Quasi Probability Decomposition

A quasi probability decomposition relaxes an axiom of probability. i.e $p_i \geq 0$, so the sum of the distribution may include negative terms. Quantum Mechanics allows for events with negative expectation values to account for phenomena like destructive interference

What does it mean to have less than 0% chance

Think of an event with N possible outcomes. We express the probabilities as a vector

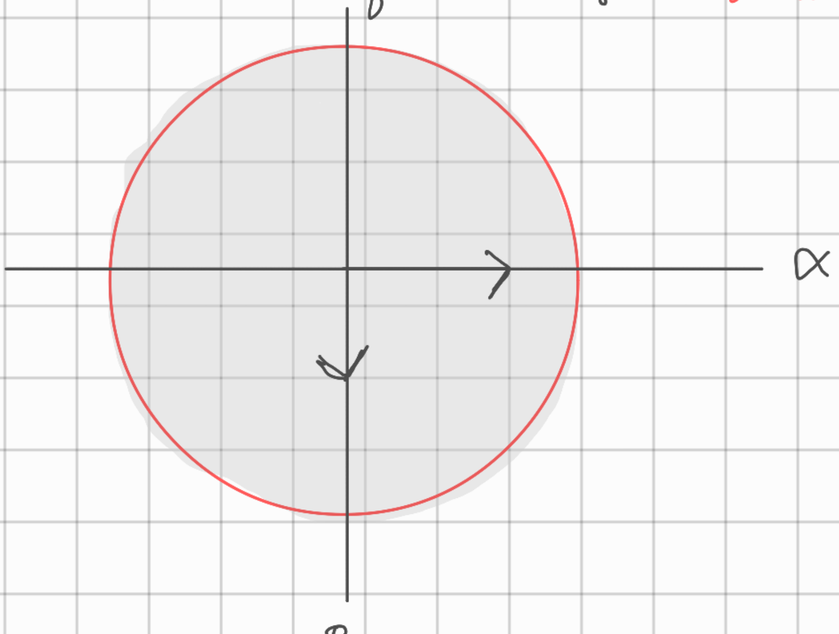
$$(p_1, p_2, \cdots p_N)$$

that follows $p_i \geq 0$ and $\sum p_i = 1$

This $\sum p_i$ is also called the 1- norm of the probability vector.

There are other norms as well, like the 2-Norm or Euclidean norm.

What happens when you come up with a "probability theory" but based on 2-Norm instead of 1-Norm? ~ Quantum Mechanics



Think of a single bit, we can describe it as having a probability p of bit 0 and $1-p$ of bit 1. But when we switch to 2-Norm instead we now want the sum of the squares of these numbers to be 1.

We want a vector (α, β) where $\alpha^2+\beta^2=1$

So, α^2 is the probability of a 0 outcome and β^2 is the probability of a 1.

but, why not just ditch α, β and simply deal with the probabilities then?

The difference comes in how the vector changes when we apply an operation to it.

$$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} p \\ 1-p \end{pmatrix} = \begin{pmatrix} 1-p \\ p \end{pmatrix}$$

Bit flip

We can represent any operation via a stochastic matrix (i.e a matrix of non-negative real numbers where every column adds up to 1)

Hence this stochastic matrix is most general sort of matrix that always maps a probability vector to another probability vector.

But why sum of squares? Why not sum of cubes, or fourth power?

The 2-norm bit that we defined is called "qubit". Here (α, β) becomes
$$\alpha|0\rangle + \beta|1\rangle$$

So given a qubit, we can transform it by applying 2x2 unitary matrix and that leads to quantum interference.

$$U = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix}$$

this matrix takes the vector in the plane and rotates it by 45° ↺

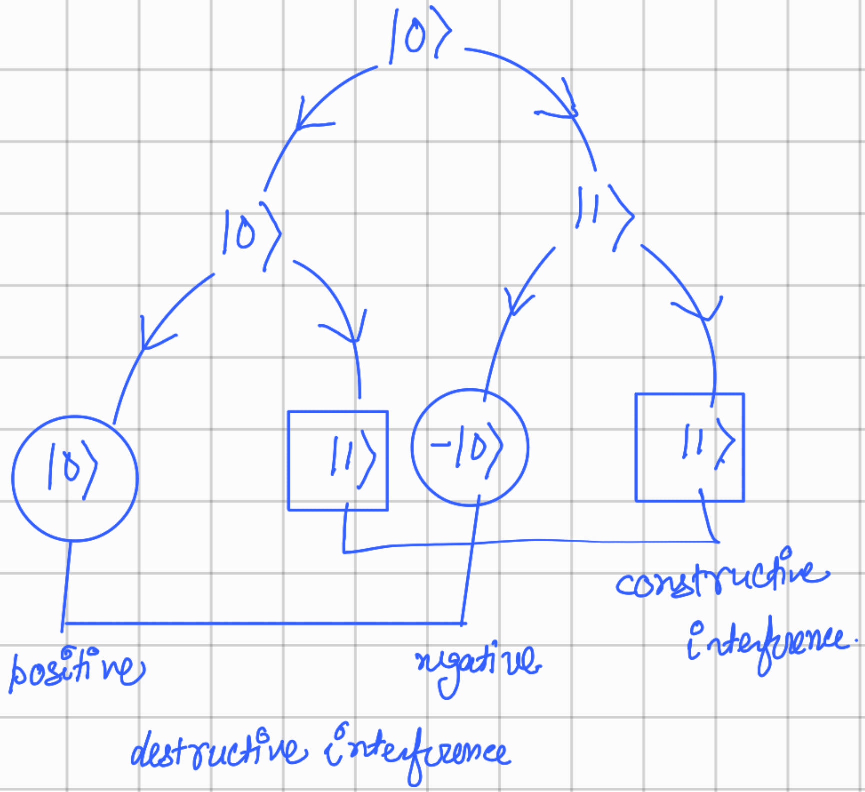
Now consider the $|0\rangle$ state, If we apply U to it we'll get

$$\begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} \Rightarrow \frac{1}{\sqrt{2}}\left(|0\rangle + |1\rangle\right)$$

its like taking a coin and flipping it. But then, if we apply the same operation U a second time, we'll get $|1\rangle$

$$\begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \Rightarrow |1\rangle$$

Applying a "randomizing" operation to a "random" state produces a deterministic outcome.



The reason that we'll never see this sort of interference in classical world is that probabilities can't be negative.

Mixed States:

Once we have these quantum states, we can always layer up "classical probability theory" on top.

what if we have $\frac{1}{2}$ probability of $\frac{1}{\sqrt{2}}(|0\rangle+|1\rangle)$ and a half probability of $\frac{1}{\sqrt{2}}(|0\rangle-|1\rangle)$?

This gives us mixed states, and we represent them using density matrix.

what is a density matrix?

Say you have a vector of N amplitudes $(\alpha_1 \dots \alpha_N)$. Then you compute the outer product of the vector with itself – $N \times N$ matrix.

$$\underbrace{\begin{pmatrix} \alpha_1\alpha_1 & \alpha_1\alpha_2 & \cdots & \alpha_1\alpha_N \\ \vdots & \vdots & & \vdots \\ \alpha_N\alpha_1 & \cdots & \cdots & \alpha_N\alpha_N \end{pmatrix}}_{N \times N}$$

and then a probability distribution over such vectors.

Quasi Probability Decomposition

It is known that certain kind of non-local operations can be simulated by sampling set of local operations with quasi-probability dist.

50 qubits $\sim$ 2^{50} complex numbers to be stored

hence we should work on decomposing the quantum circuit

"space-like" cut on quantum computer.

More specifically, a way to decompose a controlled gate into sequence of single-qubit operations

Gate Decomposition ~ Tensor Network Representation of quantum circuit

Density Operator:

Given an arbitrary state $|\psi\rangle \in \mathcal{H}$, define

$\hat{\rho} = |\psi\rangle\langle\psi|$ (projection operator for $|\psi\rangle$)

(projection operator for $|\psi\rangle$)

Suppose $\hat{O}$ is an observable with eigenvalues $\{q_u\}$ and eigenstates $\{|u\rangle\}$

$$P_u = |\langle u|\psi\rangle|^2 = \langle u|\hat{\rho}|u\rangle$$

Probability of obtaining q_u

and the Expectation value of the observable is

$$\langle O\rangle = \sum q_u P_u \Rightarrow \sum q_u \langle u|\hat{\rho}|u\rangle = \mathrm{Tr}[\hat{O}\hat{\rho}]$$

what are inner and outer products?

inner product ~ dot product

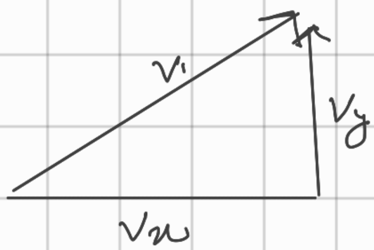
1. it defines a notion of angle.

$$\vec{v}\cdot\vec{w} = v_1 w_1 + v_2 w_2 + \cdots v_n w_n$$

$$\vec{v}\cdot\vec{w} = |\vec{v}||\vec{w}|\cos\theta$$

so when they are orthogonal. the dot product is zero.

2. it defines a notion of length



$$|\vec{v}| = \sqrt{v_x^2 + v_y^2} = \sqrt{\vec{v}\cdot\vec{v}}$$

Innerproduct takes in two vectors and gives us a number.

$$\langle\psi|\phi\rangle = c$$

$$\langle\psi|\phi + \xi\rangle = \langle\psi|\phi\rangle + \langle\psi|\xi\rangle$$

$$\langle\phi|a\psi\rangle = a\langle\phi|\psi\rangle$$

Inner product should be linear in the right hand position inner product should be linear in the right hand position

why not same for left hand side?

Suppose $\langle \varphi | \varphi \rangle = 1$

now for $\langle i\varphi | i\varphi \rangle$ we can rewrite $i \langle i\varphi | \varphi \rangle$

$i^2 \langle \varphi | \varphi \rangle$

$\langle i\varphi | i\varphi \rangle = -1$ $\qquad -1 \langle \varphi | \varphi \rangle$

and now if we want to define the length.

$\| \, | i\varphi \rangle \| = \sqrt{\langle i\varphi | i\varphi \rangle} \Rightarrow \sqrt{-1} \Rightarrow i$ doesn't make sense.

So, $\langle \psi | \varphi \rangle \neq \langle \varphi | \psi \rangle$

rather

$\langle \psi | \varphi \rangle = \langle \varphi | \psi \rangle^*$ complex conjugate

so now if we have

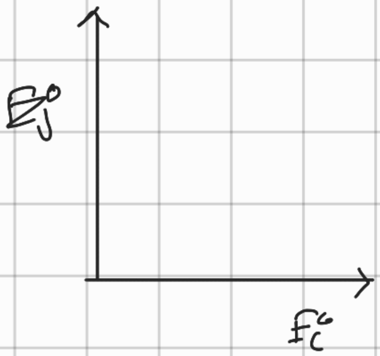
$\langle a\psi + b\xi | \phi \rangle \Rightarrow a^* \langle \psi | \varphi \rangle + b^* \langle \xi | \phi \rangle$

How does inner product works in our Hilbert space?

Orthonormal Basis $\{ |E_i\rangle \}$

$\langle E_i | E_i \rangle = 1$ (length)

$\langle E_i | E_j \rangle = 0$ (mutually orthogonal)



$\langle E_i | E_j \rangle = \delta_{ij}$

$|\psi\rangle = \sum_i c_i |E_i\rangle$

if we want the second coefficien

$\langle E_2 | \psi \rangle = \langle E_2 | \left(\sum_i c_i | E_i \rangle \right)$

linear in right slot

$$\langle E_2|\psi\rangle = \sum_i c_i \langle E_2|E_i\rangle \Rightarrow \sum_i c_i \delta_{2i}$$

$$\langle E_2|\psi\rangle = c_2$$

Let's see for two vectors

$$\langle\psi|\phi\rangle = \left(\sum_i \langle a_i E_i|\right)\left(\sum_j |b_j E_j\rangle\right)$$

$$\sum_i\sum_j \langle a_i E_i| b_j E_j\rangle \Rightarrow \sum_i\sum_j a_i^* b_j \langle E_i|E_j\rangle$$

this isn't   this one is linear

$\langle\psi_1|\psi_2\rangle$ = scalar

$\Rightarrow 1$ if same

0 if orthogonal

$$\Rightarrow \sum_{ij} a_i^* b_j \delta_{ij} \qquad \sum_i a_i^* b_i$$

Outer Product

$|\psi_1\rangle\langle\psi_2|$ = matrix

$$\langle\psi_1| = \frac{1}{\sqrt{2}}(-i\ \ 0\ \ 1) \qquad \langle\psi_2| = \frac{1}{\sqrt{11}}(1\ \ 1\ \ 3)$$

Bra state

but we need ket space

$$|\psi_1\rangle = \frac{1}{\sqrt{2}}\begin{pmatrix} i \\ 0 \\ 1 \end{pmatrix}$$

Density Matrix

One gains significant advantage using the density matrix for certain time-dependent problems.

$$\rho(t) = |\psi(t)\rangle\langle\psi(t)|$$

that also means if you have a state $|\chi\rangle$, the integral $\langle\chi|\rho|\chi\rangle$ gives the probability of finding the particle in the state $|\chi\rangle$

the state $|\lambda\rangle$

$$\langle A\rangle = \int A\, P(A)\, dA$$

$\mathrm{Tr}(\rho^2) = 1$ for pure state

≤ 1 for mixed state

$$\langle A\rangle = \langle\psi|A|\psi\rangle = \mathrm{Tr}[A\rho]$$

Gate Decomposition : Tensor network Representation

quantum circuit U, initial state ρ, observable O

Given these any quantum computation can be represented as tensor networks.

what is a Tensor ?

Tensor is a multi-dimensional array. Just like vector is 1-D

vector $\quad v = \begin{bmatrix} 2 \\ 3 \end{bmatrix} \qquad v_1 = 2,\ v_2 = 3$

matrix $\quad M = \begin{bmatrix} 5 & 7 \\ 8 & 9 \end{bmatrix} \qquad M_{11} = 5 \quad M_{12} = 7$

order-3 Tensor $\quad T = \begin{bmatrix} 3 & 4 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} 5 & 7 \\ 3 & 5 \end{bmatrix} \qquad T_{111} = 3,\ T_{112} = 5$ (cube of number)

where do Tensors occur? Naturally. even in low dimensions if you define.

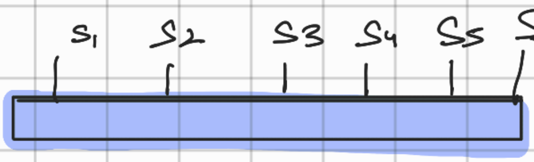
Many body quantum wavefunction

$$|\psi\rangle = \sum_{s_1 s_2 s_3 \cdots s_N} \psi^{s_1 s_2 s_3 \cdots s_N} |s_1 s_2 s_3 \cdots s_N\rangle$$

$s_i \in$ vector space / qubits

Amplitudes are a big tensor

$\psi^{s_1 s_2 s_3 \cdots s_6} =$



Tensors beyond a few indices become exponentially costly to store and manipulate.

$$\underbrace{T_{n_1 n_2 n_3 n_4 n_5 n_6}}_{10^6 \text{ enteries.}}$$

6 indices each with 10 values

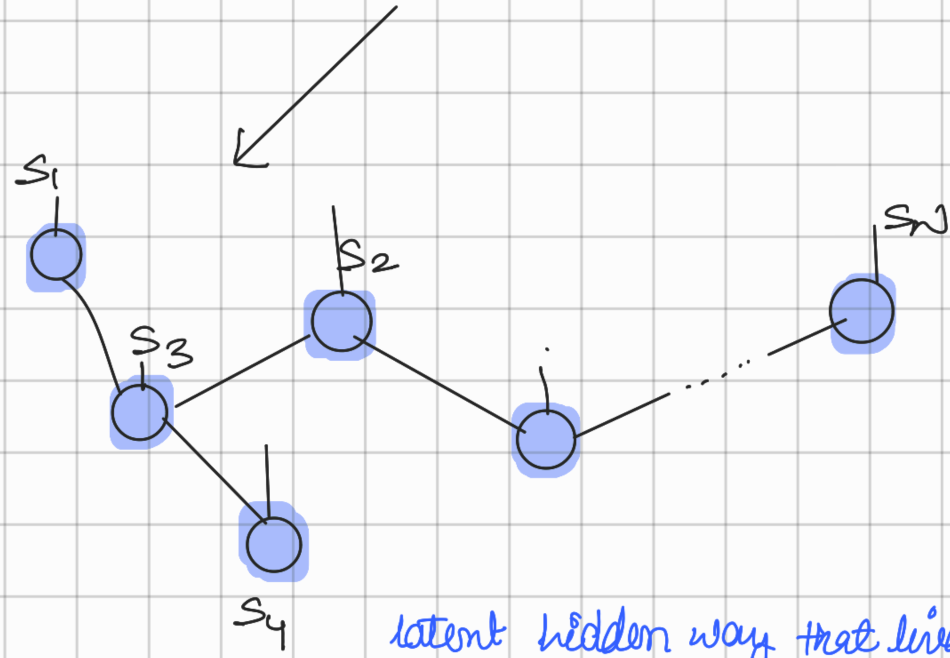
$n_j = 1, 2, .. 10$

$T_{n_1, n_2 \dots n_N}$ has 10^N enteries, exponential in N.

Tensor networks give a way to break the curse of dimensionality

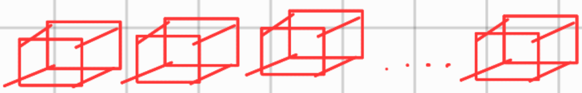
$T^{s_1 s_2 \dots s_N} =$ 

$T^{s_1 s_2 s_3 \dots s_N}$

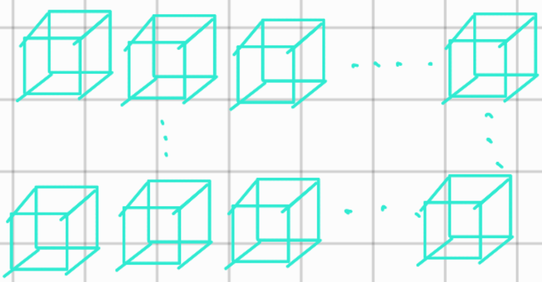


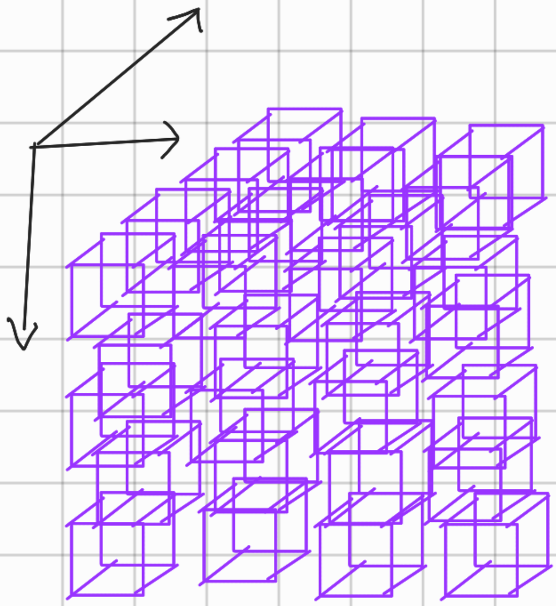
latent hidden way that lives in other tensors

Also, tensors beyond a few indices become hard to visualize



4th order tensor





6th order tensor

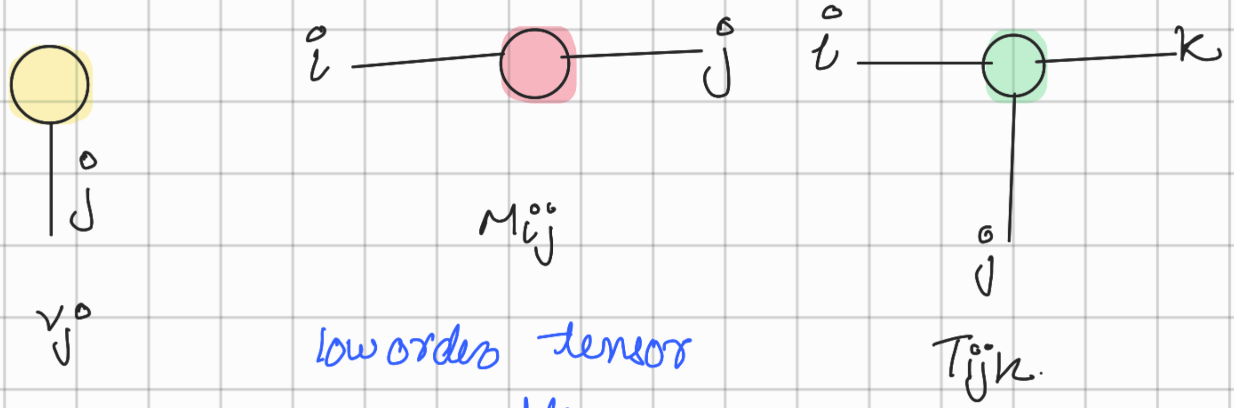
5^{th} order tensor

they are very complicated notations

$$T^{n_1 n_2 n_3 n_4 n_5 n_6} = \sum_{\alpha} A^{n_1}_{a_1} A^{n_2}_{a_1 a_2} A^{n_3}_{a_2 a_3} A^{n_4}_{a_3 a_4} A^{n_5}_{a_4 a_5} A^{n_6}_{a_5 a_6} A^{n_7}_{a_6}$$

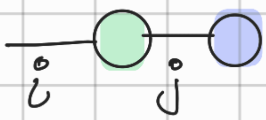
this is a matrix product state in its full glory.

so there is a nice way out, by making tensor network diagrams

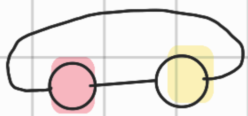


low order tensor examples.

joining lines implies contraction



$\longleftrightarrow \quad \sum_j M_{ij} V_j$

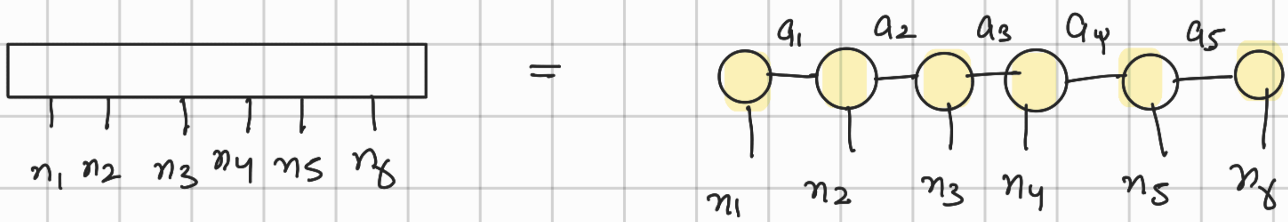


—— $A_{ij} B_{ji} \Rightarrow \mathrm{Tr}[AB]$

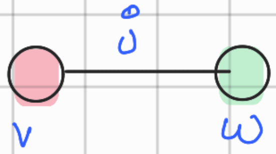
So the complicated expressions like

$$T^{n_1 n_2 n_3 n_4 n_5 n_6} = \sum_{\alpha} A^{n_1}_{a_1} A^{n_2}_{a_1 a_2} A^{n_3}_{a_2 a_3} A^{n_4}_{a_3 a_4} A^{n_5}_{a_4 a_5} A^{n_6}_{a_5 a_6} A^{n_7}_{a_6}$$

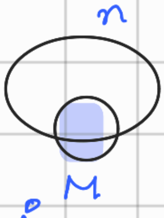
are much clearer in diagram notation



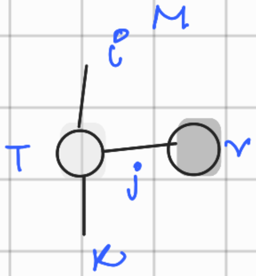
and equally rigorous



$\rightarrow \sum_j v_j w_j$ (inner product)



$\rightarrow \sum_n M_{nn}$ (trace of a matrix)



$\rightarrow \sum_j T_{ijk} v_j$ (contracting order-3 tensor with a vector)

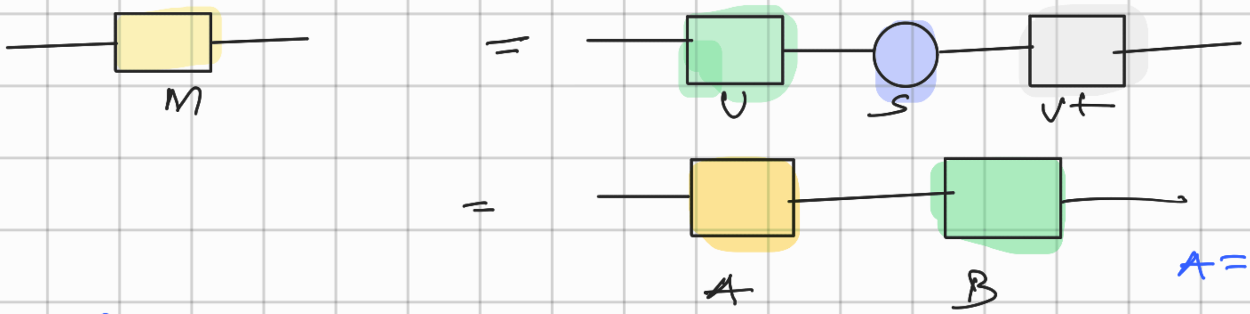
$\Rightarrow R_{ik}$

Now let's use these notations.

We can work with Tensors which are sparse, or we can do sampling or can find low rank structures.

Low Rank Structures

Uncovering low rank structures straightforward for matrices.



$A = U\sqrt{S}$

$B = \sqrt{S}\, V^{\dagger}$

Singular Value Decomposition

The idea of doing this is maybe the new index you get can be smaller than the previous one you're working on.

Eg:

$$M = \begin{bmatrix} 0.435 & 0.223 & 0.10 \\ 0.435 & 0.223 & -0.10 \\ 0.223 & 0.435 & 0.10 \end{bmatrix}$$

low rank

$$\begin{bmatrix} 0.223 & 0.435 & -0.10 \end{bmatrix}_{4\times 3}$$

Can be factorized as

$$\underbrace{\begin{bmatrix} 1/2 & -1/2 & 1/2 \\ 1/2 & -1/2 & -1/2 \\ 1/2 & 1/2 & 1/2 \\ 1/2 & 1/2 & -1/2 \end{bmatrix}_{4\times 3}}_{U} \underbrace{\begin{bmatrix} 0.933 & 0 & 0 \\ 0 & 0.300 & 0 \\ 0 & 0 & 0.200 \end{bmatrix}_{3\times 3}}_{S} \underbrace{\begin{bmatrix} 0.707 & 0.707 & 0 \\ -0.707 & 0.707 & 0 \\ 0 & 0 & 1 \end{bmatrix}_{3\times 3}}_{V^T}$$

Singular values
always real, non negative and decreasing

so we can try and keep fewer and fewer element of S

$$S_2 = \begin{bmatrix} 0.933 & 0 & 0 \\ 0 & 0.300 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

$$\|M_2 - M\|^2 = 0.04 = (0.2)^2$$

$$S_3 = \begin{bmatrix} 0.933 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \qquad \|M_3 - M\|^2 = 0.13 = (0.3)^2 + (0.2)^2$$

Controlled approximation form, and the reason why it solves some value

Hence we get:

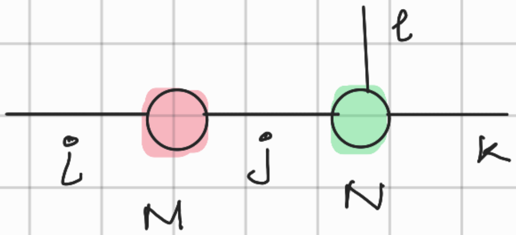
$$U = \begin{bmatrix} 1/2 \\ 1/2 \\ 1/2 \\ 1/2 \end{bmatrix} \qquad S = \begin{bmatrix} 0.933 \end{bmatrix} \qquad V^T = \begin{bmatrix} 0.707 & 0.707 & 0 \end{bmatrix}$$

lot less to store in memory

This is the idea where we can play with low rank structure and leave our rank defined at

Tensor Network Notation

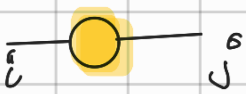
A tensor of the form $\sum_j M_{ij} N_{jkl}$ can be



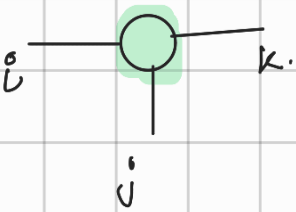
i) Tensors are notated by shape and tensor indices by lines
ii) connecting two index lines implies a contraction, or summation.

Vector V_j

matrix M_{ij}



3-index tensor T_{ijk}.

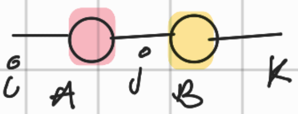


and some example of matrix like contractions:

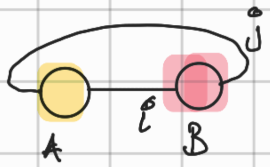
$\sum_j M_{ij} v_j$

and for tensor contractions:

$A_{ij} B_{jk}$

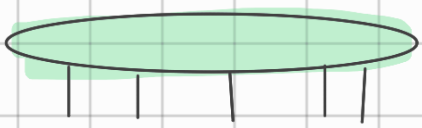


$A_{ij} B_{ji}$

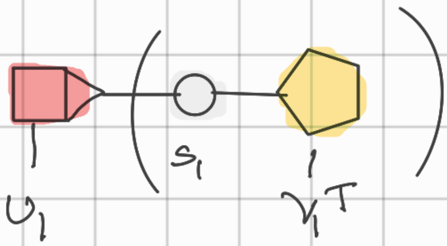


What are Tensor Networks?

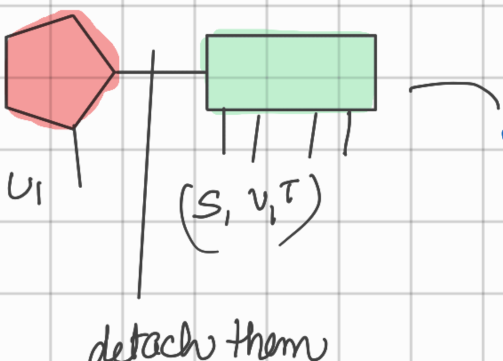
1. SVD first index from rest.



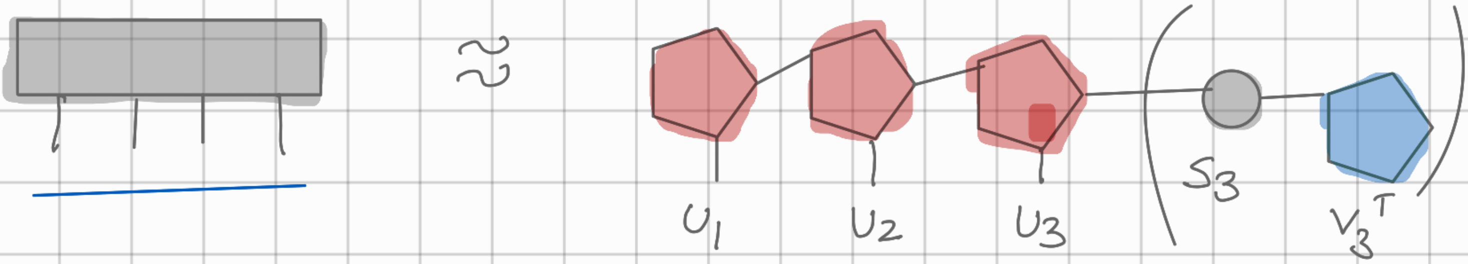
$\overset{SVD}{\approx}$

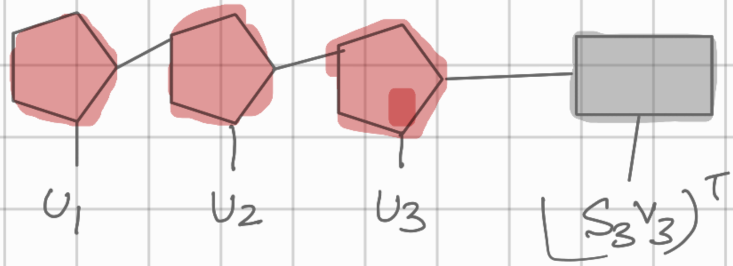


$\Rightarrow$

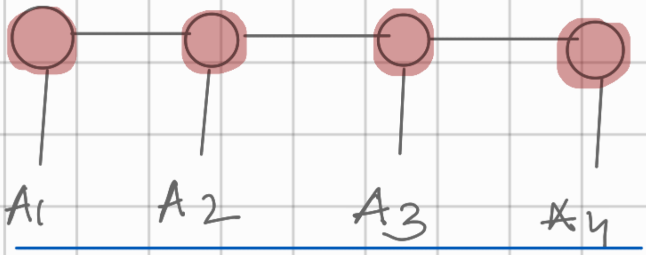


Pretend this is the original tensor and repeat this is what we'll be left with





This is an MPS



a matrix product space

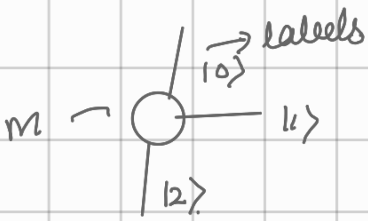
Tensor Networks and Quantum Algorithms

Quantum computation $\simeq$ approximating Tensor Networks

time $\longrightarrow$ geometry

unitary $\longrightarrow$ linear maps

Tensor



= number

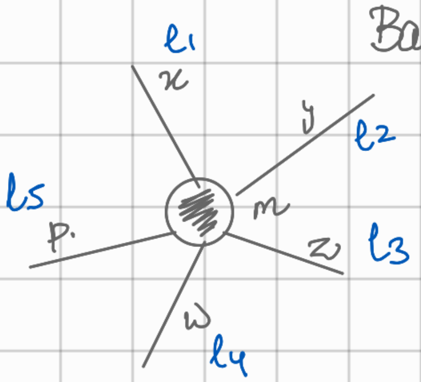
vector : $m \in (\mathbb{C}^d)^{\otimes 3}$

linear maps: $m : (\mathbb{C}^d)^{\otimes 2} \to (\mathbb{C}^d)^{\otimes 1}$

$m(|i_1\rangle |i_2\rangle) = \sum \_ |i_3\rangle$

(too advanced, start from the basic)

Basics for Tensor Networks



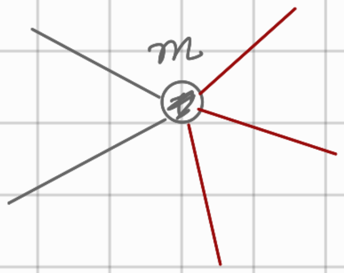
= number

$\mathcal{H} = \mathbb{C}^d$ with orthonormal basis sets.

Vector $\mathcal{H}^{\otimes 5}$

$$m = \sum_{l_1, \cdots l_5} m_{l_1 l_2 l_3 l_4 l_5} |l_1\rangle |l_2\rangle |l_3\rangle |l_4\rangle |l_5\rangle$$

other interpretation is in term of linear maps:



m from 2 copies to 3 copies

$$m : \mathcal{H}^{\otimes 2} \longrightarrow \mathcal{H}^{\otimes 3}$$

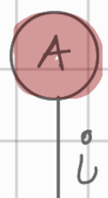
$$m(|l_1\rangle |l_2\rangle) = \sum_{l_3, l_4, l_5} m_{l_1 l_2 l_3 l_4 l_5} |l_3\rangle |l_4\rangle |l_5\rangle$$

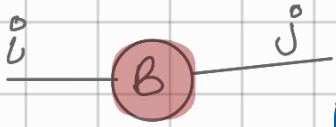
The fundamentals on Tensor Networks!

$$A = \begin{bmatrix} A_1 \\ A_2 \\ \vdots \\ A_m \end{bmatrix} \qquad B = \begin{bmatrix} B_{11} & \cdots & B_{1n} \\ \vdots & \ddots & \vdots \\ B_{m1} & \cdots & B_{mn} \end{bmatrix}$$

Vector

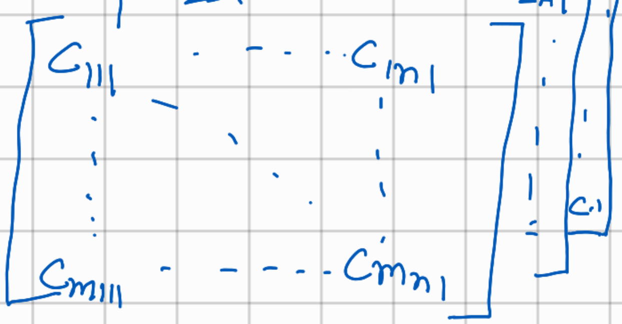
matrix

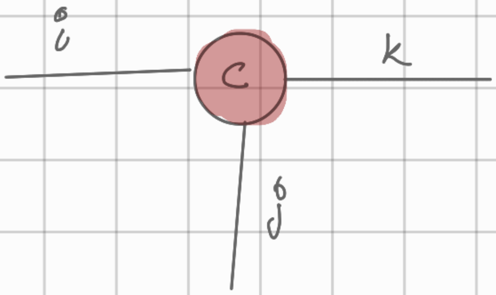




$$\begin{bmatrix} C \cdots \\ C_{221} \cdots \cdots C_{2?1} \end{bmatrix}$$

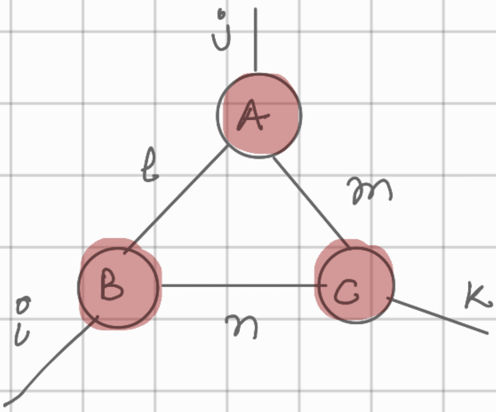
Tensor

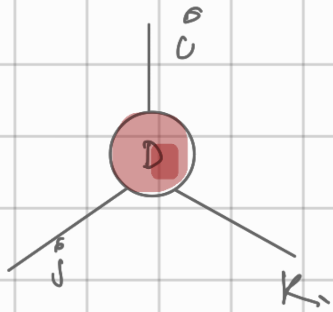




Now we can form networks comprised of multiple tensors, where an index shared by two tensors denotes a summation.

$\sum_{j} A_{ij} B_{jk}$ (this is nothing but a matrix multiplication) C_{ik}





$$\sum_{l,m,n} A_{jlm} B_{iln} C_{nmk} = D_{ijk}$$

In many applications, the goal is to approximate a single high-order tensor as a tensor network composed of many low order tensors. Since the total dimensions of a tensor grows exponentially with its order, the latter representation can be vastly more efficient.

Order- N tensor Network of low order

~ e^N parameters ~ poly (N) parameters

How?

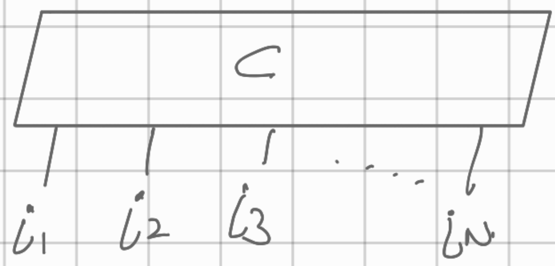
Tensor Networks on Quantum Many Body System

Consider a d-dim quantum system. A pure state $|\phi\rangle$ is represented as a vector.

$$|\phi\rangle = c_1|1\rangle + c_2|2\rangle + \cdots + c_d|d\rangle \Rightarrow \underline{\sum c_i |d_i\rangle}$$

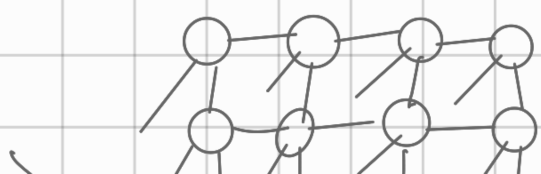
Now for a many-body system formed from a composition of N-individual systems of dimension d. (Eg: Lattice of atoms). A pure state wave function of the composite system $|\psi\rangle$ is d^n

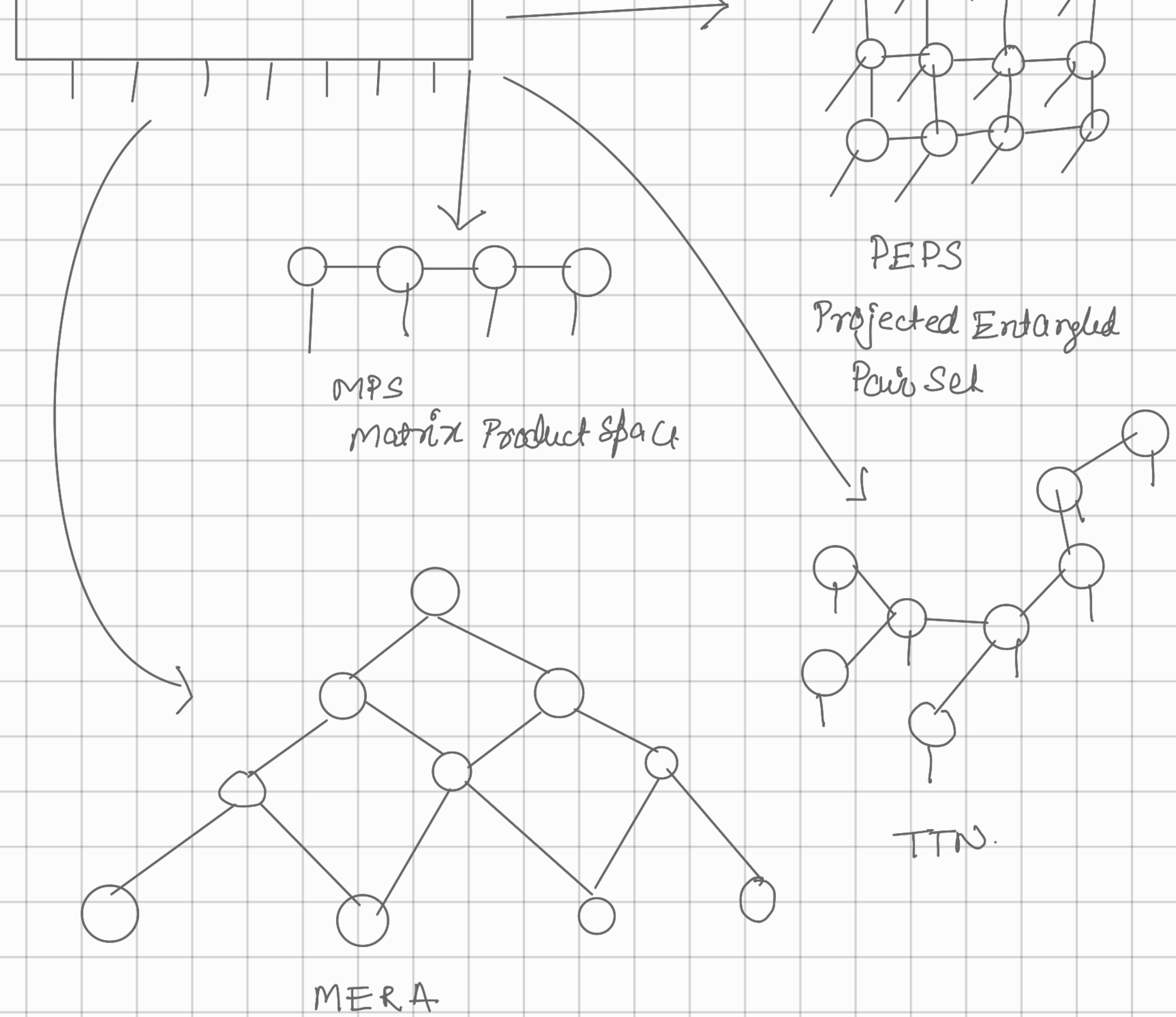
$$|\psi\rangle = \sum c_{i_1 i_2 \ldots, i_N} |i_1\rangle |i_2\rangle \ldots |i_N\rangle$$



in a many body problem we begin with a Hamiltonian which describes how the system will evolve. Our goal is to find the lowest energy eigen-state. In this setting, tensor networks are most commonly used as ansatz

State Coefficient tensor





The choice of best ansatz for a particular problem may depend on the geometry of the problem as well as its physical property.

Circuit knitting with classical communication

Various techniques have been developed to partition large quantum circuits into subcircuits that can fit smaller devices.

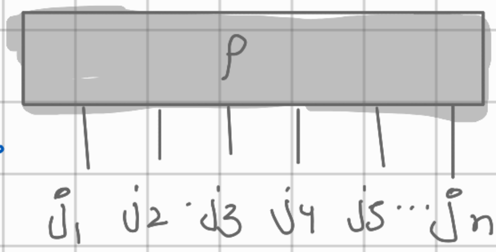
One possibility to realize circuit knitting is using the technique of quasi probability simulation.

Gate Decomposition

Quantum circuit $\rightarrow U^{n\,qubit}$ Initial state density $\rightarrow \rho$ Obs $\rightarrow O$

$$|\rho\rangle = \sum_{j_1, j_2 \cdots j_n} \rho_j \, | e_{j_1} e_{j_2} \cdots e_{j_n} \rangle$$

We can use Pauli matrices as the basis

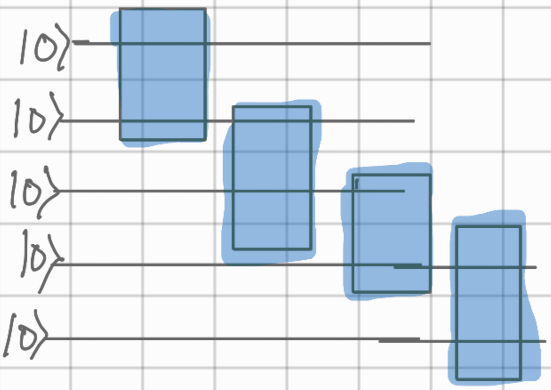


$$\rho_i, O_i \in \mathbb{R}$$

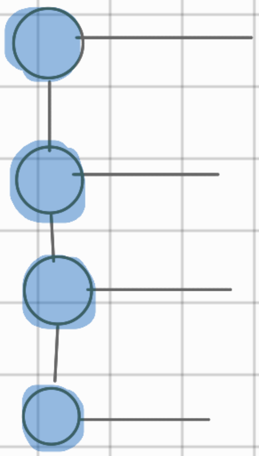
Quantum Circuit U transforms the state ρ to $U\rho U^{+}$

$$S(U) = U\rho U^{+}$$

$$S(U) = \sum_{j_1, \cdots j_n} \sum_{k_1, \cdots k_n} S(U)_{j,k} \, | e_{j_1} \cdots e_{j_n} \rangle \langle e_{k_1} \cdots e_{k_n} |$$



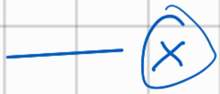
=



Matrix Product States

A generic wave function of many-body quantum system can be expressed by

$$|\psi\rangle = \sum_{k=1}^{N} c_k |k\rangle$$



$N = 2^n$

hence an exponential scaling

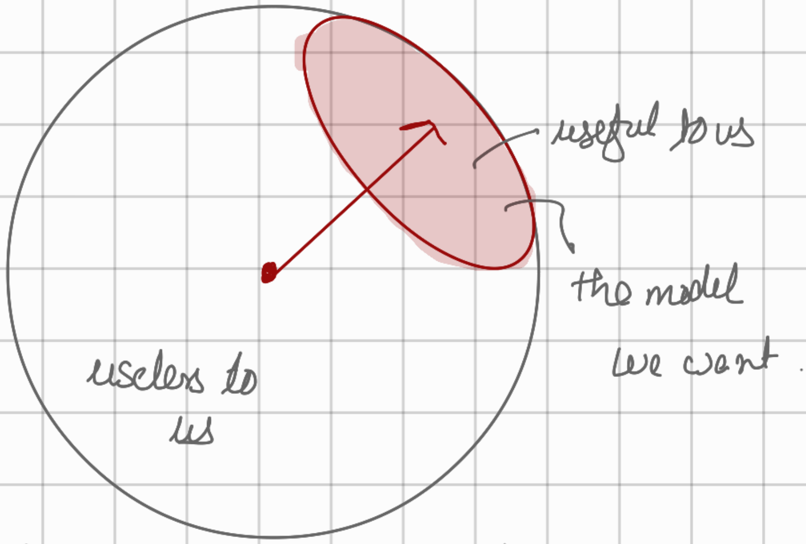
For example if we have 'n' degrees of freedom

For example, if we have 'n' degrees of freedom, ... clearly.
with local dimension 'd', the most general wave function has 'd^n' coefficients.

$$|\psi\rangle = \sum_{s_1, s_2, \cdots s_n}^{d} C_{s_1 \cdot s_2 \cdots s_n} |s_1 s_2 \cdots s_n\rangle$$

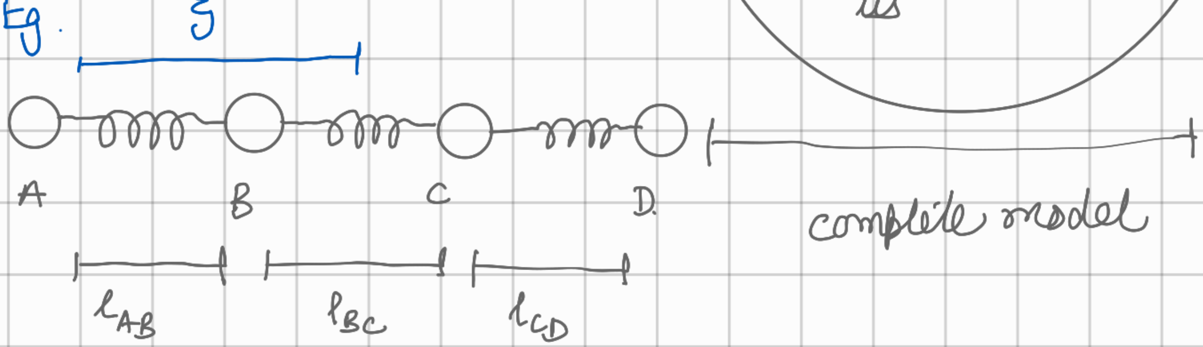
but we are only interested in the low energy solutions.

Eg (x) is very inefficient way to define a state because most of the states we are interested in belong to only a tiny subset of the whole systems Hilbert space.



Hamiltonians are local: i.e only sites that are close to each other interact significantly.

Eg.



There's a finite corelation length ξ. that means for $l_{xy} \gg \xi$ we have no effect of x on y. In the above case A does not affect D. And that's why their state can be approximated by a product state, thus requiring fewer coefficients.

Moreover, most of the Hilbert space cannot be quickly reached by time evolution under a local Hamiltonian. So, every state that can be reasonably prepared belongs to a tiny corner of the whole space of possible space.

We need a representation that is specialized to the corner of the Hilbert space. And MPS offers one such representation.

Consider an n-body system, with local dimension 'd' and open boundary conditions. MPS can be written as.

$$|\psi\rangle = \sum_{s_1,\dots s_n}^{d} \sum_{\alpha_1\dots\alpha_n}^{\chi} M^{[1]s_1}_{1\alpha_1} M^{[2]s_2}_{\alpha_2\alpha_3} \dots M^{[n-1]s_{n-1}}_{\alpha_{n-2}\alpha_{n-1}} M^{[n]s_n}_{\alpha_{n-1}1} |s_1 s_2 \dots s_n\rangle$$

Each tensor is a local description for the i^{th} site, which allows one to apply local operators

sum of basis element weighted by matrix products.

$\chi \sim$ MPS bond dimension and a sufficiently high χ is needed if we want to express truly general. However the idea is that even for low χ we can still encode all the meaningful states.

$$\chi = d^{\lfloor \frac{n}{2} \rfloor}$$

MPS are suitable to describe states with low entanglement.

Gate Decomposition

$$|e_{j_1}\rangle \otimes |e_{j_2}\rangle \otimes |e_{j_3}\rangle \dots \otimes |e_{j_n}\rangle = |e_{j_1} e_{j_2} e_{j_3} \dots e_{j_n}\rangle$$

A density matrix ρ can be decomposed into the sum of

$$|\rho\rangle = \sum_{j_1,\dots j_n} \rho_j |e_{j_1} e_{j_2} \dots e_{j_n}\rangle$$

$\underbrace{\rho_j}_{\text{elements}}$ are the tensor representation of ρ

Observable O can also be decomposed similarly.

Superoperator $S(U) = \sum_{j_1\dots j_n} \sum_{k_1\dots k_n} S(U)_{j,k} |e_{j_1} \dots e_{j_n}\rangle\langle e_{k_1} \dots e_{k_n}|$

$S(U)_{j,k} = \langle e_{j_1} \cdots e_{j_n} | S(U) | e_{k_1} \cdots e_{k_n} \rangle$ is tensor representation of $S(U)$

Pauli transfer matrix

After measuring with observable O.

$$\langle O | S(U) | \rho \rangle = \mathrm{Tr}(O U \rho U^{\dagger})$$

$$\Rightarrow \sum_{j_1 \cdots j_n} \sum_{k_1 \cdots k_n} O_j \, S(U)_{j,k} \, \rho_k$$

generally $U = U_L \cdots U_1$
product of elementary gates

$$S(U) = S(U_L) \cdots S(U_1)$$

If $\boxed{S(U) = \sum c_i S(V_i)}$ can be represented by sum of some simple operations with coefficients c_i, the expectation value of an observable O can be computed

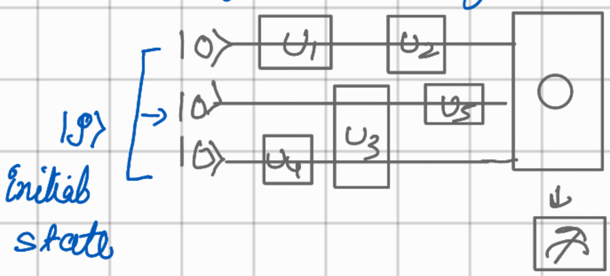
$$\langle O | \underline{\underline{S(U)}} | \rho \rangle = \sum_i c_i \langle O | \underline{\underline{S(V_i)}} | \rho \rangle$$

this is the decomposition of a circuit

What happened in this section?

$U \sim$ quantum circuit (basically collection of all gate)

$U = U_L U_{L-1} \cdots U_2 U_1$



what affect does this circuit has on state?

new state $:\Rightarrow U \rho U^{\dagger}$

Now define a super operator $S(U)$

which basically does $S(U)\rho = U \rho U^{\dagger}$

this is then made into a Tensor

$S(U) = S(U_L) \dots S(U_1)$

$S(U)$ can be decomposed into $\sum c_i S(V_i)$

As long as the tensor representation of each element is unchanged the result of overall computation is also unchanged.

the expectation value calculation

$$\langle O | S(U) | \rho \rangle = \sum c_i \langle O | S(V_i) | \rho \rangle$$

we can decompose 'U' as well, but by doing so:

$U = \sum_i c_i V_i$ we then need to calculate

$$\langle 0 | U^{\dagger} O U | 0 \rangle = \sum c_i c_j^{*} \langle 0 | V_j^{\dagger} O V_i | 0 \rangle$$

which is hard for NISQ devices to evaluate.

Virtual Two Qubit gate

(how to decompose any two-qubit gate into sequence of single qubit)

Lemma 1: A_1, A_2 are operators, such that $A_1^2 = A_2^2 = I$

$$S(e^{i\theta A_1 \otimes A_2}) = \cos^2\theta \, S(I \otimes I) + \sin^2\theta \, S(A_1 \otimes A_2) + \frac{1}{8} \cos\theta \sin\theta \sum_{\substack{(\alpha_1, \alpha_2) \\ \in \{\pm 1\}^2}} \alpha_1 \alpha_2 \Big[S((I + \alpha_1 A_1) \otimes (I + i\alpha_2 A_2)) + S((I + i\alpha_1 A_1) \otimes (I + \alpha_2 A_2)) \Big]$$

Powerful theorem!

$$\boxed{S(e^{i\theta A_1 \otimes A_2}) = \cos^2\theta \, S(I \otimes I) + \sin^2\theta \, S(A_1 \otimes A_2) + \frac{1}{8} \cos\theta \sin\theta \sum_{(\alpha_1, \alpha_2)} \alpha_1 \alpha_2 \Big[S((I + \alpha_1 A_1) \otimes (I + i\alpha_2 A_2)) + S((I + i\alpha_1 A_1) \otimes (I + \alpha_2 A_2)) \Big]}$$

Tensor representation of $S(e^{i\theta A_1 \otimes A_2})$ is .

$$\Rightarrow \langle e_i e_j | S(e^{i\theta A_1 \otimes A_2}) | e_k e_l \rangle$$

$$\Rightarrow \langle e_i e_j | S(\cos\theta + i\sin\theta A_1 \otimes A_2) | e_k e_l \rangle$$

$$\Rightarrow Tr(e_i \otimes e_j (\cos\theta I + i\sin\theta A_1 \otimes A_2) e_k \otimes e_l (\cos\theta I - i\sin\theta A_1 \otimes A_2))$$

$$\Rightarrow \cos^2\theta \, \underline{a_{1,ijkl}} + i\sin\theta\cos\theta(\underline{a_{6,ijkl}} - \underline{a_{7,ijkl}}) + \sin^2\theta \, \underline{a_{8,ijkl}}$$

$$S((I+\alpha_1 A_1) \otimes (I+\alpha_2 A_2))$$

$$\langle e_i e_j | S(I+\alpha_1 A_1) \otimes (I+\alpha_2 A_2) | e_k e_l \rangle$$

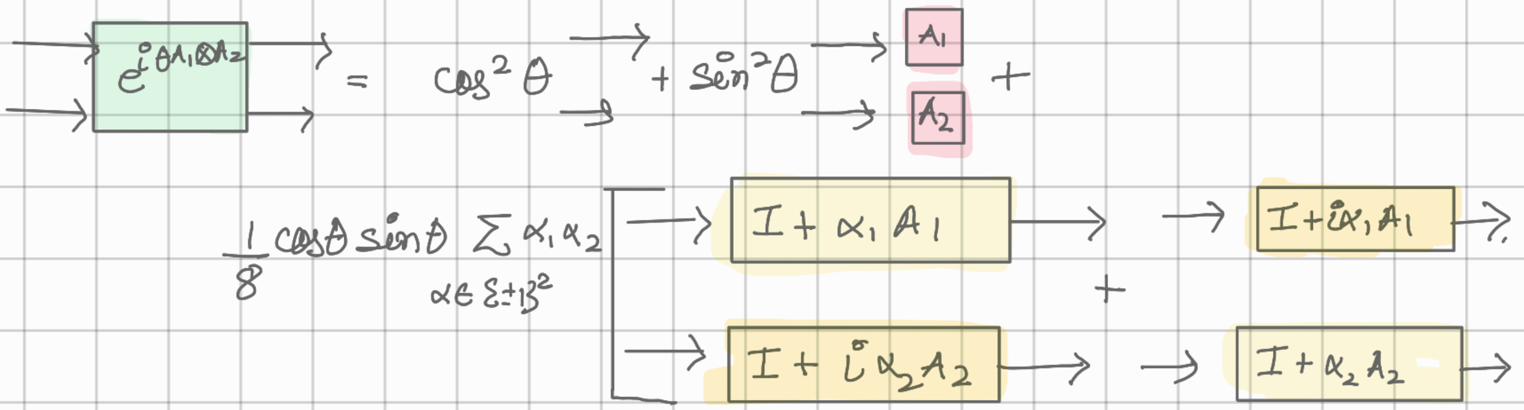
$$\Rightarrow Tr(e_i \otimes e_j (I+\alpha_1 A_1) \otimes (I+\alpha_2 A_2) e_k \otimes e_l (I+\alpha_1^* A_1) \otimes (I+\alpha_2^* A_2))$$

$$\Rightarrow Tr(e_i \otimes e_j \, e_k \otimes e_l) + \cdots\cdots$$

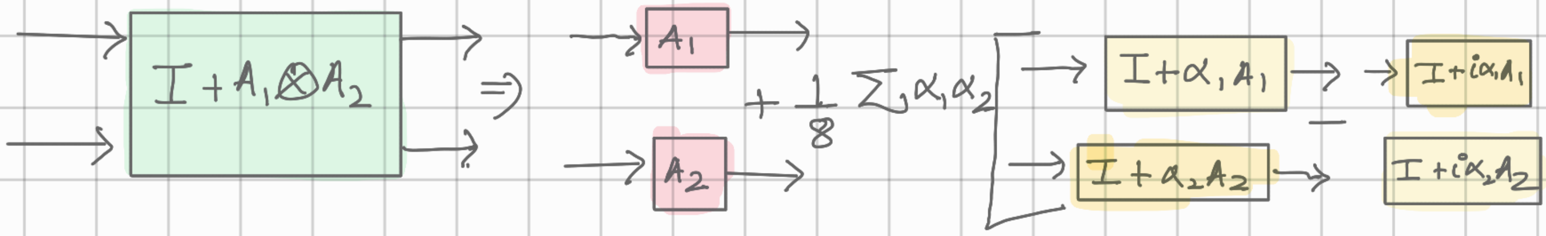
$$\{\alpha_1, \alpha_2\}_{ij,kl} = S((I+\alpha_1 A_1) \otimes (I+\alpha_2 A_2))_{ij,kl}$$

$$\alpha_{1,ijkl} = Tr(e_i \otimes e_j \, e_k \otimes e_l)$$

a)



b)

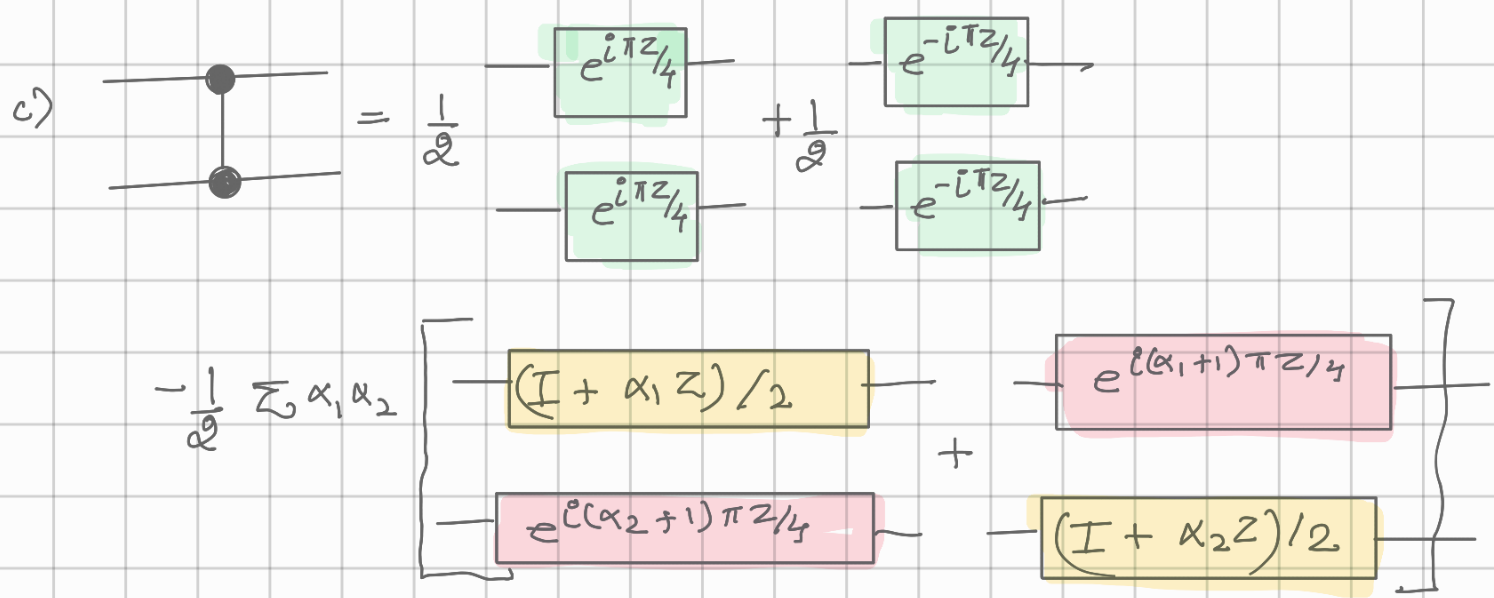


c) Decomposition of a non-local gate and

a) Decomposition of a non-Clifford

b) a non-local non-destructive measurement into a sequence of local operation.

A_1 and A_2 are operators such that $A_1^2 = I$ and $A_2^2 = I$.



c) decomposition of controlled-Z gate into a sequence of single-qubit operation.

This schematic depiction is of the Theorem mentioned above.

The operation proportional to $I \pm A$ and $I \pm iA$ for $A \in \{X, Y, Z\}$ since we are breaking the circuit using

$$S(e^{i\theta A_1 \otimes A_2}) = \cos^2\theta\,(\cdots) + \sin^2\theta\,(\cdots) + \frac{1}{8}\cos\theta\sin\theta \sum \alpha_1\alpha_2 \Big[S((I+\alpha_1 A_1) \otimes (I + i\alpha_2 A_2)) + \cdots \Big]$$

A can be $\{X, Y, Z\}$

Any of the Pauli Basis

these operations $(I \pm iA)$ can be performed by single qubit rotation and $(I \pm A)$ can be performed via projective measurement.

How? $e^{\pm i\pi A/4} = \frac{1}{\sqrt{2}}(I \pm iA)$

(rotation of angle $\pi/2$ by A axis)

for the $(I \pm A)$ part: we can do projective measure!
Let M_A be the projective measurement on the A basis $\{A \in \{x, y, z\}\}$

$$M_{A\rho} = \frac{1}{Tr\left(\rho \frac{I+\alpha A}{2}\right)} \left(\frac{I+\alpha A}{2}\right) \rho \left(\frac{I+\alpha A}{2}\right)$$

depending on the result of the measurement $\alpha \in \{1, -1\}$. This is equivalent to $S(I \pm A)$ up to the factor of $4 Tr\left(\rho \frac{I+\alpha A}{2}\right)$

$$S(I+\alpha A) = 4 Tr\left(\rho \frac{I+\alpha A}{2}\right) M_{A,\alpha}$$

$M_{A,\alpha}$ is a measurement operation post selected with the measurement outcome α.

Taking all these things together, this Lemma 1 implies that the gate $e^{i\theta A_1 \otimes A_2}$ can be decomposed into a sum of $I \otimes I$, $A_1 \otimes A_2$, $M_{A_1} \otimes e^{\pm i\pi A_2/4}$ and $e^{\pm i\pi A_1/4} \otimes M_{A_2}$

This technique can be applied for any θ, which enables us to perform continuous two-qubit gates.

Lemma 2: A quantum gate $e^{i\theta A_1 \otimes A_2}$ with operators A_1, A_2, s.t. $A_1^2 = A_2^2 = I$ can be decomposed into 6 single qubit operations. To achieve the error ϵ of the decomposition with respect to trace distance with probability at least $1-\delta$, the required number of circuit runs are $O\left(\log\left(\frac{1}{\delta}\right)/\epsilon^2\right)$

The error comes from the probabilistic part of the decomposition, where we used projective measurement to get

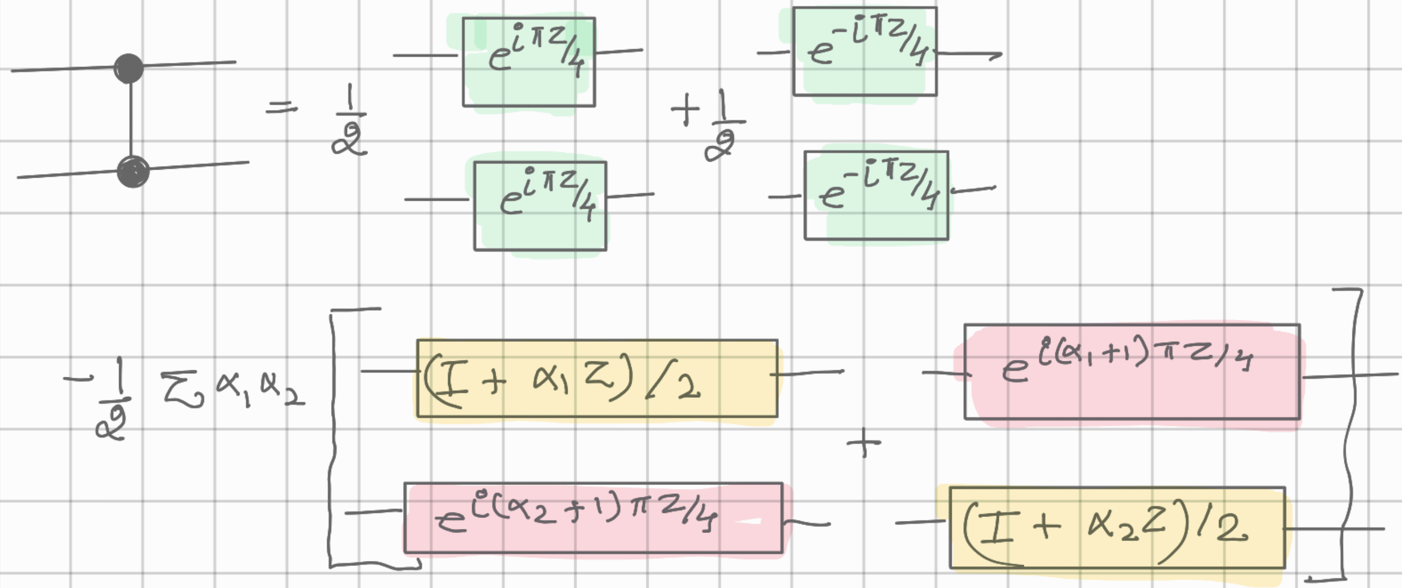
$$S(I+\alpha A) = 4\,\mathrm{Tr}\left(\rho \frac{I+\alpha A}{2}\right) M_{A,\alpha}$$

this thing induces error.

In the case of CZ gate, it can be decomposed like that:

$$CZ = e^{i\pi I\otimes Z/4}\, e^{i\pi Z\otimes I/4}\, e^{-i\pi Z\otimes Z/4}$$

c)



This is how decomposition happens. Same can be done for CNOT gate. This protocol is advantageous to the previous since number of single-qubit operations required is 6 compared to 9 in Endo et.al.

Virtual non-destructive measurement of two qubit operators.

Lemma 8: For operators A_1 and A_2 s.t $A_1^2 = A_2^2 = I$

$$S(I + A_1 \otimes A_2) = S(I\otimes I) + S(A_1\otimes A_2) + \frac{1}{8}\sum_{(\alpha_1,\alpha_2)\in\{\pm1\}^2} \alpha_1\alpha_2 \Big[S((I+\alpha_1 A_1)\otimes(I+\alpha_2 A_2))$$

$$- S((I+i\alpha_1 A_1)\otimes(I+i\alpha_2 A_2))\Big]$$

this lemma can be used to further prove another lemma

Lemma 4: A non local projection $\frac{I + A_1 \otimes A_2}{2}$ with operators A_1 and A_2 such that $A_1^2 = 1$ and $A_2^2 = 2$ can be decomposed into 6 single qubit operations. For any quantum state $|\rho\rangle$, to acheive the error ϵ of the decomposition with respect to the trace distance with prob. of at least $1-\delta$, the required number of circuit runs is $O(\log(1/\delta)/\epsilon^2)$

APPLICATION (20/01/24)

i) Simulation of large quantum circuits

Any quantum circuit 'C' can be represented by a directed graph $G = (V, E)$ with three type of vertices.

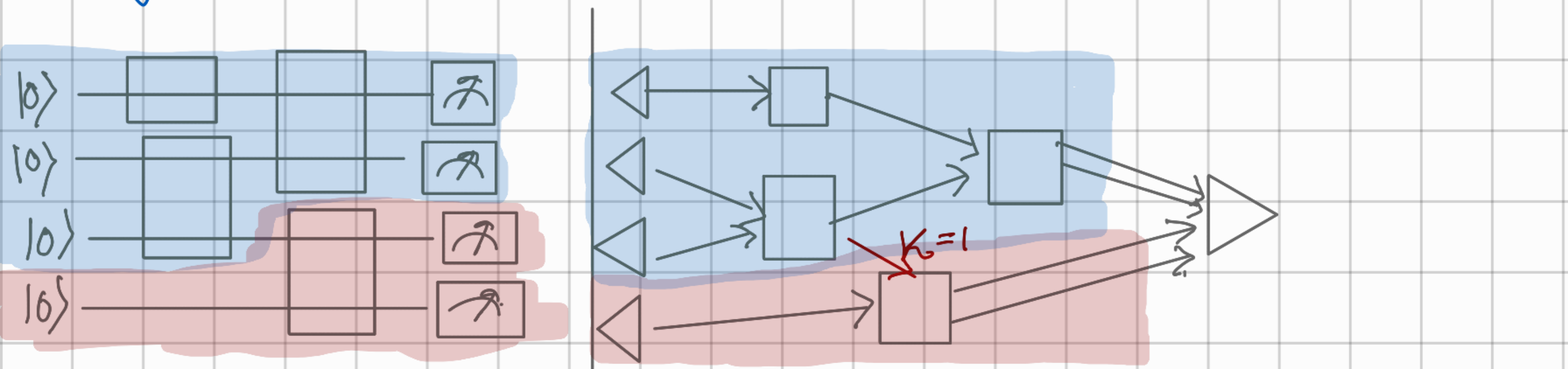
input ◁ quantum gates □ observable ▷

with directed edges showing the flow of qubits between them. Further associate each vertex $v \in V$ with a tensor $A(v)$

input state quantum gates observable

This results in a tensor network (G, A) that completely captures the algorithms described by the original quantum circuit.

Peng et. al focused much on the cluster approach, determined by a clustering parameter. They looked into decomposability of a given Tensor network. Keeping part of the computation quantum



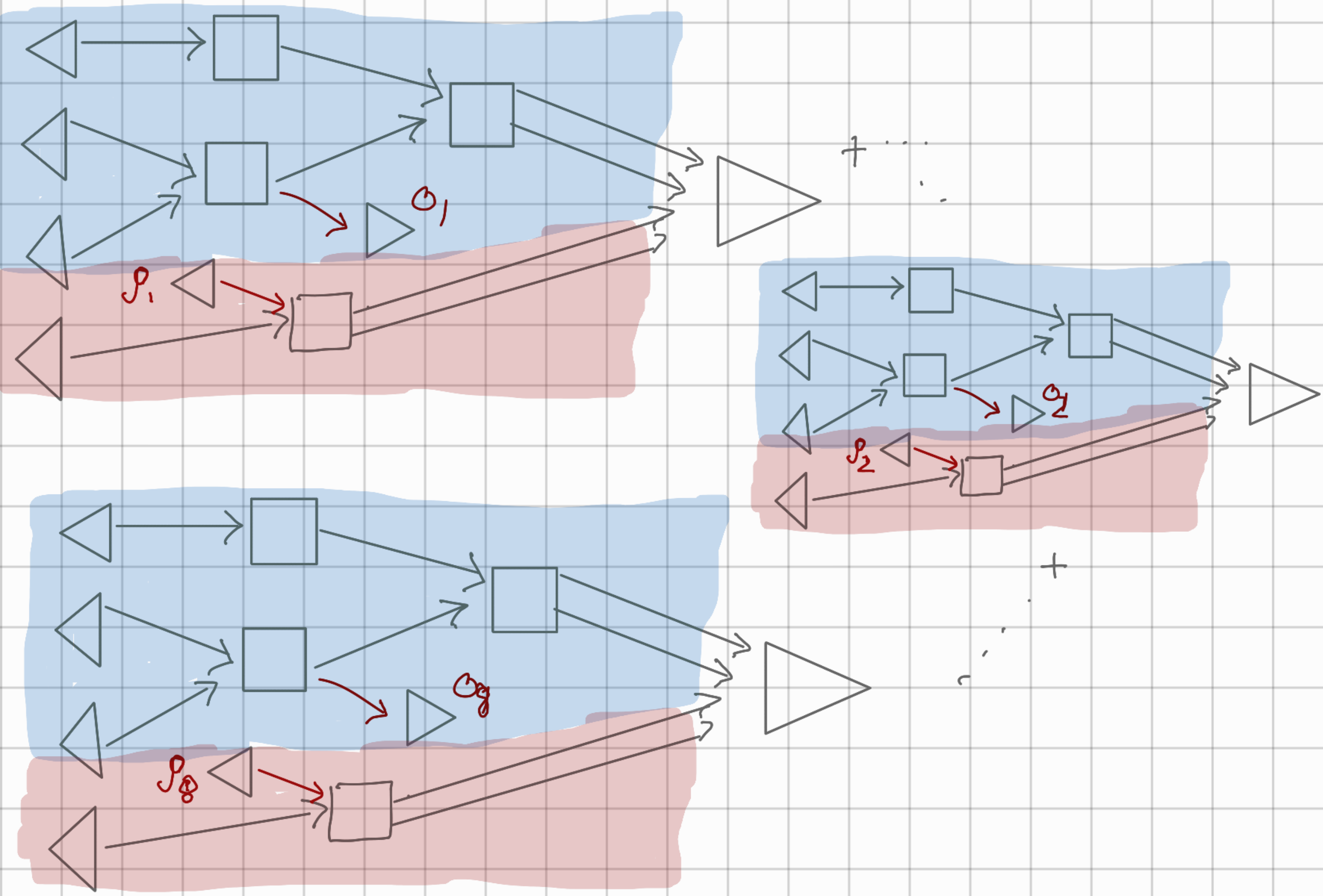
Phase 1 (C, f) the original quantum circuit

Phase 2 (G, A)

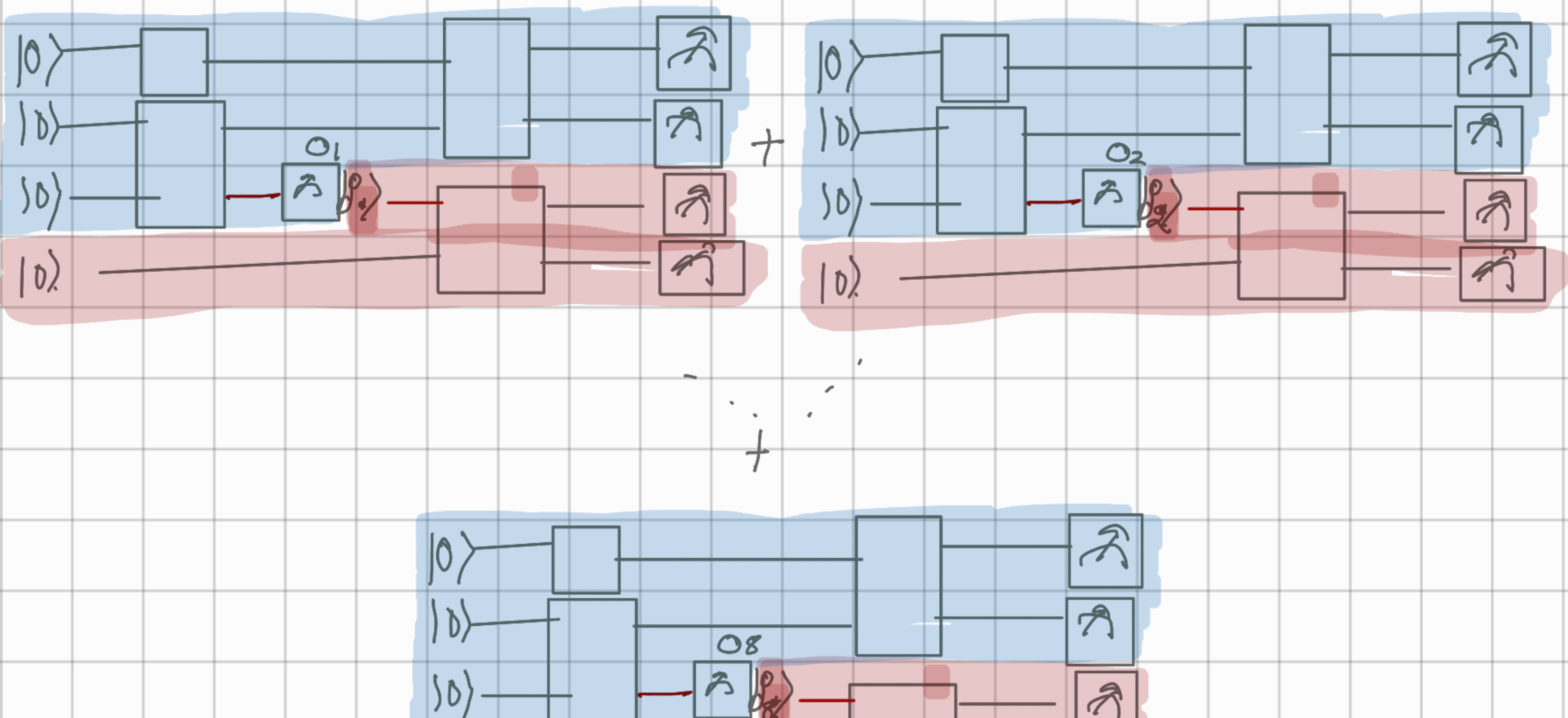
◁ state

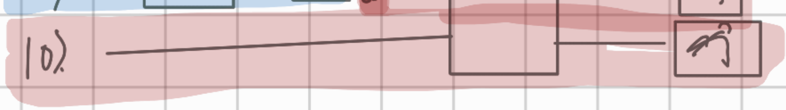
□ gate

▷ observable



Phase 3: a collection of tensor networks obtained by cutting an edge.





Phase 4: collection of smaller quantum circuits

They also make use of the fact that

Let A be any 2×2 matrix and since normalized Pauli matrices $\{I, X, Y, Z\}/\sqrt{2}$ form an orthonormal basis

$$A = \frac{Tr(AI)I + Tr(AX)X + Tr(AY)Y + Tr(AZ)Z}{2}$$

How does the math works out?

For any gate U, state ρ and observable O. entries of the tensor are computed as

$$A(\rho)_\alpha = Tr[\rho \cdot M(\alpha)^\dagger]$$
$${}_\alpha A(U)_\beta = Tr[U M(\alpha) U^\dagger \cdot M(\beta)^\dagger]$$
$${}_\beta A(O) = Tr[M(\beta) \cdot O]$$

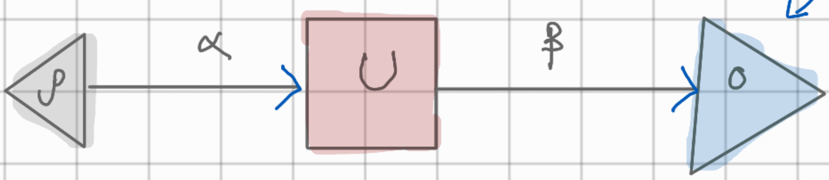
location of the subscripts indicates incoming or outgoing qubits

where $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_k)$ where each $\alpha_j = (\alpha_j^1, \alpha_j^2)$ is a pair of binary indices. Let $M(\alpha_j) = |\alpha_j^1\rangle\langle\alpha_j^2|$ and

$M(\alpha) = \bigotimes_{j=1}^{k} M(\alpha_j) = |\alpha_1^1\, \alpha_2^1, \dots \alpha_k^1\rangle\langle\alpha_1^2 \dots \alpha_k^2|$ be its k-qubit generalization

$$Tr[U\rho U^\dagger O] = \sum_{\alpha,\beta} A(\rho)_\alpha \cdot {}_\alpha A(U)_\beta \cdot {}_\beta A(O)$$

This equation is represented by this diagram.

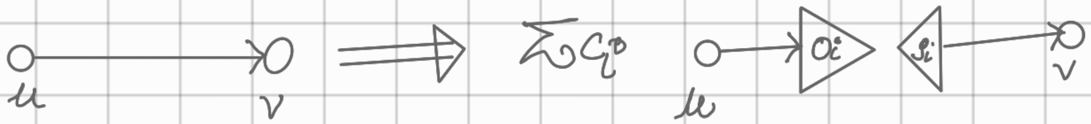


For a general tensor network (G, A), its value is

$$T(G, A) = \sum_{\alpha \in \Pi^E} \prod_{v \in V} A(v)_{\alpha(v)}$$

so a quantum circuit can be simulated by approximating the value of its tensor network.

Lemma: There is a set of eight coefficients $c_i \in \{-\frac{1}{2}, \frac{1}{2}\}$, observables O_i and states ρ_i such that the following modification of the edge uv



does not affect the value of the overall tensor network. We call this "time-like" cut. This actually yields a tensor network

$$T(G, A) = \sum_{i=1}^{8} c_i \, T(G', A_i)$$

Proof: $A \in \{I, X, Y, Z\}$

$$I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad X = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \quad Y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \quad Z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

and since normalized Pauli matrices form a orthonormal pair.

$$A = \frac{Tr(AI)I + Tr(AX)X + Tr(AY)Y + Tr(AZ)Z}{2}$$

Let's expand each Pauli matrix in its eigenbasis. If we denote Pauli matrices by O_i, their eigenprojectors by ρ_i and their corresponding eigenvalues by $2c_i$, so that $O_1 = O_2 = 2(c_1\rho_1 + c_2\rho_2)$

$$O_1 = I, \quad \rho_1 = |0\rangle\langle 0|, \quad c_1 = +1/2, \quad (12)$$
$$O_2 = I, \quad \rho_2 = |1\rangle\langle 1|, \quad c_2 = +1/2, \quad (13)$$
$$O_3 = X, \quad \rho_3 = |+\rangle\langle +|, \quad c_3 = +1/2, \quad (14)$$
$$O_4 = X, \quad \rho_4 = |-\rangle\langle -|, \quad c_4 = -1/2, \quad (15)$$
$$O_5 = Y, \quad \rho_5 = |+i\rangle\langle +i|, \quad c_5 = +1/2, \quad (16)$$
$$O_6 = Y, \quad \rho_6 = |-i\rangle\langle -i|, \quad c_6 = -1/2, \quad (17)$$
$$O_7 = Z, \quad \rho_7 = |0\rangle\langle 0|, \quad c_7 = +1/2, \quad (18)$$
$$O_8 = Z, \quad \rho_8 = |1\rangle\langle 1|, \quad c_8 = -1/2, \quad (19)$$

Measuring the qubit

in either I or Z basis physically correspond to the same circuit.

4 upstream circuits $\{I, X, Y, Z\}$ 6 downstream circuits

3 upstream circuits $\{X, Y, Z\}$ 4 downstream circuit

Plot for Cut QC

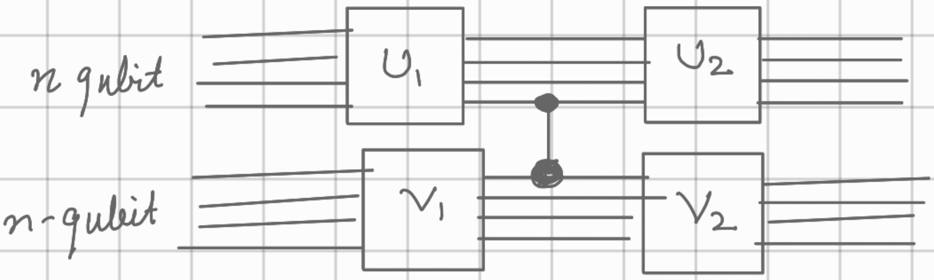
Wei Tang et. al.

(time like cut since it cut wires)

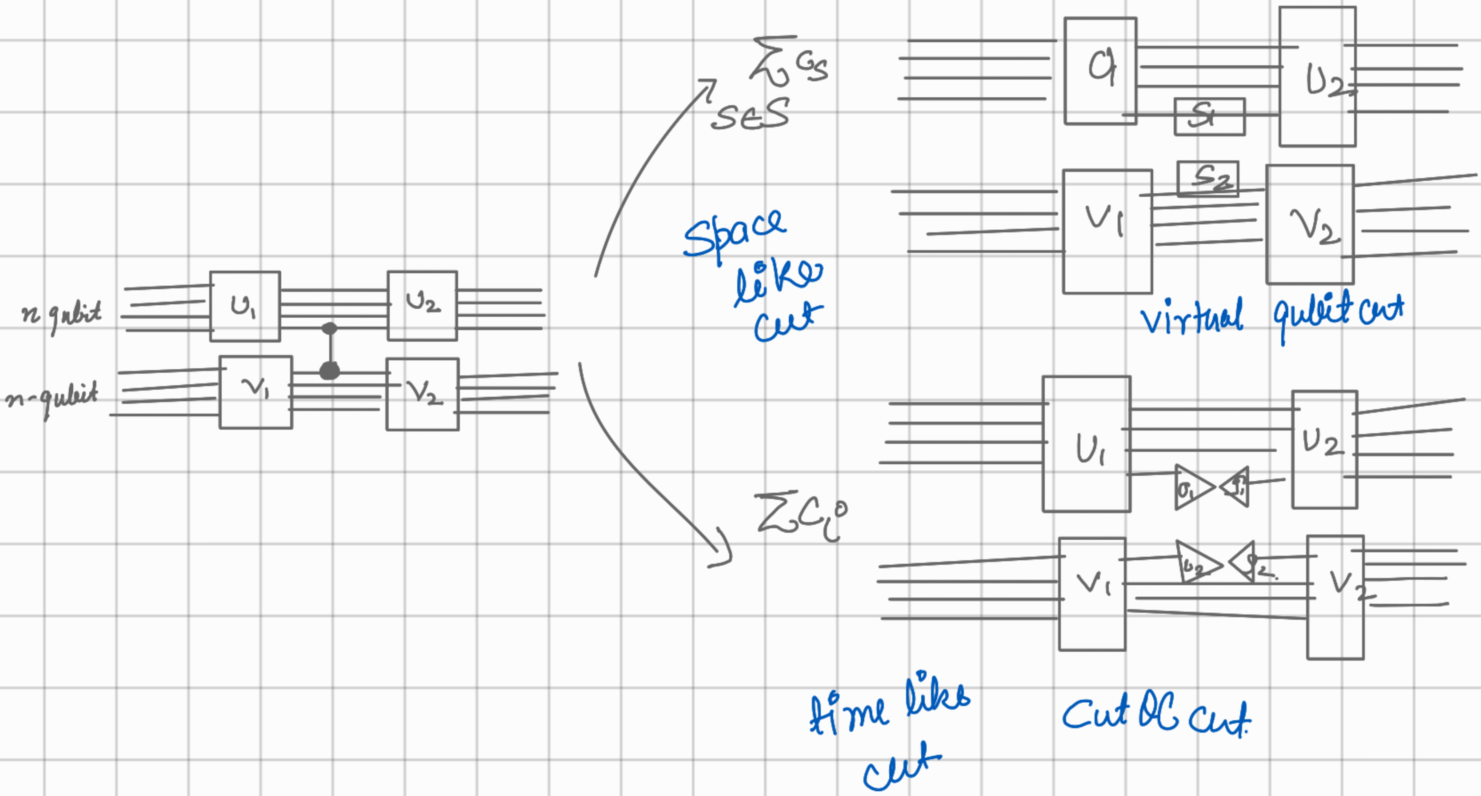
Peng et.al has exponentially higher post processing cost.

Comparing the scaling cost of decomposition

Consider we have an n-qubit quantum computer to simulate 2n qubit quantum circuit, which has only one CZ gate between n-qubit cluster



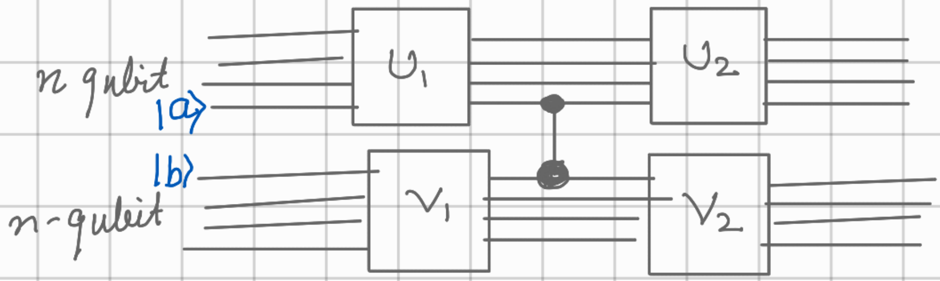
The task is to estimate the expectation value of a final observable O_f



We assume that the expectation values that we wante to measure is a string of Pauli Zs. If 'v' is the desired variance in the estimation, for an n-qubit circuit the time-like cut requires $\frac{2048}{v}$ runs while the space like cut requires $\frac{15}{2v}$ runs.

In detail

time-like cut. Let us name two-qubits on which the CZ gate acts a and b



and σ_a and σ_b be their 1-qubit reduced density matrix after the gate $U_1 \otimes V_1$.

Let estimate $\langle O_f \rangle$

We divide the allowed number of circuit runs N into $N/2$, and $N/2$ runs are then further divided into $N/128$ runs to run the circuit with O_i and ρ_j for $i,j \in \{1,2 \ldots 8\}$.

Check Appendix D: of the paper.

Although the analysis is based on a naive algorithm and there are possibilities to improve it.

General Case

We perform time-like and space-like cut simultaneously. Let the number of time-like and space-like cuts be M_t and M_s.

CZ.

$\rho \rightarrow |0\rangle\langle 0|^{\otimes m}$ and O_f $f: \{0,1\}^m \rightarrow [-1,1]$

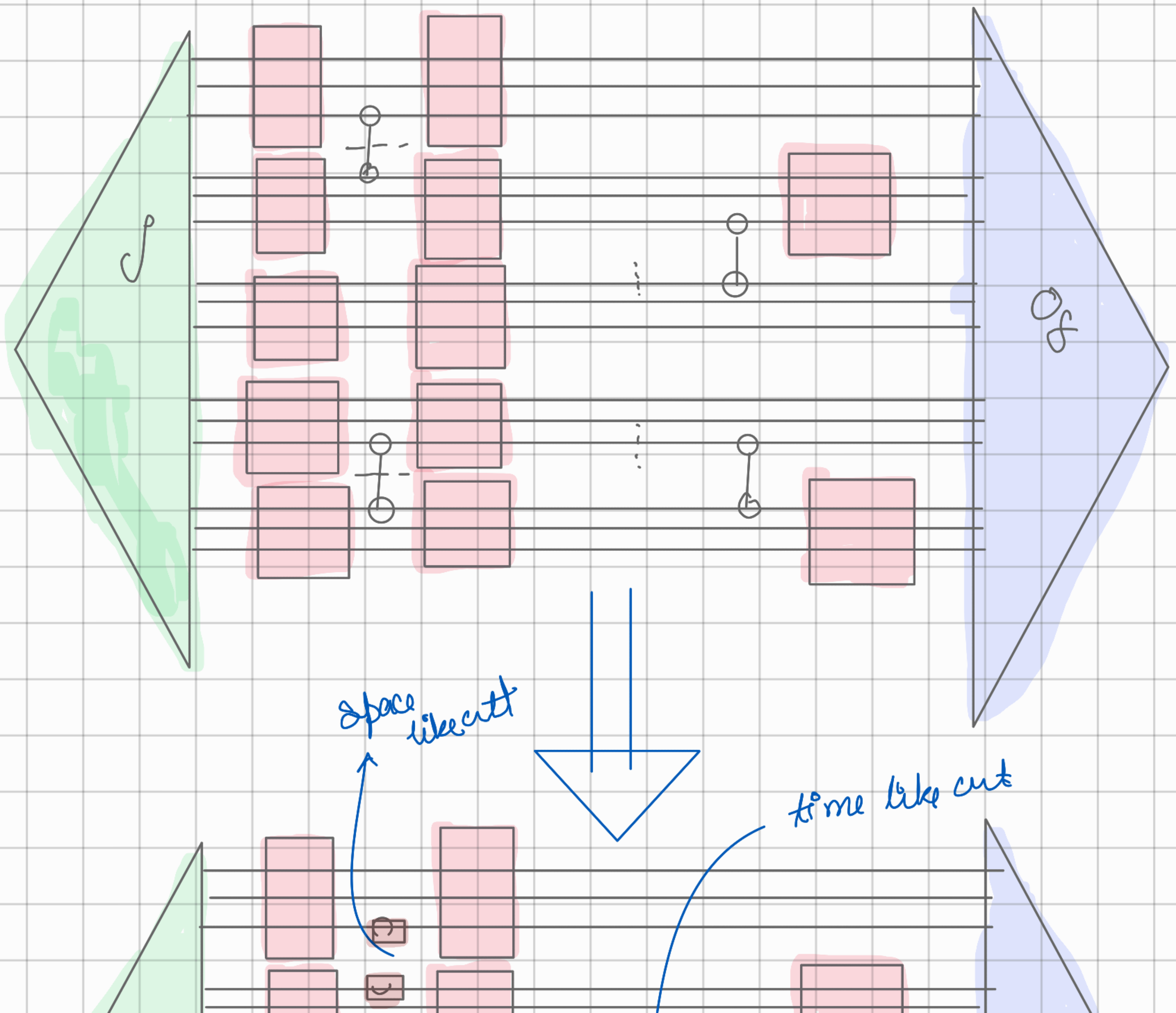
n-qubit circuit on a n-qubit device

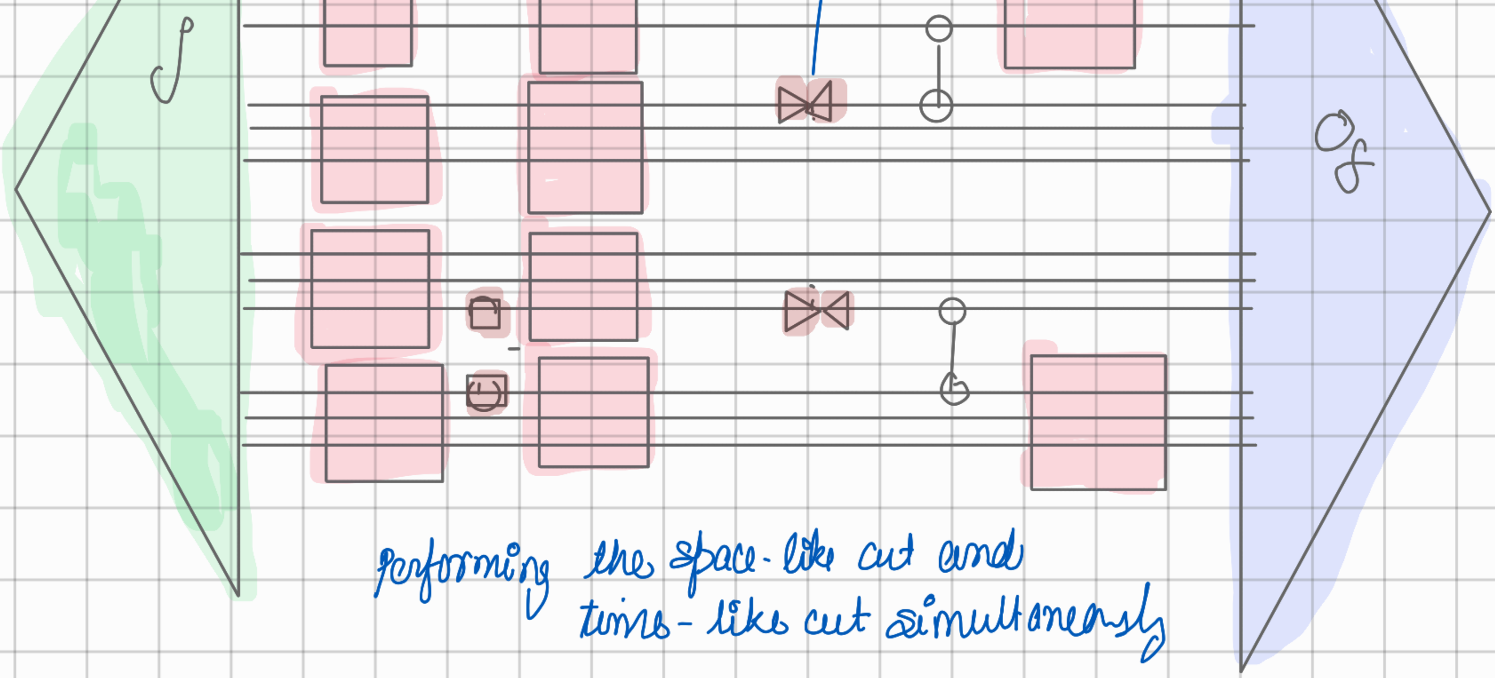
We want to estimate $E[f(y)]$

Theorem: The number of n-qubit circuit runs required to estimate $E[f(y)]$ within accuracy ϵ with some high probability $1-\delta$ is

$$O\left(\frac{9^{M_s}\,16^{M_t}}{\epsilon^2}\log\left(\frac{1}{2\delta}\right)\right)$$

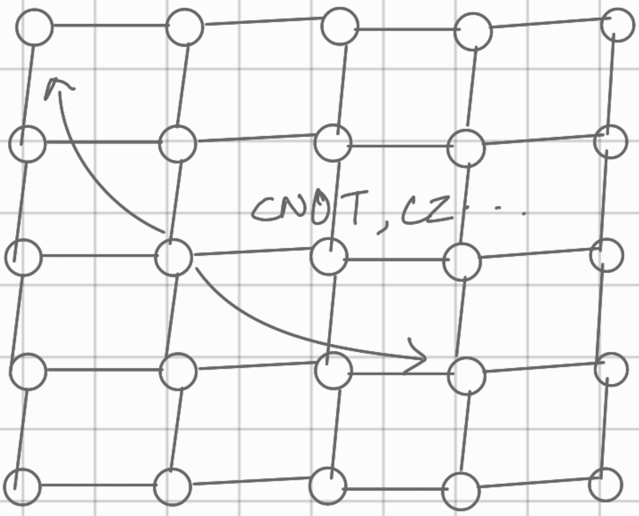
This implies decomposition of the circuit should be done to minimize $\underline{\underline{9^{M_s}\,16^{M_t}}}$

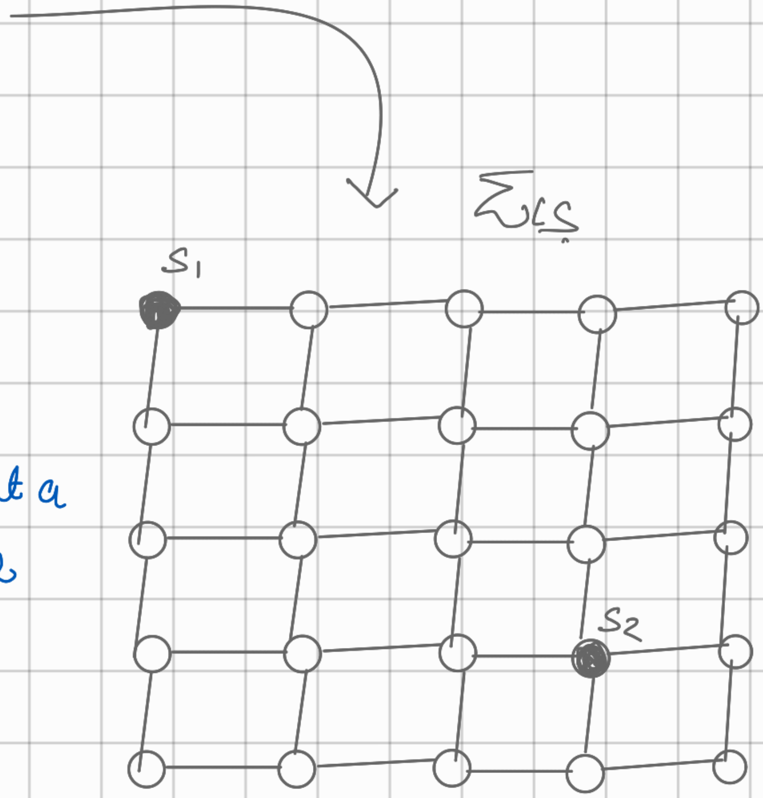




Performing the space-like cut and time-like cut simultaneously

Distant two qubit Gates





Each vertex of the graph represent a qubit and the edge represent the connectivity of the qubits.

This kind of cutting reduces the amount of SWAP gates required. This might be useful for the variational algorithms and QAOA.

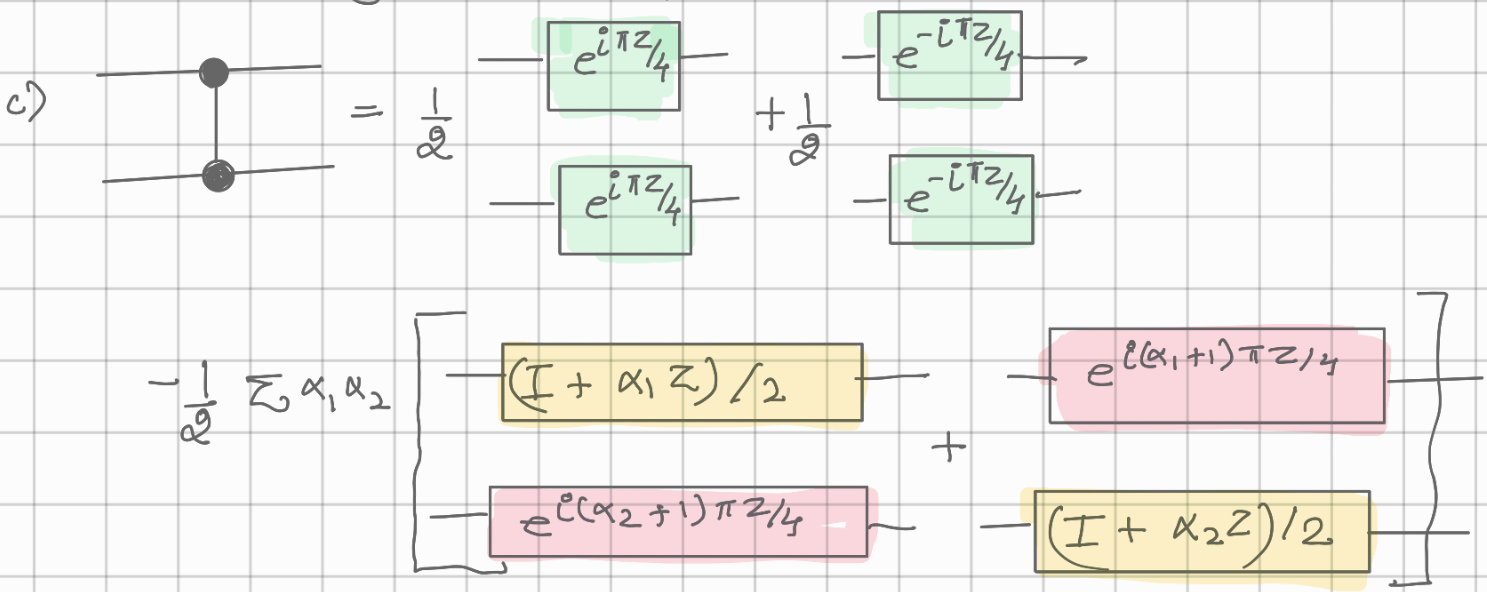
Eg

$$H = \sum J_{ij} Z_i Z_j$$

$\langle H(\beta,\gamma)\rangle = \langle +|\; U^{\dagger}(\beta,\gamma)\, H U(\beta,\gamma)|+\rangle$

CZ gate decomposition

c)



$$CZ = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix} \quad \overset{\text{decomposed}}{\Rightarrow} \quad \begin{bmatrix} 1-i & 0 & 0 & 0 \\ 0 & 1+i & 0 & 0 \\ 0 & 0 & 1+i & 0 \\ 0 & 0 & 0 & -1-i \end{bmatrix}$$

although the matrix are different, but the channel representation are the same.

what is a channel representation?

The channel representation of a quantum gate U essentially corresponds to the action of the gate on the density matrix ρ. as in

$$S(U) = U\rho U^{+}$$

instead of its action on the pure state wave function $U|\psi\rangle$

The channel representations is an appropriate description of a circuit where some of its operations are non-unitary (like the mid circuit measurement), in this case the state of quantum registers can only be fully described in terms of density operator (ρ) acted upon by superoperators $S(U)$.

Hermitian, Positive, semi-definite

$$\rho = |\psi\rangle\langle\psi|$$

Pure state

$|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$

Mixed state

(Statistical ensemble of qubits)

$$\rho = p_1|\psi_1\rangle\langle\psi_1| + p_2|\psi_2\rangle\langle\psi_2|$$

$$\rho = \frac{1}{3}|0\rangle\langle 0| + \frac{2}{3}|1\rangle\langle 1|$$

In the paper itself they mentioned that they adapted this superop. decomposition rather than standard operator decomposition

It is advantageous to use super operator formalism, since it helps us to express the expectation value that we wish to compute

$$e^{i\theta A_1 \otimes A_2} = \cos\theta\, I \otimes i\sin\theta\, A_1 \otimes A_2$$

This equality holds when we consider the operations as quantum channels.

$$e^{i\pi Z/4} = \cos\left(\frac{\pi}{4}\right) I + i\sin\left(\frac{\pi}{4}\right) Z \quad -i$$

$$\Rightarrow \frac{1}{\sqrt{2}} I + i\frac{1}{\sqrt{2}} Z$$

$$\Rightarrow \quad \frac{1}{\sqrt{2}}\left(I + iZ\right)$$

$$e^{i(\alpha_2+1)\pi Z/4} = \cos\left((\alpha_2+1)\frac{\pi}{4}\right) I + i\sin\left((\alpha_2+1)\frac{\pi}{4}\right) Z$$

$$Z = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \qquad I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

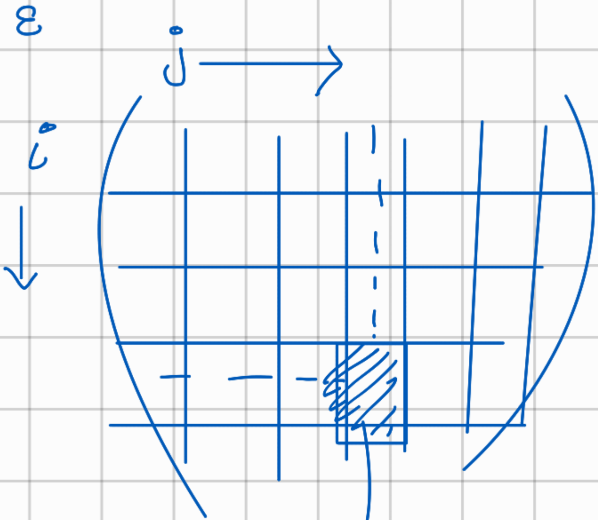
Choi Matrix

$$B(H) \longrightarrow B(H') \longleftrightarrow B(\mathcal{H}\otimes\mathcal{H}')$$

$$|\Omega\rangle\langle\Omega| = \frac{1}{d}\sum_{i,j} \underbrace{|i\rangle\langle j|}_{\text{not affected}} \otimes \underbrace{|i\rangle\langle j|}_{\text{this do}} \quad \xrightarrow{I} \quad \xrightarrow[\mathcal{E}]{}$$

$$\Rightarrow \quad \frac{1}{d}\sum_{i,j} |i\rangle\langle j| \otimes \mathcal{E}(|i\rangle\langle j|)$$

$$\mathcal{E}(|i\rangle\langle j|)$$



what is Quantum channel?

Quantum channels are little bit complicated mathematically. A density matrix is a positive semi-definite matrix ρ with $Tr(\rho)=1$. They represent the information about a quantum state. i.e the probabilities and outcomes of the state that can arise upon measurement.

A quantum channel is a linear map that maps density matrices to density matrices by preserving the fundamental properties such as positivity and trace condition. A quantum channel is a

completely positive trace-preserving (CPTP) map.