For Papers we love Bhubaneshwar. Session II

Opportunities and challenges for quantum-assisted machine learning in near-term quantum computers-

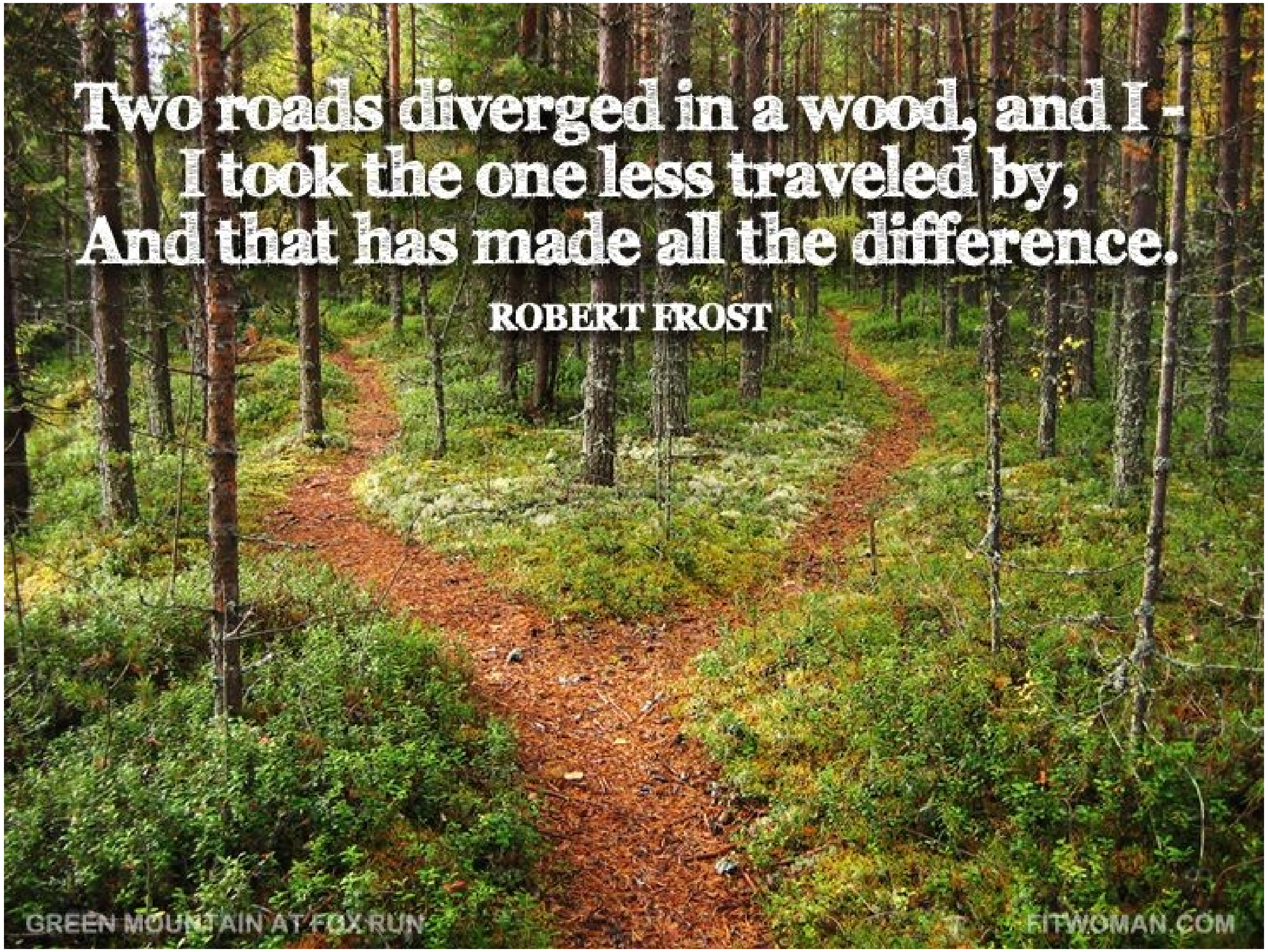
Alejandro Perdomo-Ortiz, Marcello Benedetti, John Realpe-Gómez, Rupak Biswas

whoami

My name is Ankit Jha and I’m in my penultimate year of engineering

What is a classical computer?

A turing model problem solving machine, takes binary decisions



Ternary computing. Thinkihng beyond binary

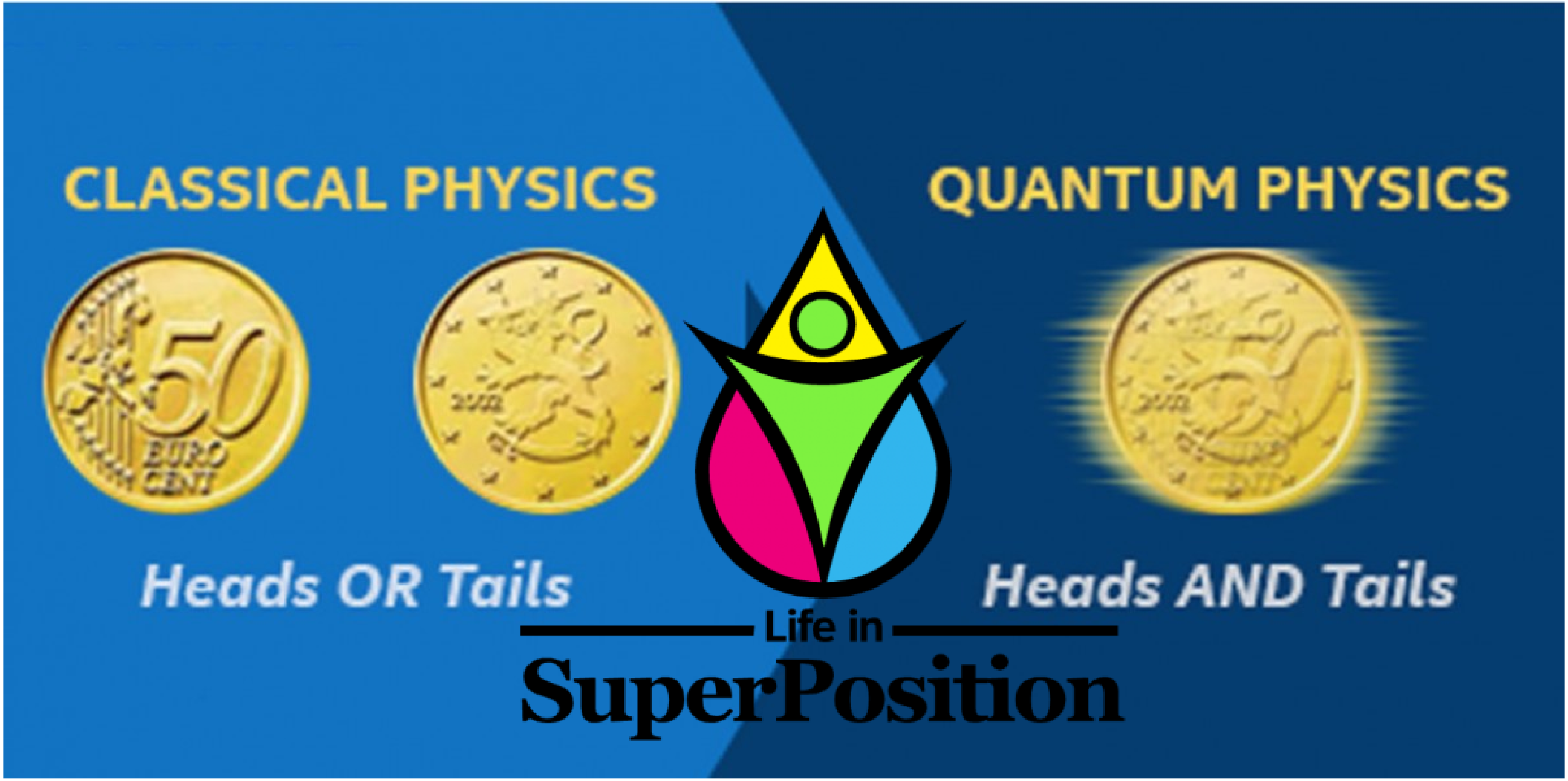
What if we had a computer that would use a base 3 instead of a base 2. (We use k maps to represent the 3 states)

Radix Economy

BASE b	Radix Value
1	infinity
e	1.0000
2	1.0615
3	1.0046

BASE b	Radix Value
4	1.0615
5	1.1429
6	1.2319
7	1.3234

Quanta, Particles and superpositions



What adds the "super" to superposition? Uncertainty Principle

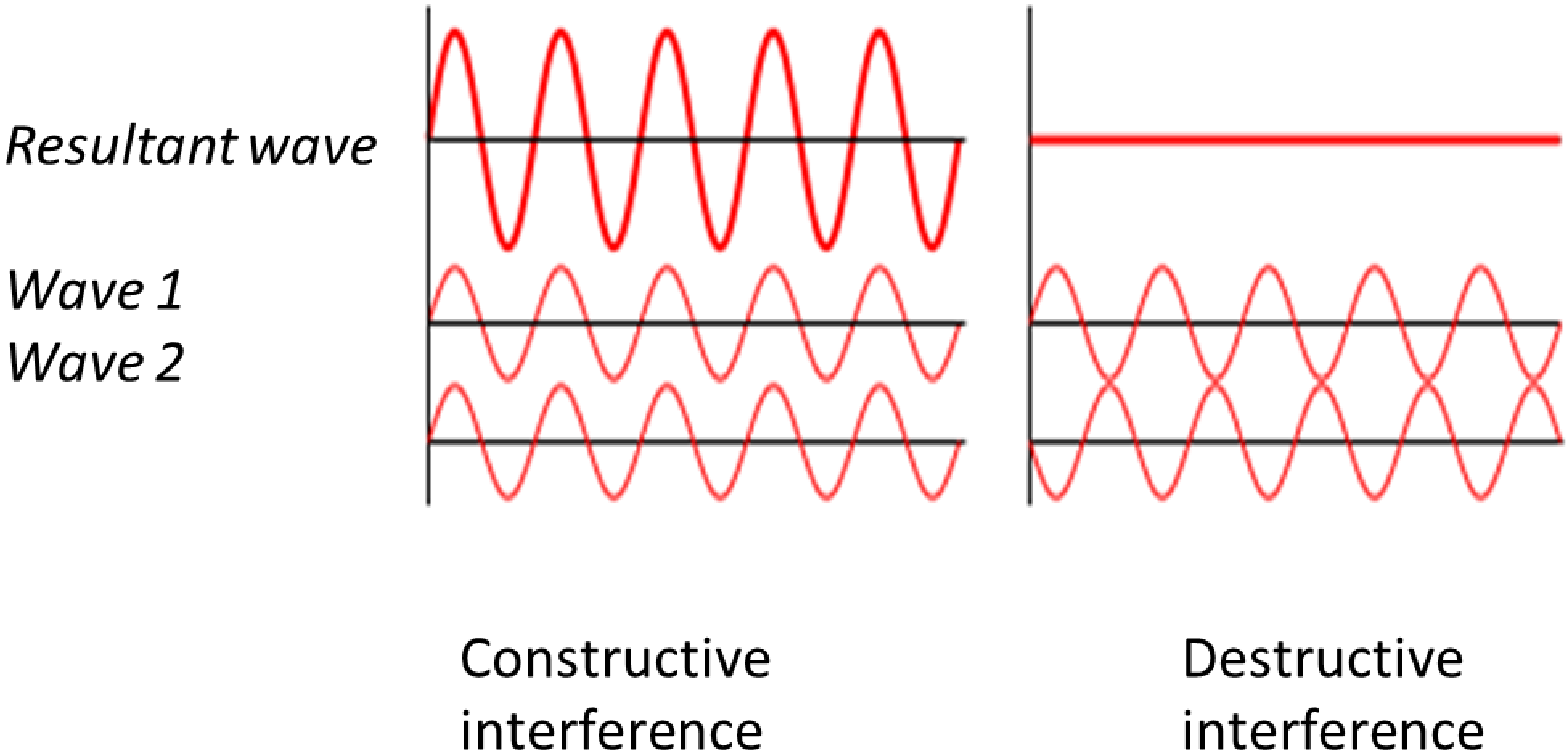
Superposition- Wanted:
Schroedinger's Cat - Dead and Alive.

Put a kitty in a box with gunpowder, let the powder blast. There are 2 outcomes- cat sees blast and dies or cat doesn’t see blast and lives but while it is in box it is both alive and dead.

Understanding entanglement of realities and quantum entanglement

All objects are waves.
Once these waves are split
and joined they behave
in the same way as
the two waves inter
with each other.
This is same as
entanglement.

Interference



What if all our bits were superpositions?

- What is data entropy? How many bits do we need to represnt data?
- What will happen if we have infinite number of bits
- Operations like addition, multiplication, blah blah.
- An array is actually a mapping function between the index and the values.

What if we entangle all these bits together

- What is a qBit or a quantum bit
- Why can’t we just measure a qbit? (Wave function collapse and revelation of entanglement)
- **Classical correlation-** a something is changed and another something changes. Like 2 envelopes problem
- **Q uantum correlation-** Let’s play a cricket match world cup which is fixed

A recap of the uncertainty principle, Schrodinger equation , Hamiltonian symbol

- **Uncertainty-** you can’t measure the position and momentum at the same time
- **Hamiltonian operator-** total energy of the system. Veerything in Quantum mechanics has a operator for itself
- **Schrodinger equation-** An equation that represents a relation between evrything else and gives a probabilistic wave fucntion for the system
- **Eignvalues-** in time dependent schrodinger waves, the wave at a specefic time. (Not exactly this but let’s keep it that way)

Quantum Parallelism

Quantum parallelism arises from the ability of a quantum memory register to exist in a superposition of base states. (0 and 1). A quantum memory register can exist in a superposition of states, each component of this superposition may be thought of as a single argument to a function. So when some a function is performed to one component among all components in superposition, we basically perform it on all the components of the component set.

Solving our first problem

Take a contigious set of numbers-

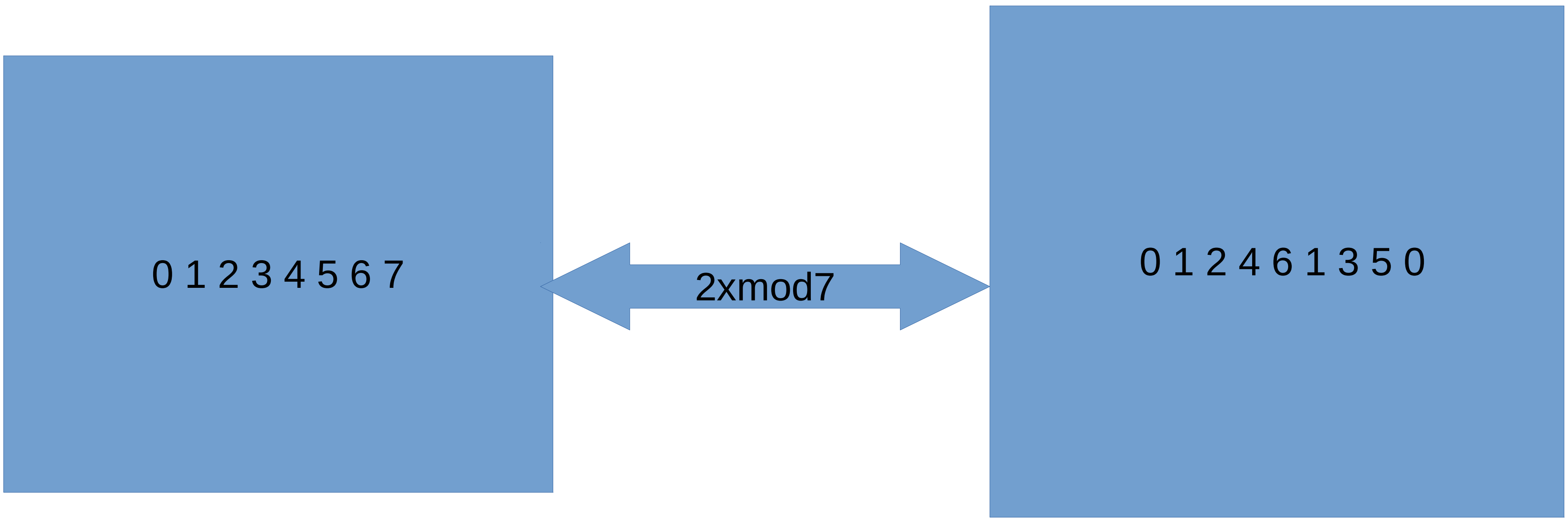
A[]=0-1-2-3-4-5-6-7

We need to take every element and calculate 2*element (mod 7) where mod is the modulus operator

```
for(i in A[]):
    A[i] <- (2*i)%7
```

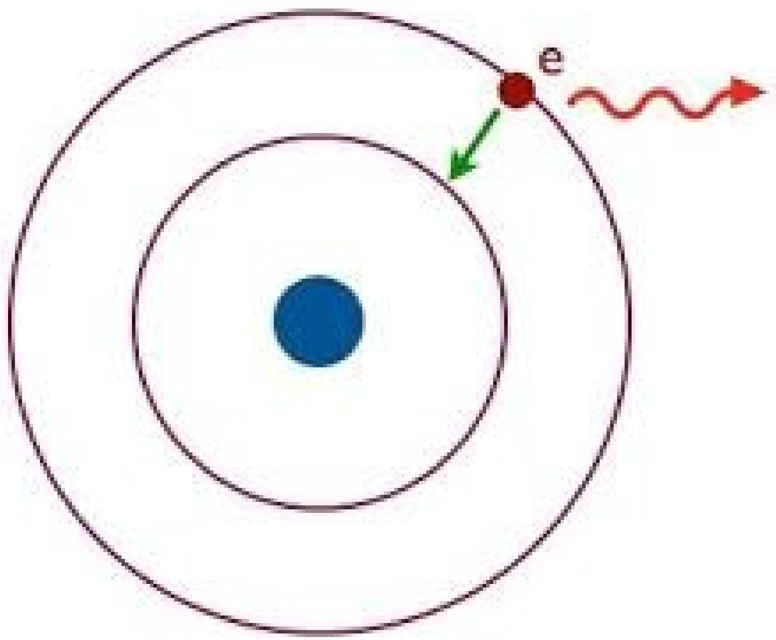
Solving via using a Quantum computer

The probability of finding a 0 and 1 is 2/8 and for everything else it is 1/8

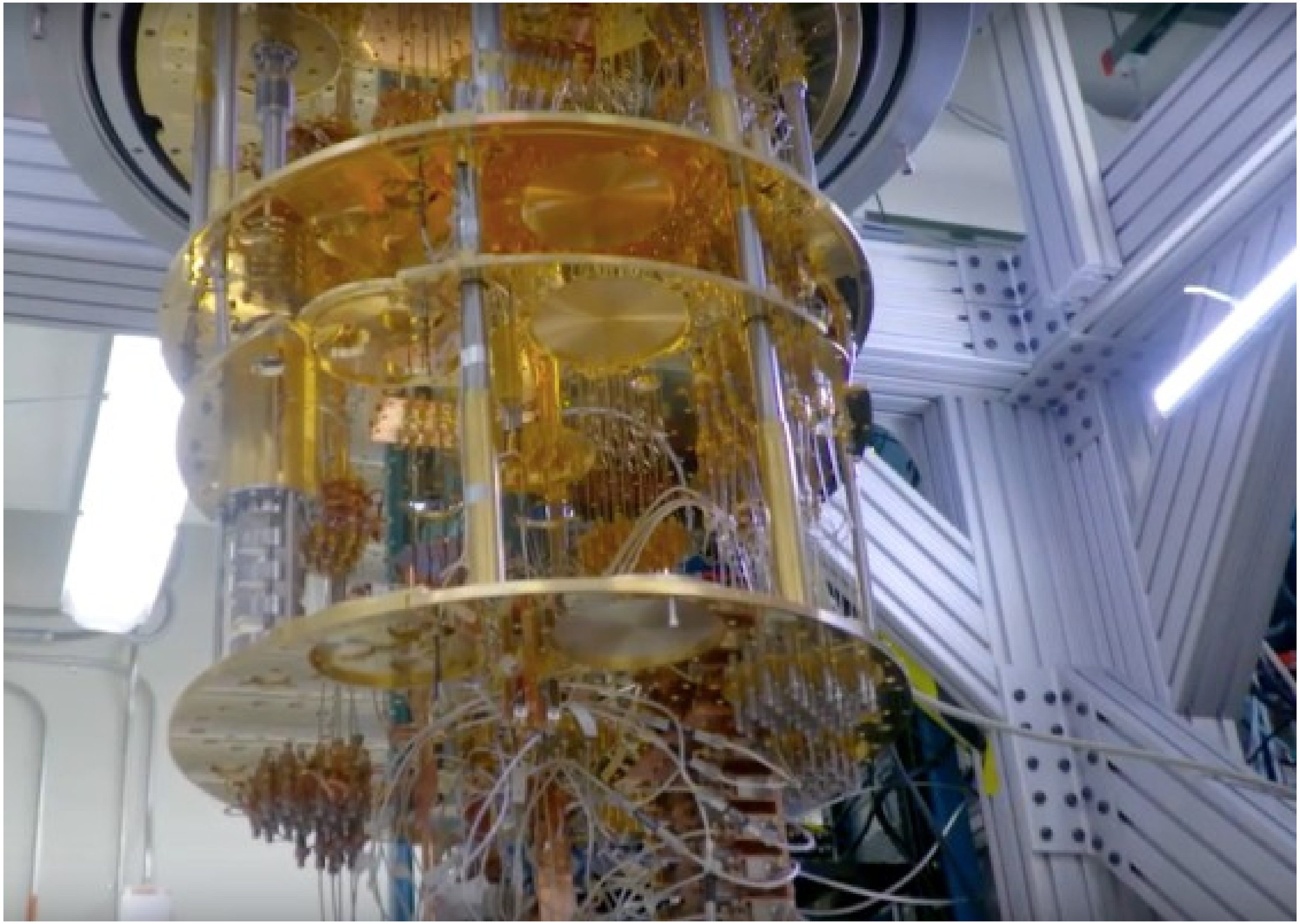


How did you do that?

- Understanding ground state and excited states
- Understanding how states get excited



Why do machines need to stay cold



The Gates of Qunatum computers

- Representation as superpostion of classical bits
- Understanding of the classical gates
- Why classical gates won’t work here
- Hadamad, Cnot, Cx, cy, cz, Pauli, etc
- All gates are unary matrices

Just getting an idea of what the Gates look like

- Qbit= Ax[0> + Bx[1>
- Here is Hadamad gate which makes probability of 0 and 1 same 0 becomes $\frac{|0\rangle + |1\rangle}{\sqrt{2}}$ and 1 becomes $\frac{|0\rangle - |1\rangle}{\sqrt{2}}$

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

Coding the ancient way

N 2 # create a new quantum bit and identify it as '2'

E 1 2 # entangle qubits '1' and '2', qubit 1 already exists and is considered input

M 1 0 # measure qubit '1' with an angle of zero (angle can be anything in [0,2pi]

qubit '1' is destroyed and the result is either True or False

operations beyond this point can be dependent on the signal of '1'

How do you code this D-Wave thing?

```
//Single Q Measurement
OPENQASM 2.0;
include "qelib1.inc";

// Register declarations
qreg q[1];
creg c[1];

// Quantum Circuit
measure q -> c;
```

Coding it with an interface

```
# single_q_measurement.py
from qiskit import QuantumCircuit, QuantumRegister, ClassicalRegister,
execute

# Define the Quantum and Classical Registers
q = QuantumRegister(1)
c = ClassicalRegister(1)

# Build the circuit
single_q_measurement = QuantumCircuit(q, c)
single_q_measurement.measure(q, c)

# Execute the circuit
job = execute(single_q_measurement, backend = 'local_qasm_simulator',
shots=1024)
result = job.result()

# Print the result
print(result.get_counts(single_q_measurement))
```

Structured code in Q

Qreg x1(); // 1-qubit quantum register with initial value 0 Qreg x2(2,0); // 2-qubit quantum register with initial value 0

Some more structured code in qcl

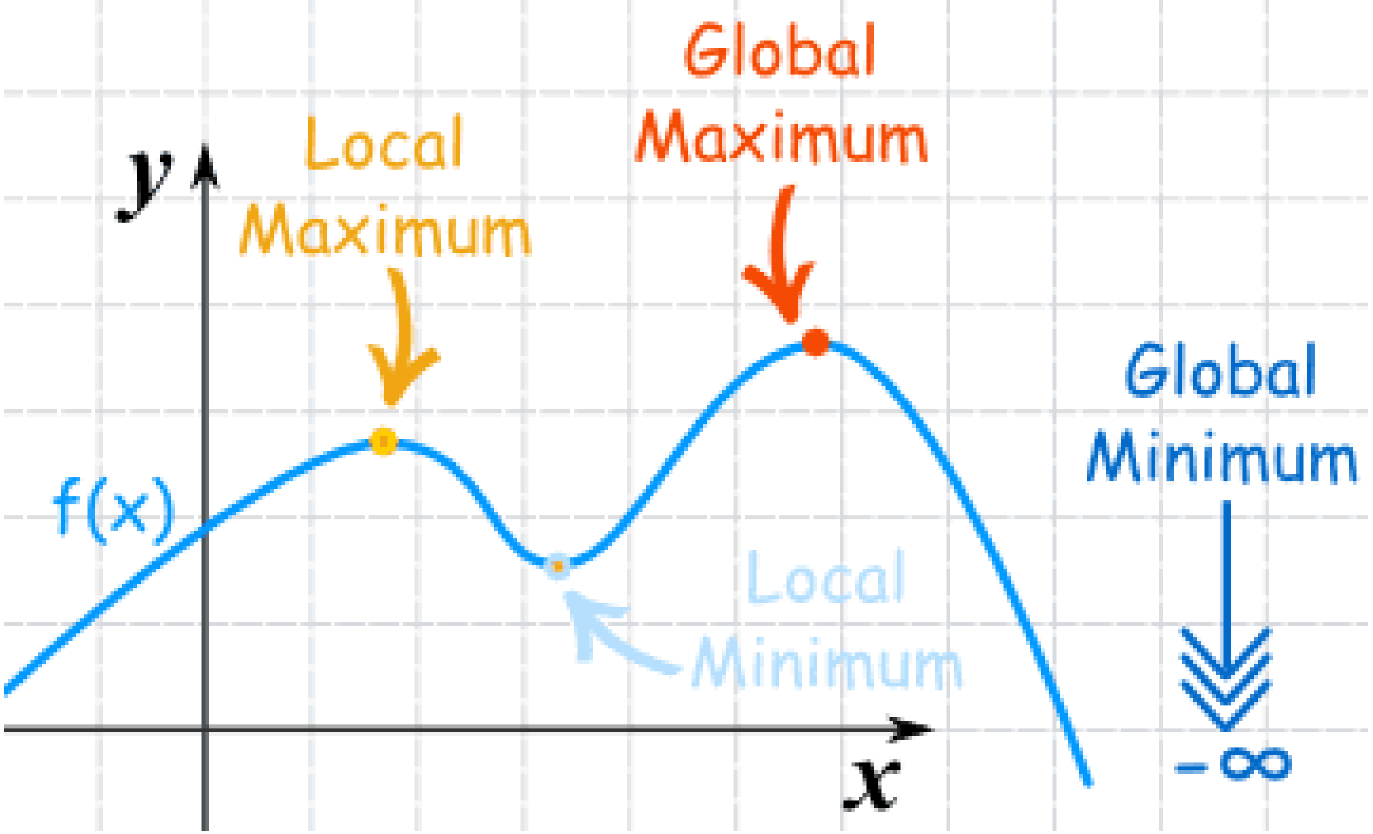
```
procedure main() {
    qureg qstate[n];
qureg ostate[1];
measure qstate, outp;
print outp;
plot qstate;
dump qstate;
}
```

A brief summary of What we actually did

- OpenQASM
- Q
- Qcl
- Qiskit
- Cirq
- Q#
- Why we won’t write any code here

Introduction to minimization and maximization

- An example of a linear growth function from the real world



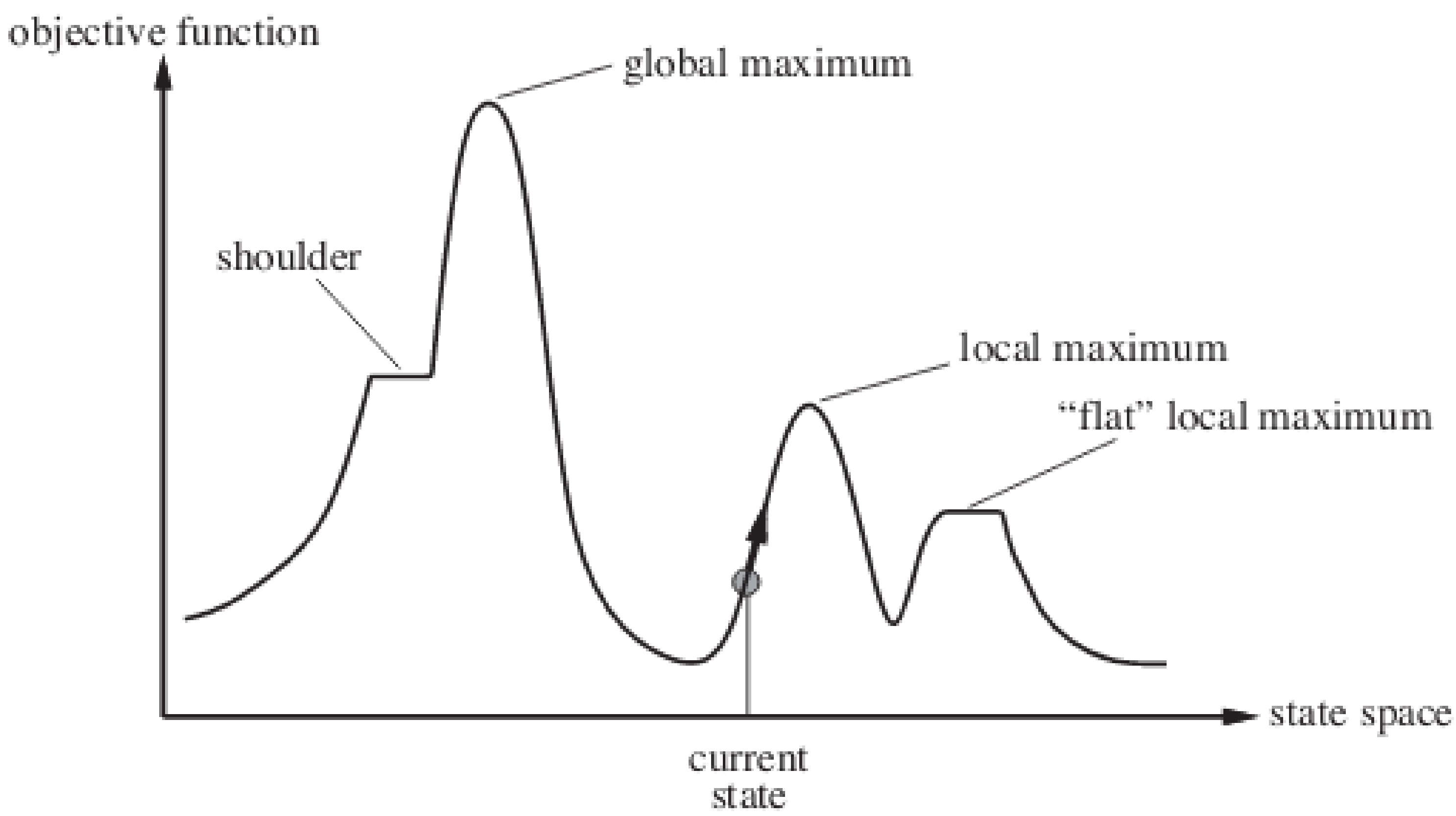
What are Heuristsics

something starts at a point and keeps moving in a direction. After every movement it checks itself and them continues moving. This is called Heuristics.

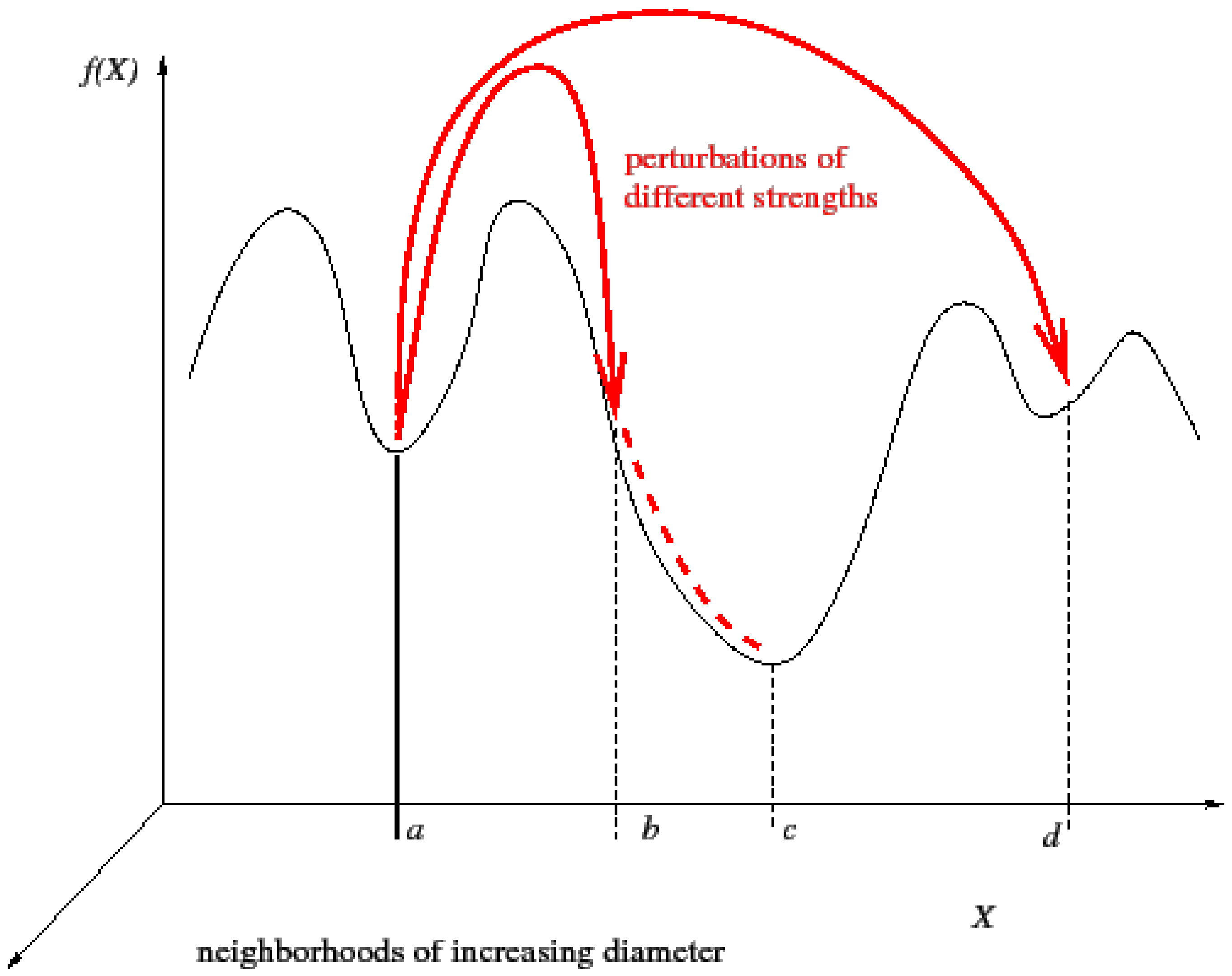
What are metaheuristics?

Searching for the Heuristics- It is a problem independent approach which means we can use the same approach to all problems (like in Tic Tac Toe we need Alpha Beta tuning and in TSP we need a search based approach but we cannot use Alpha Beta in TSP). Here we don’t know the underlying features of the problem. It is kind of a black box problem for example- Quantum annealing, genetic algorithms (Which are used for TSP or edge weight calculation for ANNs).

Example- Hill climbing maximization



Randomized Hill Climbing



Qauntum Annealing and Fluctuation

Quantum annealing- It is a metaheuristic to find the global minimum of a loss function over a given state of candidate solution (it can be discrete as well as continuous) using Quantum Fluctuation.

Quantum Fluctuation- According to the uncertainty principle the energy of the quantum states keeps changing. This change in energy is called Quantum fluctuation.

How does Qauntum annealing work

0. We take a superposition of all possible states and system evolves time dependent
1. A potentially complicated Hamiltonian is found who’s ground state can be the solution of the problem.
2. A system with a simple Hamiltonian is initialized to the ground state
3. Then the simple Hamiltonian is evolved in the direction of the complex Hamiltonian
4. The time takes in equivalent to polynomial time computing problems for any problem

Let’s get started

- **Long term Quantum computers**
- **Near term Quantum computers**

What is this paper about?

This research focussed more on what we can do with the quantum machines that we have and that we are going to have in near future than to solve current problems with Quantum computers we’ll have in the next generation or maybe later

In case of interdisciplinary QML computationally expensive subroutines are outsourced to a quantum device and to analyze quantum states instead of classical data. QAML not to be confused with AQML (Applied Quantum Machine Learning) that uses ML to analyze Quantum experiments. The problems this deals with are basically for problems with statistically hard algorithms like Image and Speech recognition problems.

Let's get started

- Annealing vs Gate computers
- Dwave
- Instructions on What is about to happen

Use case 1

What kind of problems can near term Qcomputers colve?

Focus on problems which are intractable in the ML sphere (as in problems which have more of a Brute Forcing approach than an efficient algorithmic approach to be solved and problems which can’t be optimized) for example generative models in Unsupervised and Semi Supervised learning.

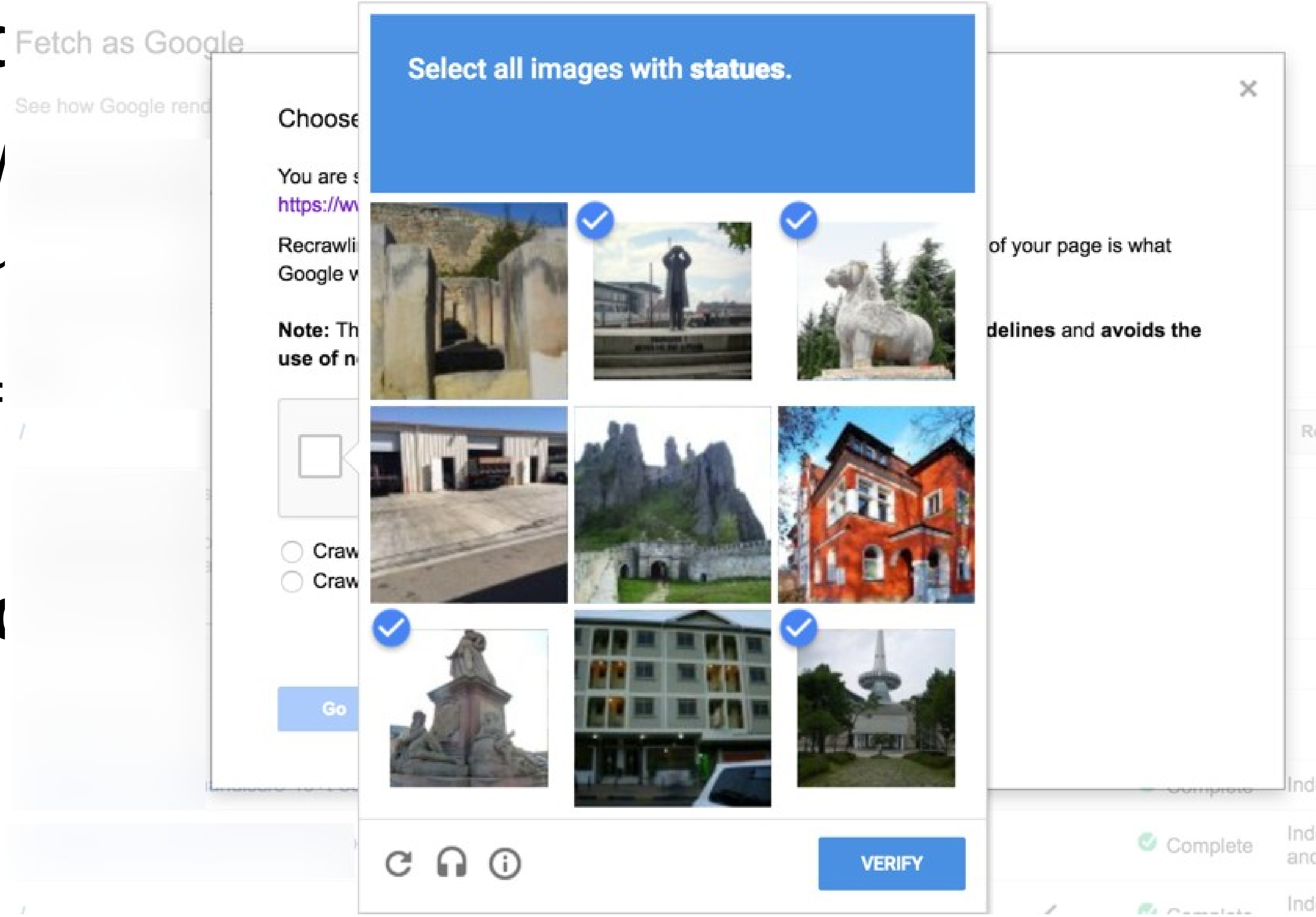
Labelled and Unlabelled data

Labelled d

- Data that w know abou example pictures of or dog

Unlabelled data

- Speech , images

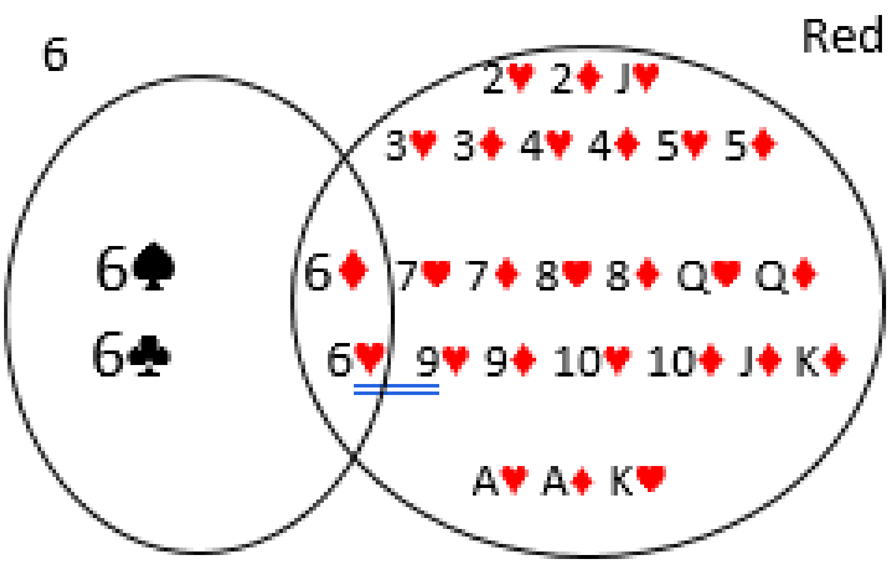


Types of Learning

- **Supervised-** Under a professor
- **Unsupervised-** Self study. Figuring out on your own
- **Semi supervised-** A coursera course with assignments duly submitted
- **Reinforcement learning-** Chosing your IIT preparation coaching

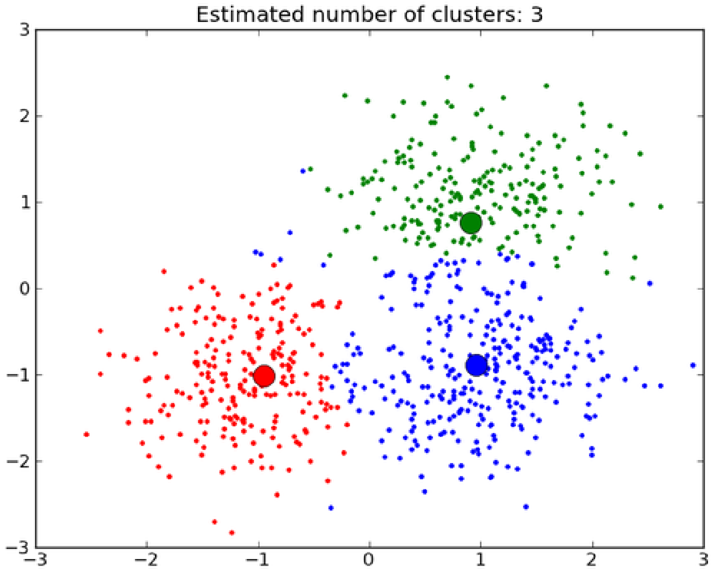
Joint probability

The probabillity of 2 events occuring together at a given time is called the joint probability – **intersecting probability**



Generative vs Discriminative models

- **Discriminative model-** you see the data and you classify it. Is it a car or a motobike?
- **Generative model-** You see the data and wonder how it came from? What is a car? Try to draw it!



Implicit and explicit density estimator

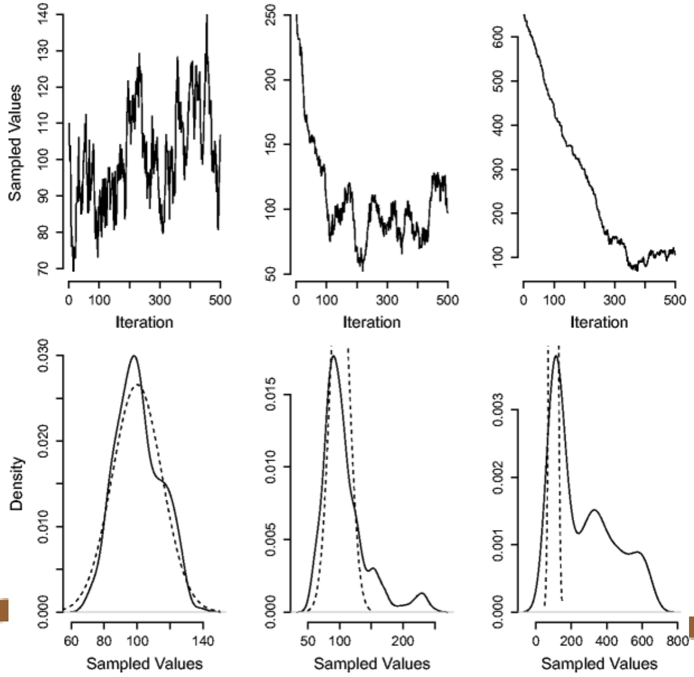
- **Explicit-** they assume a sample for example Monte Carlo sampling. They assume what kind of distribution it will be and
- **Implicit-** now let’s play a game. GANs

Monte Carlo Markov Chain (MCMC)

- We have a large data and we need to generate model
- We take a large subset of data and try to predict the pattern in the data
- We use the previous pattern as a stepping stone for next and so on
- As much deep we go, the data becomes more and more arbitrary hence improving the model

Metropolis approach

Metropolis approach- Keep taking proposals till amplitude of posterior is high or else keep adding values from noise. Generate a lot of proposals until we have enough samples. Use the last proposal to generate a chain

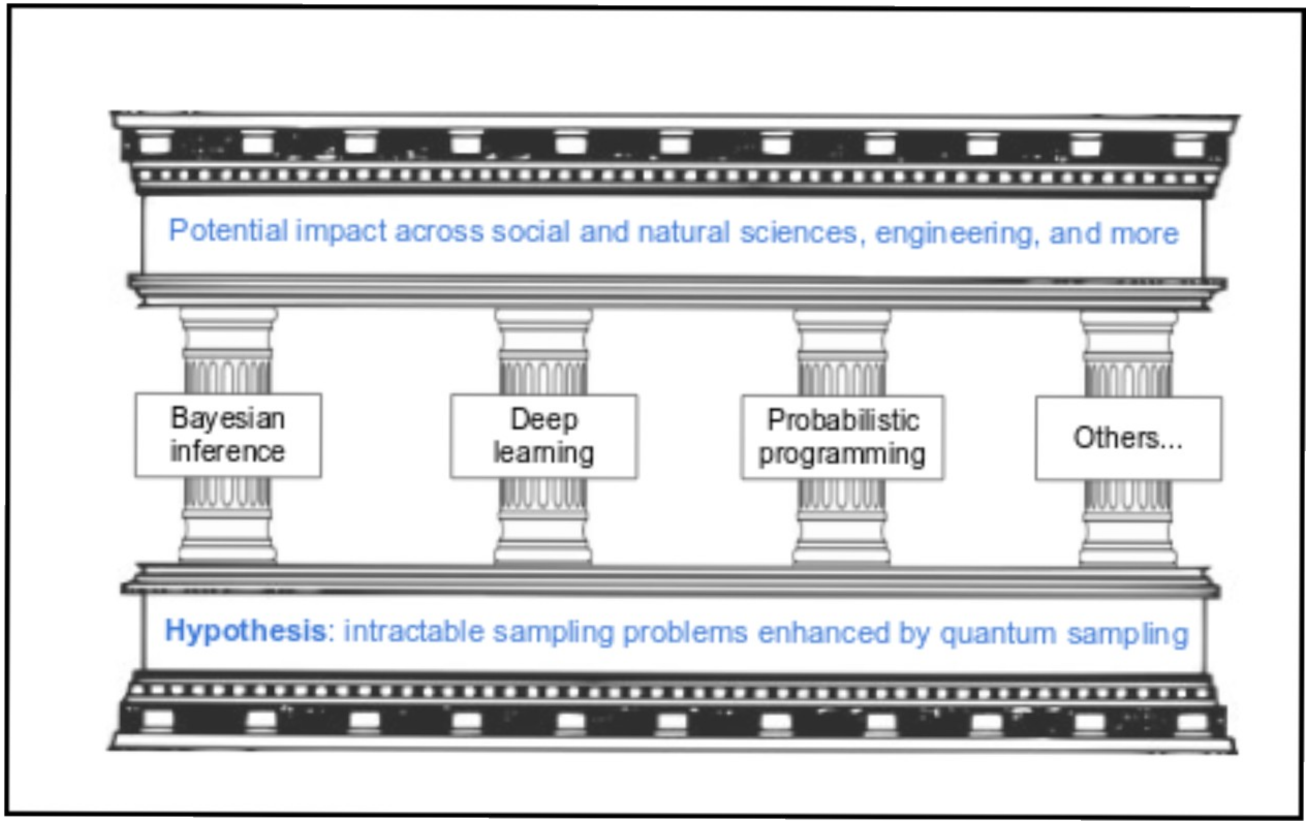


Limitations of this approach

- The data moves very slow since the mixing is very very slow. You generate a lot of random numbers and attatch them to a chain and then try to generate more randomness. Randomness becomes increasingly difficult.
- This can be solved by usig Quantum computers because jumping from one Markov Chain to another is easy because seperated by low density regions of relevant size

Impacts

We really don’t know what is happening here so this thing is called a gray box but as numbber of qBits increase, the thing will increasingly replace MCMC



Gibb’s Distribution

- It is an approach to MCMC
- It is like Metropolis Algorithm but it is a special case as in it accepts all proposals that come in
- This is more efficient for the Quantum computers

Use case II

Focusing on datasets intrinsic to quantum-like correlations,making quantum computers indispensable which can be used on machines with 50-100 Qbits. Cognitive sciences problems can give us such datasets and we can use them in these domain specific problems specifically. Also can be implemented on classical datasets where Quantum like statistical correlations can be implemented

Someone who spends less money

- **Parsimonious models-** A model that conveys a lot fo things using as few predictor variables as possible.

Akaike’s Information Criterion

- It compares quality of a set of statistical models to one another
- It uses software techniques, model fitting and relative differences
- However it doesn’t tell if a model is good or bad
- Just given a set of models, it chooses the best one.

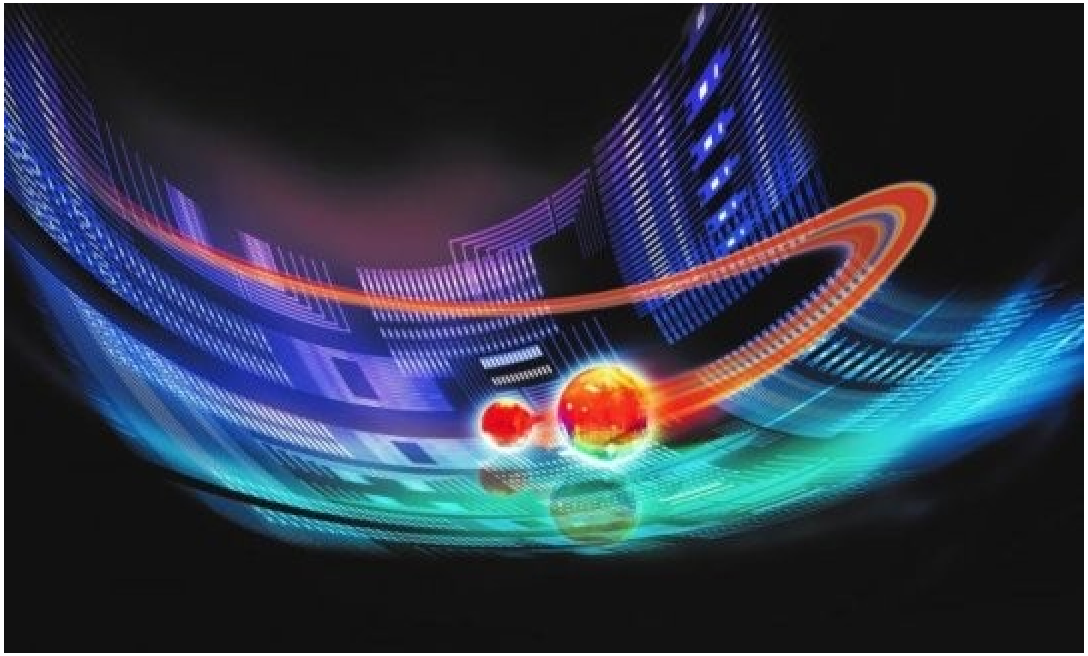
Let us get this thing clear right now

- ML algorithms can be sped up by using Quantum computers
- The usability of a Quantum computer can only be judged by it’s ability to not compete against classical computers but to work along with them
- Data represented by Quantum computers (even with limited number of Qbits) as shown by Google, leads in a superior performance to classical data modelling approaches. The data is to say Parsimonious and in accordance with Akaiske’s Information criteriation

Case study 1- Molecular Modelling

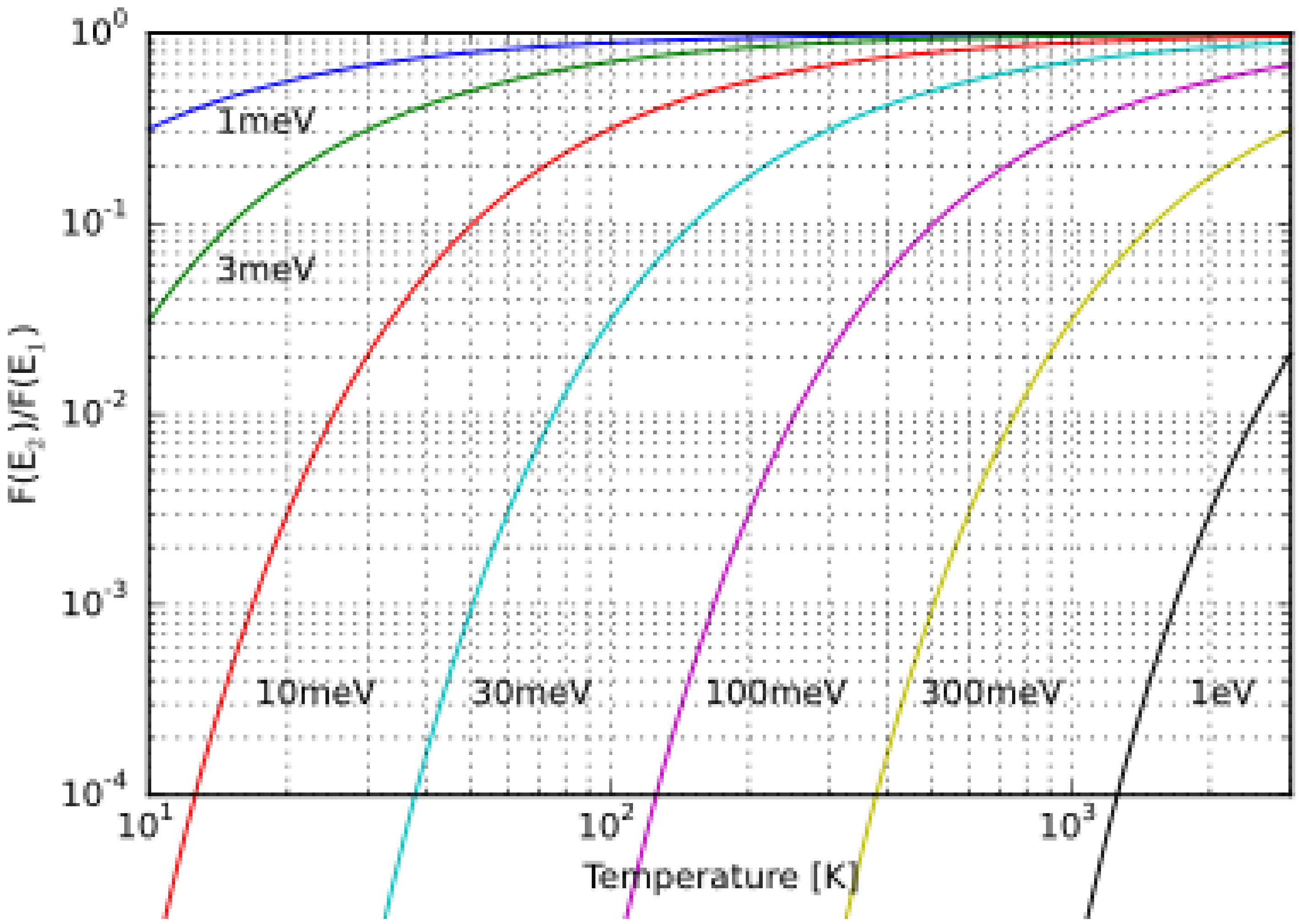
- Qauntum reactions can be modelled using Quantum computers because of entanglement property

world's first simulation of a particle using Quantum computers



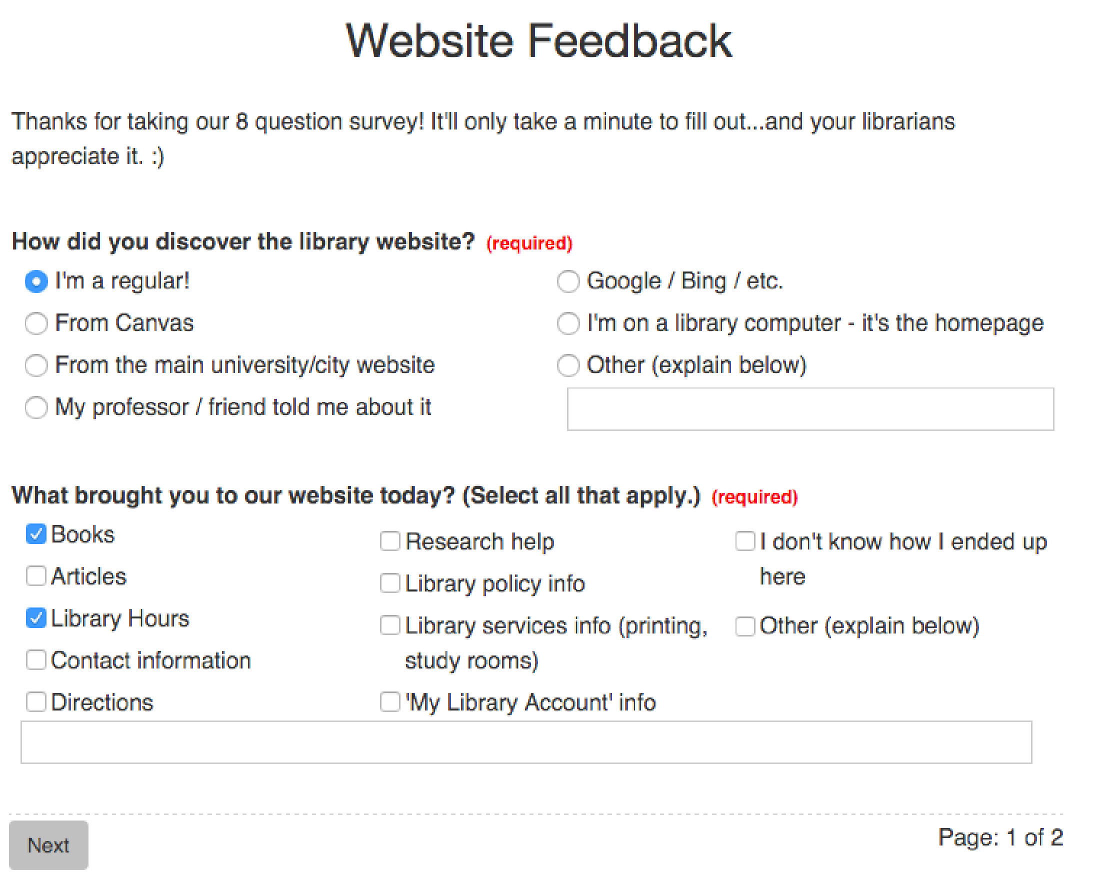
Case study 2- Boltzman Distribution and Maximum Entropy

- The principle of maximum entropy states that the probability distribution which best represents the current state of knowledge is the one with largest entropy, follows Boltzman distribution



Case study 3: Cognitive science (Order problem)

- Your answer to a question in a survey is dependent on the answer to the previous question
- In case of 100s of questions and 100s of people this can be analysed by a Quantum computer



Website Feedback

Thanks for taking our 8 question survey! It'll only take a minute to fill out...and your librarians appreciate it. :)

How did you discover the library website? (required)

- I'm a regular!
- From Canvas
- From the main university/city website
- My professor / friend told me about it
- Google / Bing / etc.
- I'm on a library computer - it's the homepage
- Other (explain below)

What brought you to our website today? (Select all that apply.) (required)

- Books
- Articles
- Library Hours
- Contact information
- Directions
- Research help
- Library policy info
- Library services info (printing, study rooms)
- 'My Library Account' info
- I don't know how I ended up here
- Other (explain below)

Next

Page: 1 of 2

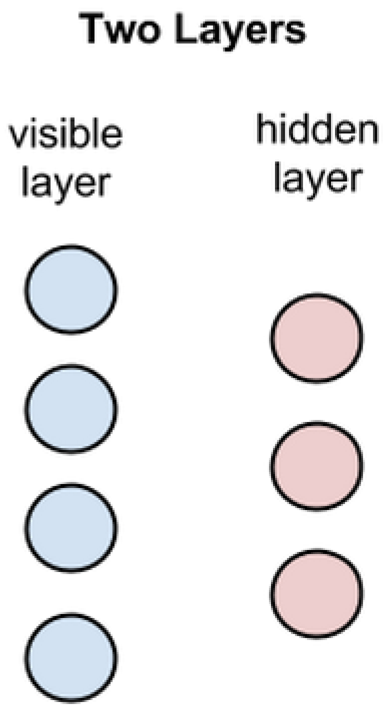
Case study 4: Cognitive biasses

- You are given a game where you win $200 or lose $100
- You play once and result is locked
- If you get your result, what is your chance to play again
- If your result is secret what is your chance to play again

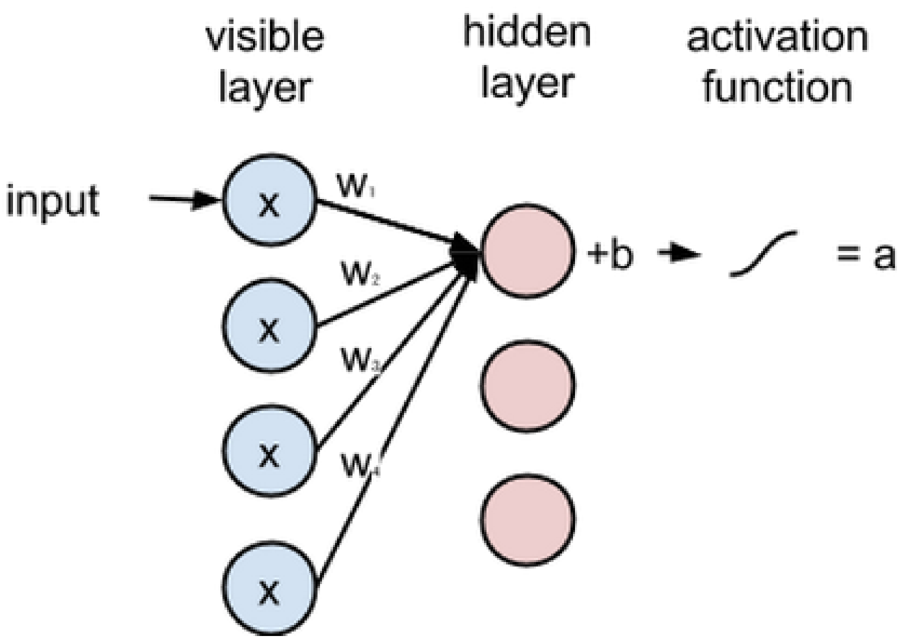
Data	10 GB	10 GB	20 GB	40 GB	60 GB
Tariff	₹ 50	₹ 50	₹ 100	₹ 200	₹ 400
Validity Plans (in days)	30	30	30	30	30
SMS	100	100	200	400	500
Voice (in minutes)	500	500	1000	1500	2500
Free Jio Apps	2	2	3	5	7
	Get the SIM	Get the SIM	Get the SIM	Get the SIM	Get the SIM

Let's get things clear. What are Restricted Boltzman Machines actually?

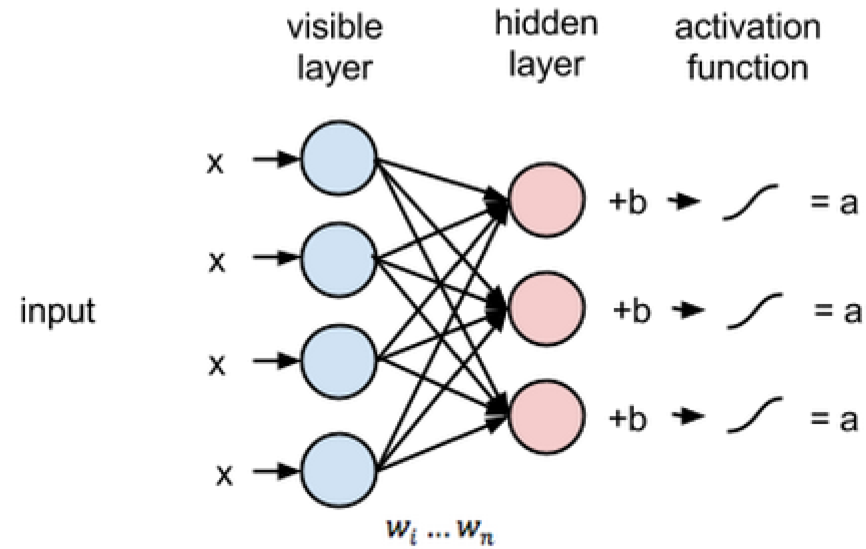
- Developed by sir Hinton they are predecesors of modern day Neural Networks
- Each of these are called nodes where calculations take place
- There are no intra layer communications going on



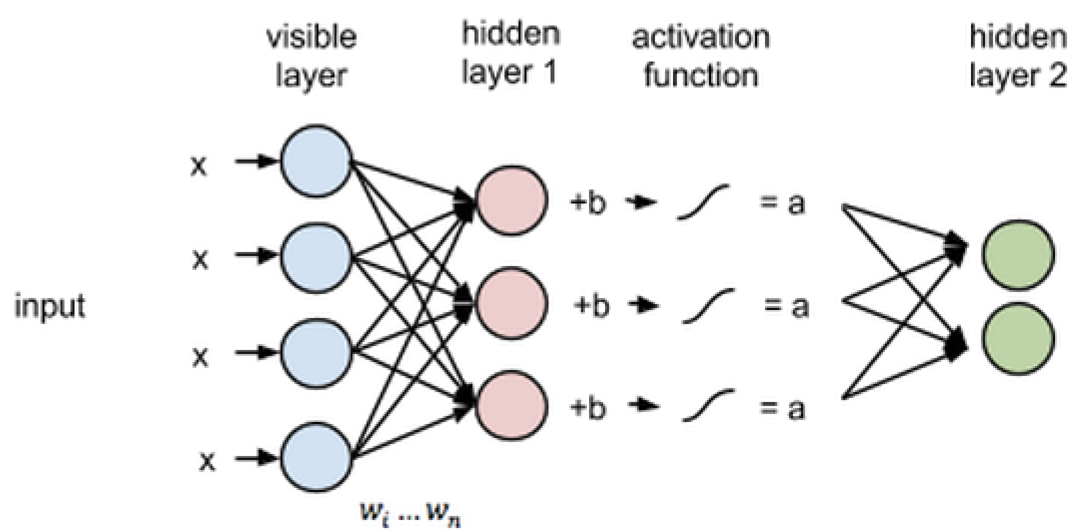
Weighted Inputs Combine @Hidden Node



Multiple Inputs



Multiple Hidden Layers



Let's actually break down how the whole thing works for everyone

- So the data is broken down into pieces
- A reconstruction of the network is carried out which tries to obtain the probability distribution of the input dataset
- Using this generative model we make a network that is called a boltzman machine
- If there is no relationship between the initial data it is a restricted Boltzman machine

Fully Visible Boltzman Machine

The parameters are learned from input data by maximizing the log likelihood (a measure of AIC better model) which means, the parameters are made into a better model by taking the input data into account.

These models can be trained on Quantum annealing hardware to imrove efficiency and they are used for convexity and tractability

Contrastive Divergence

The process to generate a chain of possible values using initial possible value (input value) while satisfying a particular constraint

Solving Linear set of Equations- HHL Algorithm

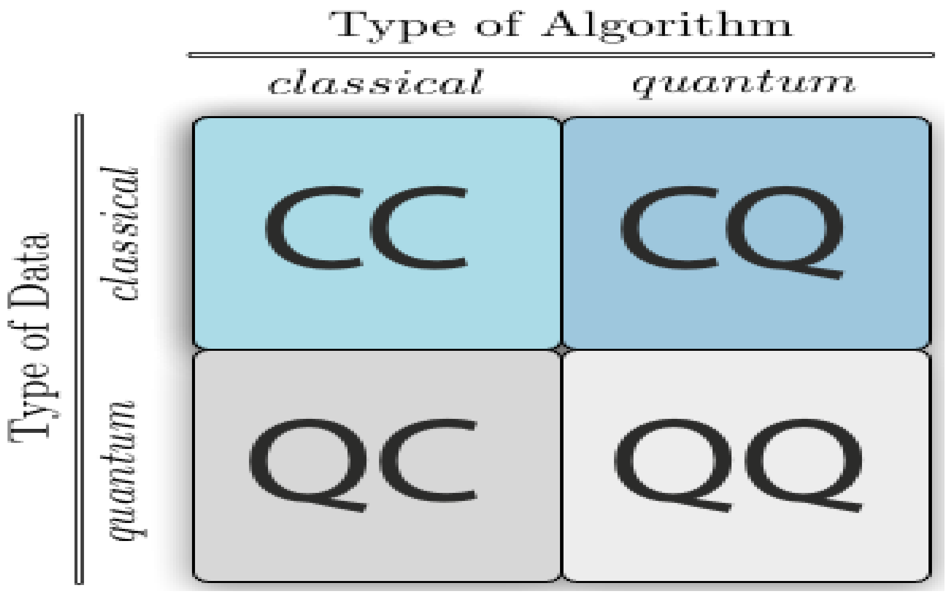
- Harrow-Hassidim-Lloyd algortihm or HHL algorithm
- If A is an NxN Hermatian matrix (Matrix which is equal to it’s own conjugate transpose), the solution for Ax=B is calculated using this process.
- The value of x is not calcultaed bu the associated scalars are calcultaed
- It’s proof gives a minimum time complexity of $O(\log(N)k^2)$ which is a significant speed up

Developing a Quantum recommendation System

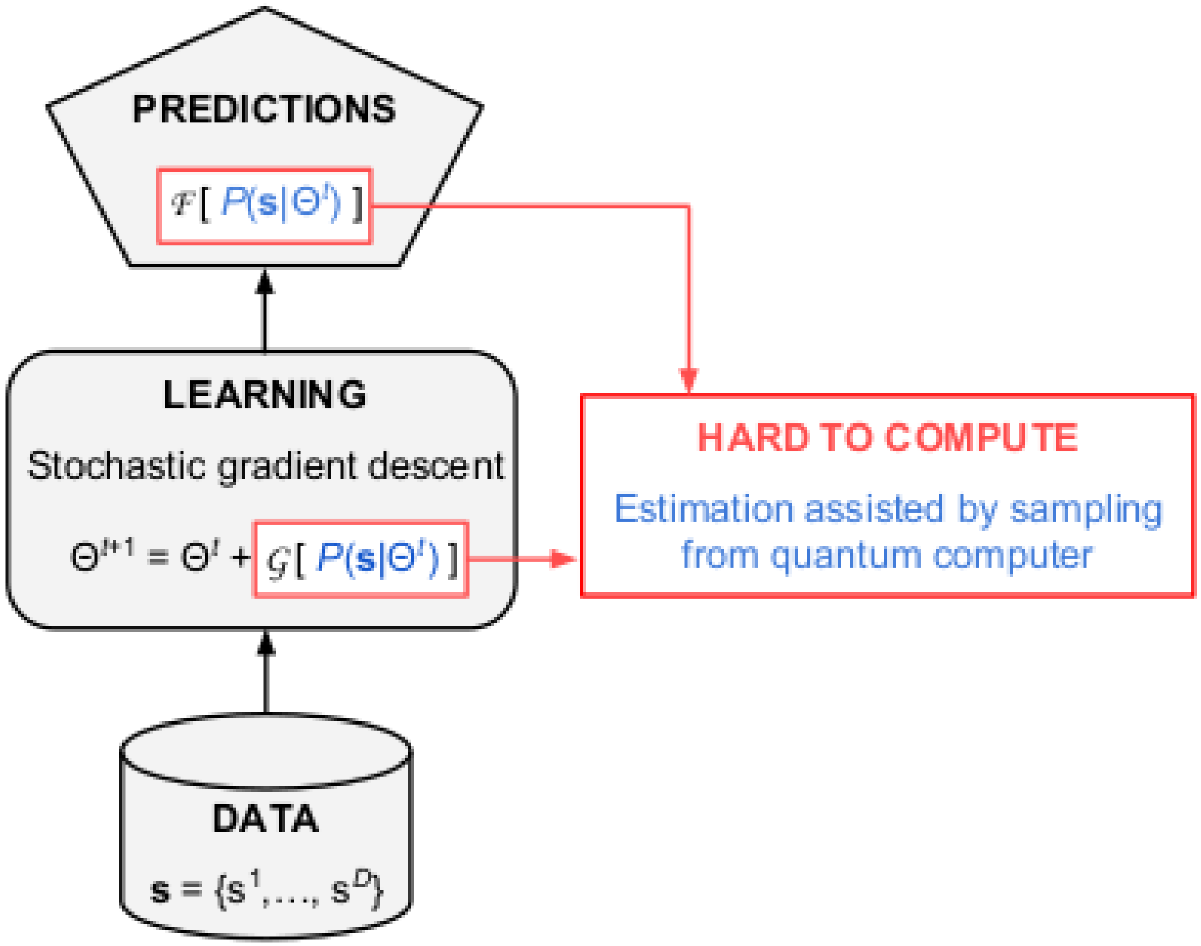
- The author develops his own data structure for matrix sampling having polynomial complexity
- Then bypasses the restrictio of HHL
- Takes in input of several million datasets and gives output

Use case 3

Some classical ML algorithms have an intractable subroutine. We can use Quantum computing to develop hybrid algorithms to execute the intractable step.



How hybrid algos work



Limitations of QAML 1: Communication

- We design a hybrid network example say Boltzman machine
- It needs 2 inputs to maintain bipartite structure
- Positive phase is better on classical computer
- Negative phase is intractable so better on Quantum computer
- So we rely on Gibb's distribution from physical process
- But we do not have any control of the temperature
- This does result in a lot of erros

Limitations of QAML 2: Noise

- Some Quantum machines have an intrinsic noise
- What that means is the freezing in Quantum annealing can create unnecessary noise in the system which can divert the data and cause unnecessary peturbation
- We can solve it by specifying a particular model and not have a random model but that would not be in line with structure of Quantum device

A very good possible solution

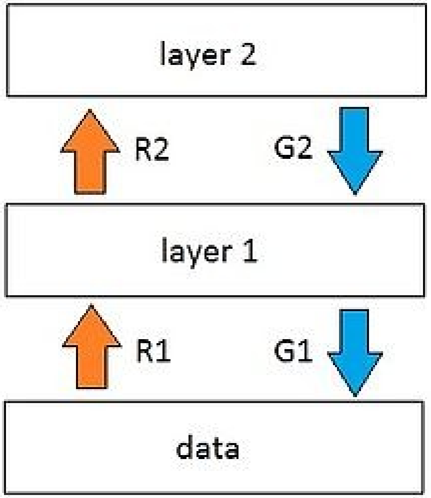
- Using a grey box approach
  Assume data come from MCMC distirbution
  Still work from initial data
- For example having a proxy to check if we are moving in the correct direction of the gradient
- But the problem is anytime we need to do a reconstruction or generation we will need to use the same Quantum device for that. No classical computing so we will have to train our models likewise

Helmholtz Machine

- It contains 2 things- a network that goes bottom up. That is, it tries to generate a pattern in the data
- The second that tries to generate the data itself which is the recognition phase
- The problem is to find the value of the hidden variables in the data.
- The training is alternate in nature and is called Wake and sleep

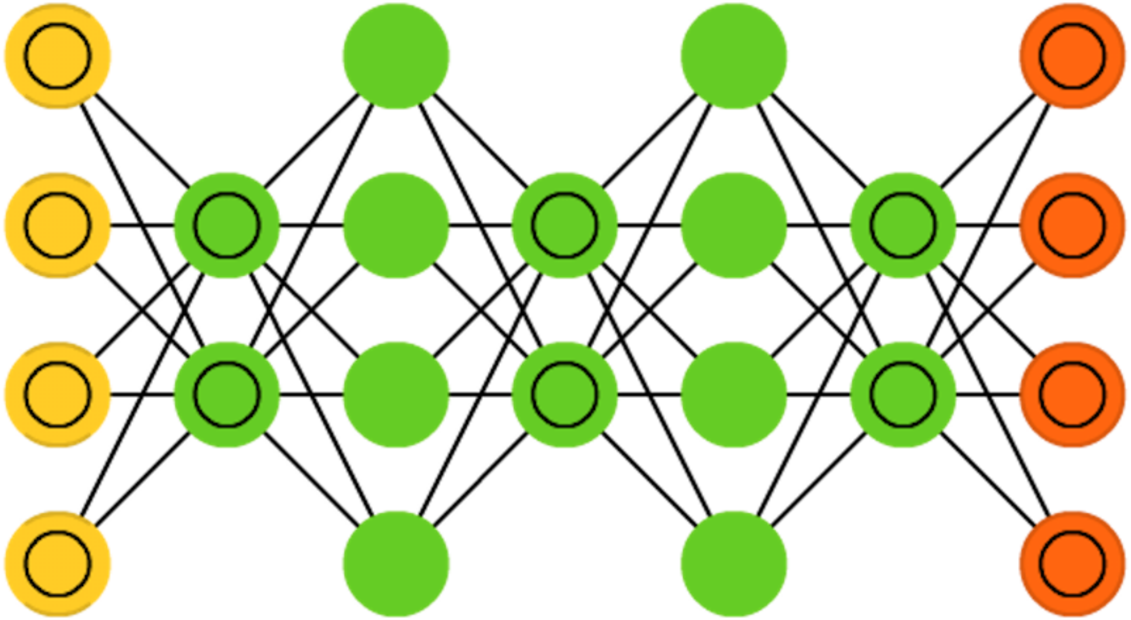
Intution

- Imagine you wake up and you decide to do your bed and everything
- And then while at night you try to minizmie the expenses for the month based o your daily activities



Getting deeper into the netwrok

- The hidden layers contain a 0 and 1
- They go deeper and deepr and as deep as we go, we try to model more abstract features than before



Solution approach 1

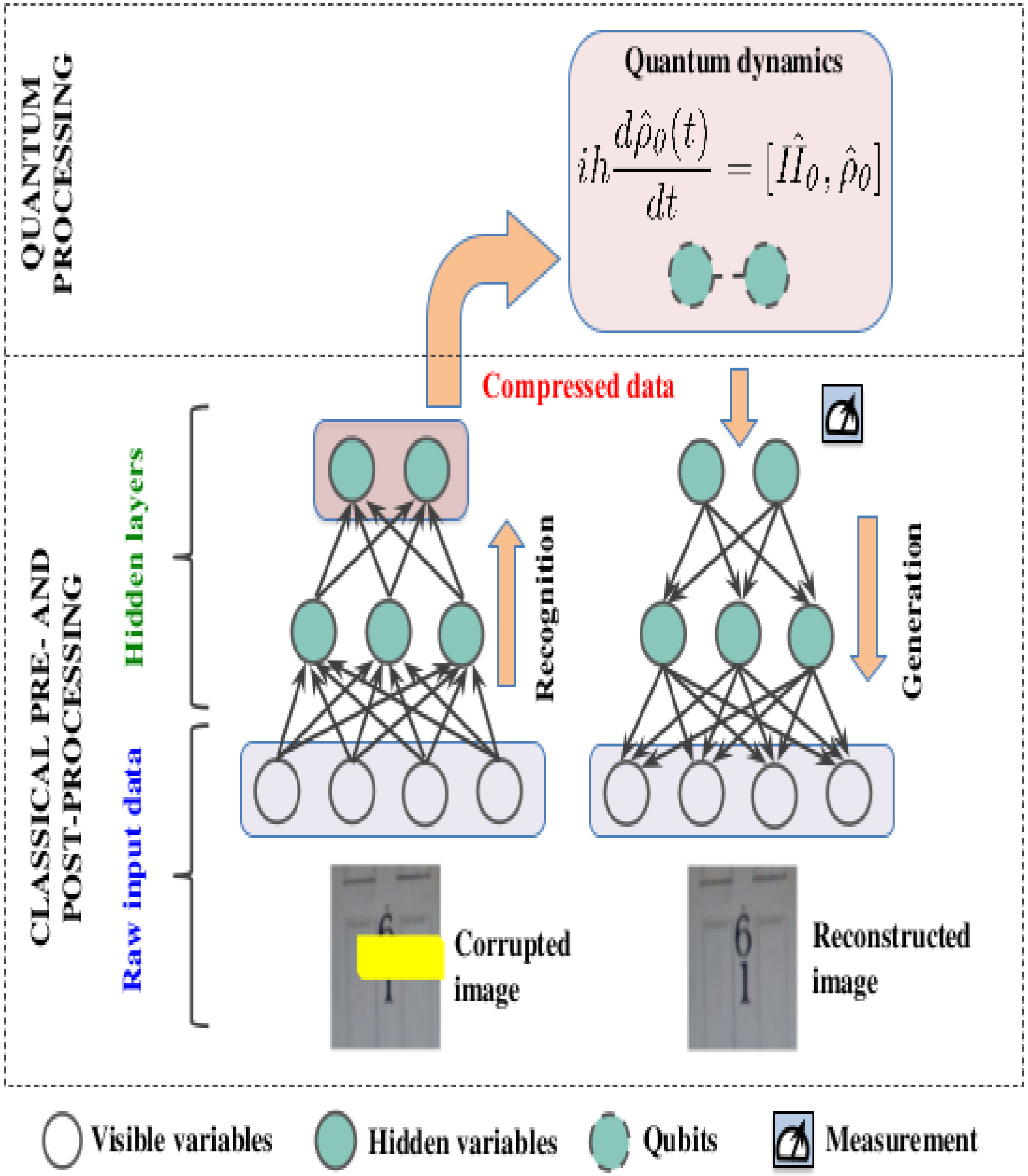
- Model the deepest layer into a Quantum sampling
- Measure hidden variables for each data point in the training set
- Now that we have a classical Neural Network we can solve it using classical computing
- We also have a reduced dimension of data and since Quantum device is only usable for binary data, we have that going on for us too

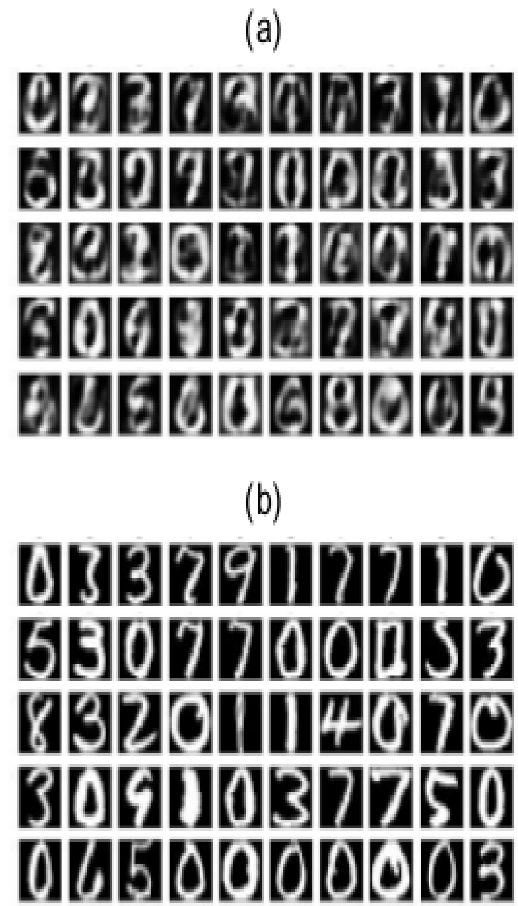
Approach 2

- Restrict the recognition to be classical
- Speed it up by using serial Quantum devices
- Use a Deep Neural Network to carry out the recognition task
- Use Quantum assisted generator network
- Use this KL divergence to find distance between the two distributions and implement the Wake-Sleep appraoch to minimize that

Example: Generation of Handwritten digits

- Complex data is data with more variables than available qbits and data can be discrete as well as continious
- The generator network is on the right and the recognizer is on the left
- The hidden variables are set by the generator network
- The model does not simply emorize the training set, but rather reproduce its statistics.


QUANTUM PROCESSING
Quantum dynamics
$i\hbar \frac{d\hat{\rho}_0(t)}{dt} = [\hat{H}_0, \hat{\rho}_0]$
Compressed data
CLASSICAL PRE- AND POST-PROCESSING
Hidden layers
Raw input data
Recognition
Generation
Corrupted image
Reconstructed image
Visible variables
Hidden variables
Qubits
Measurement


(a)
(b)

Limitations of QAML 3: Connectivity

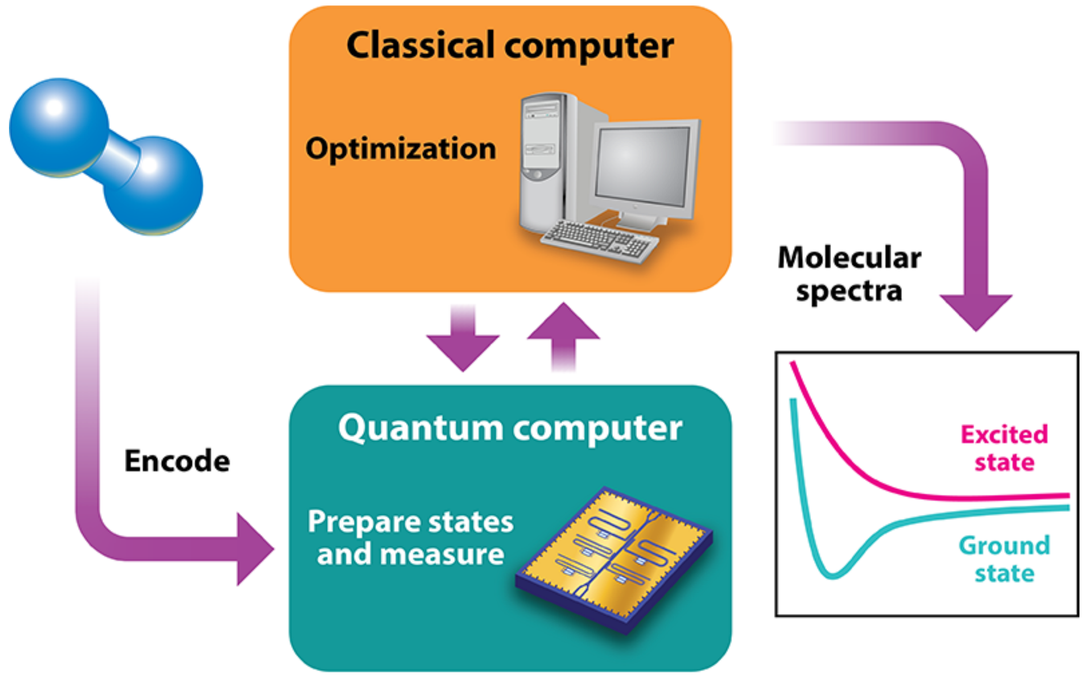
- Topological connectivity- Physical connectivity is only limitted in nature
- Qbit to qbit interaction will cause a lot of overhead expenses in the device
- In case of annealing devices a larger problem is to set the parameters of the devices
- Embedding parameters into equations can be solved by Heuristic search but will they be connected

Limitations of QAML 4: Non Binary datasets

- Quantum computers at the moment are good at performing operations on non binary dataset
- But most data like images and speech are non binary in nature
- In the long term we could use the amplitudes of the dataset to represent complex datasets or we could develop approaches to model complex ML datasets but right now we do not have the tech

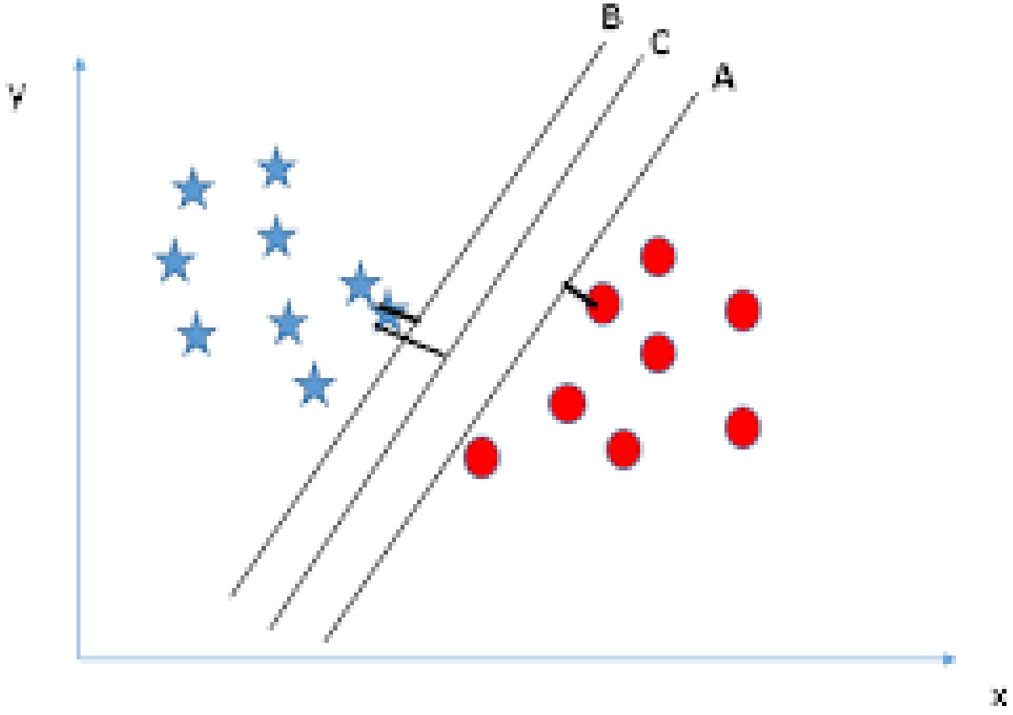
Solution

- Use Quanrum computers to model only Hidden variables which are binary in nature
- Use classical computing to model everything else



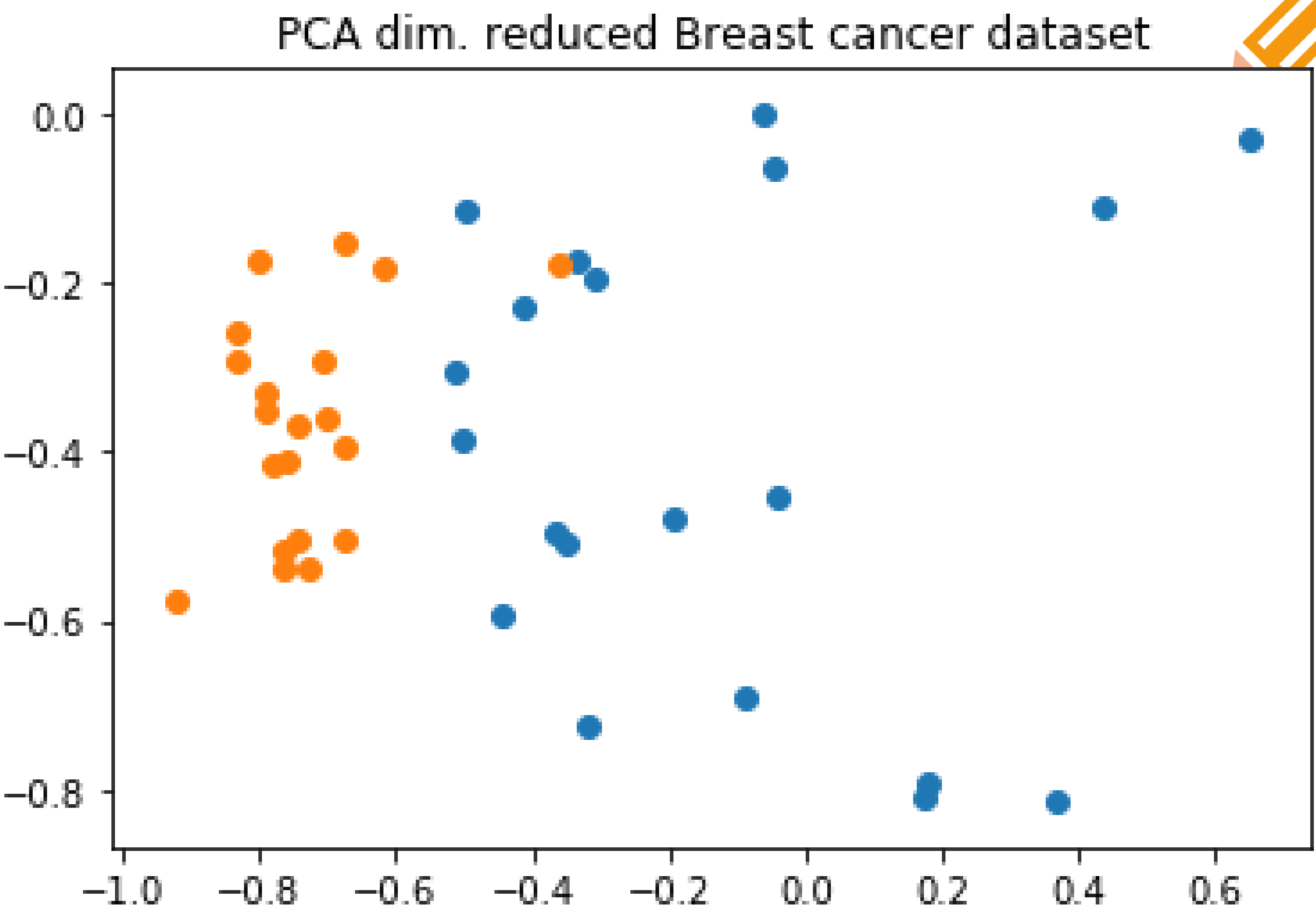
A crash course into SVM

- Stay away from me!



```
In [1]: from svm_datasets import *
        from qiskit_aqua.svm.data_preprocess import *
        from qiskit_aqua.input import get_input_instance
        from qiskit_aqua import run_algorithm

        import logging
        logger = logging.getLogger()
        # logger.setLevel(logging.DEBUG) # uncomment it to see detailed logging
```

```
In [2]: ################ import Qconfig and set APIToken and API url and  prepare backends ############
        try:
            import sys
            sys.path.append("../../") # go to parent dir
            import Qconfig
        except Exception as e:
            print(e)

        from qiskit import register, available_backends

        #set api
        APItoken=getattr(Qconfig, 'APItoken', None)
        url = Qconfig.config.get('url', None)
        hub = Qconfig.config.get('hub', None)
        group = Qconfig.config.get('group', None)
        project = Qconfig.config.get('project', None)
        try:
            register(APItoken, url, hub, group, project)
        except Exception as e:
            print(e)

        print("Backends: {}".format(available_backends()))
```


PCA dim. reduced Breast cancer dataset
0.0
−0.2
−0.4
−0.6
−0.8
−1.0
−0.8
−0.6
−0.4
−0.2
0.0
0.2
0.4
0.6

Output

```
Predicted label is  B
CORRECT

=============================================
classifying [-0.89863812 -0.44994977]
Label should be  B
Predicted label is  B
CORRECT

=============================================
classifying [-0.97648453 -0.42314792]
Label should be  B
Predicted label is  A
INCORRECT

=============================================
classifying [-0.84577403 -0.35824307]
Label should be  B
Predicted label is  B
CORRECT

=============================================
classifying [-0.76424251 -0.43948113]
Label should be  B
Predicted label is  B
CORRECT
Classification success for this set is 85.0 %

testing success ratio:  0.85
predicted labels: ['A', 'A', 'A', 'A', 'A', 'B', 'B', 'B', 'A', 'A', 'A', 'A', 'A', 'A', 'B', 'B', 'B', 'A', 'B', 'B']
```

Let’s create a human brain. The long term vision and how

Consciousness as a Quantum problem

THANK YOU