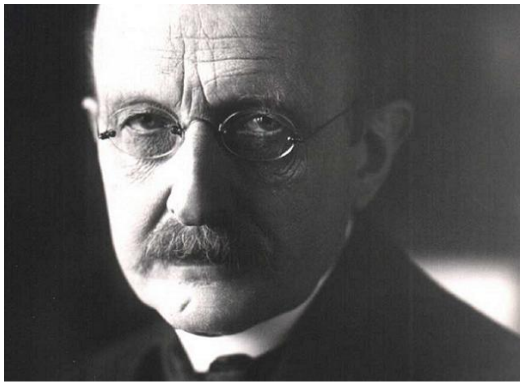
Quantum Machine Learning

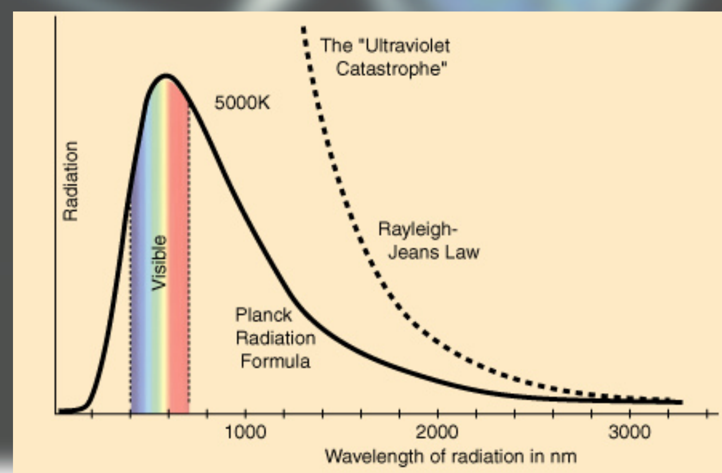
Nathan Wiebe, Researcher, Microsoft Research

Quantum Mechanics



Max Planck 1900

Classical electrodynamics fails to predict this



Solution: Assume energy is only emitted or absorbed in discrete “quanta” of energy.

$$E = nh\nu$$

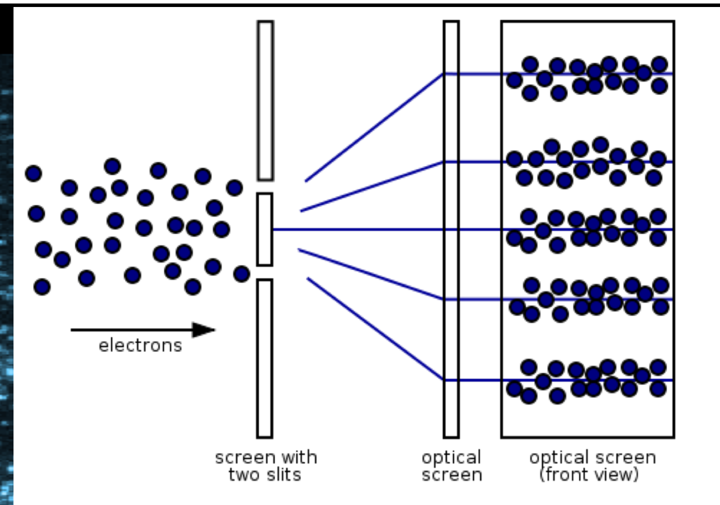
Wave Particle Duality

Einstein

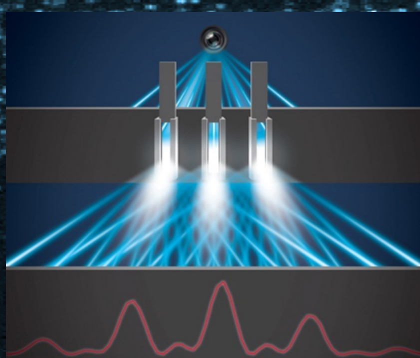
1905

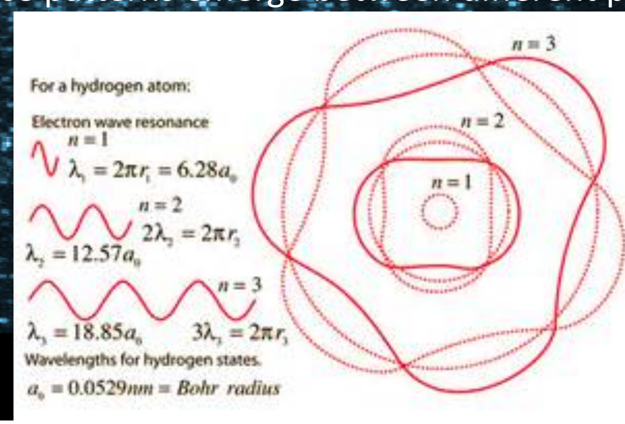
De Broglie

1924



Quantum systems can behave like waves.
Interference patterns emerge between different possible paths.

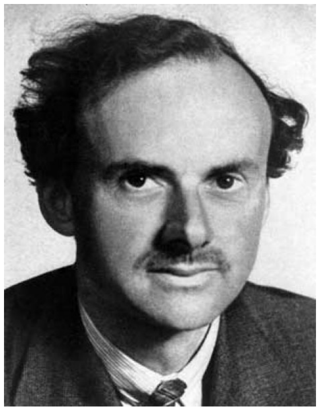




The ability to engineer quantum interference
is what gives quantum computers their power.

Formalizing Quantum Mechanics

Von Neumann



Dirac

1930

Hilbert

Quantum mechanics is just linear algebra.

States of matter

$$\psi = \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ \vdots \\ a_{2^n} \end{bmatrix}$$

Measurement

$$P(\psi \mapsto j) \;=\; \left|a_j\right|^2$$

(wave function collapse)

Transformations

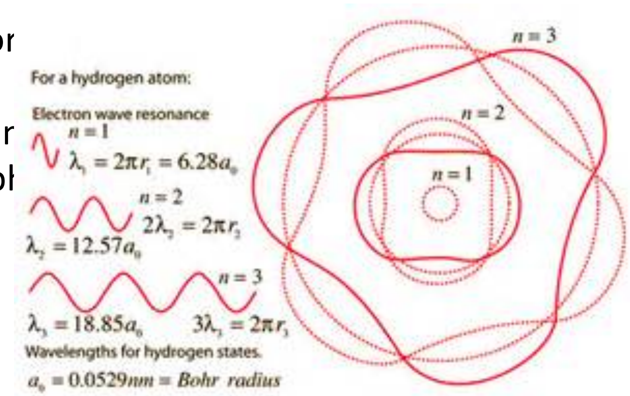
$$\psi \mapsto U\psi$$

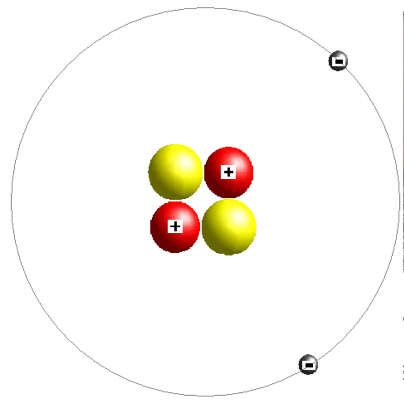
U is unitary:

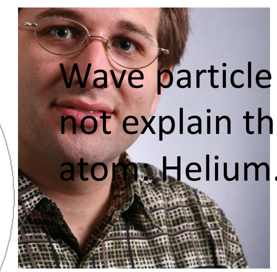
$$U^* U = 1$$

Their for nd physics.

Quantur ork upon which pl



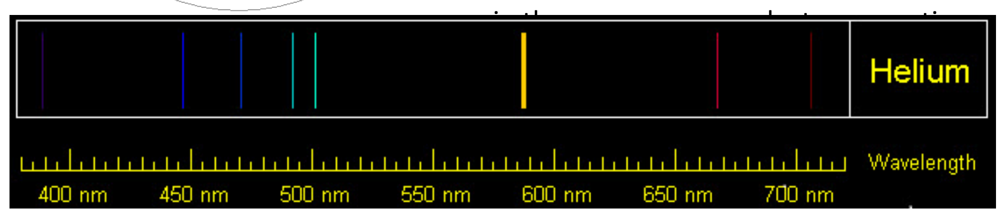




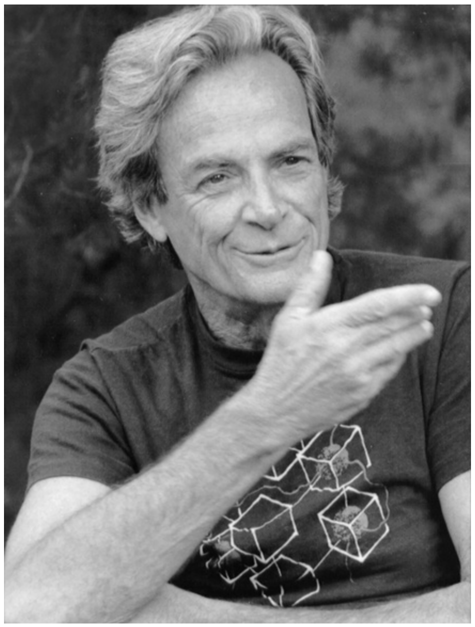
Wave particle duality alone does not explain the next most simple atom: Helium.

aronson

s? Even though it was discovered by



Quantum Mechanics and Computing



Feynman

1982

Feynman noted that the basic equations of quantum mechanics
are extremely hard to solve.

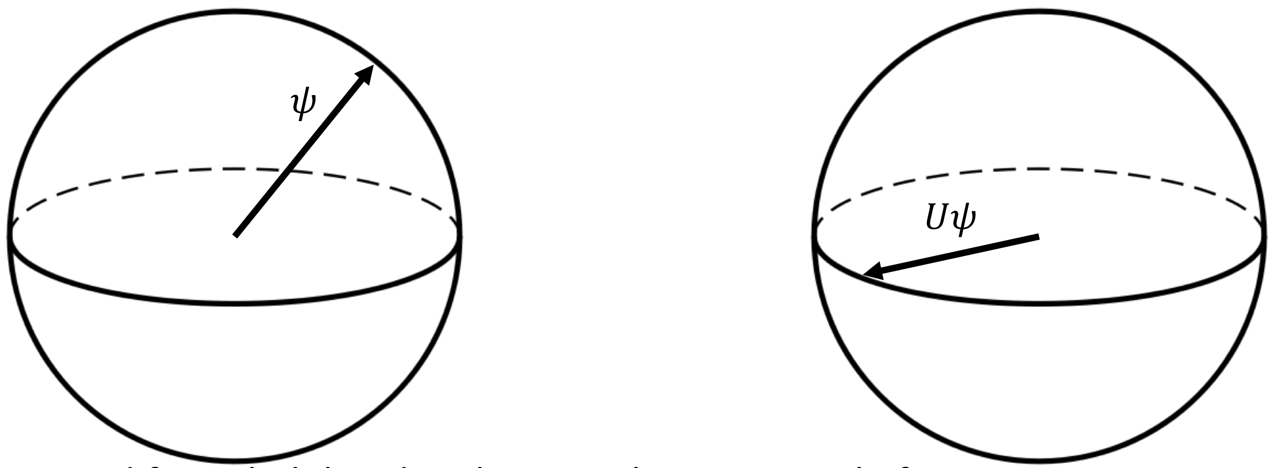
100 interacting spins require a 2^{100} dimensional wave function.

Very simple quantum systems can have extraordinary computational power.

His solution: build a computer that has quantum properties (interference and entanglement) built directly into the system.

What is quantum computing?

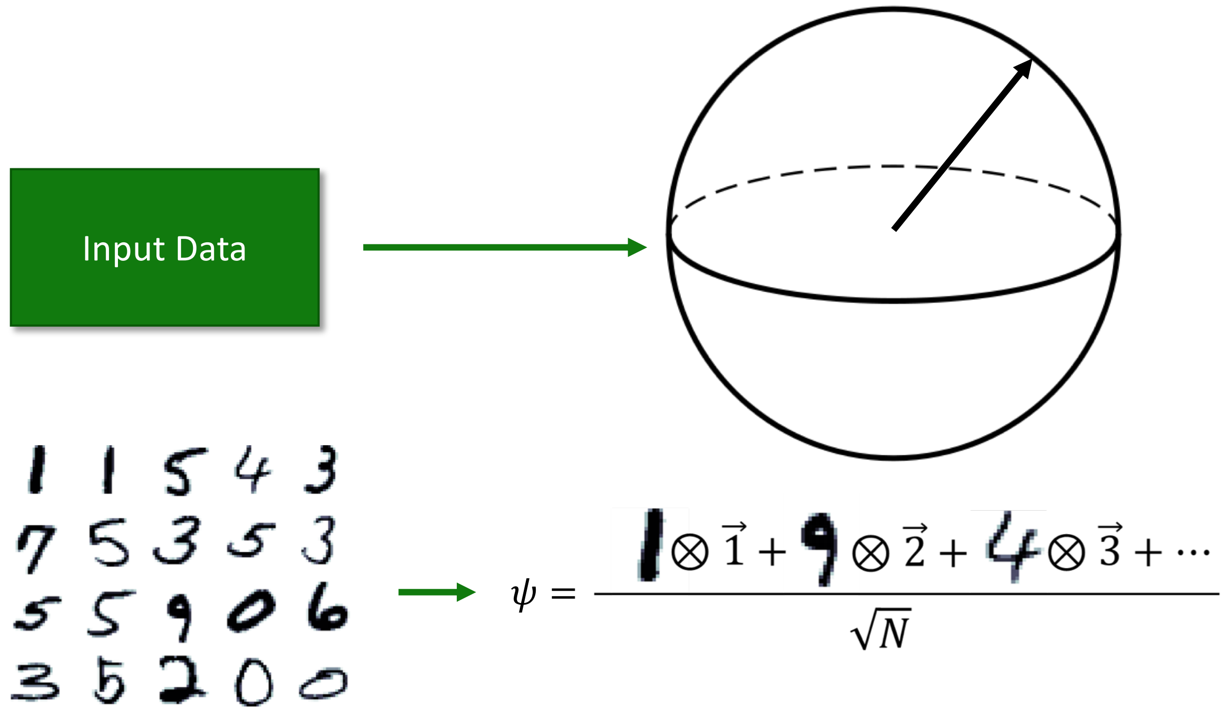
- A quantum computer is a machine that simply rotates quantum state vectors.



- This allows us to modify probability distributions that we sample from.
- Can represent a 2^n dimensional vector using only n quantum bits.
- We can only extract n bits of information from any individual computation.

Structure of quantum algorithms

- **Encoding**
- Transformation
- Measurement



Set of training vectors **Quantum state vector**

Structure of quantum algorithms

- Encoding
- **Transformation**
- Measurement

Reversible Classical Gates:

$$X = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

Not Gate

$$CNOT = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

Controlled Not

$$TOF = \begin{bmatrix} I & 0 \\ 0 & CNOT \end{bmatrix}$$

Controlled-Controlled Not

Quantum Operations:

$$\mathrm{H} = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix}$$

Hadamard Gate

$$\mathrm{P} = \begin{bmatrix} 1 & 0 \\ 0 & i \end{bmatrix}$$

Phase Gate

$$\mathrm{T} = \begin{bmatrix} 1 & 0 \\ 0 & \sqrt{i} \end{bmatrix}$$

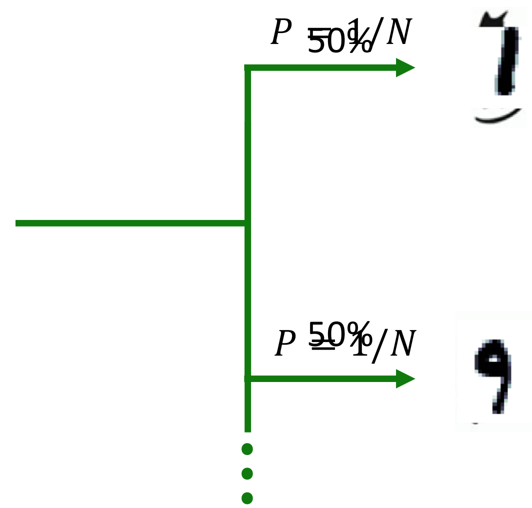
$\pi/8$ Gate

Structure of quantum algorithms

- Encoding
- Transformation
- **Measurement**

Measurement in quantum mechanics is the only non-linear operation.

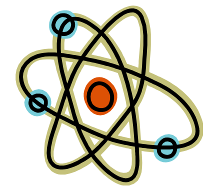
Looking at a system is in general irreversible and “collapses the state”.



Example: Quantum parallelism (aka superposition)

$$\frac{0 \;\; \cancel{\text{or}} + \;\; 1}{\sqrt{2}}$$

In quantum mechanics, we can be in many states at the same time!



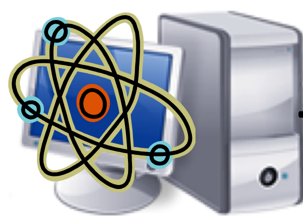
Input

P(1)

+ P(2)

⋮

+ P(2^n)



Output

Q(1)

+ Q(2)

⋮

+ Q(2^n)

Example: Grover’s search

- Interference can be used to reach target state in 3 operations.

- 1)Prepare initial state.

- 2)Reflect about space perp. to ideal.

- 3)Reflect about initial.

Target State

$$\begin{bmatrix}0\\0\\1\\0\end{bmatrix} = "2"$$

Initial State

$$\frac{1}{2}\begin{bmatrix}1\\1\\1\\1\end{bmatrix} = \frac{"0" + "1" + "2" + "3"}{2}$$

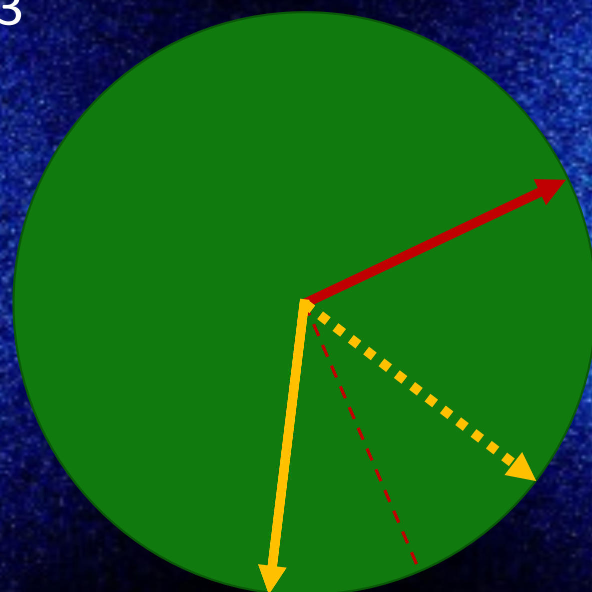
Example: Grover’s search

- Interference can be used to reach target state in 3 operations.

- 1)Prepare initial state.

- 2)Reflect about space perp. to ideal.

- 3)Reflect about initial.



Target State

$$\begin{bmatrix}0\\0\\1\\0\end{bmatrix} = \text{"2"}$$

Initial State

$$\frac{1}{2}\begin{bmatrix}1\\1\\1\\1\end{bmatrix} = \frac{\text{"0"} + \text{"1"} + \text{"2"} + \text{"3"}}{2}$$

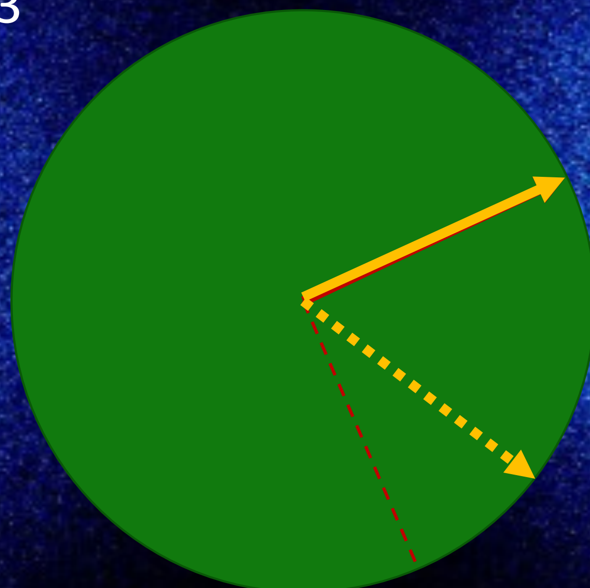
Example: Grover’s search

- Interference can be used to reach target state in 3 operations.

- 1)Prepare initial state.

- 2)Reflect about space perp. to ideal.

- 3)Reflect about initial.



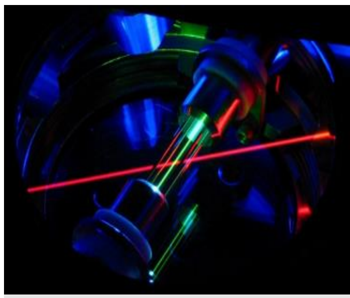
Target State

$$\begin{bmatrix}0\\0\\1\\0\end{bmatrix} = \text{"2"}$$

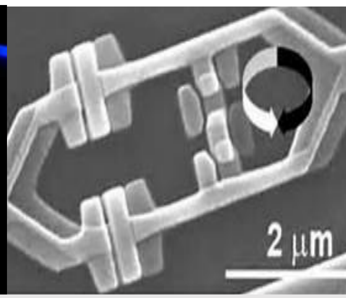
Initial State

$$\frac{1}{2}\begin{bmatrix}1\\1\\1\\1\end{bmatrix} = \frac{\text{"0"} + \text{"1"} + \text{"2"} + \text{"3"}}{2}$$

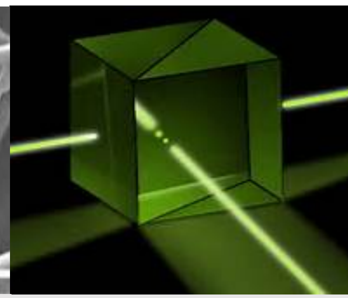
Quantum hardware technologies



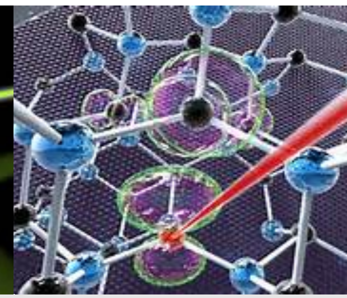
Ion traps



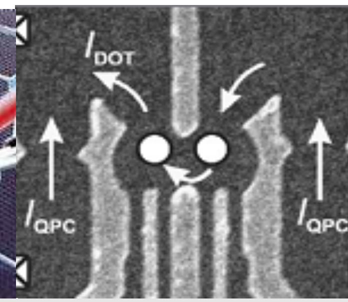
Super-conductors



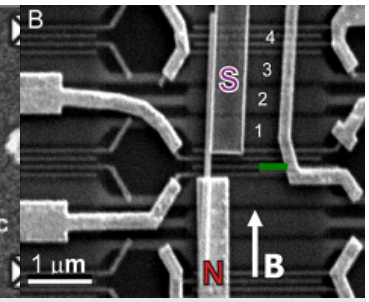
Linear optics



NV centers



Quantum dots



Topological

2000

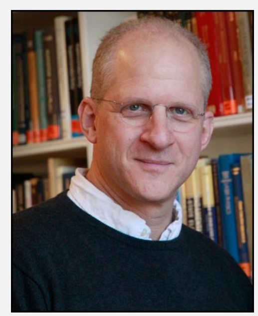
STATION Q

2006

STATION Q

REDMOND

SYDNEY, COPENHAGEN, DELFT



SANTA BARBARA

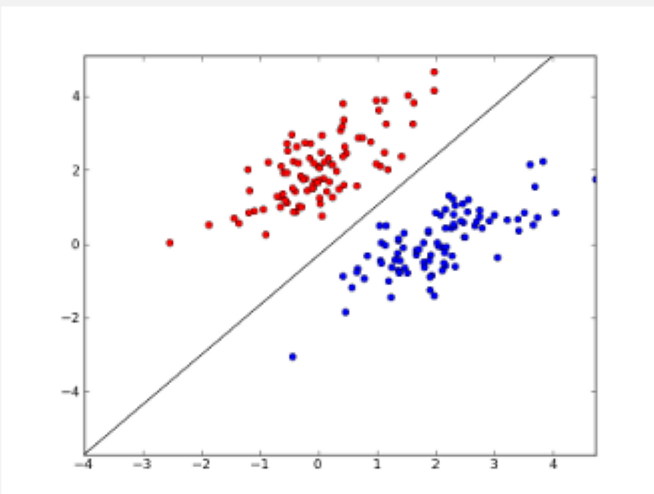
Quantum Machine Learning

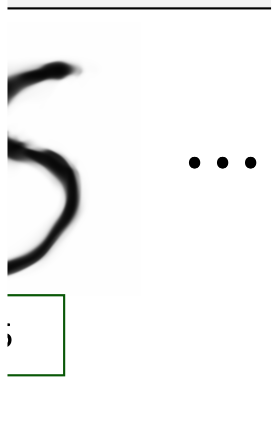
What is machine learning?

Machine learning uses computers to find patterns in data,classify data, or perform tasks.

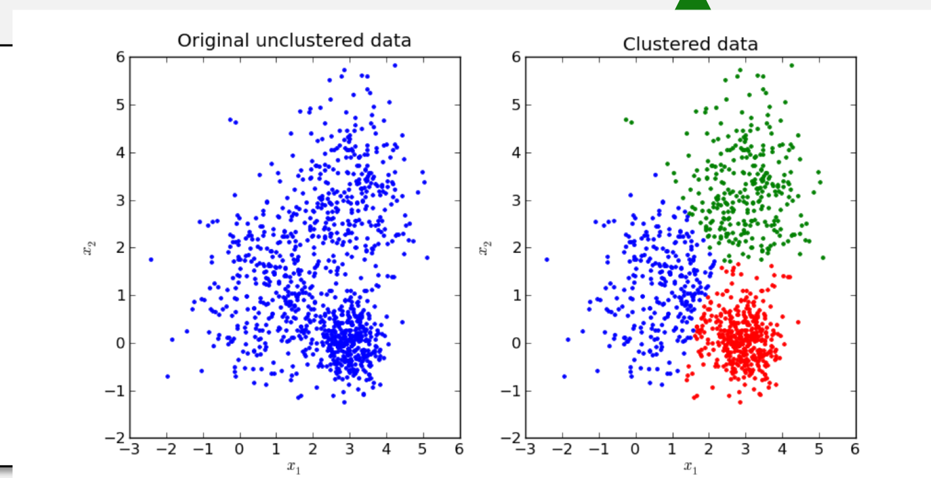
Computers are not usually told explicitly how to do this, or what *“features”* to use.

4





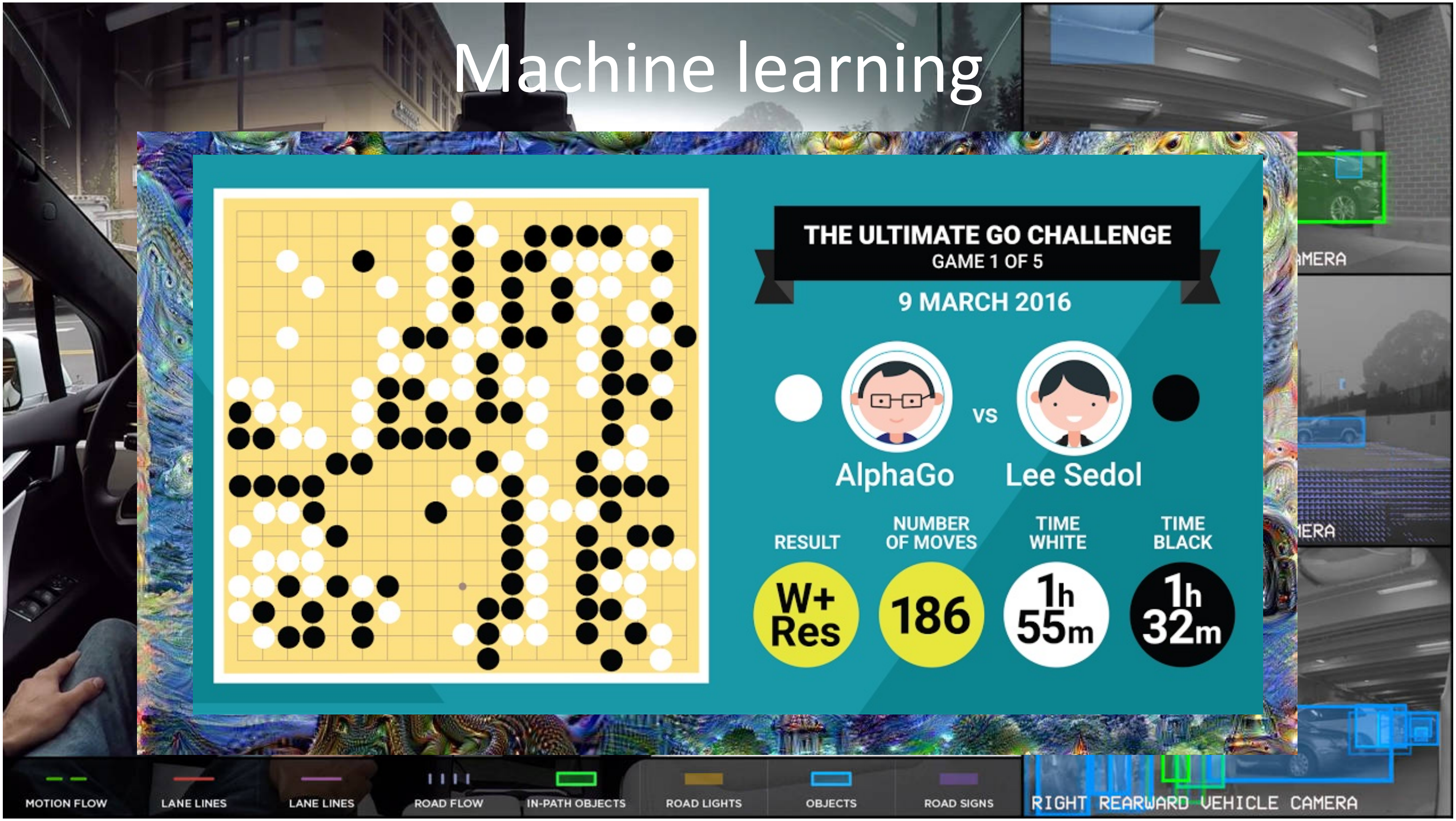
• • •



Training Data

???

Machine learning

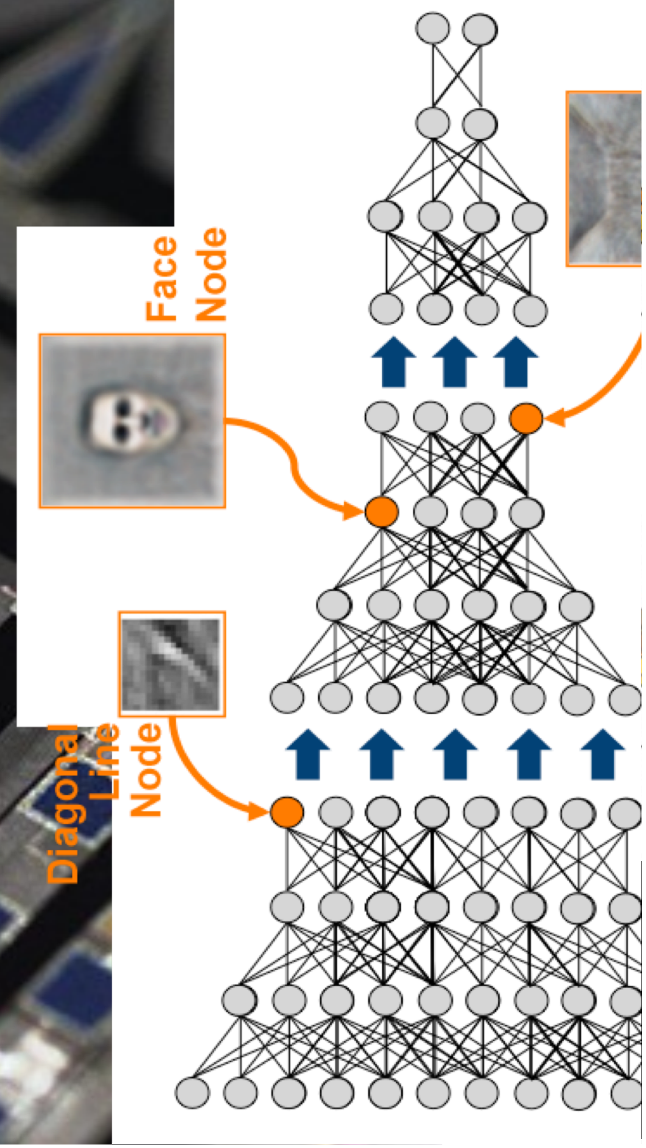


Goals of quantum ML

1. **Find new applications of quantum computers to machine learning.**

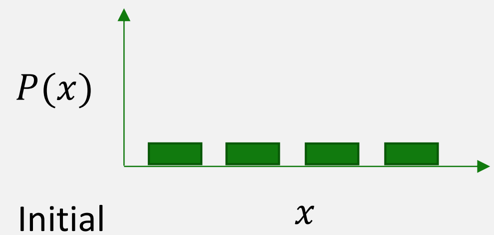
2. **Devise new methods to learn information about quantum systems.**

3. **Understand the fundamental limits physics places on learning.**

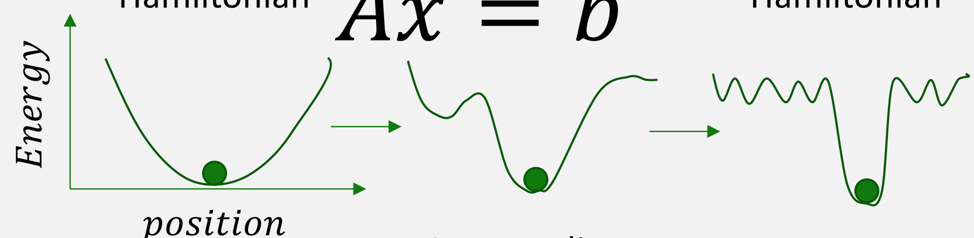


What tools are available?

- Amplitude Amplification
  - Quantum interference is used to modify underlying probabilities.
  - Quadratic speedups (Grover Search).
- Adiabatic Optimization
  - Slowly varying quantum evolution is used to solve energy minimization.
- Linear Systems Algorithms
  - Uses Hamiltonian simulation to perform matrix inversion.



$$Ax = b$$



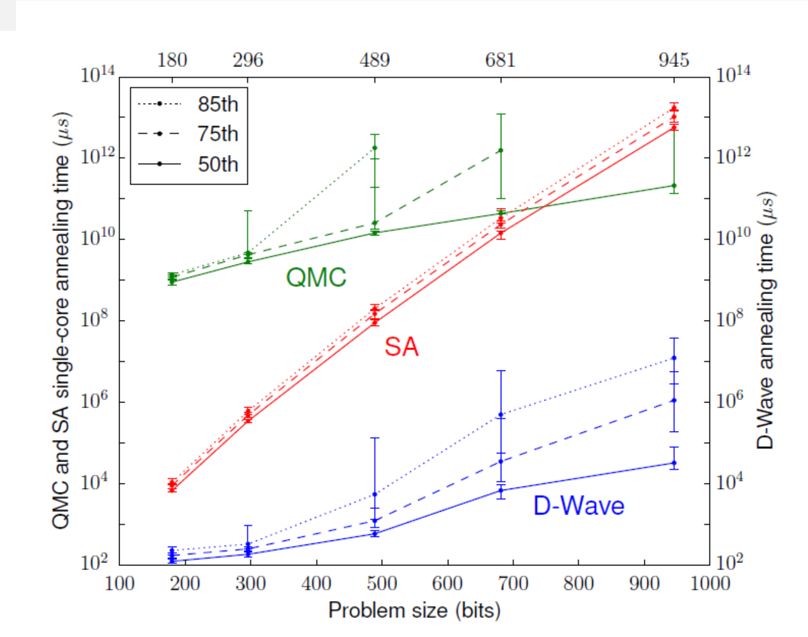
Challenges

Input problem: A huge quantum database may be needed for typical problems.

Output problem: Reading result is exponentially expensive.

Speedups: Performance advantages relative to heuristics are hard to prove.

Adiabatic Optimization



- Map optimization problem to a quantum Hamiltonian.
- Use adiabatic evolution to prepare (approximate) groundstate.
- Can be run on existing hardware.
  - Eigenvalue gap provides protection.
- Numerical evidence for constant factor speedups exists for optimization (Denchev et al 2016).

Input problem	Output problem	Speedup
N	N	Generically Unknown

Linear systems algorithm

- We want to solve Ax = b, where A is a huge but sparse matrix.

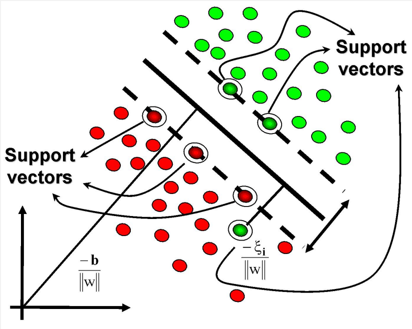
$$\begin{bmatrix} 0 & 1 & 0 & 0 & 5 & 0 \\ 1 & 0 & 0 & 4 & 0 & 0 \\ 0 & 0 & 3 & 0 & 3 & 0 \\ 0 & 4 & 0 & 0 & 0 & 4 \\ 5 & 0 & 3 & 0 & 0 & 0 \\ 0 & 0 & 0 & 4 & 0 & 6 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

- Quantum computers can output the state $A^{-1}b/||A^{-1}b||$ *exponentially faster* than inversion (Harrow, Hassidim and Lloyd 2009).
- Requires a fast way of preparing data and computing A_{xy}.

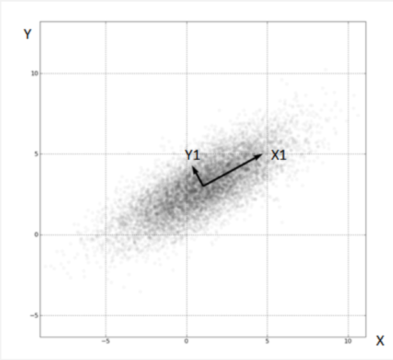
Input problem	Output problem	Speedup
Y	Y	Exponential

Quantum algorithms that use this idea

- Quantum support vector machine algorithm (Rebentrost, Lloyd Mohseni 2013)



- Quantum principal component analysis (Lloyd, Mohseni, Rebentrost 2014)



- Both can give exponential speedups but can suffer from the output problem.

Least squares fitting

- Given b, A find x such that $||Ax - b||_2$ is minimized.

- Uses a modified form of the Harrow, Hassidim and Lloyd algorithm.

- Exponential speedups for well-conditioned fitting problems with sparse A. (Wiebe, Braun, Lloyd 2013)

Input problem	Output problem	Speedup
Y	N	Exponential

Clustering / nearest neighbor classification

- K-means clustering finds clusters in a large data set.
- Nearest neighbor classifiers classify based on the closest classified vector.
- Quadratic speedups for both problems.
  (Wiebe, Kapoor, Svore, 2013)
  (Rebentrost, Mohseni, Lloyd 2013)

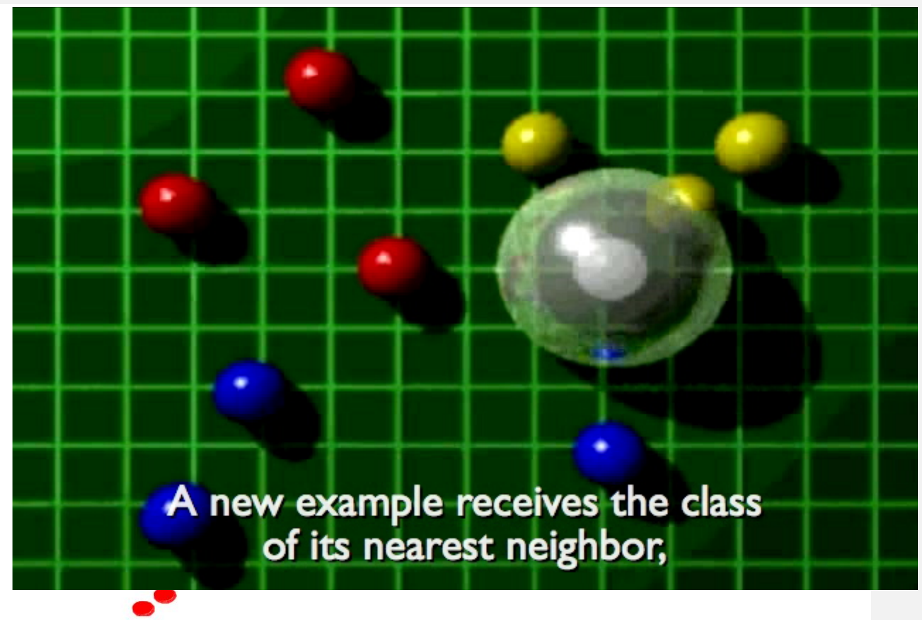
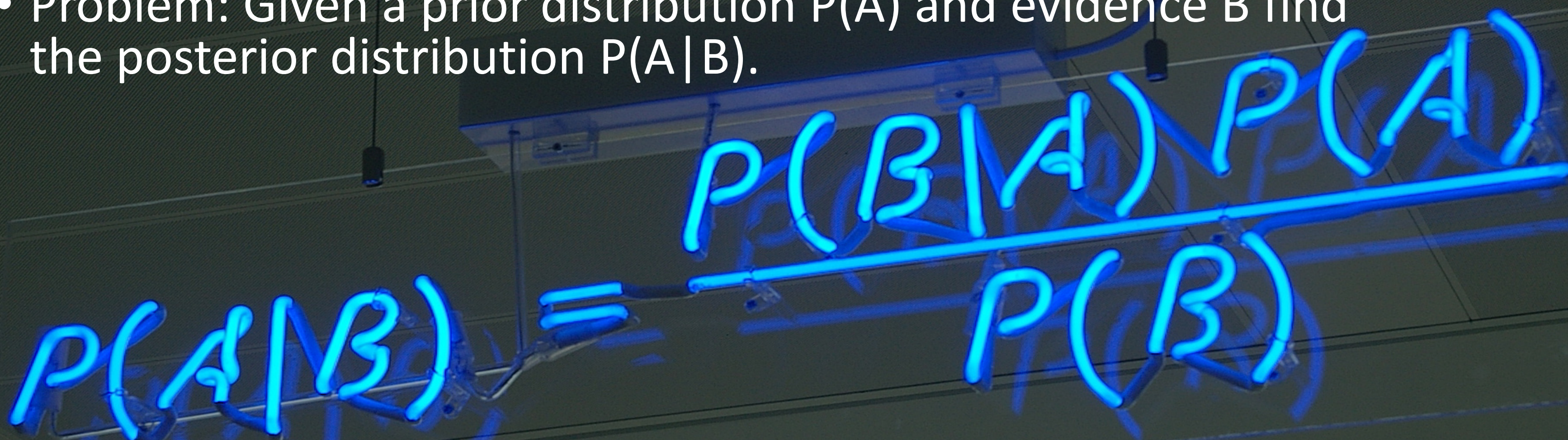


Fig. 13. Exemplary K-Means result

Input problem	Output problem	Speedup
Y	N	Quadratic

Quantum Bayesian Inference

- Problem: Given a prior distribution P(A) and evidence B find the posterior distribution P(A|B).

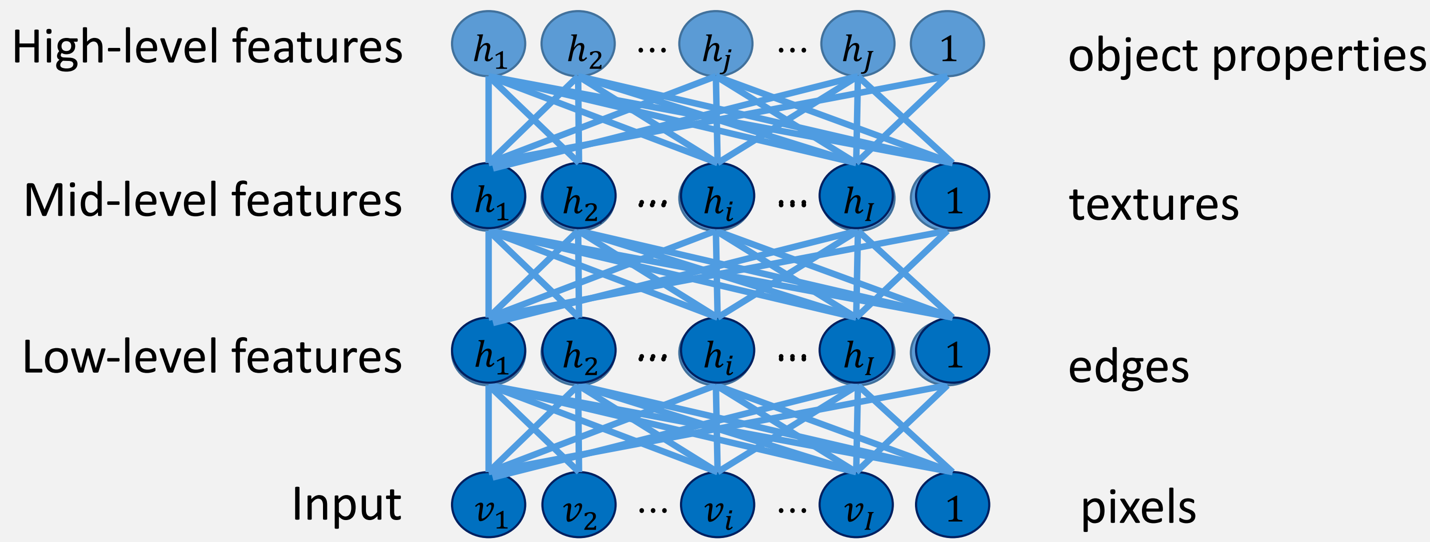


- Prior is stored in a quantum state and measurement is used to perform (non-linear) update.

Input problem	Output problem	Speedup
N	N	Quadratic

Quantum Neural Nets

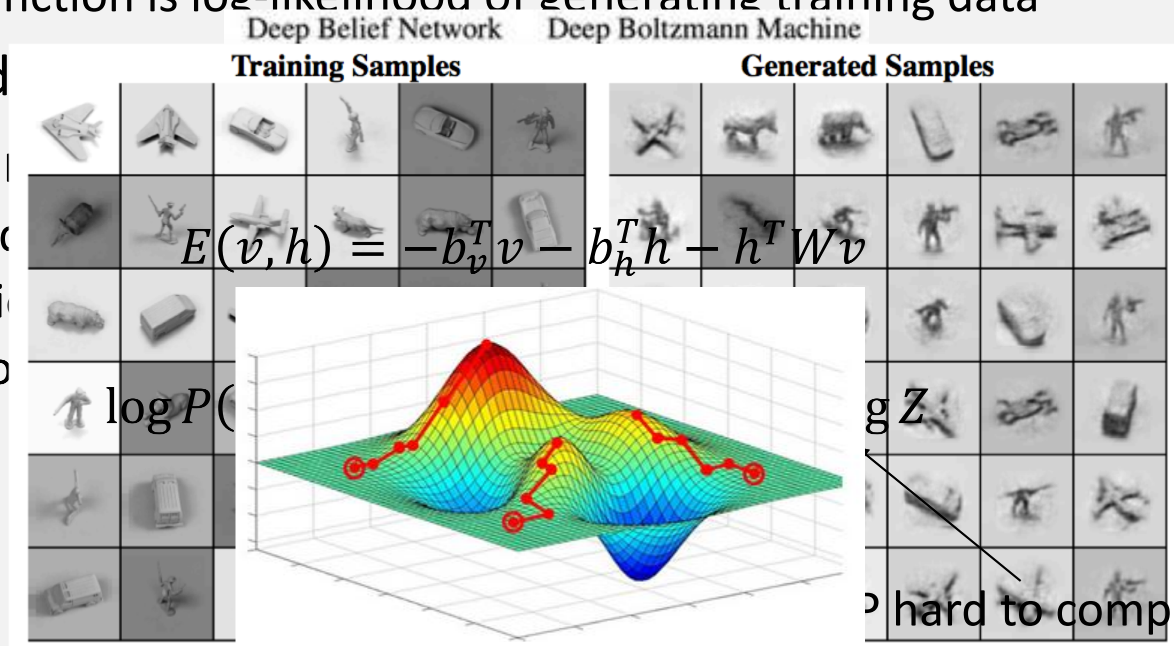
Deep learning networks



- Visible units are observable (training vector)
- Hidden units are not directly observed
- Both units take 0,1 values
- Training involves changing the interaction strengths to minimize energy for training vectors
- Probabilities set according to Gibbs distribution

Deep Boltzmann machines

- Recurrent neural networks with undirected edges.
- Boltzmann machines are generative.
  - Objective function is log-likelihood of generating training data
- Optima found
- Approximatio
  - Contrastive d
  - Approximati
  - Expensive to



$$E(v,h) = -b_v^T v - b_h^T h - h^T W v$$

$\log P($ ... $g\,Z$

P hard to compute

Salakhutdinov, Hinton (2009)

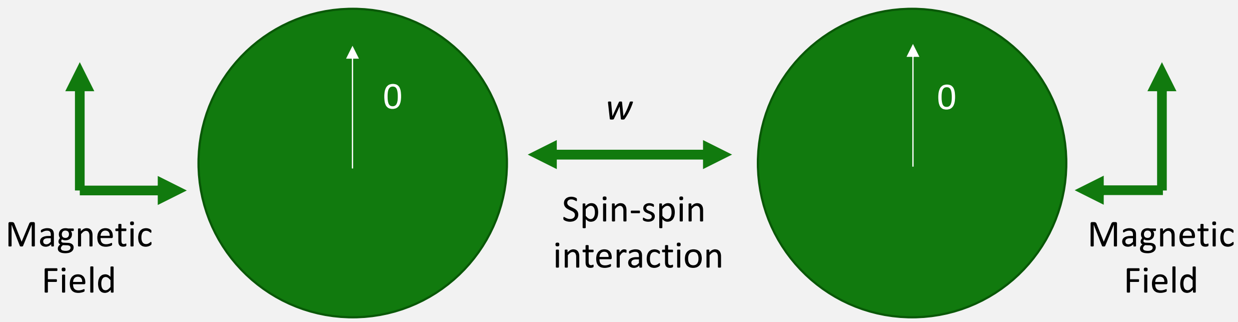
Boltzmann machine training

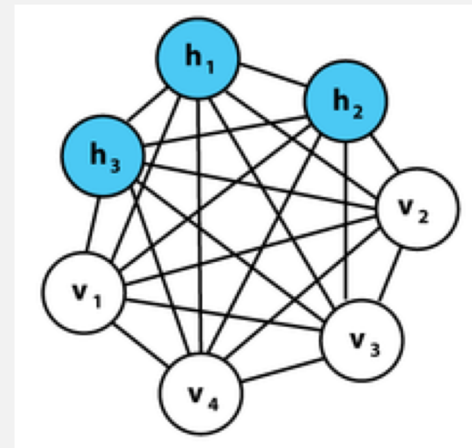
- Boltzmann machines model data using an Ising model in thermal equilibrium.
- Training these models is expensive.
- Quantum computing gives quadratic speedup, allows for more general graphical models and suggests new classical training algorithms.
  (Wiebe, Kapoor, Svore 2014)
  (Wiebe, Kapoor, Svore, Granade 2015)

Input problem	Output problem	Speedup
N/Y	N	Quadratic

Quantum Boltzmann Models

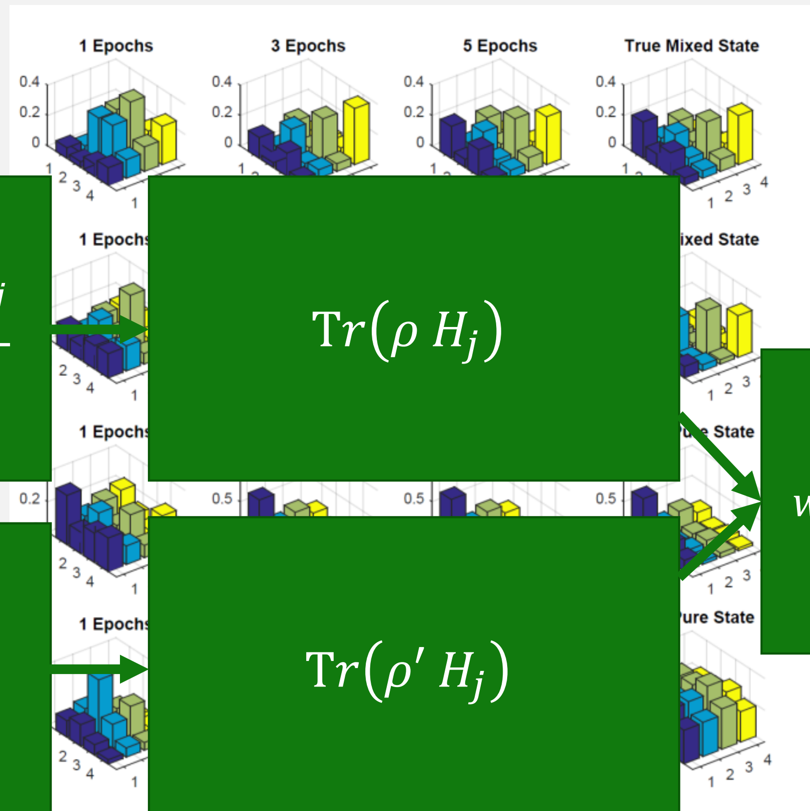
- $P(v, h) \propto \ e^{-E(v,h)}$, where E is a scalar energy.





- Energy can directly be computed from values of the spins.
- With a rotated magnetic field the model becomes “quantum”
- Creates quantum entanglement between units.
- Allows a form of efficient partial tomography.

Training quantum Boltzmann machines



$\{w_j\}$

$$\rho = \frac{e^{-H}}{Z} = \frac{e^{-\sum_j w_j H_j}}{Z}$$

$$\mathrm{Tr}\left(\rho\, H_j\right)$$

ρ'

$$\mathrm{Tr}\left(\rho'\, H_j\right)$$

Gradient descent on quantum relative entropy between p and p'.

$$w_j \rightarrow w_j + \eta\left(\mathrm{Tr}\left(\rho\, H_j\right) - \mathrm{Tr}\left(\rho'\, H_j\right)\right).$$

Conclusion

- Quantum computers can provide important speedups for many problems in machine learning.

- Scalable quantum computers can revolutionize machine learning.

- Many important questions remain:
  - What does it mean to learn for a quantum system?
  - Are quantum models qualitatively richer than deep neural networks?
  - Can quantum computers perform types of learning that are impossible for ordinary computers?

Supplementary Slides

Overlap with Gibbs state

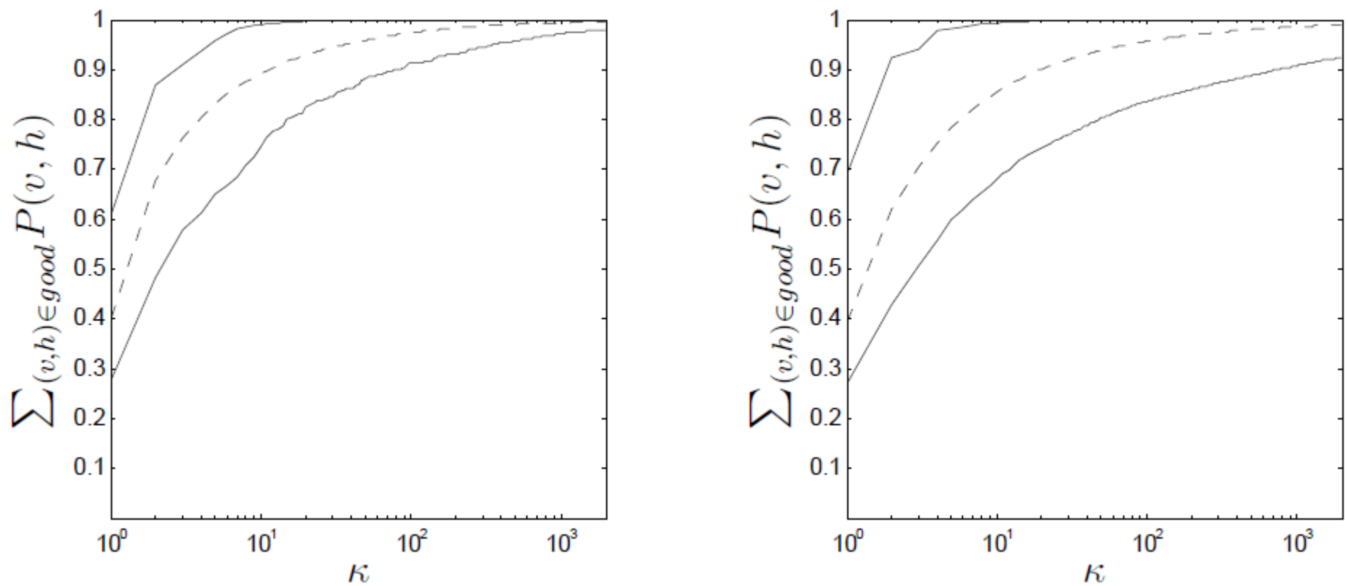


Figure 2: Probability mass such that $\mathcal{P}(v,h) \leq 1$ vs κ for RBMs trained on (22) with $n_h = 8$ and $n_v = 6$ (left) and $n_v = 12$ (right). Dashed lines give the mean value; solid lines give a 95% confidence interval.

Learning of O_{ML} objective

- Significant differences between learning the objective function with CD and ML for a 3-layer dRBM
- Complex models benefit from ML training!

n_v	n_{h1}	n_{h2}	CD	ML	% Improvement
6	2	2	−2.7623	−2.7125	1.80
6	4	4	−2.4585	−2.3541	4.25
6	6	6	−2.4180	−2.1968	9.15
8	2	2	−2.8503	−3.5125	−23.23
8	4	4	−2.8503	−2.6505	7.01
8	6	4	−2.7656	−2.4204	12.5
10	2	2	−3.8267	−4.0625	−6.16
10	4	4	−3.3329	$\boxed{-2.9537}$	11.38
10	6	4	$\boxed{-2.9997}$	−2.5978	13.40