Quantum Algorithmic Breakeven: on scaling up with noisy qubits

ICQSIM2017 : International Conference on Quantum Simulation
Paris
November 15, 2017
Daniel Lidar, USC
To scale up indefinitely we must break even first

break even
phrase of break

1. reach a point in a business venture when the profits are equal to the costs.
   "the firm will break even at the operating level this year"
What is quantum breakeven?

- Achieving the accuracy threshold for fault tolerant quantum computation
  - Performance guarantee is highly model-dependent:
    threshold = $x$ means different things for different models (e.g., Markovian vs non-Markovian noise)
- A logical qubit with higher coherence than the constituent physical qubits
- A logical gate with higher fidelity than the constituent physical gates
  - No guarantee this will continue to hold at larger scales (more qubits & gates): unanticipated errors may appear
- Can we replace with a breakeven notion that
  - is model-independent?
  - holds at large scales?
Consider Algorithmic Scaling
Algorithmic Scaling

Run an algorithm with exponential quantum speedup (e.g., quantum simulation) on quantum hardware

ideal case: no decoherence

quantum scaling advantage clear for all $N$
Algorithmic Scaling

Run an algorithm with exponential quantum speedup (e.g., quantum simulation) on quantum hardware with decoherence.

The graph shows the time-to-solution for classical and quantum algorithms as a function of problem size $N$. The quantum scaling advantage is only clear for large $N$. The classical time-to-solution grows exponentially with $N$, while the quantum time-to-solution grows more slowly, indicating a quantum speedup.
Algorithmic Scaling

Run an algorithm with exponential quantum speedup (e.g., quantum simulation) on quantum hardware with more decoherence.

quantum scaling advantage only clear for even larger $N$
Run an algorithm with exponential quantum speedup (e.g., quantum simulation) on quantum hardware with too much decoherence.

**Algorithmic Scaling**

Time-to-solution [arb. units] vs. problem size $N$:

- **Quantum**
- **Classical**

Quantum scaling disadvantage.
Run an algorithm with exponential quantum speedup (e.g., quantum simulation) on quantum hardware with too much decoherence.

Algorithmic Scaling

Quantum blowup
Run an algorithm with exponential quantum speedup (e.g., quantum simulation) on quantum hardware

with decoherence + QEC
Enter QEC

Run an algorithm with exponential quantum speedup (e.g., quantum simulation) on quantum hardware with decoherence + QEC

Algorithmic success with QEC: corrected quantum scaling is better than classical & uncorrected quantum
Algorithmic Success with QEC

Run an algorithm with exponential quantum speedup (e.g., quantum simulation) on quantum hardware with decoherence + QEC

Algorithmic success with QEC: corrected quantum scaling is better than classical & uncorrected quantum
Run an algorithm with exponential quantum speedup (e.g., quantum simulation) on quantum hardware with decoherence + QEC

**Algorithmic Breakeven with QEC**

Corrected quantum scaling is no worse than uncorrected quantum, but not necessarily better than classical

Can this be achieved with existing quantum hardware?

Model-independent holds at large scales

More modest:
Algorithmic breakeven with quantum annealing

Algorithmic breakeven with quantum annealing

Problem:
find ground state energy of

$$H_{\text{Ising}} = \sum J_{ij} \sigma^z_i \sigma^z_j$$

with random

$$J_{ij} \in \left\{ \pm \frac{1}{6}, \ldots, \pm \frac{5}{6}, \pm 1 \right\}$$

Run on a D-Wave 2
(503 qubits)

Problem:
find ground state energy of

\[ H_{\text{Ising}} = \sum J_{ij} \sigma_i^z \sigma_j^z \]

with random

\[ J_{ij} \in \left\{ \pm \frac{1}{6}, \ldots, \pm \frac{5}{6}, \pm 1 \right\} \]

Run on a D-Wave 2
(503 qubits)

Brief interlude on D-Wave processors

They are a type of quantum (and classical thermodynamics) simulator.
D-Wave 2 Connectivity Graph

512 qubits with “Chimera graph” couplings

$K_{4,4}$ unit cell

Most “natural” problems defined over complete graphs; must be (minor-)embedded

Largest complete graph embeddable in $L \times L$
Chimera graph of $L$ unit cells, each $K_{c,c}$ is $K_{cL+1}$

All D-Wave chips to date: $c = 4$
Graph degree $= c + 2 = 6$

D-Wave 1:
$L = 4, N = 128$ (USC yield: 108)
$K_{17}$ for ideal
$K_{14}$ for actual

D-Wave 2:
$L = 8, N = 512$ (USC yield: 504)
$K_{33}$ for ideal
$K_{32}$ for actual

D-Wave 2X:
$L = 12, N = 1152$ (USC yield: 1098)
$K_{49}$ for ideal
$K_{44}$ for actual

D-Wave 2000Q
$L = 16, N = 2048$ (NASA yield: 2031)
$K_{65}$ for ideal
D-Wave 2 Connectivity Graph

512 qubits with “Chimera graph” couplings

D-Wave 1:
$L = 4, N=128$ (USC yield: 108)
$K_{17}$ for ideal
$K_{14}$ for actual

D-Wave 2:
$L = 8, N=512$ (USC yield: 504)
$K_{33}$ for ideal
$K_{32}$ for actual

D-Wave 2X:
$L = 12, N=1152$ (USC yield: 1098)
$K_{49}$ for ideal
$K_{44}$ for actual

D-Wave 2000Q
$L = 16, N=2048$ (NASA yield: 2031)
$K_{65}$ for ideal
Ideally, evolve adiabatically according to $H(t) = A(t)H_X + B(t)H_P$

\[ H_X = H_{\text{transverse}} = \sum_{j \in V} \sigma^x_j \]

\[ H_P = H_{\text{Ising}} = \sum_{j \in V} h_j \sigma^z_j + \sum_{(i,j) \in E} J_{ij} \sigma^z_i \sigma^z_j \]

finds the ground state energy

\[
\min_{\{\sigma^z_i\}} H_{\text{Ising}}
\]

NP hard (Barahona 1982)

In reality: “control error”

$H$ implemented $\neq H$ intended

D-Wave 2 precision:

$\delta h \sim N(0,0.05h_{\text{max}})$,

$\delta J \sim N(0,0.04J_{\text{max}})$

$\Rightarrow 3$ bits

... 4 bits on D-Wave 2X

... 5 bits on D-Wave 2000Q

Samples from Gibbs dist.

\[
\frac{1}{Z} e^{-\beta H_{\text{Ising}}}
\]

#P-hard (=counting)

$\sim 10\text{mK}$

Designed to solve/sample Ising model problems
Note colors are flipped!
Daniel Lidar; 06/09/2017
Beyond control errors: How quantum?

Facts:
- DW Ni flux qubits have $T_2 \sim 100\text{ns}$, annealing time $t_f \geq 5\mu\text{s}$
- gap$(H)$ can be $\ll T \sim 10\text{mK}$

Properly described as an open quantum system

Governed by Markovian adiabatic master equation$^{(1)}$
- Dynamics (probably) not efficiently classically simulatable
- As an optimizer: so far no evidence of a q. speedup

A playground for testing algorithmic scaling with noisy qubits and error correction

Error Correction for Quantum Annealing

Error Correction for Quantum Annealing

1. Encode into $[n, 1, n]$ repetition code:

$$H_{\text{Ising}} = \sum_{i=1}^{N} h_i \sigma^z_i + \sum_{i<j}^{N} J_{ij} \sigma^z_i \sigma^z_j$$

$$\sigma^z = \sum_{k=1}^{n} \sigma^z_{i_k}$$

$$\sigma^z_i \sigma^z_j = \sum_{k=1}^{n} \sigma^z_{i_k} \sigma^z_{j_k}$$

Nice features of this code:

**Implementable:** Logical $Z$ and $ZZ$ operators are 1 and 2-local

**Energy Boost:** Logical operators stronger than physical by factor of $n$

1. Encode into $[n, 1, n]$ repetition code:

$$\overline{H}_{\text{Ising}} = \sum_{i=1}^{N} h_i \sigma_i^z + \sum_{i<j}^{N} J_{ij} \sigma_i^z \sigma_j^z$$

2. Add FM energy penalty:

$$H_P = -\sum_{i=1}^{N} (\sigma_{i_1}^z + \cdots + \sigma_{i_n}^z) \sigma_{i_P}^z$$

Also implementable: ZZ operators are 2-local (stabilizers of the bit-flip repetition code)

2. Add FM energy penalty:

\[ H_P = -\sum_{i=1}^{N} \left( \sigma_{i_1}^z + \cdots + \sigma_{i_n}^z \right) \sigma_i^z \]

3. Combine:

\[ \overline{H}_{\text{Ising}, P}(\alpha, \gamma) := \alpha \overline{H}_{\text{Ising}} + \gamma H_P \]

“problem scale” (controllable)  “penalty scale” (optimized)

Error Correction for Quantum Annealing

1. Encode into $[n, 1, n]$ repetition code:

\[
\overline{H}_{\text{Ising}} = \sum_{i=1}^{N} h_i \overline{\sigma_i^z} + \sum_{i<j}^{N} J_{ij} \overline{\sigma_i^z \sigma_j^z}
\]

2. Add FM energy penalty:

\[
H_P = -\sum_{i=1}^{N} \left( \sigma_{i_1}^z + \cdots + \sigma_{i_n}^z \right) \sigma_{i_P}^z
\]

3. Combine:

\[
\overline{H}_{\text{Ising}, P}(\alpha, \gamma) := \alpha \overline{H}_{\text{Ising}} + \gamma H_P
\]

"problem scale" (controllable)

"penalty scale" (optimized)

4. Run QA:

\[
\overline{H}(t) = A(t) H_X + B(t) \overline{H}_{\text{Ising}, P}(\alpha, \gamma)
\]

5. Decode at end by majority vote

**Error Correction for Quantum Annealing**

1. Encode into $[n, 1, n]$ repetition code:

$$
\overline{H}_{\text{Ising}} = \sum_{i=1}^{N} h_i \overline{\sigma}_i^z + \sum_{i<j}^{N} J_{ij} \overline{\sigma}_i^z \overline{\sigma}_j^z
$$

2. Add FM energy penalty:

$$
H_P = -\sum_{i=1}^{N} \left( \sigma_{i_1}^z + \cdots + \sigma_{i_n}^z \right) \sigma_{i_P}^z
$$

3. Combine:

$$
\overline{H}_{\text{Ising},P}(\alpha, \gamma) := \alpha \overline{H}_{\text{Ising}} + \gamma H_P
$$

"problem scale" (controllable)  "penalty scale" (optimized)

4. Run QA:

$$
\overline{H}(t) = A(t) \overline{H}_X + B(t) \overline{H}_{\text{Ising},P}(\alpha, \gamma)
$$

unencoded \[
[\overline{\sigma}^x = (\sigma^x)^\otimes n]
\]

5. Decode at end by majority vote

---

test on AFM chains

α

“unprotected”
(unencoded, no energy penalty)

Unprotected(U)

agrees with

Fair comparison for QAC:
run 4 parallel chains, take the best
(“classical” error correction)

\[ \alpha = 1.00 \]

\[ \alpha = 0.50 \]

\[ \alpha = 0.40 \]

\[ 1 - (1 - p_U)^4 \]

\[ p_U \]

From chains to random non-planar Ising ...

Chimera(DW2); degree 6; 503 functional qubits

QAC encoding
degree 3; still non-planar
119 functional logical qubits

\[
H_{\text{Ising}} = \sum_{(i,j) \in E(G)} J_{ij} \sigma_i^z \sigma_j^z
\]

random \( J_{ij} \in \left\{ \pm \frac{1}{6}, \ldots, \pm \frac{5}{6}, \pm 1 \right\} \)
Algorithmic breakeven with quantum annealing correction

"classical" (C)

best of 4 copies

\[ H_{\text{Ising}} = \sum_{(i,j) \in E(G)} J_{ij} \sigma_i^z \sigma_j^z \]

\[ J_{ij} \in \left\{ \pm \frac{1}{6}, \ldots, \pm \frac{5}{6}, \pm 1 \right\} \]

Algorithmic breakeven with quantum annealing correction

\[ H_{\text{Ising}} = \sum_{(i,j) \in E(G)} J_{ij} \sigma_i^z \sigma_j^z \]


3-qubit repetition code + energy penalty

\[ J_{ij} \in \left\{ \pm \frac{1}{6}, \ldots, \pm \frac{5}{6}, \pm 1 \right\} \]
Why does QAC work?

Main mechanism:

avoidance or modification of a quantum phase transition due to penalty term

Mean-field analysis of $p$-body ferromagnet

$N$ logical qubits each in $[K, 1, K]$ repetition code

($K = n$ in earlier notation)

Model Hamiltonian $H = H_X + H_Z$

$H_Z = -N \sum_{k=1}^{K} \left( \frac{1}{N} \sum_{i=1}^{N} \sigma_{iz}^k \right)^p - \gamma \sum_{k=1}^{K} \sum_{i=1}^{N} \sigma_{iz}^k \sigma_{iz}^0$

problem Hamiltonian

fully-connected logical qubits
Mean-field analysis of $p$-body ferromagnet

$N$ logical qubits each in $[K, 1, K]$ repetition code
($K = n$ in earlier notation)

Why does QAC work?

Model Hamiltonian $H = H_x + H_z$

$H_z = -N\sum_{k=1}^{K} \left( \frac{1}{N} \sum_{i=1}^{N} \sigma_{iz}^{k} \right)^{p} - \gamma \sum_{k=1}^{K} \sum_{i=1}^{N} \sigma_{iz}^{k} \sigma_{iz}^{0}$

$H_x = -\Gamma \sum_{k=1}^{K} \sum_{i=1}^{N} \sigma_{ix}^{k}$

use stat. mech. tricks to compute partition function

$Z = \text{Tr} e^{-\beta (H_z + H_x)}$

then free energy etc.
Free energy and magnetization for $p=2$

$\gamma = 0$

no penalty:
2nd order phase transition

$\gamma = 1$

Turning the penalty on avoids the phase transition
1st order transition for $p \geq 3$

Discontinuous transition from $m=0$ to $m=0.94$ (wide and tall tunneling barrier)

Increasing penalty $\gamma$ lowers and narrows the tunneling barrier.

Also increases the minimum gap $\Delta$:

$\gamma = 0$ at $\Gamma_c$

$\gamma = 0.5$ at $\Gamma_c$

$\Delta = C^N$

Turning the penalty on softens the phase transition.
• For models with a second order quantum phase transition:
  • QAC avoids the phase transition
• For models with a first order quantum phase transition:
  • QAC softens the closing of the gap


• QAC rescales the temperature; rescaling can be made as large as $n^2$ for [$n, 1, n$] code (using all-to-all connectivity):
  qubits can be traded for temperature reduction


⇒ Helps to achieve algorithmic breakeven
For models with a second order quantum phase transition:
• QAC avoids the phase transition

For models with a first order quantum phase transition:
• QAC softens the closing of the gap
  \[ \rightarrow \] Helps to achieve algorithmic breakeven*

Target for Error-Corrected Circuit-Model Quantum Computing

*Demonstrate error-corrected scaling that is no worse than uncorrected on a computational problem
• For models with a second order quantum phase transition:
  • QAC avoids the phase transition

• For models with a first order quantum phase transition:
  • QAC softens the closing of the gap

→ Helps to achieve algorithmic breakeven*

*Demonstrate error-corrected scaling that is no worse than uncorrected on a computational problem

Thanks!
As of June 2017, funded under IARPA’s “Quantum Enhanced Optimization” (QEO) Program we will implement all the lessons we’ve learned studying the D-Wave machines, with a mission to build a 100-qubit quantum annealer in 5 years that is
Additional slides
Surpassing breakeven with quantum annealing correction

\[ H_{\text{Ising}} = \sum_{j \in V} h_j \sigma_j^z + \sum_{(i,j) \in E} J_{ij} \sigma_i^z \sigma_j^z \]

**Random Ising**

\[ h_j = 0 \]

\[ J_{ij} \in \{ \pm \frac{1}{6}, \ldots, \pm \frac{5}{6}, \pm 1 \} \]


Begs the question: is this a quantum speedup?

Answer: We don’t know since cannot certify scaling is correct.

[Graph showing time to solution vs. number of physical/logical qubits before and after encoding.]
Certifiable speedup requires a proof of optimality

**Time-to-solution (TTS):**
the time required to find the ground state (GS) at least once
with probability 99%

Run the device with annealing time $t_a$ and measure probability $p_{GS}$ of finding the GS. Then:

$$\text{TTS} = t_a \times (\text{no. of repetitions to get to 99%}) = t_a \cdot \frac{\log(1-0.99)}{\log(1-p_{GS}(t_a))}$$

The optimal annealing time $t_a$ is that which **minimizes** the TTS for fixed problem size $N$

**Scaling cannot be trusted** unless optimal $t_a$ has been identified\(^{(1,2)}\)

---

Certifiable speedup requires a proof of optimality

Excessively long annealing time causes false flat scaling

Scaling curves without an optimality certificate cannot be trusted.
Smoking gun is a **minimum** in TTS as a function of annealing time …

Average TTS over 1000 instances at each $t_a$ and size $L$ (sqrt of number of Chimera unit cells).

Run on DW2000Q processor at Burnaby, courtesy of D-Wave Systems.

… for a new problem class: ‘logical-planted-solution instances’

1. Create a spin-glass backbone using frustrated loops with planted solutions\(^{(1)}\) on the **logical graph** of Chimera unit cells\(^{(2)}\).

\[^{(1)}\text{logical-planted-solution}\] \[^{(2)}\text{instances}\]
1. Create a spin-glass backbone using frustrated loops with planted solutions on the logical graph of Chimera unit cells.

2. Add 8-qubit ‘gadgets’ placed randomly in some fraction (10%) of all unit cells.

Can now use to demonstrate **limited** quantum speedup.
Limited quantum speedup: the evidence

Speedup relative to a `relevant’ class of classical algorithms

Start with two optimized ‘DW-like’ classical solvers:
- Simulated Annealing (SA) with single-spin updates
- Spin-vector Monte Carlo\(^{(1)}\) (SVMC)

\(a \exp(bL)\)

<table>
<thead>
<tr>
<th>Solver</th>
<th>(b) [95% CI]</th>
</tr>
</thead>
<tbody>
<tr>
<td>SA</td>
<td>1.002 ± 0.066</td>
</tr>
<tr>
<td>SVMC</td>
<td>0.867 ± 0.079</td>
</tr>
</tbody>
</table>

scaling of the median of the instance distribution

Limited quantum speedup: the evidence

Speedup relative to a `relevant’ class of classical algorithms

DW2000Q unequivocally beats the classical solvers

Fit curves to

\[ a \exp(bL) \]

<table>
<thead>
<tr>
<th>Solver</th>
<th>( b ) [95% CI]</th>
</tr>
</thead>
<tbody>
<tr>
<td>DW2KQ</td>
<td>0.760 ± 0.017</td>
</tr>
<tr>
<td>SA</td>
<td>1.002 ± 0.066</td>
</tr>
<tr>
<td>SVMC</td>
<td>0.867 ± 0.079</td>
</tr>
</tbody>
</table>

Scaling of the median of the instance distribution

Limited quantum speedup: the evidence

Speedup relative to a `relevant' class of classical algorithms

Simulated quantum annealing\(^{(1)}\) dominates

Fit curves to

\[ a \exp(bL) \]

<table>
<thead>
<tr>
<th>Solver</th>
<th>(b) [95% CI]</th>
</tr>
</thead>
<tbody>
<tr>
<td>DW2KQ</td>
<td>0.760 ± 0.017</td>
</tr>
<tr>
<td>SA</td>
<td>1.002 ± 0.066</td>
</tr>
<tr>
<td>SVM C</td>
<td>0.867 ± 0.079</td>
</tr>
<tr>
<td>SQA</td>
<td>0.370 ± 0.052</td>
</tr>
</tbody>
</table>

scaling of the median of the instance distribution

Limited quantum speedup: the evidence

Speedup relative to a `relevant' class of classical algorithms

Simulated quantum annealing\(^{(1)}\) dominates

Can we reach breakeven (and beyond) through quantum annealing correction?

Fit curves to
\[ a \exp(bL) \]

<table>
<thead>
<tr>
<th>Solver</th>
<th>(b) [95% CI]</th>
</tr>
</thead>
<tbody>
<tr>
<td>DW2KQ</td>
<td>0.760 ± 0.017</td>
</tr>
<tr>
<td>SA</td>
<td>1.002 ± 0.066</td>
</tr>
<tr>
<td>SVMC</td>
<td>0.867 ± 0.079</td>
</tr>
<tr>
<td>SQA</td>
<td>0.370 ± 0.052</td>
</tr>
</tbody>
</table>

scaling of the median of the instance distribution


Backup slides
Error/Noise Sources

- Technical/Engineering:
  - Non-ideal spin implementation (not a true 2-level system)
  - Finite digital-to-analog conversion (DAC) precision
  - Imperfect device characterization & calibration
  - Crosstalk
  - Low-frequency flux noise

- Decoherence:
  - Thermal excitations ($T_1$)
  - Dephasing in computational basis ($T_2$; problematic for instances with small gap)
  - Dephasing between energy eigenstates
    - DW2 (Vesuvius)
    - DW2X (Washington)

$h_i/J_{ij}$ control:
- DW2: 3 bits precision
- DW2X: 4 bits precision

![typical 1/f noise spectrum](image-url)
Adiabatic Markovian Master Equation


Weak coupling, work in the instantaneous energy eigenbasis of system Hamiltonian:

\[
\frac{d}{dt} \rho_S = -i \left[ H_S(t) + H_{LS}(t), \rho_S(t) \right] \quad \text{Unitary evolution with bath-induced Lamb shift}
\]

\[
+ \sum_{\omega} \gamma_{\alpha\beta}(\omega) \left( L_{\omega,\beta}(t) \rho_S L_{\omega,\alpha}^\dagger(t) + \frac{1}{2} \left\{ L_{\omega,\alpha}^\dagger(t) L_{\omega,\beta}(t), \rho_S(t) \right\} \right) \quad \text{Non-unitary dissipative dynamics}
\]

Thermal transition rate from GS:

\[
\frac{d}{dt} \rho_{00}(t) \approx \sum_{i>0} \left| \mathcal{O}_i(t) \right|^2 \gamma(\Delta_{i0}) \left( \rho_{ii}(t) - e^{-\beta \Delta_{i0}} \rho_{00}(t) \right) \quad \text{bath spectral density (FT of bath correlation function)}
\]

matrix element of the system operator (from the system-bath Hamiltonian) in the instantaneous energy eigenbasis

\[
T_2 = 1/\gamma(0) \quad \text{doesn’t appear!}
\]