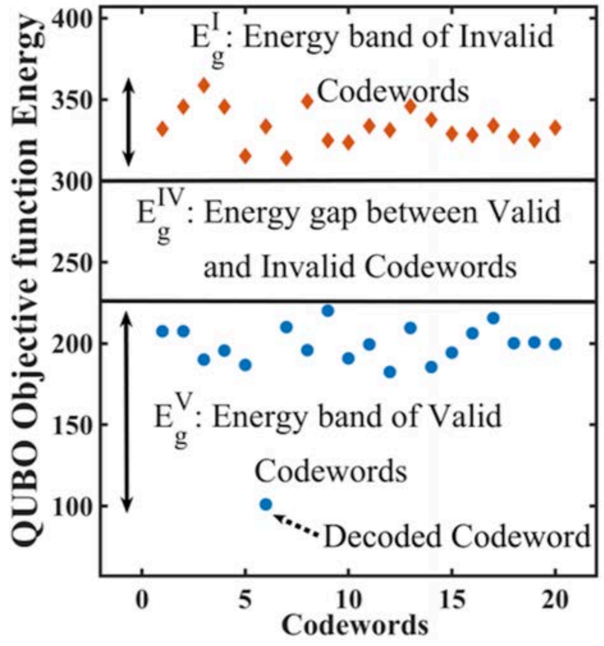
LDPC Satisfier

Ensures encoding



Distance

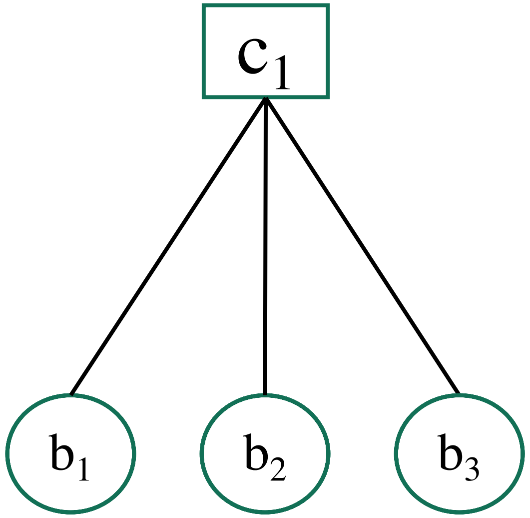
Finds correct answer



LDPC Satisfier function

- Encoding constraint : Modulo-two bit sum is zero at every check node

Example :



- c_1 checks three bits b_1, b_2, b_3
- Encoder Constraint: $b_1 \oplus b_2 \oplus b_3 = 0$ ➔ $b_1 + b_2 + b_3$ must be even
- Qubits for decoding $\{b_1, b_2, b_3\} = \{q_1, q_2, q_3\}$ respectively
- $L_{sat}(c_1) = (q_1 + q_2 + q_3 - 2q_{e1})^2$
- All q_i 's are binary variables. q_{e1} is ancillary.

Distance function

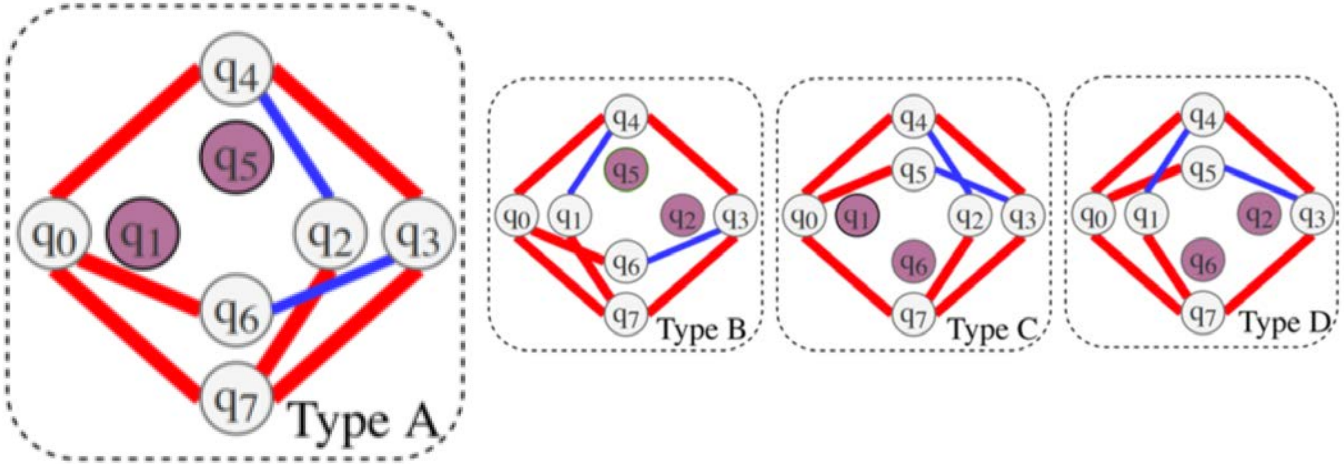
- Distance = proximity of candidate decoding to received information

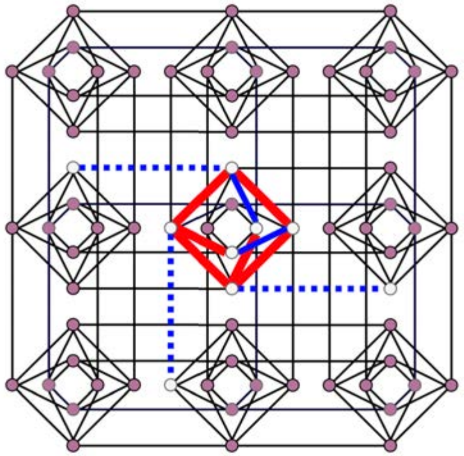
$$\Delta_i = (q_i - Pr(q_i = 1|y_i))^2$$

- qubit q_i corresponds to received bit y_i
- Δ_i → minimal for a q_i in $\{0, 1\}$ → that has greater probability of being transmitted bit
- Probability is computed after soft demapping of received symbols

QBP's Embedding (Level-I)

- Two-Level Embedding.
- Example:
  - $L_{sat}(c_i) = (q_0 + q_4 + q_7 - 2q_{e3})^2$
- Construction:
  - Types A, B, C, D
- Placement:
  - One schema per unit cell
  - Shared bits placed closer

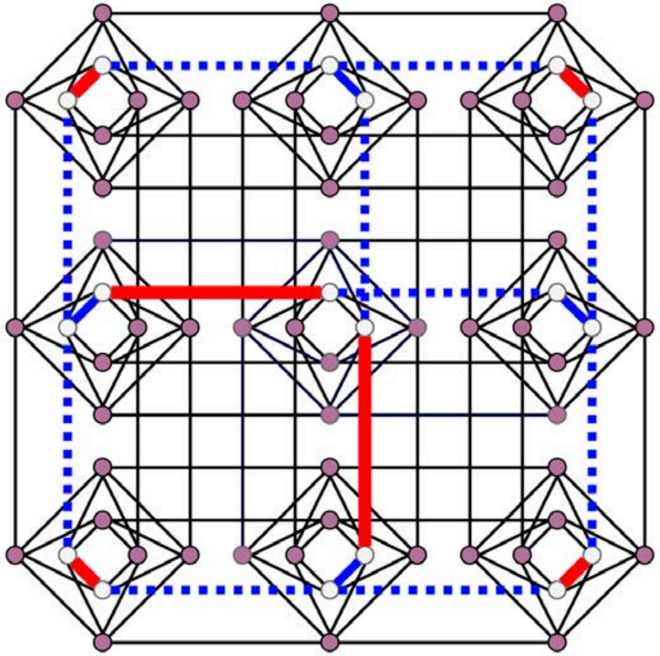




- Level-I embedding

QBP’s Embedding (Level-II)

- Construction:
  - Based on Level-I placement
- Placement:
  - Shared bits placed closer
- QBP scales over entire hardware
- Every qubit is used efficiently.

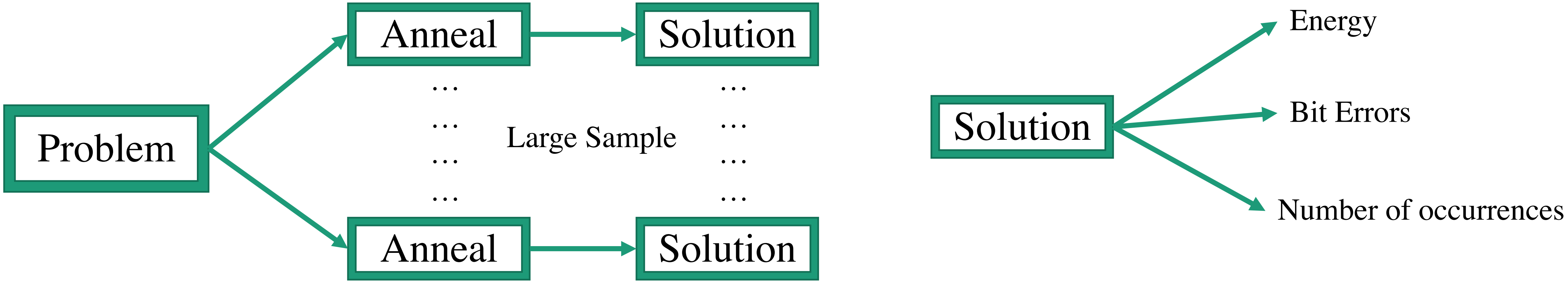


A	B	B
A	B	B
C	D	D

- Level-II embedding

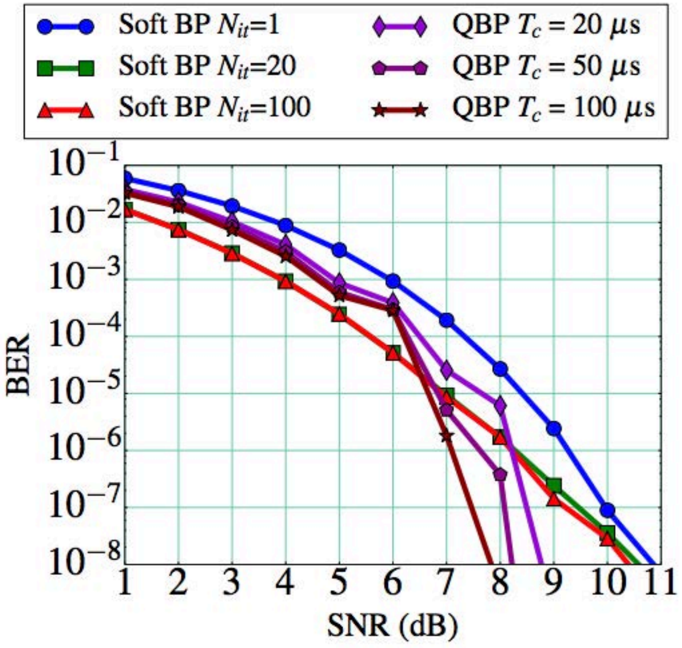
Evaluation

- Hardware: D-Wave 2000Q QA hosted at NASA Ames
- Target LDPC code: (2,3)-regular, block length 420 bits

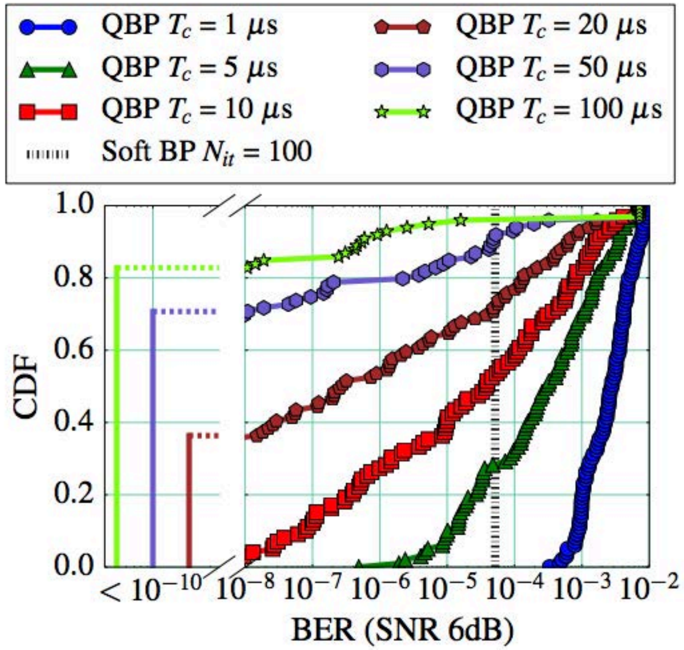


E[BER] = $\Sigma_{solutions}$ Pr (min energy = i^{th} solution) * (bit errors in i^{th} solution) / (total number of bits)

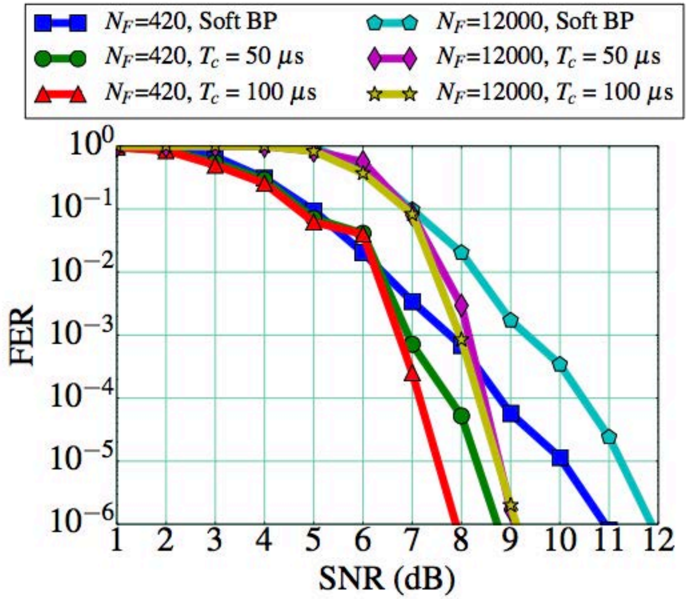
Error Performance



- Average BER



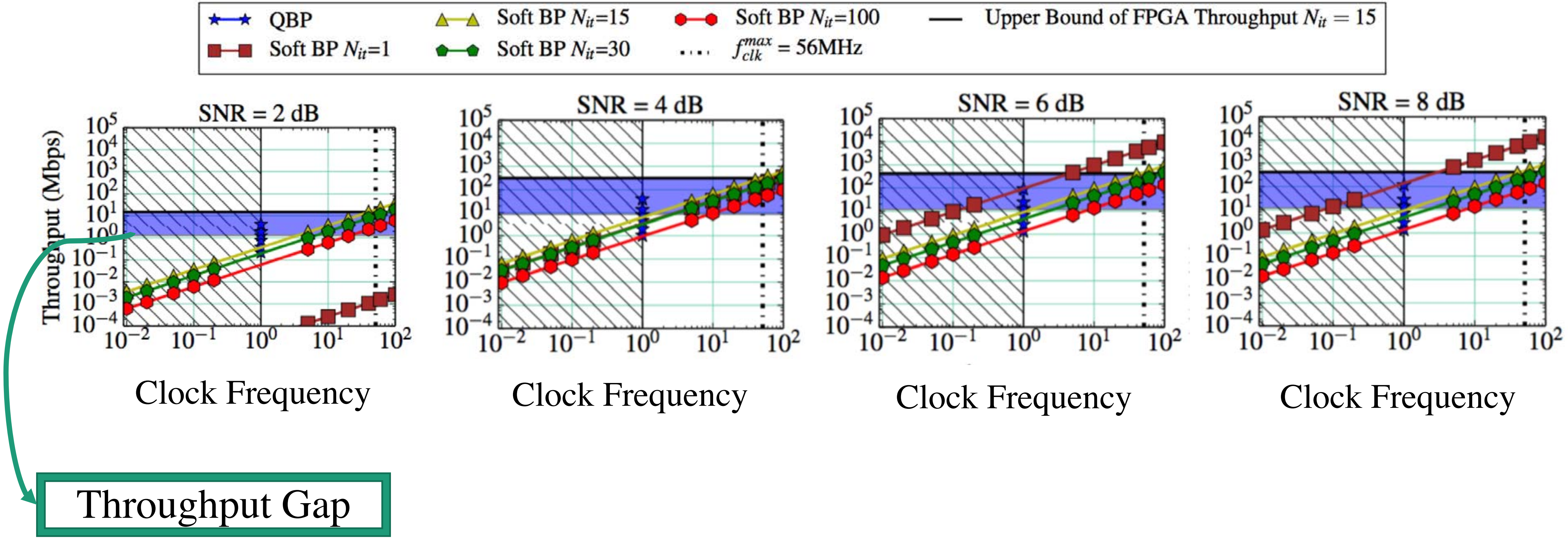
- Distribution of BERs



- Average FER

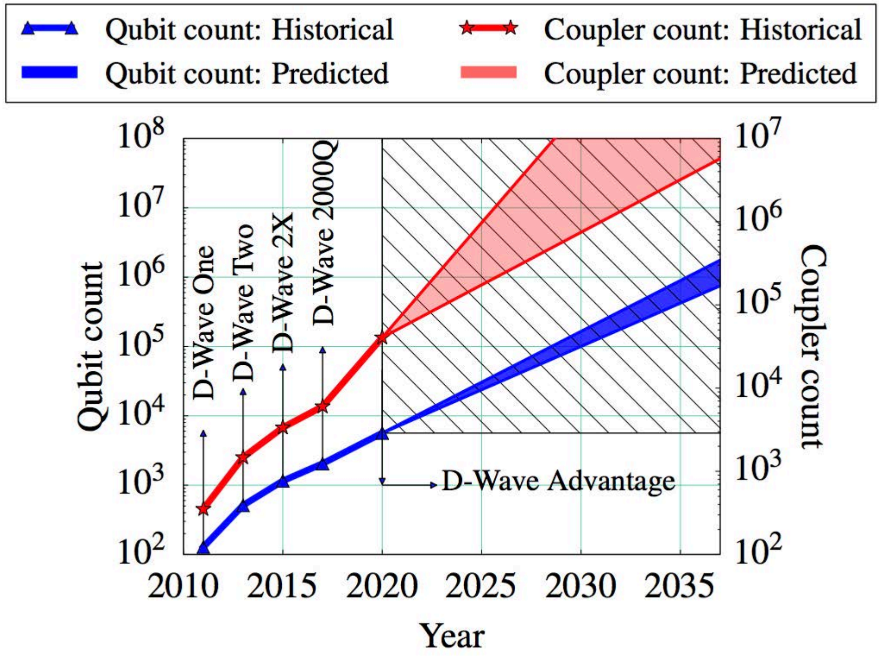
- QBP lags at SNRs < 6 dB, but reaches a 10^{-8} BER at 2-3 dB lower SNR than BP

Throughput



- Net throughput = (1 - FER)* (Processing throughput)

Looking Forward



Extrapolation of resource trends

- Expected 1M qubits by 2035
- QBP decoding block lengths upto ~200,000 bits
- QBP's peak processing throughput reaches 69.4 Gbps

Outline

1. LDPC decoding
   - Quantum LDPC decoder (MobiCom’20) [1]

2. **Large MIMO detection**
   - **Quantum detection algorithm (SIGCOMM ‘19) [2]**

1. **Srikar Kasi** and Kyle Jamieson**.** Towards Quantum Belief Propagation for LDPC Decoding in Wireless Networks. MobiCom’20.
2. **Minsung Kim, Davide Venturell, Kyle Jamieson**. Leveraging quantum annealing for large MIMO processing in centralized radio access networks. ACM SIGCOMM ‘19.

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:

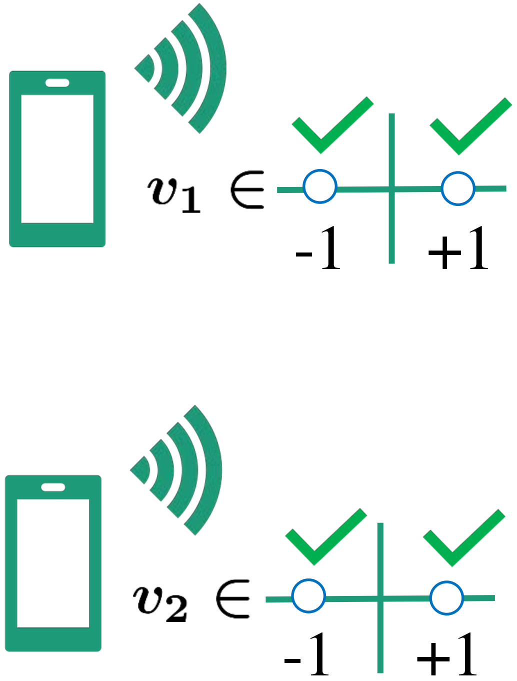
$$\hat{q}_1, \ldots, \hat{q}_N = \arg \min_{\{q_1,\ldots,q_N\}} \sum_{i \le j}^{N} Q_{ij} q_i q_j$$

QUBO Form!

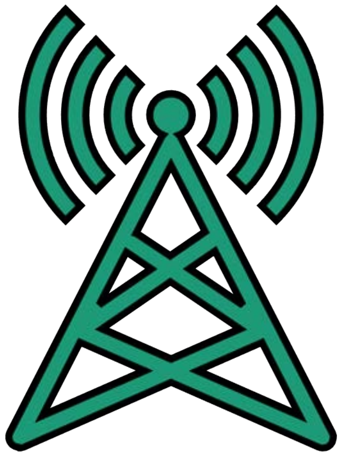
The key idea is to represent possibly-transmitted symbol v with 0,1 variables.
If this is linear, the expansion of the norm results in linear & quadratic terms.

Linear variable-to-symbol transform T

Example: 2x2 MIMO with Binary Modulation



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



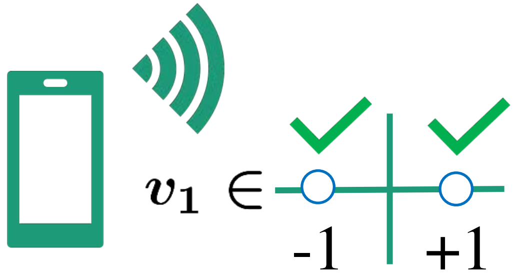
Received Signal: y

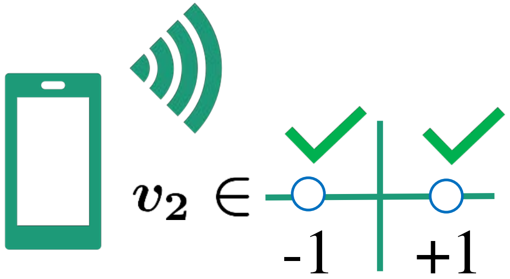
Wireless Channel: H

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

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

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 $(q^2 = q)$

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

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

QUBO Form!

QuAMax's linear variable-to-symbol Transform T

BPSK (2 symbols) $\boldsymbol{v_i \leftrightarrow 2q_i - 1}$

QPSK (4 symbols) $\boldsymbol{v_i \leftrightarrow 2q_{2i-1} - 1 + j(2q_{2i} - 1)}$

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

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

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

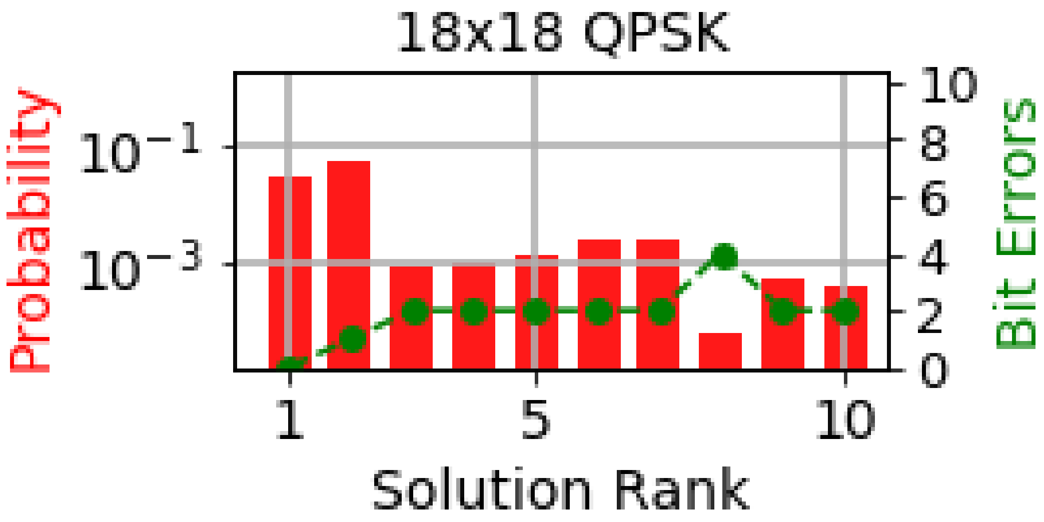
Empirical QA Results

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.



●●● Lth 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

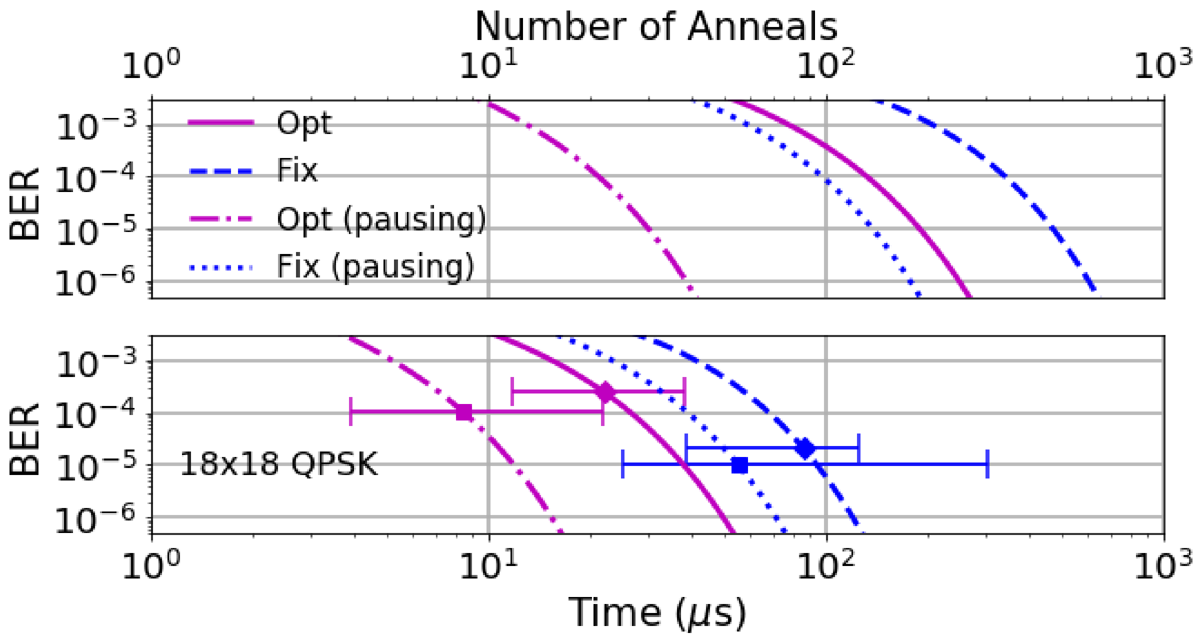
QA parameters: embedding, anneal time, pause duration, pause location, …

- 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

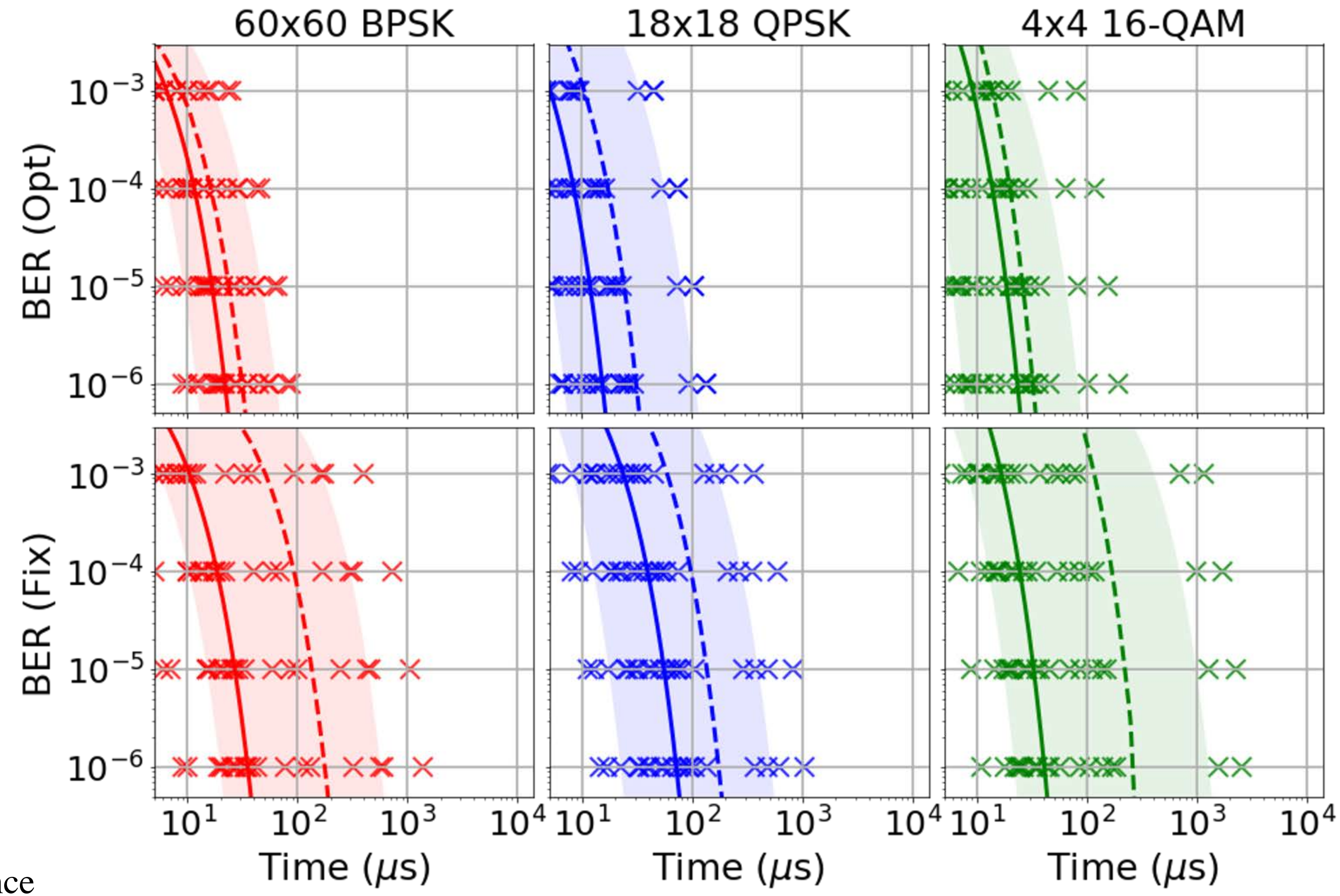
- Opt: run with optimized QA parameters per instance (oracle)
- Fix: run with fixed QA parameters per classification (QuAMax)

Expected Bit Error Rate (BER) as a Function of Number of Anneals ($\boldsymbol{N_a}$)



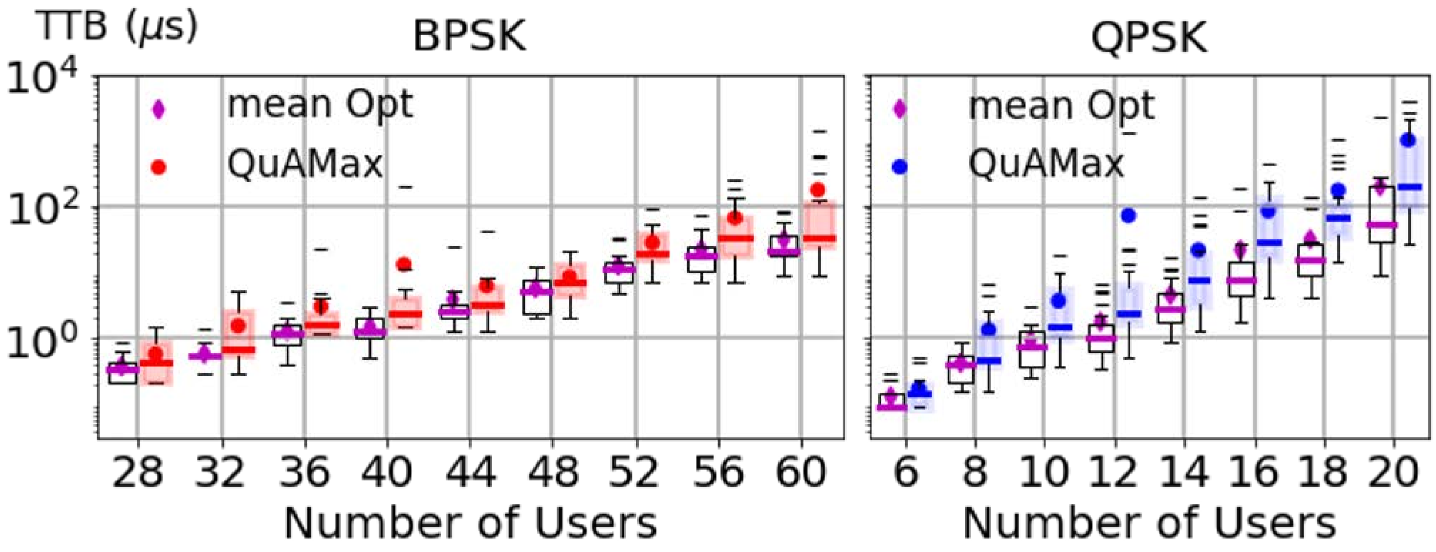
Time-to-BER (TTB)

Time-to-BER for Various Modulations



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

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

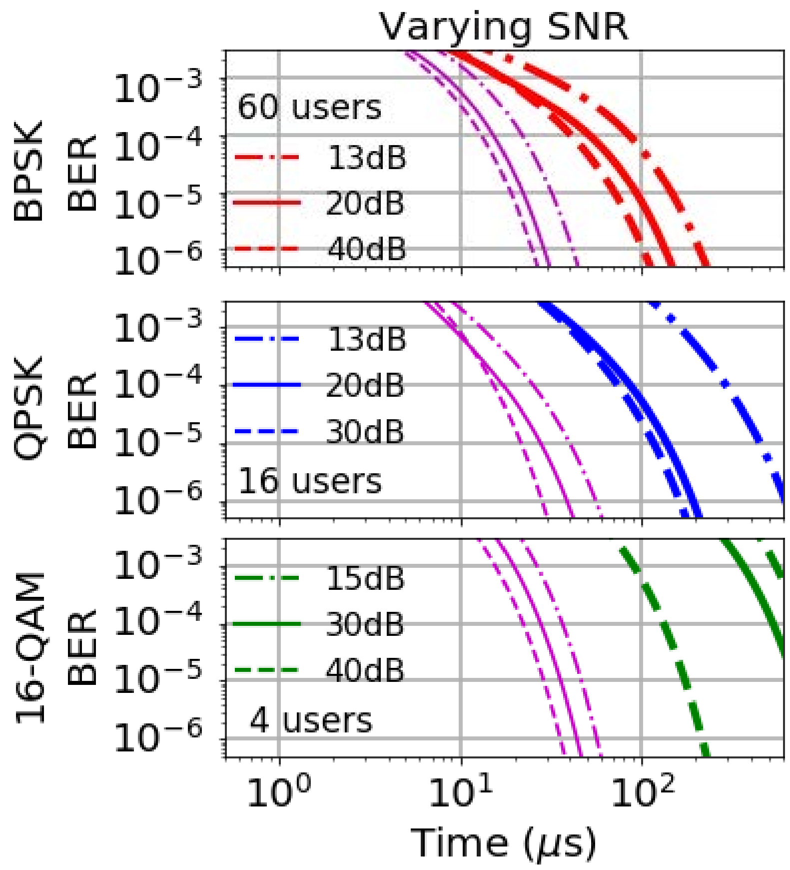


Practicality of Sphere Decoding

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 (×)

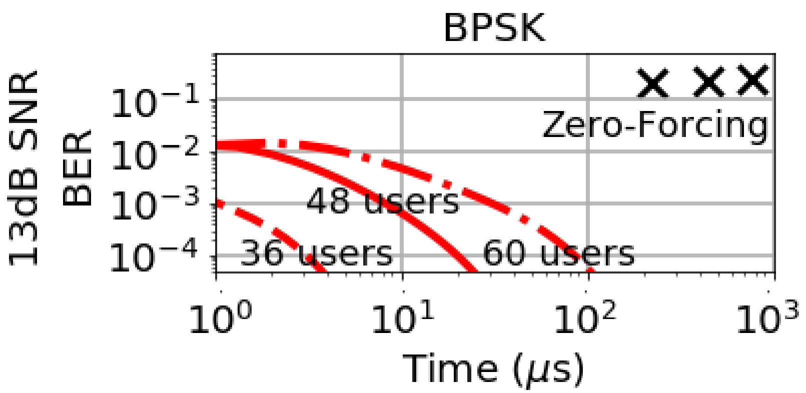
Well Beyond the Borderline of Conventional Computer

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.

Summary

1. LDPC decoding
   - Quantum LDPC decoder (MobiCom'20) [1]

2. Large MIMO detection
   - Quantum detection algorithm (SIGCOMM '19) [2]

- For further papers (hybrid classical-quantum processing) please see: **paws.cs.princeton.edu**

1. **Srikar Kasi** and Kyle Jamieson**.** Towards Quantum Belief Propagation for LDPC Decoding in Wireless Networks. MobiCom'20.
2. **Minsung Kim, Davide Venturell, Kyle Jamieson**. Leveraging quantum annealing for large MIMO processing in centralized radio access networks. ACM SIGCOMM '19.