Optimal wire cutting with classical communication

Circuit knitting is a process of partitioning large quantum circuits into smaller subcircuits such that the result of the orginal circuits can be deduced by only running the subcircuits.

The cost for any circuit knitting approach scales exponentially in the number of cuts

Circuit knitting techniques allow us to determine the outcome of the circuit by cutting it into several smaller subcircuits that fit on the quantum devices, run them individually and then combine the result in a specific manner.

Since quantum circuits consist of wires and gates, there are two possibilities for a cut - cutting wires or cutting gates

Depending on the circuit structure either wire cuts or gate cuts can be favourable

All circuit knitting idea is based on the fact of quasi probability simulation

The idea is to probabilistically simulate the outcomes of the original quantum circuit by randomly exchanging the non-local gates across the subcircuit with local operations in a careful manner - one can estimate the ideal expectation value of the oregenal circuit, though with an increased variance.

$$U = \sum_i a_i \cdot f_i$$

nonlocal / local

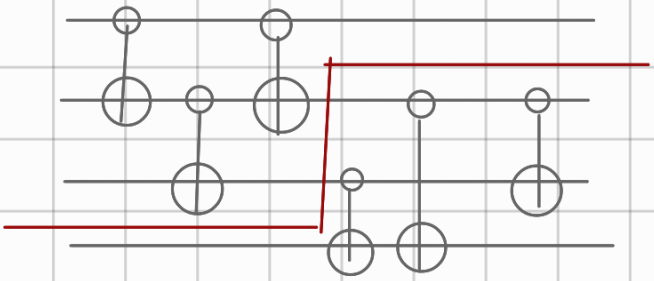
Real but can be negative.

whether classical communication between the smaller quantum devices improve the sampling overhead of the circuit cutting?
classical communication does not seem to improve the overhead sampling for a single instance of gate cut, but can be improved if multiple gate cuts are considered at once instead of separately.

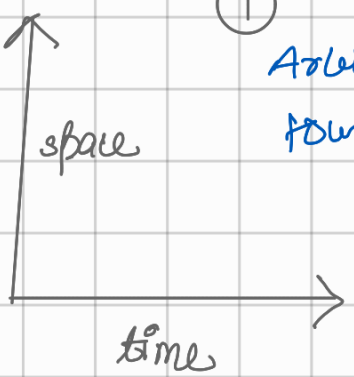
More concretely, the optimal gate cutting technique for two CNOTs gates has a strictly lower overhead than the optimal gate cutting technique for a single CNOT gate applied twice. ~ review later

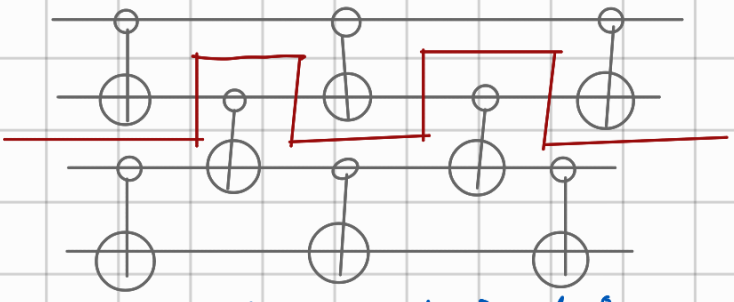
Speed up for multiple wire cuts can not only be achieved when the cuts happen in parallel, but also when they happen at arbitrary position. (at cost of few ancilla qubits)

Sampling overhead of doing multiple cuts can be lower than applying the optimal cutting procedures to individual cuts separately.



Parallel cut involving two wires.
all wires are cut at the same time

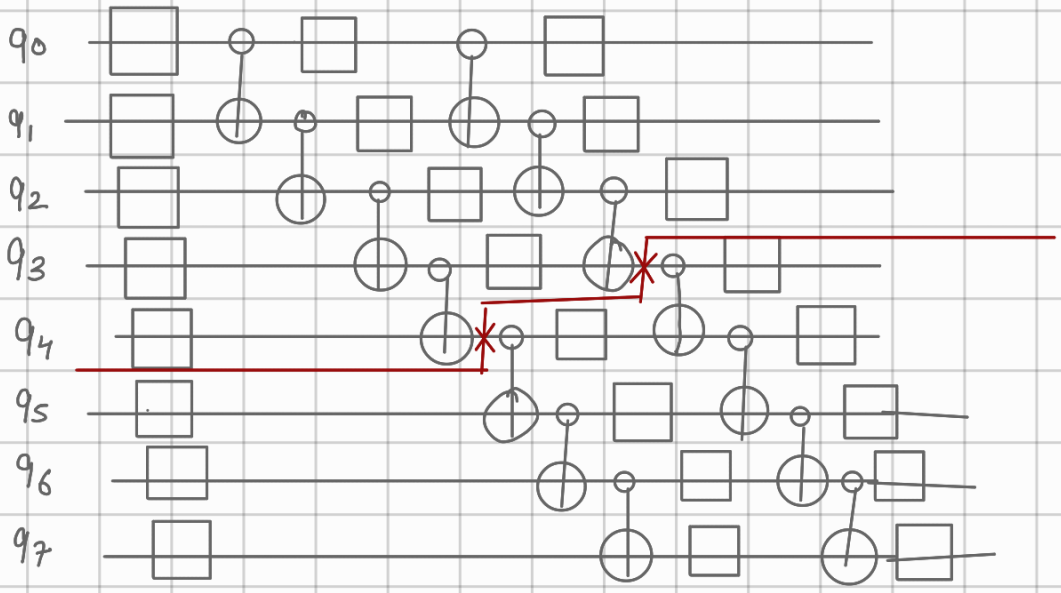


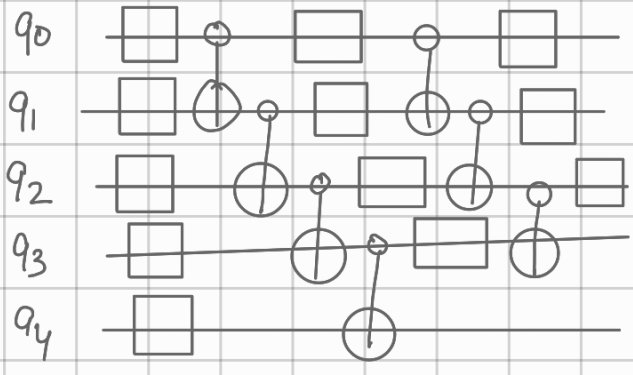


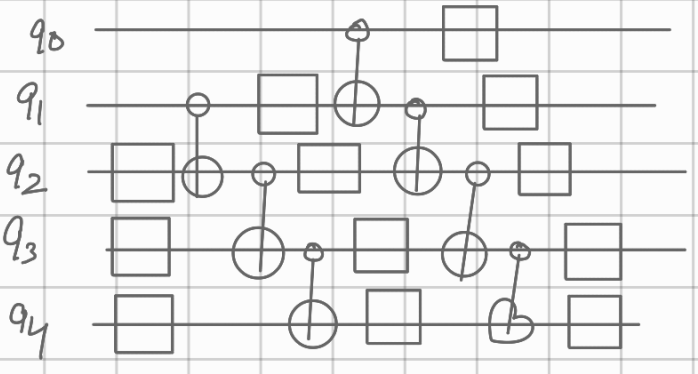
Arbitrary cut involving four qubits.
(as the name suggest, the cuts are at arbitrary position)

We explore the question, that whether including classical communication i.e two Evolved subcircuits to exchange classical communication affects the optimal sampling overhead.

To what extent classical communication can help to improve the sampling overhead in both scenarios. (parallel & arbitrary)



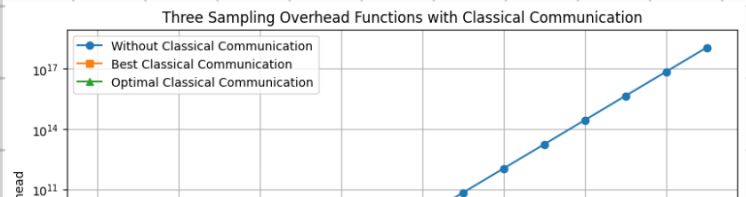


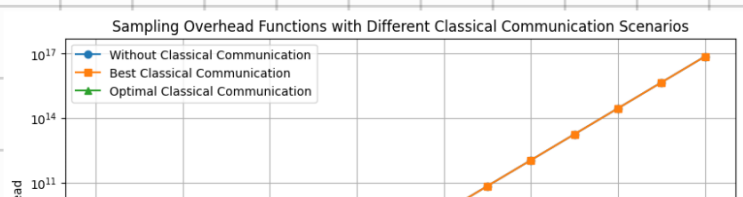


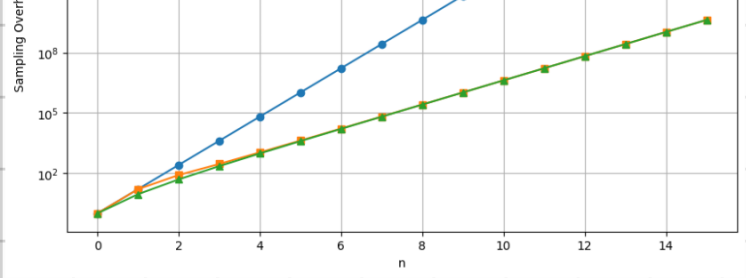
Cut QC can find both arbitrary cut and parallel cut.

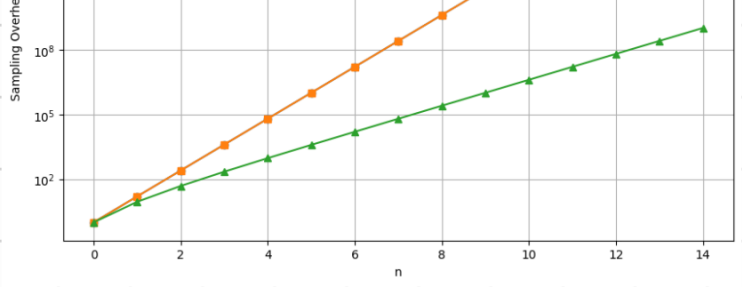
Parallel cut

Arbitrary cut









Comparison for the overhead sampling cost for both the parallel cut and arbitrary cut, with and without classical communicati

For some settings we considerably improve on the best previously known methods, for other settings we show that the best existing technique is optimal and hence can't be improved further.

For single wire cuts ($n=1$) classical communication can already strictly improve the optimal sampling overhead from 16 to 9.

without with

This kind of separation is not seen in gate cutting.
(this we will look at optimal gate cutting)

Also two separate wire cuts are more expensive than two wire cuts that are considered together. This is not seen when we do not allow classical communication

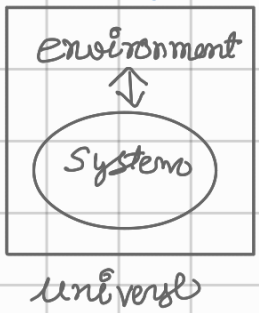
Preliminaries

linear maps for Hilbert spaces A, B : $\mathcal{L}(A, B)$ and we use the notation $\mathcal{L}(A) := \mathcal{L}(A, A)$ and thus the set of superoperators acting on the system A are given as $\mathcal{L}(\mathcal{L}(A))$

$\mathcal{L}(A, B)$	$\mathcal{L}(A, A)$	$\mathcal{L}(\mathcal{L}(A))$
linear maps from A to B	Linear maps of A onto itself.	set of superoperators acting on system A.

Density matrix / operator

developing the theory of open quantum system. Open quantum system means they can exchange energy and information with its environment.



Axioms for closed quantum systems

A state is a ray in Hilbert space.

$|\psi\rangle \in \mathcal{H}$ $\quad \langle\phi|\psi\rangle = \langle\psi|\phi\rangle^* \in \mathbb{C}$ $\quad 1 = \||\psi\rangle\|$

$\Rightarrow \sqrt{\langle\psi|\psi\rangle}$

Global phase do not matter

$$|\psi\rangle \equiv e^{i\alpha}|\psi\rangle$$

but

$$a|\phi\rangle + b|\psi\rangle \neq a|\phi\rangle + be^{i\alpha}|\psi\rangle$$

Multiplying a state vector by an overall phase has no physical effect but changing the phase in superposition of two state vectors is physically meaningful.

An observable is a self-adjoint operator on Hilbert space.

$$A : \mathcal{H} \to \mathcal{H}, \quad A = A^+ \qquad \langle\phi|A|\psi\rangle = \langle A^+\phi|\psi\rangle$$

and these operators can be diagonalized. (spectral decomposition)

$$E_n E_m = \delta_{mn} E_m$$

$$A = \sum_n a_n E_n$$

a_n: eigen value; E_n: orthogonal projector.

Probabilities of measurement outcomes are determined by the "Born rule".

$$\text{Prob}(a_n) = \|E_n|\psi\rangle\|^2 \Rightarrow \langle\psi|E_n|\psi\rangle$$

Post measurement the state is $|\psi\rangle = \dfrac{E_n|\psi\rangle}{\|E_n|\psi\rangle\|}$

If we measure it again right away we get the same outcome a second time.

Expectation value of the measurement outcome:

$$\langle A\rangle = \sum_n a_n \text{Prob}(a_n) \Rightarrow \langle\psi|H|\psi\rangle$$

Time evolution is determined by the Schroedinger eqn

$$\frac{d}{dt}|\psi(t)\rangle = -iH(t)|\psi(t)\rangle$$

The evolution proceeds via a sequence of infinitesimal unitary operators

$$\begin{aligned}|\psi(t+dt)\rangle &= (\mathbb{I} - iH(t)dt)|\psi(t)\rangle \\ &= e^{-iH(t)dt}|\psi(t)\rangle \\ &\Rightarrow U(t+dt, t)|\psi(t)\rangle\end{aligned}$$

The Hilbert space of a composite system AB is the tensor product of A and B

$$\mathcal{H}_{AB} = \mathcal{H}_A \otimes \mathcal{H}_B$$

dimension $d_A d_B$, orthonormal basis $\{|i\rangle_A \otimes |i\rangle_B\}$

These axioms make a distinction between evolution, which is deterministic, and measurement, which is probabilistic.

Is a qubit any different than a flipped coin which comes up either heads or tails? Yes, because while there is only one way to look at a bit

But there are many different ways to look at a qubit.

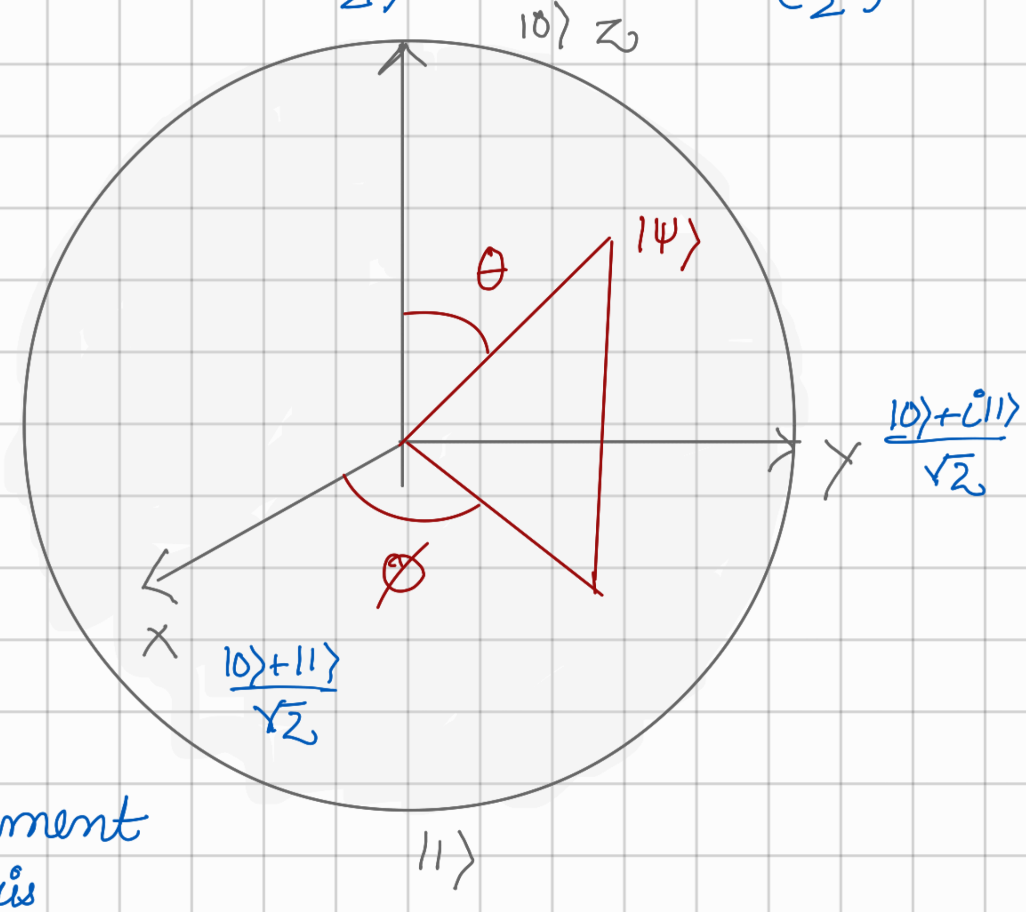
$I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ $X = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ $Y = \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix}$ $Z = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$ Eigenvalues $\{\pm 1\}$

General state of a qubit

$$|\psi(\theta, \phi)\rangle = e^{-i\phi/2} \cos\left(\frac{\theta}{2}\right)|0\rangle + e^{i\phi/2} \sin\left(\frac{\theta}{2}\right)|1\rangle$$

$\theta \in [0, \pi]$

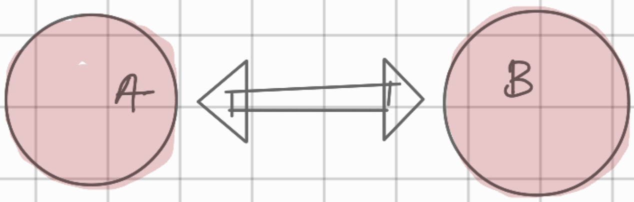
$\phi \in [0, 2\pi)$



Although a measurement along the z-axis on the Bloch sphere does not yield a definite outcome, a measurement along a different appropriately chosen axis does yield a definite outcome.

$\langle \sigma_1 \rangle = \sin\theta \cos\phi$ $\langle \sigma_2 \rangle = \sin\theta \sin\phi$ $\langle \sigma_3 \rangle = \cos\theta$

Open Quantum systems



Suppose we don't have access to system B, we can only observe A

Suppose we don't have access to system B, we can only observe A.

$$|\psi\rangle_{AB} = a|0\rangle_A \otimes |0\rangle_B + b|1\rangle_A \otimes |1\rangle_B$$

$$\Rightarrow \quad a|00\rangle + b|11\rangle$$

What if system B were measured in the z-basis. That is, the measured observable is $I_A \otimes Z_B$. According to our measurement axioms...

$$|0\rangle_A \otimes |0\rangle_B \Rightarrow \quad \text{Prob} = |a|^2$$
$$|1\rangle_A \otimes |1\rangle_B \Rightarrow \quad \text{Prob} = |b|^2$$

The post measurement state is a correlated state of A and B. If we observe only A, then

$$\text{Prob}(|0\rangle_A) = |a|^2 \qquad \text{Prob}(|1\rangle_A) = |b|^2$$

Consider a general Hermitian operator measured by $A = M_A \otimes I_B$

$${}_{AB}\langle\psi| M_A \otimes I_B |\psi\rangle_{AB} = \left(a^*\langle 00| + b^*\langle 11|\right) M_A \otimes I_B \left(a|00\rangle + b|11\rangle\right)$$

$$\Rightarrow \quad |a|^2 \langle 0| M_A |0\rangle + |b|^2 \langle 1| M_A |1\rangle$$

since $|0\rangle$ and $|1\rangle$ are orthogonal states of B.

we may write

$${}_{AB}\langle\psi| M_A \otimes I_B |\psi\rangle_{AB} = \text{tr}(M_A \rho_A)$$

where

$$\rho_A = |a|^2 |0\rangle\langle 0| + |b|^2 |1\rangle\langle 1|$$

ρ_A is a complete description

ρ_A is a characterisation of Alice's system, when that

→ density operator.

system is part of a composite system.

This has a natural "ensemble interpretation".

Example: $|\psi\rangle_{AB} = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle) \rightarrow \rho_A = \frac{1}{2}(|0\rangle\langle 0| + |1\rangle\langle 1|)$

$\Rightarrow \frac{1}{2} I$

then $\text{tr}(\vec{\sigma}\cdot\hat{n})\rho_A = 0$

Along any axis spin up and spin down occur with equal probability.

Now for a more general state of AB

$$|\psi\rangle_{AB} = \sum_{i,\mu} a_{i\mu} |i\rangle_A \otimes |\mu\rangle_B \,, \quad \sum_{i,\mu} |a_{i\mu}|^2 = 1$$

consider a general Hermitian operator measured by A: $A = M_A \otimes I_B$

$$\langle \psi | M_A \otimes I_B | \psi \rangle_{AB} \Rightarrow \sum_{j,\nu} a^*_{j\nu} (\langle j| \otimes \langle \nu|)(M_A \otimes I_B) \sum_{i,\mu} a_{i\mu} (|i\rangle \otimes |\mu\rangle)$$

$$\Rightarrow \sum_{i,j,\mu} a^*_{j\mu} a_{i\mu} \langle j | M_A | i \rangle$$

we can write this as:

$$\langle \psi | M_A \otimes I_B | \psi \rangle_{AB} = \text{tr}(\rho_A M_A) \quad \text{where}$$

$$\rho_A = \sum_{i,j,\mu} a^*_{j\mu} a_{i\mu} |i\rangle\langle j| \equiv \text{tr}_B(|\psi\rangle\langle\psi|)$$

The density operator of A is obtained by taking the partial trace over B of the projector $(|\psi\rangle\langle\psi|)_{AB}$

Think of a dual vector for B as a map from AB to A

$${}_B\langle\mu|: \mathcal{H}_B \to \mathcal{H}_A, \qquad |\mu\rangle_B : \mathcal{H}_B \to \mathcal{H}_A$$

$${}_B\langle\mu|\nu\rangle_{AB} = \delta_{\mu\nu}|i\rangle_A \cdot {}_{AB}\langle i,\nu|\mu\rangle_B = \delta_{\mu\nu}{}_A\langle i|$$

In general

$$tr_B(M_{AB}) = \sum_\mu {}_B\langle\mu|M_{AB}|\mu\rangle_B$$

we can evaluate the partial trace in any basis we like

In particular $\rho_A = \sum_\mu {}_B\langle\mu|\psi\rangle\langle\psi|\mu\rangle_B$

Properties of density operators

$$\rho_A = \sum_{i,j,\mu} a^*_{i\mu} a_{j\mu} |i\rangle\langle j| \equiv tr_B(|\psi\rangle\langle\psi|)$$

where $|\psi\rangle_{AB} = \sum a_{i\mu}|i,\mu\rangle_{AB}$

The density operator is Hermitian $\rho = \rho^\dagger$

The density operator is non negative $\langle\phi|\rho|\phi\rangle$

$$\Rightarrow \sum_{i,j,\mu} a^*_{j,\mu} a_{i,\mu} \langle\phi|i\rangle\langle j|\phi\rangle = \sum_\mu \sum_i |a_{i\mu}\langle\phi|i\rangle|^2 \geq 0$$

The density operator has unit trace

$$tr\,\rho = \sum |a_{i,\mu}|^2 = \||\psi\rangle_{AB}\|^2 = 1$$

Hence there is orthonormal basis in which the density operator is diagonal. The eigen values are non negative real numbers $\sum = 1$

$$\rho = \sum p_a |a\rangle\langle a| \cdot p_a \geq 0, \quad \sum_a p_a = 1$$

In the ensemble interpretation, the eigenvalue is probability that the corresponding basis state has been prepared in system A.

If there is just one non zero eigenvalue, we say that the state of A is pure. Otherwise it is mixed.

Schmidt decomposition of Bipartite system

Using the basis in which the density operator of A is diagonal. we can put the bipartite pure state in a standard form.

$$|\psi\rangle_{AB} = \sum a_{iu} |i\rangle_A \otimes |u\rangle_B \quad \text{and} \quad \rho_A = \sum p_i |i\rangle\langle i|$$

Let's write

$$|\psi\rangle_{AB} = \sum |i\rangle_A \otimes |\tilde{i}\rangle_B, \text{ where } |\tilde{i}\rangle = \sum_u a_{iu} |u\rangle_B$$

Then

$$\rho_A = \sum_i p_i |i\rangle\langle i| = \sum_{i,j} (|i\rangle\langle j|)_A \, tr_B(|\tilde{i}\rangle\langle \tilde{j}|)$$

$$\Rightarrow \sum_{i,j} (|i\rangle\langle j|)_A \langle \tilde{j}|\tilde{i}\rangle$$

Hence $\langle \tilde{j}|\tilde{i}\rangle = \delta_{ij} p_i \rightarrow |i\rangle_B = \frac{1}{\sqrt{p_i}} |\tilde{i}\rangle_B$ $(p_i > 0)$ are orthonormal vectors

and

$$|\psi\rangle_{AB} = \sum_i \sqrt{p_i} \, |i\rangle_A \otimes |i\rangle_B$$

This is the Schmidt decomposition of the bipartite state. The orthonormal bases are called Schmidt bases and the coefficient are called Schmidt coeffs.

Quasi probability Simulation

$\mathcal{L}(A,B)$ ~ set of linear maps, that map from A to B

$\mathcal{L}(\mathcal{L}(A))$ ~ set of superoperators acting on the system A

$CPTP(A) \subset \mathcal{L}(\mathcal{L}(A))$

$H(A)$ ~ set of Hermitian op.

$CPTN(A) \subset \mathcal{L}(\mathcal{L}(A))$ on A

Complete positive non decreasing

$S(A)$: set of density op on A

$\|\cdot\|$ ~ op norm

$\|\cdot\|_1$ trace norm

Let's think of a quantum computer (A) which is not universal, hence it has a unitary (U) that it cannot execute.

The channel $\mathcal{U}$ induced by the gate U lies outside the set of super operators $S \subset \mathcal{L}(\mathcal{L}(A))$ that can be achieved by the computer.

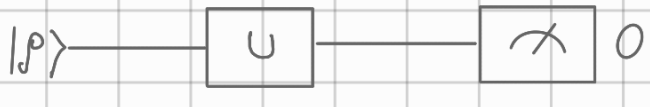
The goal of quasi probability simulation is to simulate the execution of the gate U while only having access to the operations that lie in S. It can do it at the cost of an additional sampling overhead.

The central ingredient for the quasi probability simulation is a quasi probability decomposition (QPD)

$$\mathcal{U} = \sum_i a_i F_i \quad (1)$$

$a_i \in \mathbb{R}$ F_i are local operations.

U gets randomly replaced by one of the gates F_i. The sampling overhead of this approach is $K^2 = (\sum_i |a_i|^2)$, so desirable to find a decomposition in the above eq. form that minimizes K



the ideal outcome from this circuit is the Expectation Value

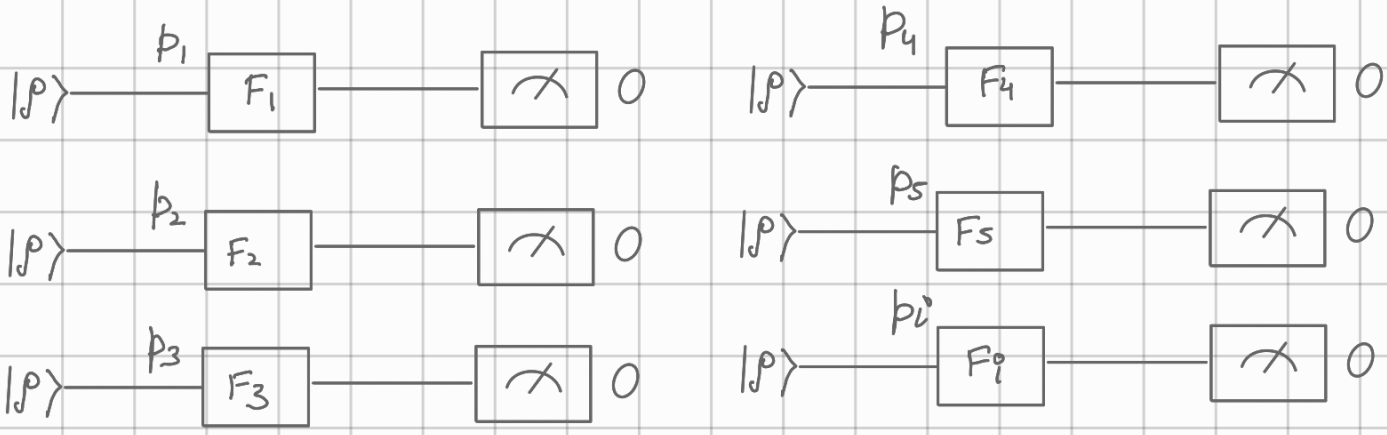
$$tr[O\,U(\rho)]$$

but using QPD

$$tr[O\,U(\rho)] = \sum_i p_i\, tr[O F_i(\rho)]\, \kappa\, sign(a_i) \quad -(ii)$$

where $\kappa := \sum_i |a_i|$ and $p_i := \frac{|a_i|}{\kappa}$

Eq(ii) give us a Monte Carlo approach to estimate the expectation value:



For each shot of the circuit, we randomly replace the unachievable gate U, with one of the achievable gates F_i with probability p_i and weight of the outcome with $\kappa\, sign(a_i)$. Here κ^2 is the sampling overhead.

To minimize the sampling overhead it is crucial to find an optimal QPD that minimizes the one-norm of the coefficients a_i.

$$\gamma_S(U) = \min\left\{ \sum_{i=1}^{m} |a_i| : m \geq 1,\ U = \sum_{i=1}^{m} a_i F_i,\ a_i \in \mathbb{R},\ F_i \in S \right\}$$

This γ-factor exactly captures the optimal sampling overhead required to simulate U using the operations in S.

When U is close to S, the $\gamma_S(U) \cong 1$ and thus sampling overhead is small. The farther away it is, the more complex and costly

it is to simulate $\mathcal{U}$ using operations in $\mathcal{S}$.

Using the individual optimal QPD for each gate leads to a simulating overhead of $\prod_{i=1}^{n} \gamma_{\mathcal{S}}(\mathcal{U}_i)^2$ for the whole circuit.

There are instances where combining optimal QPDs for the individual gates does not result into an optimal QPD for the complete circuit, i.e. local optimality does not necessarily imply global optimality.

Trace-non increasing and non-positive operations

The goal of quasi probability simulation is to simulate the execution of the gate $\mathcal{U}$ while only having access to the operations that lie in $\mathcal{S}$.

Quasi probability simulation can recover the expected value of the measurement outcomes of the circuit at the cost of an additional sampling overhead.

The central ingredient for the quasi probability simulation is a quasi probability decomposition (QPD)

$$\mathcal{U} = \sum_i a_i \mathcal{F}_i$$

The set of achievable operations $\mathcal{S}$ has to be a subset of trace-preserving completely positive maps. $\mathcal{S} \subset CPTP(A)$

but one can also make use of trace-non increasing maps

$\mathcal{S} \subset CPTN(A)$

because any trace-non increasing map can effectively be simulated by some measurement process and post selection of the corresponding measurement outcome.

$\mathcal{E} \in CPTN(A) \quad \exists\, \mathcal{F} \in CPTP(A)$

s.t

$\mathcal{E} + \mathcal{F} \in CPTP(A)$

To simulate $\mathcal{E}$ one can perform the trace preserving CP map

$$\rho_A \longmapsto \mathcal{E}(\rho)_A \otimes |0\rangle\langle 0|_E + \mathcal{F}(\rho)_A \otimes |1\rangle\langle 1|_E$$

This is done by multiplying the final outcome of the circuit by 0 in case the measurement outcome 1 is obtained.

Gate cutting and state preparation via Quasi probability simulation