QUANTUM COMPUTING, MACHINE LEARNING, AND THE D-WAVE ARCHITECTURE

Paul Pham
4 December 2012
OVERVIEW

- Review of discrete optimization as search
- Quantum annealing, Hamiltonians, energy
- D-Wave architecture, protein folding, image search
- Analogy with alchemy or radio
We are given a cost function $f : \{0, 1\}^n \rightarrow \mathbb{R}$
We are given a cost function $f : \{0, 1\}^n \rightarrow \mathbb{R}$

- Example: layover time in airline schedules
We are given a cost function $f : \{0, 1\}^n \rightarrow \mathbb{R}$

- Example: layover time in airline schedules

We want to find its minimum (and the corresponding input) with the fewest queries to $f$ as possible.
We are given a cost function \( f : \{0, 1\}^n \rightarrow \mathbb{R} \)

- Example: layover time in airline schedules

We want to find its **minimum**
(and the corresponding input)
with the fewest queries to \( f \) as possible

Naive search would take **exponential** time.
We are given a cost function $f : \{0, 1\}^n \to \mathbb{R}$.

- Example: layover time in airline schedules

We want to find its **minimum** (and the corresponding input) with the fewest queries to $f$ as possible.

Naive search would take exponential time.

Use **local heuristics** instead.
SEARCH FOR MINIMUM OVER A LANDSCAPE
LOCAL SEARCH
(HILLCLIMBING)
LOCAL SEARCH (HILLCLIMBING)

- Start at a random point.
LOCAL SEARCH (HILLCLIMBING)

- Start at a random point.
- While not done:
LOCAL SEARCH
(HILLCLIMBING)

- Start at a random point.
- While not done:
  - Choose $y$ a random neighbor of $x$
LOCAL SEARCH (HILLCLIMBING)

- Start at a random point.
- While not done:
  - Choose $y$ a random neighbor of $x$
  - if $f(y) < f(x)$:
LOCAL SEARCH
(HILLCLIMBING)

- Start at a random point.
- While not done:
  - Choose \( y \) a random neighbor of \( x \)
  - if \( f(y) < f(x) \):
    - set \( x \) to \( y \)
LOCAL SEARCH
(HILLCLIMBING)

- Start at a random point.
- While not done:
  - Choose a random neighbor of \( x \)
  - if \( f(y) < f(x) \):
    - set \( x \) to \( y \)
- Shortcomings: gets stuck in local minima
ANNEALING
ANNEALING

- A physically-inspired technique
ANNEALING

- A physically-inspired technique
- Reach equilibrium at a high temperature, then slowly cool to stay in the lowest energy configuration.
ANNEALING

- A physically-inspired technique
- Reach equilibrium at a high temperature, then slowly cool to stay in the lowest energy configuration.
- Example: metallurgy, rock candy
SIMULATED ANNEALING
SIMULATED ANNEALING

- [Metropolis-Hastings algorithm]
SIMULATED ANNEALING

- [Metropolis-Hastings algorithm]
- Uses **temperature** to get out of local minima
SIMULATED ANNEALING

- [Metropolis-Hastings algorithm]
- Uses **temperature** to get out of local minima
  - hot objects move due to thermal fluctuations
SIMULATED ANNEALING

- [Metropolis-Hastings algorithm]
- Uses **temperature** to get out of local minima
  - hot objects move due to thermal fluctuations
- Randomly explores nearby configurations
SIMULATED ANNEALING
SIMULATED ANNEALING

• Start at a random point (same as local search)
SIMULATED ANNEALING

- Start at a random point (same as local search)
- if $f(y) < f(x)$:
SIMULATED ANNEALING

- Start at a random point (same as local search)
- if $f(y) < f(x)$:
  - set $x$ to $y$
SIMULATED ANNEALING

- Start at a random point (same as local search)
- if \( f(y) < f(x) \):
  - set \( x \) to \( y \)
- else with probability \( e^{-(f(y) - f(x))/T} \)
SIMULATED ANNEALING

- Start at a random point (same as local search)
  - if \( f(y) < f(x) \):
    - set \( x \) to \( y \)
  - else with probability \( e^{-\frac{(f(y) - f(x))}{T}} \)
    - set \( x \) to \( y \)
How to choose temperature? When are we done?

Various strategies:

- Start at an initial temperature, repeat procedure and lower temperature. At $T=0$, same as local search.
- Stop when improvements slow down.
SIMULATED ANNEALING: WHERE IT FAILS

- From [0201031]

- Hamming weight function with a spike

\[ f(x) = |x| + n^{1/3} \mathbb{1}_{|x|=n/4} \]

- Bush of implications

\[ f(x) = x_1(x_2 + \ldots + x_n) - \frac{x_1}{10} \]
PHYSICAL QUBITS
PHYSICAL QUBITS

- Leading candidates (here at UW) include:
Leading candidates (here at UW) include:

- trapped atomic ions (Boris Blinov)
Leading candidates (here at UW) include:

- trapped atomic ions (Boris Blinov)
- nitrogen-vacancy centers in diamond (Kai-Mei Fu)
OVERVIEW

- Review of discrete optimization as search
- Quantum annealing, Hamiltonians, energy
- D-Wave architecture, protein folding, image search
- Future of quantum computing & machine learning
THE HAMILTONIAN
THE HAMILTONIAN

- A Hermitian operator on quantum states which corresponds to observable quantities.

\[ H |\psi\rangle = \lambda |\psi\rangle \]
A Hermitian operator on quantum states which corresponds to observable quantities.

\[ H |\psi\rangle = \lambda |\psi\rangle \]

Eigenvalues determine discrete energies
THE HAMILTONIAN

- A Hermitian operator on quantum states which corresponds to observable quantities.

\[ H |\psi\rangle = \lambda |\psi\rangle \]

- Eigenvalues determine discrete energies

- Determines time evolution of quantum state

\[ |\psi(t)\rangle = e^{-Ht} |\psi(0)\rangle \]
GROUND STATE
GROUND STATE

- Lowest eigenvalue is the ground state energy
GROUND STATE

- Lowest eigenvalue is the ground state energy
- Systems naturally tend to this state over time
GROUND STATE

- Lowest eigenvalue is the ground state energy
- Systems **naturally tend** to this state over time
- Like a minimum of function
GROUND STATE

- Lowest eigenvalue is the ground state energy
- Systems **naturally tend** to this state over time
- Like a minimum of function
- Corresponding state (input) is **ground state**
CLASSICAL EXAMPLE:
MAGNETIC HARD DRIVE

- Each bit location on a magnetic hard drive is a cluster of tiny magnets which share an alignment.
- The ground state is when all spins are aligned.
- This encodes a 0 or a 1.
CLASSICAL EXAMPLE: MAGNETIC HARD DRIVE

\[ H(s) = - \sum_{j \sim k} J_{jk} s_j s_k \quad s_j \in \{+1, -1\} \]

- When a spin flips, it disagrees with its neighbors, which increases the energy of the system, taking it away from the ground state.

- It takes a lot of energy to completely disorder the system.
ADIABATIC QUANTUM COMPUTING
ADIABATIC QUANTUM COMPUTING

- AQC, an alternative to the quantum circuit model
AQC, an alternative to the quantum circuit model

We begin with a Hamiltonian whose ground state is easy to prepare (usually an equal superposition of all classical states).
ADIABATIC QUANTUM COMPUTING

- **AQC**, an alternative to the quantum circuit model
- We begin with a Hamiltonian whose ground state is easy to prepare (usually an equal superposition of all classical states).
- We slowly change this over time to a Hamiltonian whose ground state encodes the answer to our problem.
If we change it slowly enough, we will stay in the ground state the whole time [Adiabatic Theorem]

\[ H(\tau) = \tau H_1 + (1 - \tau) H_2 \quad \tau \in [0, 1] \]
Quantum states can **tunnel** through energy barriers in the landscape with some probability.

When used for optimization problems, AQC is often called **quantum annealing**.

Tunneling **takes the place of temperature** in simulated annealing.
OVERVIEW

- Review of discrete optimization as search
- Quantum annealing, Hamiltonians, energy
- D-Wave architecture, protein folding, image search
- Analogy with alchemy or radio
D-WAVE SYSTEMS
D-WAVE IN THE NEWS

- Google image search collaboration, NIPS 2009
D-WAVE IN THE NEWS

- Google image search collaboration, NIPS 2009
- Initial purchase of a D-Wave One by Lockheed Martin for $16 million, given to USC
D-WAVE IN THE NEWS

- Google image search collaboration, NIPS 2009
- Initial purchase of a D-Wave One by Lockheed Martin for $16 million, given to USC
- Recent funding round by Amazon, CIA
D-WAVE IN THE NEWS

- Google image search collaboration, NIPS 2009
- Initial purchase of a D-Wave One by Lockheed Martin for $16 million, given to USC
- Recent funding round by Amazon, CIA
- Nature paper on protein folding
D-WAVE ONE ARCHITECTURE

- 128 qubit Chimera chip (4x4 grid of 8 qubit cells)
LIGHT-SWITCH GAME
(1-LOCAL)

\[ H(s) = \sum_{i} h_i s_i \]

\[ h_i \in [-1, +1] \]

\[ s \in \{-1, +1\} \]
LIGHT-SWITCH GAME (1-LOCAL)

\[ H(s) = \sum_i h_i s_i \]

\[ h_i \in [-1, +1] \]

\[ s \in \{-1, +1\} \]

- The light switches (spins) are uncorrelated.
The light switches (spins) are uncorrelated.

We want to minimize their weighted sum.

\[ H(s) = \sum_i h_i s_i \]

\[ h_i \in [-1, +1] \]

\[ s \in \{-1, +1\} \]
The light switches (spins) are uncorrelated.

We want to minimize their weighted sum.

There is a simple solution (in $\mathbb{P}$)

$$H(s) = \sum_i h_i s_i$$

$$h_i \in [-1, +1]$$

$$s \in \{-1, +1\}$$
LIGHT-SWITCH GAME
(1-LOCAL)

\[ H(s) = \sum_{i} h_i s_i \]

\[ h_i \in [-1, +1] \]

\[ s \in \{-1, +1\} \]
LIGHT-SWITCH GAME (1-LOCAL)

\[ H(s) = \sum_{i} h_i s_i \]
\[ h_i \in [-1, +1] \]
\[ s \in \{-1, +1\} \]

- In-class exercise
LIGHT-SWITCH GAME
(2-LOCAL)

\[ H(s) = \sum h_i s_i + \sum_{j,k} J_{j,k} s_j s_k \]

\[ h_i, J_{j,k} \in [-1, +1] \]

\[ s_i \in \{-1, +1\} \]
Now light-switches are correlated (2-local terms)

\[
H(s) = \sum_i h_i s_i + \sum_{j,k} J_{j,k} s_j s_k
\]

\[
h_i, J_{j,k} \in [-1, +1]
\]

\[
s_i \in \{-1, +1\}
\]
LIGHT-SWITCH GAME
(2-LOCAL)

Now light-switches are correlated (2-local terms)

Minimizing this is in fact \( \textbf{NP} \)-complete

\[
H(s) = \sum_i h_i s_i + \sum_{j,k} J_{j,k} s_j s_k
\]

\( h_i, J_{j,k} \in [-1, +1] \)

\( s_i \in \{-1, +1\} \)
LIGHT-SWITCH GAME
(2-LOCAL)

Now light-switches are correlated (2-local terms)

Minimizing this is in fact \textbf{NP-complete}

\[
H(s) = \sum_i h_i s_i + \sum_{j,k} J_{j,k} s_j s_k
\]

\(h_i, J_{j,k} \in [-1, +1]\)

\(s_i \in \{-1, +1\}\)
LIGHT-SWITCH GAME
(2-LOCAL)

\[ H(s) = \sum_i h_i s_i + \sum_{j,k} J_{j,k} s_j s_k \]

\[ h_i, J_{j,k} \in [-1, +1] \]

\[ s_i \in \{-1, +1\} \]
LIGHT-SWITCH GAME (2-LOCAL)

In-class exercise, 64 different possible inputs

\[ H(s) = \sum_i h_i s_i + \sum_{j,k} J_{j,k} s_j s_k \]

\[ h_i, J_{j,k} \in [-1, +1] \]

\[ s_i \in \{-1, +1\} \]
Nature, August 2012

“Finding low-energy conformations of lattice protein models by quantum annealing”

<table>
<thead>
<tr>
<th>Amino-acid sequence</th>
<th>Interaction</th>
<th>ΔE</th>
</tr>
</thead>
<tbody>
<tr>
<td>P-S-V-K-M-A</td>
<td>P III K</td>
<td>-1</td>
</tr>
<tr>
<td></td>
<td>P III A</td>
<td>-2</td>
</tr>
<tr>
<td></td>
<td>S III M</td>
<td>-3</td>
</tr>
<tr>
<td></td>
<td>V III A</td>
<td>-4</td>
</tr>
</tbody>
</table>
PROTEIN FOLDING

Graph of the D-Wave architecture, and the subset used for protein folding experiments.
**PROTEIN FOLDING**

- Lowest-energy configuration found 17.67% of the time.
- Still slower than a classical computer.
- Can it scale?
GOOGLE IMAGE SEARCH
Collaboration between Google and D-Wave to use quantum hardware for machine learning tasks.
Collaboration between Google and D-Wave to use quantum hardware for machine learning tasks.

Trains strong binary classifier (SC) from weak binary classifiers (WC) for detecting cars in digital images.
[H. Neven, V. Denchev, G. Rose, W.G. Macready]
[H. Neven, V. Denchev, G. Rose, W.G. Macready]

[0811.0416] software training SC from thresholded linear superposition of WC.

\[ y = H(x) = \text{sign} \left( \sum_{i=1}^{N} w_i h_i(x) \right) \]
[H. Neven, V. Denchev, G. Rose, W.G. Macready]

[0811.0416] software training SC from thresholded linear superposition of WC.

\[ y = H(x) = \text{sign} \left( \sum_{i=1}^{N} w_i h_i(x) \right) \]

[0912.0779] software training of a large-scale SC by iterating over subsets of WCs.
[H. Neven, V. Denchev, G. Rose, W.G. Macready]

[0811.0416] software training SC from thresholded linear superposition of WC.

\[ y = H(x) = \text{sign} \left( \sum_{i=1}^{N} w_i h_i(x) \right) \]

[0912.0779] software training of a large-scale SC by iterating over subsets of WCs.

NIPS 2009 hardware demonstration
GOOGLE IMAGE SEARCH
Results of hardware quantum annealing compared favorably with Adaboost running on Google MapReduce cluster.
GOOGLE IMAGE SEARCH

- Results of hardware quantum annealing compared favorably with Adaboost running on Google MapReduce cluster.

- No comparison of running time, power consumption, or cost of hardware given.
D-WAVE LIMITATIONS:
APPROXIMATE AQC
D-WAVE LIMITATIONS: APPROXIMATE AQC

- Qubit connectivity: cannot represent an Ising problem with arbitrary coupling terms, must map to a constrained graph.
Qubit connectivity: cannot represent an Ising problem with arbitrary coupling terms, must map to a constrained graph.

Non-zero temperature: 20-40 mK of surrounding hardware significant compared to chip.
D-WAVE LIMITATIONS: APPROXIMATE AQC

- Qubit connectivity: cannot represent an Ising problem with arbitrary coupling terms, must map to a constrained graph.
- Non-zero temperature: 20-40 mK of surrounding hardware significant compared to chip.
- Parameter variability: coupling values samples from a Gaussian centered on desired value.
D-WAVE SKEPTICISM
D-WAVE SKEPTICISM

- Summarized in this Quora answer (May 2011) by Dave Bacon.
D-WAVE SKEPTICISM

- Summarized in this Quora answer (May 2011) by Dave Bacon.
- Also in this blog post (May 2011) and this one (February 2012) by Scott Aaronson.
D-WAVE SKEPTICISM

- Summarized in this Quora answer (May 2011) by Dave Bacon.
- Also in this blog post (May 2011) and this one (February 2012) by Scott Aaronson.
- Also, Scott has bet $100,000 that no one can prove scalable quantum computers are impossible.
D-WAVE OBJECTION #1: NOT UNIVERSAL?
D-WAVE OBJECTION #1: NOT UNIVERSAL?

- AQC is equivalent to quantum circuit model for arbitrary 2-local Hamiltonians.
D-WAVE OBJECTION #1: NOT UNIVERSAL?

- AQC is equivalent to quantum circuit model for arbitrary 2-local Hamiltonians.
- D-Wave only claims to solve stoquastic Hamiltonians, which is in POST-BPP. [0606140]
D-WAVE OBJECTION #1: NOT UNIVERSAL?

- AQC is equivalent to quantum circuit model for arbitrary 2-local Hamiltonians.
- D-Wave only claims to solve stoquastic Hamiltonians, which is in POST-BPP. [0606140]
- Are stoquastic Hamiltonians universal? **Open question!**
D-WAVE OBJECTION #2: NOT EVEN QUANTUM?
D-WAVE OBJECTION #2: NOT EVEN QUANTUM?

- Coherence time is much shorter than algorithm running time.
D-WAVE OBJECTION #2: NOT EVEN QUANTUM?

- Coherence time is much shorter than algorithm running time.
- Could just be running classical simulated annealing (but in a direct physical way).
D-WAVE OBJECTION #2: NOT EVEN QUANTUM?

- Coherence time is much shorter than algorithm running time.
- Could just be running classical simulated annealing (but in a direct physical way).
- Runs at a finite temperature, not a closed system.
NATURE PAPER 2011: QUANTUM EFFECTS?
NATURE PAPER 2011: QUANTUM EFFECTS?

- D-Wave One restricted to 8 qubits
D-Wave One restricted to 8 qubits

“Tunneling” between states could be explained by either temperature or quantum effects.
D-Wave One restricted to 8 qubits

“Tunneling” between states could be explained by either temperature or quantum effects.

Temperature was lowered, and tunneling decreased, until a “freezeout” point.
D-Wave One restricted to 8 qubits

“Tunneling” between states could be explained by either temperature or quantum effects.

Temperature was lowered, and tunneling decreased, until a “freezeout” point.

Below the freezeout point, there was no temperature dependence.
D-WAVE OBJECTION #3: NOT ENTANGLED?
D-WAVE OBJECTION #3: NOT ENTANGLED?

- Difficult for hardware reasons to test for entanglement.
D-WAVE OBJECTION #3: NOT ENTANGLED?

- Difficult for hardware reasons to test for entanglement.
- It is not known whether separable (mixed) states demonstrate quantum power.
D-WAVE OBJECTION #3: NOT ENTANGLED?

- Difficult for hardware reasons to test for entanglement.
- It is not known whether separable (mixed) states demonstrate quantum power.
- Noise rate of DWave qubits is quite high.
HOW SHOULD D-WAVE PROCEED?
HOW SHOULD D-WAVE PROCEED?

Scott Aaronson observes two alternatives:
HOW SHOULD D-WAVE PROCEED?

Scott Aaronson observes two alternatives:

• Continue with high-profile public demonstrations for larger problem sizes and different applications to gain funding and support.
Scott Aaronson observes two alternatives:

- Continue with high-profile public demonstrations for larger problem sizes and different applications to gain funding and support.
- Characterize qubit better and improve noise, to ensure that we have a truly quantum machine.
TO BELIVE OR NOT TO BELIEVE?
TO BELIVE OR NOT TO BELIEVE?

- Most radical new technologies faces skepticism from scientific community (that’s their job).
Most radical new technologies faces skepticism from scientific community (that’s their job).

Even if the D-Wave is not quantum, it has a huge effect on quantum computing and any future machines.
TO BELIVE OR NOT TO BELIEVE?

- Most radical new technologies faces skepticism from scientific community (that’s their job).
- Even if the D-Wave is not quantum, it has a huge effect on quantum computing and any future machines.
- Do you believe that D-Wave has a quantum computer? Do you think that claiming to have a quantum computer can still be beneficial?
OVERVIEW

- Review of discrete optimization as search
- Quantum annealing, Hamiltonians, energy
- D-Wave architecture, protein folding, image search
- Analogy with alchemy or radio
ALCHEMY: LEAD INTO GOLD?
WIRELESS COMMUNICATION
RADIO WAVES ACROSS THE ATLANTIC?
EARLY AERIAL STATIONS
Radio Waves Across the Atlantic?

The Largest Circulation

10 Pages

THE HAL

Vol. XXXIV

The Largest Circulation

MARCONI SAYS HE SENT 10,000 ACROSS THE ATLANTIC

Messages Alleged to Have Been Transmitted Without Slightest Difficulty.


No Definite Public Demonstration That Messages are Being Forwarded.

OFFICE, RAY, October 10-11:

A cablegram from Port Morden when the international wireless service between Char-Who and Ireland was announced by Lord Melchon, the British Consul-General, and other notable men, and the first newspaper to get a story from the other side was the New York Times. A cable

The Largest Circulation

The Largest Circulation

MARCONI SAYS HE SENT 10,000 ACROSS THE ATLANTIC

Messages Alleged to Have Been Transmitted Without Slightest Difficulty.


No Definite Public Demonstration That Messages are Being Forwarded.

OFFICE, RAY, October 10-11:

A cablegram from Port Morden when the international wireless service between Char-Who and Ireland was announced by Lord Melchon, the British Consul-General, and other notable men, and the first newspaper to get a story from the other side was the New York Times. A cable
WIRELESS COMPETES WITH OCEANIC CABLE
MARCONI FACES SCIENTIFIC SKEPTICISM
MARCONI CORPORATION
AND RADIO
FUTURE OF D-WAVE

- 512 qubit processor (Vesuvius) unveiled in early 2012
FUTURE OF D-WAVE

- 512 qubit processor (Vesuvius) unveiled in early 2012
- At this scale, hardware will be faster than simulator
FUTURE OF D-WAVE

- 512 qubit processor (Vesuvius) unveiled in early 2012
- At this scale, hardware will be faster than simulator
- Time will tell whether it scales for large machine learning problems.
ACKNOWLEDGEMENTS
ACKNOWLEDGEMENTS

- Thanks to Aram Harrow and Lukas Svec, for providing many of the examples in this talk