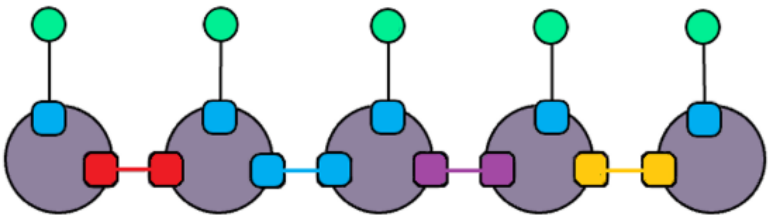
Entanglement Entropy of Many-body Quantum Systems in the Age of Quantum Information and Computing with Tensor Networks



Hisham Ashraf Amer

Physics department

University of Science and Tehcnology at Zewail City

Submitted on June 2020

in partial fulfillment of the requirements for B.Sc in Physics

Supervisor: Dr.Ali Nassar

Entanglement Entropy of Many-body Quantum Systems in the Age of Quantum Information and Computing with Tensor Networks

Hisham Ashraf Amer

Abstract

Quantum information has seen a recent surge in its popularity with the strides being made in upgrading the computational capacity of quantum computers. One concept at the helm of this movement is entanglement, which is widely considered the cornerstone of quantum information. However, since the Hilbert space in many-body quantum systems experiences exponential growth with system size, entanglement quantification and utilization become a difficult task. In this study, we started by investigating how one can quantify entanglement using concepts like von Neumann entropy and mutual information. Next, we looked at how one can circumvent this limitation of an exponentially growing Hilbert space, through referring to area laws of entanglement entropy and correlation length. We discuss the relevance of these laws here, and eventually introduced tensor networks, a formulation which is exceptionally well suited for their realization. Tensor networks, specifically matrix product states (MPS's), offer a means for breaking down the exponentially large wavefunction state vectors into manageable smaller tensors. Following the introduction of some basic mathematical concepts needed for tackling entanglement entropy, including density matrices; partial traces; pure and mixed bipartite and multipartite quantum systems; we ease the reader into identifying and quantizing entanglement in multipartite quantum systems. Accordingly, we were able to utilize an advanced tensor network library called ITensor, which we used to implement the Density Matrix Renormalization Group (DMRG) on a 20-body half-spin MPS chain, under the influence of the Heisenberg Hamiltonian, to obtain a ground state MPS approximation. Single value decomposition (SVD) was used on the resulting 20-unit MPS after it was gauged to the 4^{th} bond, which in turn sets up a bipartite system around the bond center. The decomposition was used to extract Schmidt coefficients for the von Neumann entanglement entropy calculation. Finally, the one-to-one mapping between tensor networks and quantum circuits was addressed, where we highlight how MPS's with MPO's are analogous to quantum circuits. For this, we design a trivial MPS quantum circuit for generating a 4-qubit GHZ state. The MPS-inspired circuit is then Implemented on a 5-qubit IBM quantum computer, then on the qasm simulator with noise and error correction, after which quantum state tomography and error correction circuits were used to infer the pre-measured density matrix. The tomography generated density matrix was later directly compared to the ideal GHZ state results through extracting entanglement entropies and mutual information from their reduced density matrices, in an attempt to infer the loss of information following decoherence. The mutual information of the noisy, quantum circuit generated, GHZ state experienced $\sim 10.6\%$ reduction in correlations with error correction circuits.

Content:

Introduction:

With the advent of increasingly efficient quantum devices and the ongoing success in utilizing quantum effects, quantum information has become one of the most promising areas of research in physics, being a synthesis of three key themes of the century: quantum physics, computer science, and information theory [1], [2]. An indispensable quantum phenomenon that governs many of the models in quantum information is entanglement. Entanglement is widely considered the cornerstone of quantum information, being a vital resource for most revolutionary quantum effects, including quantum teleportation[3] , quantum cryptography[4] , quantum computing[5], and has successfully bridged quantum information with the application oriented field of many-body condensed matter physics[1]. Entanglement in many-body quantum systems [1], [6] gets exponentially challenging to study, both experimentally and theoretically, due to the exponential growth of quantum degrees of freedom. Experimentally, detecting and quantifying entanglement is difficult to achieve, whilst theoretically, the language used to fully describe quantum systems is based on state functions, which get exponentially complicated to base models on, build algorithms and run computations for, or even simulate. One way to overcome these obstacles, has been to incorporate quantum information techniques in condensed matter physics, in particular, techniques that can overcome the inherent sophistication of exponentially growing degrees of freedom. When successful, these techniques will usher in a new revolution in physics, by unifying the practicality of many-body quantum systems that govern matter phenomena with the theoretical aptitude of information theory[1].

An immediate question when dealing with quantum information is how can one quantify, classify and assess the physical quantum phenomenon of entanglement using some mathematical equivalence. One way is through relating it to entropy, and in turn we project informational value onto entangled states. This is achieved through extending the well-established classical Shannon entropy to quantum microstates, giving rise to von Neumann entropy[2], [5], one of the most information-oriented approaches to quantifying entanglement, which allows one to mathematically reference an “entanglement entropy”.

As one delves into entanglement entropy in quantum matter, they will quickly need to address the exponentially growing quantum many-body Hilbert space spanned by the system’s possible microstates.[7] Dirac, being as insightful as he was, was able to pinpoint the inherent limitation of this groundbreaking wavefunction description of the quantum system as he pointed our quite clearly that:

“The underlying physical laws necessary for the mathematical theory of a large part of physics and the whole of chemistry are thus completely known, and the difficulty is only that the exact application of these laws leads to equations much too complicated to be soluble. It therefore becomes desirable that approximate practical methods of applying quantum mechanics should be developed, which can lead to an explanation of the main features of complex atomic systems without too much computation.”[8]

A quantum wavefunction of spin-half particles assigns an amplitude to each classical state, so we get 2^N states given N sites, say sites on a lattice. This gives rise to a spin state at every site, luckily spin states being binary, are realizations of qubits, which are either, spin up ↑, or spin down ↓, mimicking a computational basis $\{0,1\}$. To better grasp the extent of this limitation, we’ll emphasize that for $N > 265$, the number of amplitudes might exceed the number of atoms in the known universe [9], which some estimates make to be around 10^{80}. This makes storing this much info practically impossible, as we

would lack the number of particles to code for the information, even if we were to efficiently carry a single bit of info on a single atom. So, we are forced to ask the question, must we know all the amplitudes of a wavefunction to describe the physics of a system? Which would be a huge problem, or can we somehow identify some underlying aspects of the system we can target, ones that are "more relevant" to setting up most of the physics of a system; or in other words, ones that encode enough of the information in a system to allow one to generate acceptable approximate models? Restating this in computational terms, the problem is that when calculating certain functions of ***multipartite quantum systems***, the running times of classical computers will increase exponentially with growing partitions. A workaround is to identify alternatives to the brute forcing of increasingly complex calculations. The Alternatives must allow us to extract approximate information about quantum systems within the physical confines of classical computing power[5], [10], [11]. Among the methods used to approximate solutions today, are quantum Monte Carlo (MC) simulations, mean field theory approximations and density functional theory (DFT).

That said, there do seem to be laws, specifically area laws of entanglement entropy, that govern the way particles usually interact in a lattice. A large branch of quantum-matter in nature, in particular ones with finite range interactions, can be effectively modelled in accordance with these laws. This locality of interactions forces the physics of this matter into being governed by a small part of its Hilbert space. This manifold is of states which are the low energy eigenstates of gapped local Hamiltonians, states which abide by the area law of entanglement entropy. Consequently, it seems that the nature of these quantum many-body systems, is lurking in some small corner of Hilbert space, one that can be essentially targeted if we are to efficiently study these groups of quantum matter occurring regularly in Nature. This is the motivation behind one of the most promising tools developed to date, and the one that relates most to quantum information theory, Tensor Network states (TNS), the most common of which being the Matrix Product States (MPS's) for 1-D models and projected entangled pair states (PEPS) for 2-D models. These tools utilize area laws of entanglement entropy to target the "important states", then expresses itself in a mathematical tensorial language perfectly suited for computation. This makes TNS effective tools for bridging experimental and physically driven quantum matter with the computationally demanding and flavored quantum information theory. Tensor networks have proven to be fundamental to understanding quantum states of many-body systems, and form the groundwork for many specialized techniques and algorithms, including the Density Matrix Renormalization Group (DMRG), MERA networks, PEPS and others[7], [9], [12], [13].

That said, our goal here is to help the reader bridge the worlds of many-body quantum systems and information theory. We first discuss the mathematical basis needed to approach the topic, including density matrices and the purity/mixedness of states; Multipartite systems and partial traces in compound Hilbert spaces; and the mathematical realization of correlations and entanglement. We also intend on easing the reader into recognizing the relevance of entanglement and multipartite quantum systems, (focusing mainly on pure bipartite states) to quantum information. Secondly, we discuss the quantification of entanglement, starting with the transition from classical Shannon entropy onto the von Neumann entropy, followed by addressing several quantifiers of the informational value of entanglement and entanglement entropy, including conditional, relative and mutual entropies. Thirdly we introduce tensor networks, and emphasize the utility and advantages of their graphical and mathematical tensorial nature. Emphasis will be given to MPS's owing to its current prominence and success in modelling 1-D systems. An advanced tensor network library called ITensor [14] was used to implement DMRG on a 20-body spin-half MPS chain under the Heisenberg Hamiltonian to obtain a ground state MPS approximation. Single

value decomposition (SVD) was used on the resulting 20-unit MPS after it was gauged to the 4^{th} bond, which in turn sets up a bipartite system around the bond center. The decomposition was used to extract Schmidt coefficients for the von Neumann entanglement entropy calculation. Finally, the one-to-one mapping between tensor networks and quantum circuits will be addressed, where we illustrate the how MPS's and MPO's are analogous to quantum circuits, and apply this to a quantum circuit design for generating the MPS realizable 4-qubit GHZ state. The MPS-inspired circuit is then Implemented on a 5-qubit IBM quantum computer, then simulated with noise and exposed to quantum state tomography and error correction circuits for inferring a pre-measurement density matrix. In turn, the density matrices are directly compared to the ideal GHZ state through extracting entanglement entropies and mutual information from their reduced density matrices as a final gesture to the realizability and relevance of the theories and models we've discussed.

1.0.0 Section 1: Some Mathematical pillars

To study entanglement entropy, we first need to erect some mathematical scaffolding on which we can mold the topic. This involves familiarizing oneself with qubit representations and formulation, density matrices; mixedness of states; partial traces and multipartite systems; coherences; reduced density matrices and Schmidt decompositions.

1.1.0 A qubit and the Bloch Sphere [1]:

We begin by introducing the simplest quantum system, a ***qubit***. A qubit is a normalized, ***2-D complex vector*** in a Hilbert space, and so $\in \mathbb{C}^2$. We can therefore define any qubit state with 2 basis vectors, usually $|0\rangle$ and $|1\rangle$, called the "***computational basis***", giving an arbitrary qubit state $|\psi\rangle$ the form

$$|\psi\rangle = a|0\rangle + b|1\rangle = \begin{pmatrix} a \\ b \end{pmatrix} \tag{1}$$

With normalization $|a|^2 + |b|^2 = 1$. One should have immediately noticed that an obvious candidate for realizing a qubit is the electron spin state with its 2 spins. The advantage of making this relation, to a physicist new to quantum information theory, is that one whom is familiar with spin algebra will quickly familiarize with the mathematics surrounding qubits. Taking inspiration from spin, we refer to the identity

$$\hat{S}_n = \bar{\sigma} \cdot \bar{n} \tag{2}$$

With $\hat{S}_n$ being an arbitrary spin operator, $\bar{\sigma}$ the 3-component vector with the 3 Pauli matrices as its x, y and z components, and $\bar{n}$ being an arbitrary spin direction. Now, given all spin operators are related by unitary transformations, which are known to preserve eigenvalues, we know the eigenvalues for any spin

matrix will be $\frac{\hbar}{2}$. By solving typical eigenvalue problems, we identify that the eigenvectors for the most general spin operator in equation (2) are

$$|n_+\rangle = \begin{pmatrix} \cos\left(\frac{\theta}{2}\right) \\ e^{i\varphi}\sin\left(\frac{\theta}{2}\right) \end{pmatrix}, \ |n_-\rangle = \begin{pmatrix} -e^{-i\varphi}\sin\left(\frac{\theta}{2}\right) \\ \cos\left(\frac{\theta}{2}\right) \end{pmatrix} \tag{3}$$

The 2 basis vectors in equation (3), present us with the opportunity to achieve a ***normalized parametrization***, of arbitrary 2 component spinors, into 3-D real space, with emphasis given to their normalization, since normalization will be what fixes this representation onto the surface of a unit sphere. With this representation we are able to visualize, the single qubit state despite it having imaginary components. Using the identity in equation (3) we can identify potential parametrization in θ and φ. Accordingly, we can visualize ANY single qubit state by defining a point $\bar{n}$ (3 component vector) in $\mathbb{R}^3$ on a unit sphere. For this to work, the qubit will be parametrized by 2 angles θ and φ by assigning

$$a = \cos\left(\frac{\theta}{2}\right), b = e^{i\varphi}\sin\left(\frac{\theta}{2}\right) \tag{4}$$

This applies up to any global phase $e^{i\chi}$, since it leaves the normalization unchanged, and the state physically equivalent for any quantum mechanical observation. Now if we take $|n_+\rangle$ to be the z_+ axis, this means the z_- is simply $|n_-\rangle$ since the trig parameter is $\frac{\theta}{2}$, and therefore orthogonal axes will be separated by π, not the usual $\frac{\pi}{2}$, as is the usual case with cartesian x, y, z axes. This makes this sphere very non intuitive, since we can clearly see by design that the state $|n_+\rangle$ and $|n_-\rangle$ are orthogonal, but are visualized on the sphere as parallel, this is just something we have to live with. That said, this representation is called a "Bloch Sphere".

One consequence of the Bloch sphere representation, is that the vector $\bar{n}$ has components

$$n_x = sin\theta cos\varphi, \qquad n_y = sin\theta sin\varphi, \qquad n_z = cos\theta \tag{5}$$

Knowing our spin half operators, we clearly see that these are just the expectation values of the spin operators in the state $|n\rangle$ so

$$n_x = \langle\sigma_x\rangle, \qquad n_y = \langle\sigma_y\rangle, \qquad n_z = \langle\sigma_z\rangle \tag{6}$$

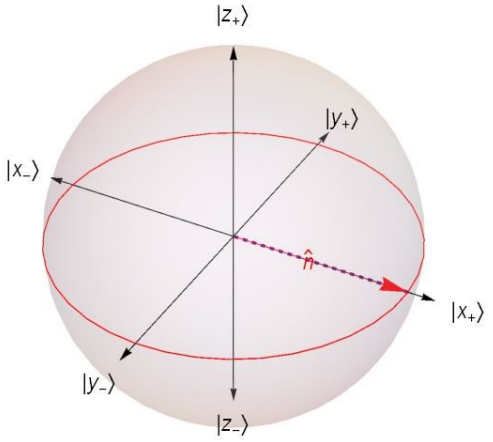
This implies we can use the same Pauli matrices spin algebra to work with $\bar{n}$ in a Bloch sphere as we are used to doing with spin algebra, this will come in handy in mixed state representations on the Bloch sphere later on.

Why Density matrices? It might be surprising to some that the state vector on its own, expressed by a ket $|\psi_i\rangle$, is not the most instructive way of describing a quantum state. To illustrate further, consider how the spin half particle pointing in an arbitrary direction has a spin state always points somewhere, as is evident in the "***Bloch sphere***" formalism, such that the state points in the direction $\bar{n}$ of the sphere. Therefore, we can always match a specific position to a quantum spin state. Even when a global state exists as a superposition of our 2 available states, it is still a definite superposed state with a clear direction on the Bloch sphere. Take the $|x_+\rangle$ state, which has many possible representations in the z basis:

$$|x_+\rangle \equiv \frac{1}{\sqrt{2}}\binom{1}{1} \equiv \frac{1}{\sqrt{2}}|z_+\rangle + \frac{1}{\sqrt{2}}|z_-\rangle \equiv \frac{1}{\sqrt{2}}|0\rangle + \frac{1}{\sqrt{2}}|1\rangle \ , \tag{7}$$

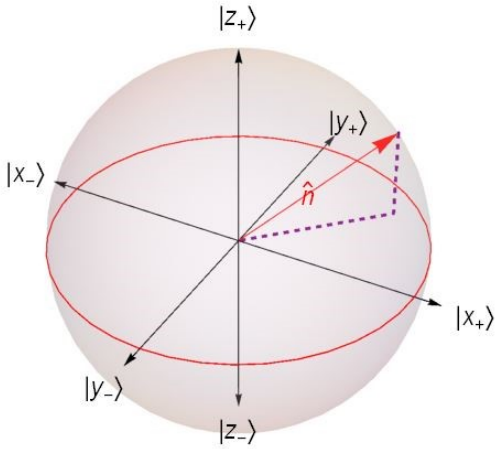
Notice how it illustrates a superposition between $|z_+\rangle$ and $|z_-\rangle$, yet is still specifically parametrized on a Bloch sphere by the angles $\theta = \frac{\pi}{2};\ \varphi = 0, 2\pi$.

A short simple Mathematica code, inspired by reference [15] , was written to generate a dynamic Bloch sphere that can be placed into any possible qubit state through manipulating θ and φ. It will be demonstrated with 2 examples in the figures (1) and (2). My full code can be found in Appendix A.



, $\psi = 0.7071\,|0\rangle + 0.7071\,|1\rangle$

Figure (1): an $|x_+\rangle = \frac{1}{\sqrt{2}}|z_+\rangle + \frac{1}{\sqrt{2}}|z_-\rangle$ state depicted on the Bloch sphere.



, $\psi = 0.8686\,|0\rangle + (0.3503 + 0.3503\,i)\,|1\rangle$

Figure (2): a more arbitrary state $0.8686\ |z_+\rangle + (0.3503 + 0.3503\ i)\ |z_-\rangle$ on the Bloch sphere

Now consider putting the spin in a thermal bath at equilibrium, that should eliminate the tendency for the spin to align along one preferred magnetization direction, i.e. it will randomize the spin, but didn't we just state that it was impossible to identify a spin state that did not point in a specific direction, so how would we explain the ambiguity introduced by this randomization? Simply put, coming in contact with an environment, in this case the thermal bath, adds a "***classical uncertainty***" to the problem, such that given a ket and nothing else, we would fail to give a complete description of our system, since a single ket has no means of conveying classical uncertainty, as a lone ket can only encompass "***quantum uncertainty***". Consequently, it was worthwhile to devise a new tool that more generally describes a quantum system. This motivated the creation of the "***density matrix" $\widehat{\boldsymbol{\rho}}$***.

1.2.0 *The Density Matrix Formulation* [5], [10], [11], [16]*:*

One way to introduce density matrices, is through looking at an apparatus producing an ensemble of random states. Take a look at an instrument preparing quantum systems. Since the apparatus is imperfect, it will tend to produce a "***mixture***" of states. To illustrate this, consider a 3 qubit ensemble for simplicity, the source produces state $|0\rangle$ with a probability of 0.5, state $|1\rangle$ with a probability 0f 0.25 and state $|2\rangle$ with probability 0.25, (this can be experimentally determined through monitoring the instrument as it functions). It must be emphasized that these are not superpositions, rather classical mixtures, which impart "***classical uncertainty***" to a system. Now, since these are classical probabilities, say q_i, so for any number N of q_i states:

$$\sum_i q_i = 1 \qquad ; \qquad where\ q_i \in [0,1] \tag{8}$$

But how do we define expectation values, i.e. make measurements, now that we've incorporated both classical and quantum uncertainties? Given an observable A and eigenstates $|0\rangle$, $|1\rangle$, and $|2\rangle$, we know that the expectation value of A will be $\langle 0|A|0\rangle$ if the state is $|0\rangle$, $\langle 1|A|1\rangle$ if it's $|1\rangle$, and $\langle 2|A|2\rangle$ if $|2\rangle$, but what about the fact that our system is a "classical mixture" of $|0\rangle$, $|1\rangle$, and $|2\rangle$, with predefined classical probabilities. We therefore re-define the value of observable A as a "classical average of quantum averages", so given the different pure states $|S_i\rangle$ constituting the full mixture, the average value of observable A becomes

$$\langle \hat{A} \rangle = \sum_i q_i \langle S_i|A|S_i\rangle \tag{9}$$

By referring to the trace of an operator expressed in the orthogonal full set of eigenbasis $\{\psi_i\}$, the trace is given by

$$tr(\hat{A}) = \sum_i \langle \psi_i | A | \psi_i \rangle \tag{10}$$

and given traces are basis independent, we can choose the basis which diagonalizes the normal matrix $\hat{A}$, so the trace simply becomes the sum of diagonal entries. Realizing that diagonal entries of diagonal matrix are also the eigenvalues λ_i of $\hat{A}$ the trace becomes

$$tr(\hat{A}) = \sum_i \lambda_i \tag{11}$$

Making it unique for a given normal matrix $\hat{A}$ regardless of representation. Now Remember that an arbitrary normal matrix $\hat{A}$ can be written as some $|k\rangle\langle b|$, so

$$\mathrm{tr}\,(\hat{A}) = \mathrm{tr}\,(|k\rangle\langle b|) = \sum_i \langle \psi_i | k \rangle \langle b | \psi_i \rangle \tag{12}$$

Since expectation values are numbers, they commute, so

$$\mathrm{tr}\,(|k\rangle\langle b|) = \sum_i \langle b | \psi_i \rangle \langle \psi_i | k \rangle \tag{13}$$

From completeness we know that

$$\sum_i |\psi_i\rangle\langle\psi_i| = \mathbb{1} \tag{14}$$

Using equation (13) and the completeness in equation (14) we can identify that

$$tr\,(|k\rangle\langle b|) = \langle b | k \rangle \tag{15}$$

One way to naively cheat through this mathematical rigor is by considering cyclicity such that $tr\,(|k\rangle\langle b|) = tr\langle b|k\rangle$ which is just the trace of a number $\langle b|k\rangle$, which returns that number again.

This motivates us to write

$$\langle S_i | A | S_i \rangle = tr(A | S_i \rangle\langle S_i|) \tag{16}$$

This holds even if the states $|S_i\rangle$ are superpositions of appropriately weighted orthogonal eigenstates, and since the trace is a linear operator and so $tr\,(\,|a\rangle\,\langle a| + \,|b\rangle\langle b|) = tr\,(\,|a\rangle\,\langle a|) + \,tr(\,|b\rangle\langle b|)$, Applying this to equation (9) gives:

$$\langle \hat{A} \rangle = \sum_i q_i \, tr(\,A|S_i\rangle\langle S_i|) = tr\left(A \sum_i q_i \,\, |S_i\rangle\langle S_i|\right) \tag{17}$$

We can now identify the object $\hat{\boldsymbol{\rho}}$, the ***Density Matrix***:

$$\hat{\rho} = \sum_i q_i \,\, |S_i\rangle\langle S_i| \tag{18}$$

Such that

$$\langle \hat{A} \rangle = tr(A\rho) = tr(\rho A) \tag{19}$$

Using this relation, we have identified a means of making ***"measurements"*** in the context of a system with both classical and quantum uncertainty. Using the density matrix, we can recast all of quantum mechanics in this language instead of kets.

It is important to note that this was just one way of perceiving the density matrices, casting it as a bridge between classical and quantum pictures, however there is one other way to further apply it to quantum information theory, and it turns out that it is through the purely quantum phenomenon of entanglement. In a few sections we will introduce entanglement and define what role density matrices play there. To address the most basic utility of density matrices, we cast the concepts of mixed and pure states in terms of them [10].

1.2.1 The projection operator and expectation values in Pure states:

A pure state is one which can be represented by a definite wave function, meaning it has a precise location in some D-dimensional Hilbert space. Although pure states can still be expressed through a density matrix, a ket can still hold all the info regarding the state of the entire ensemble. A mixed state however cannot be described by a standalone wave function, since there will always be some ambiguity regarding the proportions of different states in the mixture, and to deal with the ambiguity we need density matrices.

First, we define the closely related operator, the ***"i^{th} eigenbasis projection operator"*** on the $\hat{\mathrm{P}}_i$, that extracts the projection of the arbitrary state on the eigenbasis $\{|\psi_i\}$

$$\hat{\mathrm{P}}_i = |\psi_i\rangle\langle\psi_i| \tag{20}$$

This is the quantum mechanical (QM) mathematical equivalent of the probability of "observing" or "measuring" some eigenstate, and as one can see from equation (20) ***measurements must take place in some basis*** $\{|\psi_i\rangle\}$, and so defining a measurement in QM dictates we specify the basis for it.

To set up an example, let $|S\rangle$ be a pure state, and the state space have a set of eigenbasis $\{|\psi_i\rangle\}$, so $|S\rangle$ can be written as

$$|S\rangle = \sum_i c_i|\psi_i\rangle \tag{21}$$

Let our operator/observable in question, be $\hat{A}$ with eigenvalues a_i in the $|\psi_i\rangle$ basis, so

$$\hat{A}|\psi_i\rangle = a_i|\psi_i\rangle \quad \text{and} \quad c_i = \langle\psi_i|S\rangle \tag{22}$$

The probability $\mathscr{p}_i$ of measuring the eigenvalue a_i, by observing $\hat{A}$ on our pure ensemble, would be

$$\mathscr{p}_i = |c_i|^2 \tag{23}$$

Using equations (23), (22) and (20) we can show that

$$\mathscr{p}_i = |c_i|^2 = \langle S|\psi_i\rangle\langle\psi_i|S\rangle = \langle S|\mathrm{P}_i|S\rangle = \langle\hat{\mathrm{P}}_i\rangle \tag{24}$$

Using the reverse operations on $\langle\hat{A}\rangle$ where $\langle\hat{A}\rangle = \sum_i\langle\hat{\mathrm{P}}_i\rangle\, a_i$ we find that

$$\langle\hat{A}\rangle = \sum_i\langle S|\psi_i\rangle\langle\psi_i|S\rangle\, a_i = \sum_i\langle\psi_i|S\rangle\langle S|\psi_i\rangle\, a_i = \sum_i\langle\psi_i|S\rangle\,\langle S|A|\psi_i\rangle \tag{25}$$

By letting $T = |S\rangle\langle S|$, we can write equation (25) as

$$\langle\hat{A}\rangle = \sum_i\langle\psi_i|TA|\psi_i\rangle \tag{26}$$

Which from equation (9) we can identify as trace of (TA) and is thus also cyclic so

$$\langle \hat{A} \rangle = tr(TA) = tr(AT) \tag{27}$$

This is where the utility of the projection operator comes in. Since we set no conditions on $\hat{A}$, it's an arbitrary operator, we can choose this operator to be $\hat{\mathrm{P}}_i$, so

$$\langle \hat{\mathrm{P}}_i \rangle = tr(T\mathrm{P}_i) = tr(\mathrm{P}_i T) = \mathscr{p}_i \tag{28}$$

This means we can find the probability of the state $|S\rangle$ being in the i^{th} eigenstate by using the T operator $\hat{T} = |S\rangle\langle S|$ and the projection operator of the arbitrary state on the i^{th} eigenbasis, $\hat{\mathrm{P}}_i = |\psi_i\rangle\langle\psi_i|$, and calculating $tr(T\mathrm{P}_i)$. ***Diagonal terms***, which hold the info regarding the probabilities, are conveniently called ***populations***, the distinction between them and off diagonal terms is to be discussed in a moment.

1.2.2 Mixed states and the utility of density matrices in defining expectation values

We can repeat the same analysis for a "mixed ensemble" of particles where each one has the probability $\mathscr{p}_{S^m}$ of being in a different state $|S^m\rangle$, we get that for mixed states the operator/observable $\langle \hat{A} \rangle$ has a unique expectation $\langle \hat{A} \rangle^i$ value for every unique member of the mixed states:

$$\langle \hat{A} \rangle = \sum_m \mathscr{p}_{S^m} \langle \hat{A} \rangle^m \tag{29}$$

And similar to equation (21), we still identify that

$$|S^m\rangle = \sum_i (c_i)^m |\psi_i\rangle \tag{30}$$

$$\langle \hat{A} \rangle^m = \langle S^m | A | S^m \rangle \tag{31}$$

$$\mathscr{p}_i = \sum_m \mathscr{p}_{S^m} \langle \mathrm{P}_i \rangle^m$$

Where

$$\langle \mathrm{P}_i \rangle^m = (\mathscr{p}_i)^m = |(c_i)^m|^2 = \langle \psi_i | S^m \rangle \langle S^m | \psi_i \rangle \tag{32}$$

So, we weight the probability of observing an i^{th} eigenstate in every single m^{th} compound state in the mixture and sum up all the individual averages over all available total states to get the average of the mixture. We now define $\langle \hat{A} \rangle^m$ as we did for pure states, to find the expectation value of $\hat{F}$ in the state $|S^m\rangle$ explicitly, taking equation (31):

$$\langle \hat{A} \rangle^m = \langle S^m|A|S^m\rangle = \sum_i |(c_i)^m|^2 \, a_i = \sum_i \langle \psi_i|S^m\rangle\langle S^m|\psi_i\rangle \, a_i \tag{33}$$

Subbing back into equation (29) gives

$$\langle \hat{A} \rangle = \sum_m \mathscr{p}_{s^m} \langle \hat{A} \rangle^m = \sum_m \mathscr{p}_{s^m} \sum_i \langle \psi_i|S^m\rangle\langle S^m|\psi_i\rangle \, a_i = \sum_i \langle \psi_i| \sum_m \mathscr{p}_{s^m} |S^m\rangle\langle S^m|A|\psi_i\rangle \tag{34}$$

Like we did in the pure state example we can identify the matrix $T = \sum_m \mathscr{p}_{s^m} |S\rangle\langle S|$, which is exactly the ***density matrix*** we discussed in section (…x…….) using our simpler argument. So, we can now safely say that the ***most general form of the density matrix*** is

$$\hat{\rho} = \sum_m \mathscr{p}_{s^m} |S^m\rangle\langle S^m| \tag{35}$$

We can now deduce that the T matrix we extracted for the pure state, is actually the most trivial density matrix, one whose index m runs over a single value, and cuts the probabilities $\{\mathscr{p}_{s^m}\}$ of finding different mixed states into a single certainty $\{\mathscr{p}_{s^m}\} = \{1\}$, and we find that the $\hat{T}$ we identified was actually ρ_{pure}. Now we get

$$\langle \hat{A} \rangle = \sum_i \langle \psi_i|\rho A|\psi_i\rangle = tr(\rho A) = tr(A\rho) \tag{36}$$

What we have here in equation (36) is a powerful tool, in the ***density matrix***, to evaluate the expectation value of an arbitrary operator $\hat{A}$ for the most general quantum system, by simply finding the trace given any $\hat{A}$ as long as we know the system's $\hat{\rho}$ in the same basis. This includes the probability of observing a given state, which we can compute by using the projection operator $\hat{\mathrm{P}}_i = |\psi_i\rangle\langle\psi_i|$ introduced in equation (29).

Before we move on, it is crucial to mention that for pure states, one could always locate them on a Bloch sphere, and by adopting that direction when picking our basis, we can always express the pure state such that it is a single ket $|S^1\rangle$, with $\{\mathscr{p}_s\} = \{1\}$. This happens when we pick a basis in which $|S^1\rangle$ is diagonal, for example using $|x_+\rangle$ to represent a state $|S^1\rangle = \frac{1}{\sqrt{2}}(|z_+\rangle + |z_-\rangle)$, which would've led to off diagonal elements. This is also clearly illustrated by the fact that one can always choose a ***single vector $\overline{\boldsymbol{n}}$***, on a

Bloch sphere to point to ANY ***pure qubit state***. In contrast, a mixed state is a mixture of different $|S^m\rangle$ states, where each $|S^m\rangle$ state is its own separate sum of eigenstates, as shown in equation (30), with each its own vector on a Bloch sphere. This brings up very important behavior in density matrices, which is that ***changing the basis can change the density matrix***, so how do we deal with this problem? Density matrices carry certain information such as the purity, which are independent of the basis. One viable basis independent candidate is the ***trace***, which also hints as to why the trace and density matrices usually go hand in hand in most analyses.

The important thing to note is that no matter what the pure state density matrix looks like, following a certain choice of basis, basis independent properties still hold, one of which is that a pure state can always be written as $\hat{\rho}_{pure} = |S\rangle\langle S|$.

Keep in mind that the only way to practically differentiate between a mixed state, which encompasses a classical uncertainty, and a superposition state encompassing quantum uncertainty, is through the unique property of ***quantum interference***. To probe this contrast further, consider a $50/50$ mixed state and a superposition of 2 equally likely states. Both systems would give the same experimental results if our aim was to simply observe the state, they'd both result in a 50 % chance of getting either qubit state, however the underlying physics couldn't be more different. In fact, unlike the mixed state, the superposition can be manipulated through interference properties of quantum superposition. A simple example would be a source that has a $50/50$ chance of producing ***either*** $|Z_+\rangle$ ***or*** $|Z_-\rangle$ electrons, i.e. $50/50$ ***mixture*** of $|Z_+\rangle$, $|Z_-\rangle$. Observation has a $50/50$ chance of yielding $|Z_+\rangle$ or $|Z_-\rangle$. Now compare this situation to light in superposed state $\frac{1}{\sqrt{2}}(|Z_+\rangle+|Z_-\rangle)$, i.e. $|Z_+\rangle$ ***+*** $|Z_-\rangle$, (Notice the ***+*** replacing the ***or***). A measurement would collapse the superposed state into one of its constituents, with a $50/50$ chance of yielding $|Z_+\rangle$ or $|Z_-\rangle$ as well! Here's the difference however, measuring the mixed state along $|X_+\rangle$, we would still give $50/50$ ***mixture*** of $|Z_+\rangle$, $|Z_-\rangle$, however the superposed state is certain to give $|Z_+\rangle$. Accordingly, a mixed has what we call a subjective probability. A ***subjective probability*** depends on the state of information of the observer, whom is in an uninformed state, while in actuality, the state of the particle as we mentioned is ***either*** $|Z_+\rangle$ ***or*** $|Z_-\rangle$, but since this information is only available to someone on the outside preparing the experiment from the source, the observer cannot remove this probability. In contrast, a superposed state has an ***objective probability***, which is intrinsic to the system and no amount of information can remove it. However, if a suitable observable is chosen, the objective probability can be bypassed which is what we did when we chose the $|X_+\rangle$ state. Therefore, we see that although both mixed and superposed states on observing spin along the Z direction gave the same practical results, they are inherently very different situations, in terms of both their physics and the way information is available in a system.

To better investigate mixed and pure states, we need to consider several properties of density matrices:

1.2.3 Purity and other Important Properties of Density Matrices

1. From $\hat{\rho} = \sum_m \mathcal{P}_{S^m}|S^m\rangle\langle S^m|$, with $\mathcal{P}_{S^m}$ being a probability, so a real number, and the symmetry of $|S^m\rangle\langle S^m|$ whose adjoint is simply the same term we deduce that $\hat{\rho}$ is Hermitian since

$$\hat{\rho}^{\dagger} = \hat{\rho}$$

This also means that the eigenvalues are positive definite.

2. The trace of density matrices has a unique property

$$tr(\hat{\rho}) = \sum_i \langle\psi_i| \sum_m \mathcal{p}_{s^m} |S^m\rangle\langle S^m|\psi_i\rangle$$
$$= \sum_m \mathcal{p}_{s^m} \sum_i \langle\psi_i|S^m\rangle \langle S^m|\psi_i\rangle$$
$$= \sum_m \mathcal{p}_{s^m} \sum_i |\langle\psi_i|S^m\rangle|^2 = \sum_m \mathcal{p}_{s^m} \sum_i |c_i^m|^2 = \sum_m \mathcal{p}_{s^m} = 1$$

And since traces are independent of basis, we can choose the $|\psi_i\rangle$'s to be the ones that diagonalize the density matrix, call them $|\lambda_i\rangle$, which means the diagonal entries will be the eigenvalues λ_i , so the trace turns out to be

$$tr(\hat{\rho}) = \sum_m \mathcal{p}_{s^m} = \sum_i \lambda_i = 1 \tag{37}$$

3. From equation (37), the sum of λ_i's must be 1, and the fact that λ_i's are positive definite in property (1), we can deduce that if one eigenvalue λ_i is equal to 1, all the others are 0. We mentioned before that a pure state has $\mathcal{p}_{s^m}$, equal to 1, so from (25) we can see that this case represents the pure state's density matrix. We can thus express the spectral representation of the pure state by starting from the most general spectral representation

$$\hat{\rho} = \sum_i \lambda_i \, |\lambda_i\rangle\langle\lambda_i| \tag{38}$$

For pure states this is reduced to

$$\hat{\rho}_{pure} = |\lambda_i\rangle\langle\lambda_i| \tag{39}$$

We get in equation (39) the density matrix definition of a pure state

4. The following property gives us a tool to identify how close states are to being pure just by looking at their density matrices, conveniently called ***"purity"*** of the quantum state, defined as $tr(\hat{\rho}^2)$. Since squaring doesn't change eigenbasis and only affects eigenvalues, and given traces are basis independent, squaring the diagonalized $\hat{\rho}$ would just result in a diagonal matrix of all λ_i^2 so

$$tr(\hat{\rho}^2) = tr\left(\sum_i |\lambda_i\rangle\langle\lambda_i| \ \lambda_i^2\right) = \sum_i \lambda_i^2 \tag{40}$$

Now using equation (40) and property (3), where we deduced that "unless we have a pure state with a single λ_i=1, we get a sum of λ_i's that add up to 1", meaning each $\lambda_i < 1$, therefore squaring and summing these values means we definitely get a sum that is smaller than 1 so

$$tr(\hat{\rho}^2) \leq 1 \tag{41}$$

Equality is only realized in pure states. It is also easy to prove that, given a "d" dimensional Hilbert space, we can identify a lower limit as well so the purity becomes bounded between

$$\frac{1}{d} \leq tr(\hat{\rho}^2) \leq 1 \tag{42}$$

With the lower limit thus representing the ***"maximally mixed state"***

$$\hat{\rho}_{max\ disorder} = \frac{\mathbb{I}_d}{d} \tag{43}$$

5. Finally, an important quality of Density matrices of mixed states, is their indifference towards what states are in the mixture, which in turn signifies some sort of "lost information", as one loses knowledge of the original states used in the mixture. To illustrate let's compare a 50/50 mixture of $|x_+\rangle$, $|x_-\rangle$ with a 50/50 mixture of $|z_+\rangle$, $|z_-\rangle$. We can check how this ambiguity property holds.

First consider the 50/50 mixture of $|z_+\rangle$, $|z_-\rangle$:

$$\hat{\rho}_{|x_+\rangle,|x_-\rangle} = \sum_{i=1}^{2} \mathscr{p}_{s^i} |S^i\rangle\langle S^i|$$

Here

$$\mathscr{p}_{s^1} = \mathscr{p}_{s^2} = 0.5,$$

$$|S^1\rangle = |z_+\rangle \equiv |0\rangle \equiv \begin{pmatrix}1\\0\end{pmatrix}, \qquad |S^2\rangle = |z_-\rangle \equiv |1\rangle \equiv \begin{pmatrix}0\\1\end{pmatrix},$$

so, there are no superpositions her both states are pure states but the ensemble is a mixture of pure states so

$$\hat{\rho}_{|z_+\rangle,|z_-\rangle} = 0.5|0\rangle\langle 0| + 0.5|1\rangle\langle 1| = \begin{pmatrix}0.5 & 0\\ 0 & 0.5\end{pmatrix} \tag{44}$$

Next, the 50/50 mixture of $|x_+\rangle$, $|x_-\rangle$:

To express the $|x_+\rangle$, $|x_-\rangle$ states, we use their projection on the $|z_+\rangle$, $|z_-\rangle$ basis [*]:

Here

$$\mathscr{p}_{S^1} = \mathscr{p}_{S^2} = 0.5$$

$$|S^1\rangle = |x_+\rangle \equiv |0\rangle^{(x)} \equiv \binom{1}{0}^{(x)} \equiv \frac{1}{\sqrt{2}}|z_+\rangle + \frac{1}{\sqrt{2}}|z_-\rangle \equiv \frac{1}{\sqrt{2}}|0\rangle^z + \frac{1}{\sqrt{2}}|1\rangle^z \equiv \frac{1}{\sqrt{2}}\binom{1}{1}^z,$$

$$|S^2\rangle = |x_-\rangle \equiv |1\rangle^{(x)} \equiv \binom{0}{1}^{(x)} \equiv \frac{1}{\sqrt{2}}|z_+\rangle + \frac{1}{\sqrt{2}}|z_-\rangle \equiv \frac{1}{\sqrt{2}}|0\rangle^z + \frac{1}{\sqrt{2}}|1\rangle^z \equiv \frac{1}{\sqrt{2}}\binom{1}{-1}^z$$

Using the x basis:

$$\hat{\rho}_{|x_+\rangle,|x_-\rangle} = 0.5|x_+\rangle\langle x_+| + 0.5|x_-\rangle\langle x_-|$$

$$= \frac{0.5}{\sqrt{2}*\sqrt{2}}\binom{1}{1}(1\ \ 1) + \frac{0.5}{\sqrt{2}*\sqrt{2}}\binom{1}{-1}(1\ \ -1) = \frac{1}{4}\left(\begin{pmatrix}1 & 1\\ 1 & 1\end{pmatrix} + \begin{pmatrix}1 & 1\\ 1 & 1\end{pmatrix}\right) = \begin{pmatrix}0.5 & 0\\ 0 & 0.5\end{pmatrix}$$

Supporting our original claim. This loss of information is referred to as the "**ambiguity of mixtures**".

It is important to emphasize that we could always find a direction on the Bloch sphere (i.e. representing a unique state) which leaves the pure density matrix with one diagonal entry

$$\hat{\rho}_{pure\ e.g\ ensemble\ of\ pure\ |z_+\rangle} = \begin{pmatrix}1 & 0\\ 0 & 0\end{pmatrix} \tag{45}$$

However, by changing basis we could introduce off diagonal terms, while still preserving the purity properties coded in the trace. An example where $\hat{\rho}_{pure}$ is not diagonal is a superposition into a state like

$$|S\rangle = \sqrt{3/4}\,|z_+\rangle + \sqrt{1/4}\,|z_-\rangle$$

Whose density matrix gives off diagonal elements

$$\hat{\rho}_{pure\ with\ unequal\ superposition\ in|z_+\rangle} = \begin{pmatrix}3/4 & \sqrt{3}/4\\ \sqrt{3}/4 & 1/4\end{pmatrix}$$

Which still satisfies both $tr(\hat{\rho}) = 1$ and the purity condition $tr(\hat{\rho}^2) = 1$ making it by definition a pure state. We will also add that considering the state above with equal mixing would have led to the $|x_+\rangle$ state, which in the $|z\rangle$ basis we are using, is $|x_+\rangle = \frac{1}{\sqrt{2}}|z_+\rangle + \frac{1}{\sqrt{2}}|z_-\rangle$, with a density matrix of

$$\hat{\rho}_{pure\ with\ equal\ superposition\ in|z_+\rangle} = \begin{pmatrix}0.5 & 0.5\\ 0.5 & 0.5\end{pmatrix}$$

[*] NB: if we do not mention the basis explicitly, they the convention is we are referring to the $|z_+\rangle$, $|z_-\rangle$

Now if we had picked our eigenbasis to be $|x_+\rangle$ and $|x_-\rangle$ instead of $|z_+\rangle$ and $|z_-\rangle$, we would get

$$\hat{\rho}_{pure\ in\ |x_+\rangle} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$$

Finally, the most general single qubit density matrix can be written by parametrizing the diagonal entries p and $(1-p)$, since $tr(\hat{\rho}) = 1$. Now Since $\hat{\rho}^\dagger = \hat{\rho}$, we have the off diagonals as q and q^* making

$$\hat{\rho} = \begin{pmatrix} p & q \\ q^* & (1-p) \end{pmatrix} \tag{46}$$

For the pure state we have the extra condition $tr(\hat{\rho}^2) = 1$ the most general state of superposition is $\psi\rangle = a|0\rangle + b|1\rangle$ so the qubit is in **both** states $|0\rangle$ **and** $|1\rangle$, then matrix becomes

$$\hat{\rho} = \begin{pmatrix} |a|^2 & a^*b \\ ab^* & |b|^2 \end{pmatrix}$$

For mixed states, the qubit can be in **either** $|0\rangle$ **or** $|1\rangle$, so the density matrix for a single qubit is

$$\hat{\rho} = \begin{pmatrix} p & 0 \\ 0 & (1-p) \end{pmatrix} \tag{47}$$

Since most of the properties we have discussed lie in the trace, it might not be clear what the role of ***off diagonal entries*** in density matrices is. This vagueness prompts us to directly address ***off diagonal entries***, the info they might hold, if any, how they ***respond to basis changes***, and their relation to populations. This is where the concept of "***coherence***" comes in.

1.2.4 Von Neumann Equation

Before introducing coherence, we might need to quickly go over time dependence in the density matrix formulation. Why? Because we stated that all of our usual quantum mechanical basics can be transferred to the density matrix formulation, and we need an equation that reflects time evolution. To achieve this, consider the time evolution of a typical ket $|\Psi^m\rangle$, which we know is governed by the Schrödinger equation, where $\Psi^m = \sum_i c_i|\psi_i\rangle$ and $|\psi_i\rangle$ is the i^{th} eigenstate of the state space containing the wave function Ψ^m. We want to extend this to $\hat{\rho}$, given $\hat{\rho} = \sum_m \mathcal{p}_{\Psi^m}|\Psi^m\rangle\langle\Psi^m|$. For the purposes of quantum info and, it is not only warranted, but desirable to have our ensemble isolated, so $\mathcal{p}_{\Psi^m}$ are time-independent, while Ψ^m evolve according to a time independent Hamiltonian, so $\hat{\rho}$ follows the time dependent states and so is a time dependent operator. We know, from standard quantum mechanics, that states evolve like

$$i\hbar\frac{d|\Psi^m(t)\rangle}{dt} = H|\Psi^m(t)\rangle \tag{48}$$

so

$$\frac{d|\Psi^m(t)\rangle}{dt} = \frac{H}{i\hbar}|\Psi^m(t)\rangle$$

By taking the complex conjugate

$$\frac{d\langle\Psi^m(t)|}{dt} = -\langle\Psi^m(t)|\frac{H}{i\hbar}$$

Introducing $\hat{\rho}$ and taking the time derivative yields

$$\frac{d\hat{\rho}}{dt} = \frac{d}{dt}\sum_m \mathcal{p}_{\Psi^m}|\Psi^m(t)\rangle\langle\Psi^m(t)| = \sum_m \mathcal{p}_{\Psi^m}\left(\frac{d|\Psi^m(t)\rangle}{dt}\langle\Psi^m(t)| + |\Psi^m(t)\rangle\frac{d\langle\Psi^m(t)|}{dt}\right)$$

$$= \sum_m \mathcal{p}_{\Psi^m}\left(\frac{H}{i\hbar}|\Psi^m(t)\rangle\langle\Psi^m(t)| - |\Psi^m(t)\rangle\langle\Psi^m(t)|\frac{H}{i\hbar}\right)$$

$$= \sum_m \mathcal{p}_{\Psi^m}\left(\frac{H}{i\hbar}\hat{\rho} - \hat{\rho}\frac{H}{i\hbar}\right)$$

We finally get the relation

$$\frac{d\hat{\rho}}{dt} = \frac{1}{i\hbar}[H, \hat{\rho}] \tag{49}$$

This is the ***von Neumann equation***

The von Neumann equation can also be expressed in its integral form, we can simply refer to the simplest case, which is for time independent Hamiltonians of **non-superposed states**, then we can always separate the spatial and time variables and get

$$|\Psi^m(t)\rangle = e^{-iHt/\hbar}|\Psi^m(0)\rangle$$

Subbing this relation into $\hat{\rho}$ gives

$$\hat{\rho}(t) = \sum_m \mathcal{p}_{\Psi^m}\left(e^{-iHt/\hbar}|\Psi^m(0)\rangle\langle\Psi^m(t)|e^{-iHt/\hbar}\right)$$

$$\hat{\rho}(t) = e^{-iHt/\hbar}\hat{\rho}(0)e^{-iHt/\hbar} \tag{50}$$

1.2.5 ***Coherence, decoherence and dephasing noise***[10], [16]***:***

The concept of coherence is central to quantum information theory, and has a direct relation to entanglement. It would be naïve, yet true, to say that any state with ***off diagonal elements***, is said to

contain a certain amount of ***coherence* with respect to the chosen basis**[10]. Off diagonal elements $\hat{\rho}_{kl}$ express ***interference effects*** between the pairs $|\psi_k\rangle$ and $|\psi_l\rangle$ in the superposition pure state $|S^m\rangle$.

When it comes to coherence and working with quantum effects, one's goal is usually to maintain a coherent system without succumbing to interreference from unknown factors "outside" the system. Environments outside the system are uncontrolled and ambiguous, and the observer doing a controlled experiment probably lacks all the info concerning the different ways the outside environment can interact with their system, which accounts for the imperfection of a quantum experiment or apparatus. This forces one to account for some "randomness" in their quantum calculations. If this sounds familiar, it is because we touched on this imperfection in the introduction. We can put this more technically by stating that creating this randomness is the act of ***"transitioning" from a pure quantum state to a mixture*** of states, which introduces classical randomness in our system. One can therefore see why we spent so much time introducing density matrices and purity. This transitioning is referred to as **decoherence**. Decoherence is best described by the following tautology "it is the process through which the system loses its coherence". In terms of density matrix elements, it is the disruption of the "interaction terms" between members of the superposition, and the information held within, in favor of a mixed state.

This concept is critical given that ALL systems are in contact with some surrounding, making decoherence an inevitability, this shifts the question from, how one can eliminate decoherence, to how long a system can stay within some coherent threshold, and whether we can get something done with the quantum states, within an acceptable error margin.

Putting decoherence into a better perspective we discuss one of the simplest models, the **dephasing noise**[10], and describe it by the equation

$$\frac{d\hat{\rho}}{dt} = -i[H, \hat{\rho}] + \frac{\gamma}{2}(\sigma_z \rho \sigma_z - \rho) \tag{51}$$

With the Hamiltonian $H = \frac{\Omega}{2}\sigma_z$ and γ being the "coupling strength" so the second term describes the action of the environment, which means making $\gamma = 0$ we get total isolation recovering the Neumann Eq.

Solving the above equation, we get

$$\frac{d\hat{\rho}}{dt} = \frac{-i\Omega}{2}\left[\begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix}\begin{pmatrix}p & q\\ q^* & 1-p\end{pmatrix} - \begin{pmatrix}p & q\\ q^* & 1-p\end{pmatrix}\begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix}\right] + \frac{\gamma}{2}\left[\begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix}\begin{pmatrix}p & q\\ q^* & 1-p\end{pmatrix}\begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix} - \begin{pmatrix}p & q\\ q^* & 1-p\end{pmatrix}\right]$$

$$= \frac{-i\Omega}{2}\left[\begin{pmatrix}p & q\\ -q^* & -1+p\end{pmatrix} - \begin{pmatrix}p & -q\\ q^* & -1+p\end{pmatrix}\right] + \frac{\gamma}{2}\left[\begin{pmatrix}p & -q\\ -q^* & 1-p\end{pmatrix} - \begin{pmatrix}p & q\\ q^* & 1-p\end{pmatrix}\right]$$

$$= \frac{-i\Omega}{2}\left[\begin{pmatrix}0 & 2q\\ -2q^* & 0\end{pmatrix}\right] + \frac{\gamma}{2}\left[\begin{pmatrix}0 & -2q\\ -2q^* & 0\end{pmatrix}\right]$$

Subbing equation (46) in the L.H.S Leaves

$$\frac{d}{dt}\begin{pmatrix} p & q \\ q^* & (1-p) \end{pmatrix} = \left[\begin{pmatrix} 0 & -q(i\Omega+\gamma) \\ q^*(i\Omega-\gamma) & 0 \end{pmatrix}\right] \tag{52}$$

The 2 equations from the diagonals are basically the same and the 2 from the off diagonals are complex conjugates of one another and thus basically the same this leaves us with 2 simple ODEs:

$$\frac{dp}{dt} = 0 \tag{53}$$

$$\frac{dq}{dt} = -q(i\Omega+\gamma) \tag{54}$$

The solution to ODE (42) is simply

$$q(t) = e^{-(i\Omega+\gamma)t}q(0) \tag{55}$$

We can clearly see form equation (53) that the p dependent terms i.e. the Diagonals, i.e. the elements holding the trace information are unchanged, which carry the probabilities and thus the population info of the ensemble, whilst from equation (54), only the q dependent off-diagonal terms change, and it is CRUCIAL that we point out that it is **exponential decay**.

From the above analysis we can rewrite the general qubit pure state to also include the effects of the environment, such that If we start in a pure state in identity (41.1) we can identify the new state after a time t by using

$$\hat{\rho} = \begin{pmatrix} |a|^2 & a^*be^{-(i\Omega+\gamma)t} \\ ab^*e^{(i\Omega-\gamma)t} & |b|^2 \end{pmatrix} \tag{56}$$

And most importantly after an infinite time we will reach a mixed state

$$\lim_{t\to\infty}\hat{\rho}(t) = \begin{pmatrix} |a|^2 & 0 \\ 0 & |b|^2 \end{pmatrix} \tag{57}$$

Therefore, we have shown how the action of the environment is to change quantum probabilities into classical probabilities, destroying quantum features in the process. Finally, it is important to mention that coming in contact with the environment does not always cause decoherence while preserving populations, and that the dephasing model is a simple one in that regard, but generally interactions will induce both decoherence (change in off-diagonal terms) and changes in populations (diagonal terms).

One interesting property of the $\hat{\rho}$, arises when we map it to the form taken by the most general 2×2 Hermitian matrix utilizing the identity and the 3 σ_i Pauli matrices as our basis gives

$$\hat{\rho} = \frac{1}{2}(\mathbb{1} + \bar{n} \cdot \bar{\sigma}) \tag{58}$$

This form coupled with how the Bloch sphere's components of vector $\bar{n}$ mimic $\langle \sigma_i \rangle$ the following relation can be built

$$\hat{\rho} = \frac{1}{2}(\mathbb{1} + \bar{n} \cdot \bar{\sigma}) \tag{59}$$

With what we've formulated till now we know that given the $\hat{\rho}$ we can get the expectation value of any operators using equation (36), so we can get $\langle \sigma_i \rangle$

$$n_i = \langle \hat{\sigma}_i \rangle = tr(\sigma_i \rho) \tag{60}$$

$$n_x = \langle \sigma_x \rangle = tr(\sigma_x \rho) = tr\left(\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} p & q \\ q^* & 1-p \end{pmatrix}\right) = tr\left(\begin{pmatrix} q^* & 1-p \\ p & q \end{pmatrix}\right) = q + q^*$$

$$n_y = \langle \sigma_y \rangle = tr(\sigma_y \rho) = tr\left(\begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}\begin{pmatrix} p & q \\ q^* & 1-p \end{pmatrix}\right) = tr\left(\begin{pmatrix} -iq^* & -i+ip \\ ip & iq \end{pmatrix}\right) = i(q - q^*)$$

$$n_z = \langle \sigma_z \rangle = tr(\sigma_z \rho) = tr\left(\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}\begin{pmatrix} p & q \\ q^* & 1-p \end{pmatrix}\right) = tr\left(\begin{pmatrix} p & q \\ -q^* & -1+p \end{pmatrix}\right) = 2p - 1$$

So, given these n_i and we already know σ_i so we can use equation (59) to get $\hat{\rho}$, so

$$\hat{\rho} = \frac{1}{2}(\mathbb{1} + \bar{n} \cdot \bar{\sigma}) = \frac{1}{2}\begin{pmatrix} 1+n_z & n_x - in_y \\ n_x + in_y & 1-n_z \end{pmatrix} \tag{61}$$

Getting purity

$$tr(\hat{\rho}^2) = tr\left(\frac{1}{4}\begin{pmatrix} 1+n_z & n_x - in_y \\ n_x + in_y & 1-n_z \end{pmatrix}\begin{pmatrix} 1+n_z & n_x - in_y \\ n_x + in_y & 1-n_z \end{pmatrix}\right)$$

$$= tr\left(\frac{1}{4}\begin{pmatrix} 1 + 2n_z + n_z^2 + n_x^2 + n_y^2 & \dots \\ \dots & 1 - 2n_z + n_z^2 + n_x^2 + n_y^2 \end{pmatrix}\right)$$

$$= \frac{1}{2}\left(1 + n_x^2 + n_y^2 + n_z^2\right)$$

So

$$tr(\hat{\rho}^2) = \frac{1}{2}(1 + n^2) \tag{62}$$

Noticing $n^2 = n_x^2 + n_y^2 + n_z^2$ which is basically the length of the radius of the Bloch sphere!!

What equation (62) means is that we get the parametrization of purity in terms of the radius n^2 of the sphere and it should immediately jump at us that we proved that the Purity followed the relation

$$tr(\hat{\rho}^2) \leq 1$$

So, the maximum value of n^2 is 1, which makes $tr(\hat{\rho}^2) = \frac{1}{2}(1+1) = 1$, which is a pure state, consequently we deduce that

$$radius\ of\ Bloch\ sphere = n^2 = n_x^2 + n_y^2 + n_z^2 \leq 1$$

So, we showed that Bloch spheres are of radius 1 for pure states and can also represent mixed states by a vector inside the Bloch sphere. In fact, the most mixed state would be

$$\hat{\rho} = \frac{1}{2}\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

Since

$$tr(\hat{\rho}^2) = \frac{1}{2}(1+0)$$

Meaning $n^2 = 0$

We can also get the general equation for the eigenvalue of a qubit's density matrix

$$\hat{\rho}\ eigenvalues = \frac{(1 \pm n)}{2} \tag{63}$$

One final note, we stated that both populations and coherences are basis dependent, but are they related? Yes, it turns out populations can set an ***upper limit for off diagonal terms***. Where

$$|\rho_{lm}|^2 \leq \rho_{ll}\rho_{mm} \tag{64}$$

It might be confusing that although coherence is a basis dependent concept, we still identified decoherence as a physical effect, which by definition has to be independent of basis. So how can one explain this? Well it was shown that a certain basis can be given preferential treatment based on the environment and interaction, where the ***system-environment interaction singles out a specific basis***, but this is beyond the scope of this work.

2.0.0 Section 2: Composite/ Multipartite system and product states [6], [10], [16]

2.1.0 Compound Hilbert space and product states

The easiest way to introduce entanglement is by giving a concrete example of the concept, and for that we will need to go through compound Hilbert spaces and product states. If qubit A is in lab A and qubit B in lab B, separated by some large distance, there could be something relating the 2 qubits if A and B were prepared together into what we call a ***Global state*** $|\psi^A\rangle \otimes |\psi^B\rangle$, under the global basis $|i,j\rangle^{AB}$ defining the Hilbert space for any $|0\rangle \otimes |1\rangle$, iterating indices $i \; and \; j$ over 1 and 2.

Keep in mind that the notation $|i,j\rangle^{AB}$ means that we are in the product state defined by particle A being in the i^{th} state while particle B is in the j^{th} state, with this product state existing in the Hilbert space $\mathcal{H}^{AB}$ where

$$\mathcal{H}^{AB} = \mathcal{H}^{A} \otimes \mathcal{H}^{A} \,, \tag{65}$$

and $\mathcal{H}^{AB}$ is defined by the set of 4 basis states

$$\{|0,0\rangle, |0,1\rangle, |1,0\rangle, |1,1\rangle\}. \tag{66}$$

We can generalize this to any number N of qubits by considering the different number of permutations one can get from choosing from 2 eigenstates N times, which is 2^N. The two-qubit system is what we call a ***Bipartite System*** and we will consider **Multipartite Systems** later on.

To form a global state, we have to consider the general states

$$|\psi^A\rangle = a|0\rangle + b|1\rangle \,, \;\; |\psi^B\rangle = c|0\rangle + d|1\rangle \,, \tag{67}$$

so that

$$|\psi^{AB}\rangle = |\psi^A\rangle \otimes |\psi^B\rangle = (a|0\rangle + b|1\rangle) \otimes (c|0\rangle + d|1\rangle)$$

$$= ac|0,0\rangle + ad|0,1\rangle + bc|1,0\rangle + bd|1,1\rangle \tag{68}$$

By construction we have the state $ac|0,0\rangle + ad|0,1\rangle + bc|1,0\rangle + bd|1,1\rangle$ contractable into a product $|\psi^A\rangle \otimes |\psi^B\rangle$ Global state given they are related by $|\psi^A\rangle = a|0\rangle + b|1\rangle \,, \;\; |\psi^B\rangle = c|0\rangle + d|1\rangle$. It is this

unique situation that defines ***Product States***. However, if we start with some general describing the most general state in the Hilbert space $|0\rangle \otimes |1\rangle$, we would write it as

$$C_{00}|0,0\rangle + C_{01}|0,1\rangle + C_{10}|1,0\rangle + C_{11}|1,1\rangle \tag{69}$$

We immediately notice that C_{ij} is a tensor (in tensor we mean a multidimensional array of elements), and that its 4 components are not necessarily $\{ac, ad, bc, bd\}$. Meaning that if the state

$$|\psi^{AB}\rangle = \sum_{i,j} C_{ij}|i,j\rangle^{AB} \tag{70}$$

does not somehow expand to the form

$$ac|0,0\rangle + ad|0,1\rangle + bc|1,0\rangle + bd|1,1\rangle \text{ for some}$$

$$|\psi^{A}\rangle = a|0\rangle + b|1\rangle\ ,\ |\psi^{B}\rangle = c|0\rangle + d|1\rangle,$$

we would never be able to ***separate*** the state into $|\psi^{A}\rangle \otimes |\psi^{B}\rangle$, the state resulting from the ***Tensor product*** of the two states, conveniently referred to as a ***Product State***.

So, if C_{ij} can be written as

$$C_{ij} = f_{(A)}g_{(B)} \tag{71}$$

where $f_{(A)} \in \{a, b\}$ and $g_{(B)} \in \{c, d\}$

We can then deduce from equation (68) that $|\psi^{AB}\rangle$ will factor into a tensor product $|\psi^{A}\rangle \otimes |\psi^{B}\rangle$. We therefore identify 2 mutually exclusive cases, one where the state is ***separable*** into a ***Product State***, and the other of being ***inseparable***, the latter defining an ***Entangled State***.

Next, we discuss the significance of being entangled and how it might relate to quantum information theory by considering the state

$$|\psi\rangle = \frac{1}{\sqrt{2}}(|\uparrow,\downarrow\rangle + |\downarrow,\uparrow\rangle) \tag{72}$$

We will use spins states $|\uparrow,\downarrow\rangle$ rather than the more generic $|0,1\rangle$ to more conveniently reference eigenvalues. If the observer measures the spin of particle A and obtains $\sigma_z = \frac{1}{2}$, the spin of the particle B is automatically known to be $\downarrow$ so must have $\sigma_z = -\frac{1}{2}$, owing to the system's collapse on observation into the 1^{st} eigenstate $|\uparrow,\downarrow\rangle$ in the superposition and we can safely cross out any possibility of particle B being $\uparrow$ since we ***gain additional information*** on observing one particle due to knowing that the outcome would be either $|\uparrow,\downarrow\rangle$ or $|\downarrow,\uparrow\rangle$ on collapsing the superposition. This means that the particles are **entangled**, in particular, ***Maximally Entangled*** since we gain all the info from observation (we will later show that this is actually one of the ***4 Bell states***, a set of 4 unique maximally entangled states). This also hints at why

entanglement in quantum information theory is so important, since entanglement seems to be related to the amount of information attainable from measuring a state. Contrast this to a state

$$|\psi\rangle = |\downarrow, \uparrow\rangle \tag{73}$$

There is no ambiguity here, so we gain no ***additional information*** from measuring either particle states, and we can be sure that these states are ***separable***, which if we check using the definition we used above, turns out to be true and we can write

$$|\psi^{AB}\rangle = |\downarrow\rangle \otimes |\uparrow\rangle \tag{74}$$

where

$$|\psi^{A}\rangle = |1\rangle\,, |\psi^{B}\rangle = |0\rangle \tag{75}$$

We can immediately think of the possibility of an observation giving only partial information regarding the states of other particles in the system, and we therefore need a way to quantify the amount of this information, and as we know from statistical mechanics, a well-established framework for dealing with the amount of information in a system is ***entropy***. One can vaguely claim that in one of its simplest interpretations, entropy is a ***measure of the disorder (or mixedness) of a density matrix***, which also happens to be the direct result, among other things, of entanglement.

Entanglement can be defined for both pure and mixed states, although we focus in this work on pure states. Correlation is a closely related concept to entanglement, and is where the measurement of one state gives information about the others. Entanglement in pure states will always translate to a "***Correlation***" between members of the superposition. However non-entangled mixed states may contain correlations, so correlations do not directly imply entanglement. In addition, in bipartite states, entanglement can be easily quantized, in contrast to how it is expressed in multipartite states, where it has no clear definition or quantization, we discuss these possibilities later.

2.2.0 Bipartite states

Before we tackle how one might quantize or even identify entanglement, besides brute forcing state decompositions to check for it, we need to address how measurements are expressed in Multipartite systems, obviously starting with the bipartite state.

2.2.1 Measurements as expressed in bipartite systems

Recall that measurements in a certain single qubit basis $\{|0\rangle, |1\rangle\}$ required we define an "i^{th} eigenbasis projection operator" $\hat{\mathrm{P}}_i$, which we can use to extract the probability of measuring a certain state by finding the expectation value of $\hat{\mathrm{P}}_i$, on the i^{th} eigenbasis. However, the eigenbasis for bipartite systems, as discussed prior, are not of the form

$$\{|i\rangle\} = \{|0\rangle, |1\rangle\}, \tag{76}$$

Rather look like this

$$\{|i,j\rangle\} = \{|0,0\rangle, |0,1\rangle, |1,0\rangle, |1,1\rangle\} \tag{77}$$

The projection operator still measures either particle A or B, and the measurement still uses one of the 2 single particle eigenbasis $\{|0\rangle, |1\rangle\}$, so we need to modify how we write the projection operator to accommodate the bipartite Hilbert space, and one way to extend it to a larger space, is to tensor product it with the identity matrix so

$$\hat{\mathrm{P}}_i^A = |i\rangle\langle i| \otimes \mathbb{I} \tag{78}$$

While

$$\hat{\mathrm{P}}_i^B = \mathbb{I} \otimes |i\rangle\langle i| \tag{79}$$

with i having just 2 indices, 0 or 1

For an example on how this works, consider the Bell state

$$|\psi_{bell}\rangle = \frac{1}{\sqrt{2}}(|\uparrow,\downarrow\rangle + |\downarrow,\uparrow\rangle) \tag{80}$$

If one measures A in the basis $\{|\uparrow\rangle, |\downarrow\rangle\}$, we can get the probabilities of our 2 potential outcomes for A, which are either $\uparrow$ or $\downarrow$, since we have a superposition of 2 states $(|\uparrow,\downarrow\rangle + |\downarrow,\uparrow\rangle)$, with A being $\uparrow$ in the first state and $\downarrow$ in the second. To calculate the probability of A being in a certain i^{th} state $\mathscr{p}_{A^i}$ we need $\langle\hat{\mathrm{P}}_i^A\rangle$, so

$$\mathscr{p}_{A^\uparrow} = \langle\hat{\mathrm{P}}_i^A\rangle = \langle\psi_{bell}|\hat{\mathrm{P}}_\uparrow^A|\psi_{bell}\rangle = \frac{1}{2}[\,(\langle\uparrow,\downarrow| + \langle\downarrow,\uparrow|)\;\; |\uparrow\rangle\langle\uparrow| \otimes \mathbb{I}\;\; (|\uparrow,\downarrow\rangle + |\downarrow,\uparrow\rangle)\,]$$

$$= \frac{1}{2}[\,\langle\uparrow\,|\,\uparrow\rangle\langle\uparrow|\,\langle\downarrow| + \langle\downarrow\,|\,\uparrow\rangle\langle\uparrow|\,\langle\uparrow| \quad (|\uparrow,\downarrow\rangle + |\downarrow,\uparrow\rangle)\,]$$

The red terms are identical eigenstates and so 1, while the blue ones are orthogonal and cancel out

$$= \frac{1}{2}[\,\langle\uparrow|\,\langle\downarrow|\,(|\uparrow,\downarrow\rangle + |\downarrow,\uparrow\rangle)]$$

$$= \frac{1}{2}[\,\langle\uparrow,\downarrow|\;\;(|\uparrow,\downarrow\rangle + |\downarrow,\uparrow\rangle)]$$

$$= \frac{1}{2}[\,\langle\uparrow,\downarrow\,|\,\uparrow,\downarrow\rangle + \langle\uparrow,\downarrow\,|\,\downarrow,\uparrow\rangle]$$

$$= \frac{1}{2}\langle\uparrow,\downarrow\,|\,\uparrow,\downarrow\rangle$$

so

$$\mathscr{p}_{A^{\uparrow}} = \langle \hat{\mathrm{P}}_i^A \rangle = \frac{1}{2} \tag{81}$$

2.2.2 ***Reduced Density Matrices and the partial Trace operation*** [6], [10], [11], [16]

The work we've done thus far has been strictly with kets, so the formulations we covered were suitable for pure states. However, there is a deeper relation between mixedness and entanglement that appears in multipartite systems, one that ties quantum mechanics with information theory. To tap into this, we have to introduce the concept of reduced density matrices. When a composite system, (in this case a ***Global bipartite state*** AB with its ***substates***/subparts A and B), is a product state, we can attribute a ket to each state and we have covered this formulation. However, for the most general entangled states, where the state is inseparable, it turns out that some density matrix formulation must be used to describe states A and B in what we call their ***reduced state***. This formulation is one of ***reduced density matrices***, and those represent mixed states[10], [11], [16]. To successfully present this we need to go over reduced density matrices, reduced states and the ***partial trace*** function.

We start by considering a ***bipartite pure system*** AB of 2 identical parts A and B, then generalizing the modification we made to the projection operator so it can be used on a composite Hilbert space of AB.

Before we dwell any further it is important to emphasize that entanglement in multipartite mixed states is very complicated, and not all the methods described here are directly transferable to mixed states. That said our main focus is on entanglement in pure states, since this is directly related to our eventual discussion of Tensor Networks and MPS.

In a bipartite system AB, if we specify that the basis of A is $\{|\psi_\alpha\rangle\}$, the basis of B is $\{|\phi_\beta\rangle\}$, and the basis of AB is their tensor product $\{|\tau_{\alpha\beta}\rangle\}$, we can construct an operator $\widehat{Fm}$ ***operating in*** the entire bipartite Hilbert space, while only ***operating on*** part m, where m can be either subsystem A or B. If we instead consider subspace A separately, we can identify $\hat{F}^{(A)}$, which ***operating in*** the Hilbert subspace A. Using tensor products, we can effectively construct from $\hat{F}^{(A)}$, the operator $\widehat{FA}^{(:AB)}$ by the following definition

$$\widehat{FA}^{(:AB)} = \hat{F}^{(A)} \otimes \mathbb{I}^{(B)} \tag{82}$$

And similarly

$$\widehat{FB}^{(:AB)} = \mathbb{I}^{(A)} \otimes \hat{F}^{(B)} \tag{83}$$

Such that the expectation value of $\widehat{FA}^{(:AB)}$ on composite state AB is

$$\langle \hat{F}^{(A)} \rangle^{(:AB)} = \langle \widehat{FA} \rangle^{(:AB)} = tr\left(\rho \widehat{FA}\right) = \sum_{\alpha\beta} \langle \tau_{\alpha\beta} | \rho \widehat{FA} | \tau_{\alpha\beta} \rangle \tag{84}$$

We can use the completeness of the AB basis $\{|\tau_{\alpha\beta}\rangle\}$ to add an identity matrix $\mathbb{1} = \sum_{\omega\upsilon} |\tau_{\omega\upsilon}\rangle\langle\tau_{\omega\upsilon}|$ between $\hat{\rho}$ and $\widehat{FA}^{(:AB)}$, then use the definition $\widehat{FA}^{(:AB)} = \hat{F}^{(A)} \otimes \mathbb{I}^{(B)}$, we get

$$tr^{(:AB)}\left(\rho^{(:AB)}\widehat{FA}\right) = \sum_{\omega\upsilon}\sum_{\alpha\beta}\langle\tau_{\alpha\beta}|\rho^{(:AB)}|\tau_{\omega\upsilon}\rangle\langle\tau_{\omega\upsilon}|\hat{F}^{(A)} \otimes \mathbb{I}^{(B)}|\tau_{\alpha\beta}\rangle \tag{85}$$

$$= \sum_{\omega\upsilon}\sum_{\alpha\beta}\langle\psi_\alpha|\langle\phi_\beta|\rho^{(:AB)}|\phi_\upsilon\rangle|\psi_\omega\rangle\langle\phi_\upsilon|\phi_\beta\rangle\langle\psi_\omega|\hat{F}^{(A)}|\psi_\alpha\rangle \tag{86}$$

The purple term is just the $\delta_{\upsilon\beta}$ which replaces all υ with β

$$= \sum_{\alpha\omega}\langle\psi_\alpha|\sum_{\beta}\langle\phi_\beta|\rho^{(:AB)}|\phi_\beta\rangle|\psi_\omega\rangle\langle\psi_\omega|\hat{F}^{(A)}|\psi_\alpha\rangle \tag{87}$$

If we identify the term $\sum_\beta\langle\phi_\beta|\rho^{(:AB)}|\phi_\beta\rangle$ as $tr^{(/B)}\left(\rho^{(:AB)}\right)$ which traces out state B from $\rho^{(:AB)}$, and call the resulting term $\hat{\rho}_A$ (reason to be clear later) we can write

$$\langle\hat{F}^{(A)}\rangle^{(:AB)} = \sum_{\alpha\omega}\langle\psi_\alpha|\hat{\rho}_A|\psi_\omega\rangle\langle\psi_\omega|\hat{F}^{(A)}|\psi_\alpha\rangle \tag{88}$$

$$= \sum_{\alpha}\langle\psi_\alpha|\hat{\rho}_A\,\hat{F}^{(A)}|\psi_\alpha\rangle \tag{89}$$

so

$$\langle\hat{F}^{(A)}\rangle^{(:AB)} = tr^{(/A)}(\hat{\rho}_A\,\hat{F}^{(A)}) \tag{90}$$

This result is exactly what we wanted, a modification to the density matrix of a global system that allows one to measure the expectation value of an operator which acts only on a subsystem, in this case subsystem A, of a global bipartite state AB. We achieved this by introducing new definitions, $\hat{\rho}_m$ and $tr^{(/m)}$.

We can now generalize this to ***any operator on AB***, by considering the trace operation described above within the full AB space. Given any operator on AB can be written as

$$\mathcal{O}^{(:AB)} = \sum_{\alpha} A_\alpha \otimes B_\alpha \tag{91}$$

2.2.3 Bipartite operators of the simplest case, $\mathcal{O}^{(:AB)} = A\otimes B$

We used in the previous example an $A_\alpha = A = \hat{F}^{(A)}$ and $B_\alpha = B = \mathbb{I}$, meaning our operator acted only subsystem A, but we expressed it successfully in the entire space AB using a tensor product. By similarly considering the trace for a slightly more general operator $\mathcal{O}^{(:AB)} = A\otimes B$ in the $|a,b\rangle$ tensor product basis we get

$$tr^{(:AB)}(\mathcal{O}) = \sum_{a,b} \langle a,b|\mathcal{O}|a,b\rangle \tag{92}$$

$$= \sum_{a,b} \langle a|\otimes\langle b|\,(A\otimes B)\,|a\rangle\otimes|b\rangle \tag{93}$$

$$= \sum_{a} \langle a|A|a\rangle \otimes \sum_{b} \langle b|B|b\rangle \tag{94}$$

$$= \sum_{a} \langle a|A|a\rangle \sum_{b} \langle b|B|b\rangle \tag{95}$$

This result is simply the trace of the operators A and B in their respective Hilbert spaces, so we just showed that

$$tr^{(:AB)}(A\otimes B) = tr^{(/A)}(A) * tr^{(/B)}(B) \tag{96}$$

And by choosing that, we trace out subsystem A exclusively, thereby defining a ***partial operation*** on global system AB which goes as

$$tr^{(/A)}(A\otimes B) = tr^{(/A)}(A) * B\,, \qquad tr^{(/B)}(A\otimes B) = A * tr^{(/B)}(B) \tag{97}$$

These are the infamous ***Partial traces***, tracing over A and B respectively in the bipartite system AB.

Please notice that the original trace, introduced at the beginning, acted on operators to generate a number. The partial trace however was defined such that it receives a tensor product of operators built on complex Hilbert spaces, in this example $\mathcal{H}_{AB}$, to generate another operator, but one living in a smaller Hilbert space, $\mathcal{H}_A$ or $\mathcal{H}_B$. We can clearly see that ***"Tracing out $\hat{A}$"***, while it does map $\hat{A}$ to a number, it also leaves us with operator $\hat{B}$ that lives in $\mathcal{H}_B$.

We can now write the components of the partial trace, of an $A\otimes B$ operator, explicitly using equation (93)

$$tr^{(/A)}(\mathcal{O}) = \sum_{a} \langle a|\otimes\langle \mathbb{I}|\,(A\otimes B)\,|a\rangle\otimes|\mathbb{I}\rangle \tag{98}$$

And to check the validity quick we compute

$$\sum_{a} \langle a|\otimes\langle \mathbb{I}|\,(A\otimes B)\,|a\rangle\otimes|\mathbb{I}\rangle = \sum_{a} \langle a|A|a\rangle \otimes \sum_{b} \langle \mathbb{I}|B|\mathbb{I}\rangle \tag{99}$$

$$= tr^{(/A)}(A) * B \qquad (Q.E.D)$$

Getting back to the identity we specified in the partial trace derivation as $tr^{(/B)}\left(\rho^{(:AB)}\right)$ we find that

$$\hat{\rho}^{(/A)} = tr^{(/B)}\left(\rho^{(:AB)}\right) = \sum_{\beta} \langle \phi_\beta | \rho^{(:AB)} | \phi_\beta \rangle$$

allowed us to compute

$$\langle \hat{F}^{(A)} \rangle^{(:AB)} = tr^{(/A)}(\rho^{(/A)}\ \hat{F}^{(A)}) \tag{100}$$

Which means that It acts like a density matrix, but for getting expectation values on compound global states where $A \in$ some global state, and we therefore call it a ***reduced density matrix***, much like the operator that traces out A is called a ***partial trace***.

We now have the opportunity to take a better look at Quantum correlations and separability. For pure states, the difference when looking at expectation values, between entangled or product states is that entanglement gives rise to correlations. If we have a product state $\left|\psi^{(:AB)}\right\rangle = |\varphi_A, \varphi_B\rangle$, and an operator $\sigma^{(:AB)} = \sigma_A \otimes \sigma_B$, we explained how we would get an expectation value of

$$\langle \sigma^{(:AB)} \rangle = \left\langle \psi^{(:AB)} \middle| \sigma_A \otimes \sigma_B \middle| \psi^{(:AB)} \right\rangle$$

This form allows us to write

$$\langle \sigma^{(:AB)} \rangle = \langle \varphi_A \otimes \varphi_B | \sigma_A \otimes \sigma_B | \varphi_A \otimes \varphi_B \rangle$$

$$= \langle \varphi_A | \sigma_A | \varphi_A \rangle \langle \varphi_B | \sigma_B | \varphi_B \rangle \tag{101}$$

This shows that in the case of separability, the results of any measurements will also be separable, and thus there is no correlation since measurements on A and B happen independently. However, if we are provided with an entangled state, we lose this separability, and we can never measure one subsystem without acting on the other, hence the "***correlation***".

We can now consider the most general bipartite operator, $\mathcal{O}^{(:\mathrm{AB})}$, we get this by adding two $\mathbb{I}$ terms, indexed over a, b and a', b' respectively, to decompose $\mathcal{O}^{(:\mathrm{AB})}$ into

$$\mathcal{O}^{(:AB)} = \sum_{a,b,a',b'} |a,b\rangle\langle a,b|\ \mathcal{O}\ |a',b'\rangle\langle a',b'| \tag{102}$$

To get the trace over a subsystem, say B, we set $b = b'$ and do the sum, this leaves behind the terms

$$tr^{(/B)}(\mathcal{O}^{(:AB)}) = \sum_{a,a',b'} \sum_{b} (\mathbb{I}\otimes\langle b|)|a\rangle|b\rangle\ \langle a,b|\ \mathcal{O}\ |a',b'\rangle\ \langle a'|\langle b'|\ (\mathbb{I}\otimes|b\rangle)$$

$$tr^{(/B)}(\mathcal{O}^{(:AB)}) = \sum_{a,b,a'} |a\rangle\langle a,b|\,\mathcal{O}\,|a',b\rangle\langle a'|$$

And since expectation values are just numbers so

$$tr^{(/B)}(\mathcal{O}^{(:AB)}) = \sum_{aa'} |a\rangle\langle a'| \sum_{b} \langle a,b|\,\mathcal{O}\,|a',b\rangle \tag{103}$$

The $\sum_{aa'}|a\rangle\langle a'|$ term is simply a sum of outer products, like in equation (91), meaning that $tr^{(/B)}(\mathcal{O}^{(:AB)})$, the partial trace of B, is an operator acting on A

2.2.4 *Special case of a bipartite operator of the form $\mathcal{O} = |a,b\rangle\langle a',b'|$*

We can consider a special case of this when $\mathcal{O} = |a,b\rangle\langle a',b'|$ to reach a very convenient simplification.

$$\mathcal{O} = |a,b\rangle\langle a',b'| = (|a\rangle\otimes|b\rangle) \otimes (\langle a'| \otimes \langle b'|) = |a\rangle\langle a'| \otimes |b\rangle\langle b'| \tag{104}$$

We can apply the partial trace over B, by simply applying the trace to the $|b\rangle\langle b'|$ term such that

$$tr^{(/B)}\left(\mathcal{O}^{(:AB)}\right) = |a\rangle\langle a'|\ tr^{(/B)}|b\rangle\langle b'| \tag{105}$$

Using equation (15), where $tr\,(|k\rangle\langle b|) = \langle b|k\rangle$ we get

$$tr^{(B/)}\left(\mathcal{O}^{(:AB)}\right) = |a\rangle\langle a'|\ \langle b'|b\rangle$$

And by analogy

$$tr^{(/A)}\left(\mathcal{O}^{(:AB)}\right) = |b\rangle\langle b'|\ \langle a'|a\rangle$$

so, we get the very useful relations for the $\mathcal{O} = |a,b\rangle\langle a',b'|$ cases

$$tr^{(/B)}\left(\mathcal{O}^{(:AB)}\right) = |a\rangle\langle a'|\ \delta_{b,b'} \qquad tr^{(/A)}\left(\mathcal{O}^{(:AB)}\right) = |b\rangle\langle b'|\ \delta_{a,a'} \tag{106}$$

An example of when this result is useful, is when ***getting the density matrix of a pure state*** !! since ***bipartite pure states***, analogous to single particle density matrices shown in equation (39), have density matrices,

$$\hat{\rho}_{pure}{}^{(:AB)} = |a,b\rangle\langle a',b'| \tag{107}$$

And you can already guess when we'll need to partially trace this, $\hat{\rho}^{(:AB)}{}_{pure}$, to get the reduced density matrix $\hat{\rho}^{(/A)}$, so

$$\hat{\rho}_{pure}{}^{(/A)} = tr^{(/B)}\left(\rho_{pure}{}^{(:AB)}\right) = tr^{(/B)}(|a,b\rangle\langle a',b'|)$$

so, given a bipartite system AB

$$\hat{\rho}_{pure}{}^{(/A)} = |a\rangle\langle a'|\ \delta_{b,b'} \tag{108}$$

Example:

Consider one of the Bell states, similar to the one we've mentioned in equation (80), a bipartite pure state AB

$$|\psi_{bell}\rangle = \frac{1}{\sqrt{2}}(|0,0\rangle + |1,1\rangle)$$

We want to get the partial trace on A, $\hat{\rho}_{pure}{}^{(/A)}$,

$$\hat{\rho}_{pure}{}^{(:AB)} = |\psi_{bell}\rangle\langle\psi_{bell}| = \frac{1}{2}\{(|0,0\rangle + |1,1\rangle) \otimes (\langle 0,0| + \langle 1,1|)\}$$

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

Using equation (108)

$$\hat{\rho}_{pure}{}^{(/A)} = tr^{(/B)}\left(\rho_{pure}{}^{(:AB)}\right) = tr^{(/B)}(|a,b\rangle\langle a',b'|) = |a\rangle\langle a'|\ \delta_{b,b'}$$

From $\delta_{b,b'}$ we will only keep the terms with similar b indices, by checking the terms in $\frac{1}{2}\{|0,0\rangle\langle 0,0| + |0,0\rangle\langle 1,1| + |1,1\rangle\langle 0,0| + |1,1\rangle\langle 1,1|\}$, where we've colored the tensor components pertaining to substate B, if they're similar, they'll fulfill condition $\delta_{b,b'}$ $(terms\ we\ highlighted\ in\ gray)$, and we trace them out, and keep their corresponding A substates, otherwise we omit the entire term, this gives

$$\hat{\rho}_{pure}{}^{(/A)} = |a\rangle\langle a'|\ \delta_{b,b'} = \frac{1}{2}\{|0\rangle\langle 0| + |1\rangle\langle 1|\} = \begin{pmatrix} 0.5 & 0 \\ 0 & 0.5 \end{pmatrix} \tag{109}$$

As we predicted this is a ***maximally mixed state*** for a state in 2-D Hilbert space, a condition we introduced in equation (43)

$$\hat{\rho}_{max\ disorder-2D} = \frac{\mathbb{I}_d}{d} = \frac{\mathbb{I}_2}{2} = \begin{pmatrix} 0.5 & 0 \\ 0 & 0.5 \end{pmatrix} \tag{110}$$

This is a remarkable result, since we are seeing that the reduced state of A, is actually a ***maximally mixed state***. This turns out to be a feature of all Bell states, and the reason we refer to them as ***maximally entangled states***. This result is quite remarkable as it turns out that whenever a compound AB state is ***entangled***, the ***reduced density matrix*** of its substates will be a mixed state, the only exception being if the ***compound state/Global state*** was originally a product state, then it's just the

$$\hat{\rho}_{pure\ product\ state}{}^{(/A)} = |a\rangle\langle a|. \tag{111}$$

Meaning that if the global state was a product state, the reduced density matrices of its substates are pure states, and therefore exhibit no mixedness.

This approach places density matrices in a whole new perspective, where we seem to have generated a density matrix without any classical input, in contrast to when we first introduced them for single particle systems. There is no faulty apparatus here, the ***mixedness*** seems to be of quantum origin, originating from having an entangled multipartite system, then performing measurements on a subsystem by one of two observers, say Alice and Bob. Let's say we have an ensemble of bipartite pairs each pair in the bell state discussed in the previous example, and Alice and Bob get one or the other from every pair. Only Bob is in charge of measuring particles in his ensemble. If Bob measures one of his particles, the global state of the pair collapses into one of certain states $|0,0\rangle$ or $|1,1\rangle$, and both Bob's and Alice's particle in the pair is determined. Repeating this for the whole ensemble, Bob gets a mixture of spins with density matrix $\begin{pmatrix} 0.5 & 0 \\ 0 & 0.5 \end{pmatrix}$, which in turn sets Alice's ensemble to the same mixture. However, if Bob performed no measurement, he knows his ensemble is still in the pure maximally entangled Bell state, while Alice has no way of knowing if Bob has performed his measurement or what his states were, unless he specifically tells her about his measurement through a classical channel. If Alice gets no signal from Bob regarding his measurements, she can at best say that her ensemble is in a state $|0,0\rangle$ or $|1,1\rangle$, both occurring with probability $\frac{1}{2}$, so this can be illustrated in a density matrix $\frac{1}{2}\{|0\rangle\langle 0| + |1\rangle\langle 1|\} = \begin{pmatrix} 0.5 & 0 \\ 0 & 0.5 \end{pmatrix}$. This introduces another source of ambiguity to density matrices, where it is not clear to Alice if the mixedness in the density matrix is due to still having entangled particles, or due to her ensemble's collapse into a mixed state, following Bob's measurements. This is another example of the ***ambiguity of mixtures*** we mentioned before when tackling single particle states and seems to be an unavoidable property of the density matrix formalism.

The Alice-Bob dilemma and the ambiguity generated by forming density matrices, reflect the ***loss of information*** inherent to taking the partial trace of a subsystem, and shows why, in general

$$\hat{\rho}^{(:AB)} \neq \hat{\rho}^{(/A)} \otimes \hat{\rho}^{(/B)} \tag{112}$$

irrespective of whether the state is pure or mixed

Unless $\hat{\rho}^{(:AB)}$ was originally ***uncorrelated*** with all zero off-diagonal entries. This loss of info makes the partial trace an ***irreversible*** operation, such that one loses info concerning the original state, and can never ***"retrace"*** their steps. This loss is a manifestation of the ambiguity of mixtures.

2.2.5 Extending the Bipartite entanglement model to Multipartite systems

Entanglement becomes exponentially difficult the larger a multipartite system becomes, and till this day an open research problem that has seen extensive activity over the past decade. One way to bypass the difficulty is through delegating the complexity of a Global system to bipartitions. To do this, we can consider the general N-partite pure state $\sum_s \psi^{s_1 s_2 s_3 \dots s_N} |s_1 s_2 s_3 \dots s_N\rangle$ we can cast this problem into a collection of bipartite cases of the form $|\psi^{AB}\rangle = \sum_{i,j} C_{ij} |i,j\rangle^{AB}$ using the idea of ***collective indices***.

An example would we breaking down the index set $\{s_1, s_2, s_3, \dots s_N\}$ into 2 subsets A and B, where

$$A = \{s_1, s_2, \dots s_k\} \ , \quad B = \{s_{k+1}, s_{k+2}, \dots s_N\} \tag{113}$$

This reduces the global system's tensor of coefficients $\psi^{s_1 s_2 s_3 \dots s_N}$ into

$$\psi^{s_1 s_2 s_3 \dots s_N} = \psi^{\{s_1, s_2, \dots s_k\}, \{s_{k+1}, s_{k+2}, \dots s_N\}} = \psi^{AB} \tag{114}$$

Although the Global state $|\psi^{AB}\rangle$ is generally entangled, it still can be broken down into subsystems, where it may happen that some of the subsystems are still disentangled with respect to one another. We can therefore consider all possible Bipartite pairings and use our bipartite model for each.

Let's Apply this to the next multipartite system in complexity, 3-qubits $\{A, B, C\}$. Here we have $3\ choose 2 = \binom{3}{2} = 3$ pair combinations, with possible entanglement between (A, B), between (A, C), or between (B, C) and an extra 4th and 5^{th} possibility where all are entangled (A,B,C) or the system is in a product state with no entanglement whatsoever. We then treat these systems as we do bipartites, and apply our Schmidt Decomposition approach to the pairs. This is not a perfect solution, but we will see that at best we can tackle multipartite systems through computational methods that try and extract as much info about a Global system from as few select states and subsystems as possible. Some of the best methods at doing this at the moment have been ones based on Tensor Networks and MPS, and they depend heavily on isolating the most "favorable" Schmidt coefficients in systems after breaking them down into multiple partitions, more on that in Section 4.

3.0.0 Section 3: Entanglement Entropy and quantizing entanglement

So far, we've only very generally touched on what entanglement is, now we need to be more specific in our definition, identification, and quantification of entanglement. We therefore start by addressing entanglement in the specific case of ***bipartite pure states***. To do so, we will need to use an indispensable tool when one considers multipartite systems, and that's the Schmidt decomposition.

3.1.0 Singular Value Decomposition (SVD), Schmidt decomposition, and their role in addressing entanglement in bipartite pure states [10], [12]

Given a bipartite pure state, meaning one that can be written as $\left|\psi^{(AB)}\right\rangle = \sum_{i,j} C_{ij} |i, j\rangle^{(AB)}$, C_{ij} are the components of a rank 2 tensor (matrix) whose dimensions are determined by d_A, the dimension of substate A, and d_B, the dimension of substate B, so most generally a $d_A \times d_B$ matrix. Therefore, using Single Value Decomposition (SVD), we know any $d_A \times d_B$ matrix can be decomposed into

$$\hat{C}^{(AB)} = USV^{\dagger}$$

Such that if we carry SVD on the matrix $\hat{C}_{ij}$, we get

$$\hat{C}_{ij}{}^{(AB)} = \sum_{m=1}^{r} U_{im}\sigma_m V_{jm}^* \tag{115}$$

Where **S is a diagonal matrix** of dimension $(min[d_A, d_B] \times min[d_A, d_B])$,

U is $d_A \times \min[d_A, d_B]$; is unitary, so $UU^\dagger = \mathbb{I}$; with orthonormal column vectors, ***left singular vectors***

V is $\min[d_A, d_B] \times d_B$; is also unitary, so $VV^\dagger = \mathbb{I}$ with orthonormal row vectors, ***right singular vectors***

$\sigma_m = S_{mm}$, the diagonal elements of matrix S, also called the ***Singular values*** of matrix $\hat{C}^{(AB)}$

And $r =$ ***Shmidt Rank*** of matrix $\hat{C}_{ij}$, which is the number of non-zero σ_m

We also see that since U and V are unitary, $\hat{C}^{(AB)}\left(\hat{C}^{(AB)}\right)^\dagger = US^2U^\dagger$ while $\left(\hat{C}^{(AB)}\right)^\dagger \hat{C}^{(AB)} = VS^2V^\dagger$, and since unitary matrices do not affect eigenvalues, both $\hat{C}^{(AB)}\left(\hat{C}^{(AB)}\right)^\dagger$and $\left(\hat{C}^{(AB)}\right)^\dagger \hat{C}^{(AB)}$, despite having different dimensions (since U and V have different dimensions), share the ***same eigenvalues*** $\boldsymbol{\sigma_m^2}$, but do ***not share in the eigenstates***. The point is that the largest of the two will have the same eigenvalues as the smaller one, plus extra zero eigenvalues. We will emphasize the importance of this when we find that getting the entanglement entropy of a bipartite system AB can be obtained by referencing ***either of its subsystem*** A or B, since they will share all non-zero σ_m^2.

Now the special thing about the coefficient matrix elements $\hat{C}_{ij}{}^{(AB)}$, lies in the normalization of the quantum state described by these coefficients making $\sum_{ij}\left|\hat{C}_{ij}{}^{(AB)}\right|^2 = 1$, which directly implies that

$$\sum_{m=1}^{r} |\sigma_m|^2 = 1 \tag{116}$$

Using equation (115), and ***given we are dealing with a pure state***, we can derive the following

$$\left|\psi^{(AB)}\right\rangle = \sum_{i,j,m} U_{im}\sigma_m V_{jm}^* |i,j\rangle^{(AB)} = \sum_m \sigma_m \left(\sum_i U_{im}\,|i\rangle^{(A)}\right) \otimes \left(\sum_j V_{jm}^*\,|j\rangle^{(B)}\right)$$

We notice that the blue and red terms define separate states for A and B respectively, so we were able to "Decompose" the Global bipartite state AB, into a tensor product of column vectors state A and B.

$$\left|m^{(A)}\right\rangle = \sum_i U_{im}\,|i\rangle \tag{117}$$

$$\left|m^{(B)}\right\rangle = \sum_j V_{jm}^*\,|j\rangle \tag{118}$$

This gives us the relation

$$|\psi^{(AB)}\rangle = \sum_{m}^{r} \sigma_m |m^{(A)}\rangle \otimes |m^{(B)}\rangle \tag{119}$$

this is called the ***Schmidt Decomposition***, with $r \leq \min[d_A, d_B]$, and the basis sets $\{|m^{(A)}\rangle\}$ and $\{|m^{(B)}\rangle\}$ are orthonormal as the orthonormality of U *and* $V^\dagger$ will extend to them.

How is this formulation related to entanglement one might ask? Well, we will show that the square of the singular values contains all the information about entanglement when we introduce reduced density matrices and partial traces. Notice that we have a key term

$$\sigma_m^2 \textit{ called the } \boldsymbol{Schmidt\ Coefficients\ \lambda_m}$$

With this decomposition in hand, we have reduced the process of identifying whether a bipartite system is product or not, or in other words ***if the state is entangled***, to simply finding its Schmidt values.

$$\text{If } \boldsymbol{\sigma_1 = 1}$$

Then based on equation (116) this means all the others are zero, we have a state with ***Schmidt rank 1 and*** our state is therefore ***a Product state.***

One of the most practical applications of SVD and Schmidt decomposition will be discussed in detail in section (4.4.2) where it is used to construct and resized Matrix Product States. This practicality is due to being able to adjust the Schmidt rank to generate approximations of the matrix $\hat{C}$, (which is by construction defaulted to rank-r) the approximate matrix $\hat{C}'$, by choosing rank $r' < r$. This approximation retains only the first $\mathbf{r'}$ **singular values**, while setting the remaining $(r - r')$ to zero. The r' cut-off shrinks the column dimension of U and the row dimension of $V^\dagger$ to r′ and we get

$$\hat{C}^{(AB)'} = US'V^\dagger$$

Therefore, this functionality is of the utmost importance when it comes to approximating exponentially large Hilbert spaces, since we can ***find the best approximation*** (with respect to some closeness metric for example the 2-Norm) for matrix $\hat{C}$ of rank r, by some $\hat{C}'$ of smaller rank r', spanned over state spaces of A and B ***if and only if the basis sets*** $\{|\boldsymbol{m^{(A)}}\rangle\}$ ***and*** $\{|\boldsymbol{m^{(B)}}\rangle\}$ ***are orthonormal***, which is exactly the case with the Schmidt decomposition, which gave us an orthonormal basis in terms of SVD for some subset of the Hilbert space. A rigorous proof of this approximation is directly related to the Frobenius norm and can be found in almost any advanced linear algebra textbook, the reader is free to look this up for themselves. Consequently, we can get the optimal approximation of $\hat{C}$ by $\hat{C}'$ where $\hat{C}'$ is a matrix of rank r′, where r' entails targeting the largest singular values of $\hat{C}$, giving us the approximate state $|\tilde{\psi}^{(AB)}\rangle$. Finally, we can say that the Schmidt decomposition of the optimal approximation reads

$$|\tilde{\psi}^{(AB)}\rangle = \sum_{m}^{r'} \sigma_m |m^{(A)}\rangle \otimes |m^{(B)}\rangle \tag{120}$$

We will revisit this technique in section (4.4.3), when discussing MPS approximations for approximating large Hilbert spaces.

3.2.0 *Purity/mixedness of reduced matrices and maximally entangled pure bipartite states* [1], [6], [10], [16]

Finally, it turns out that the ***Schmidt coefficients*** are vital to quantifying the degree of entanglement, and we get to display the utility of reduced density matrices here, since the way to approach this for a bipartite system is by computing the reduced density matrices $\hat{\rho}_{pure}{}^{(/A)}$ and $\hat{\rho}_{pure}{}^{(/B)}$. Using the Schmidt Decomposition $\sum_m \sigma_m |m^{(A)}\rangle \otimes |m^{(B)}\rangle$, this is straightforward, since $|m^{(A)}\rangle$ and $|m^{(B)}\rangle$ are by construction ***orthogonal,*** and can therefore be targeted independently, because the Schmidt decomposition separates the global state into its substates. Therefore, to get $\hat{\rho}_{pure}{}^{(/A)}$, we can omit the $|m^{(B)}\rangle$ part living in B's Hilbert subspace, and keep the $|m^{(A)}\rangle$ term which contains all the info regarding the Subsystem A, this gives us

$$\hat{\rho}_{pure}{}^{(/A)} = \sum_m \sigma_m^2 |m^{(A)}\rangle\langle m^{(A)}|$$

And by analogy

$$\hat{\rho}_{pure}{}^{(/B)} = \sum_m \sigma_m^2 |m^{(B)}\rangle\langle m^{(B)}|$$

With the reduced density matrix, we can calculate the purity similar to the single qubit states, such that

$$tr\left(\left(\hat{\rho}^{(/A)}\right)^2\right) = tr\left(\sum_m |m^{(A)}\rangle\langle m^{(A)}|\ \sigma_m^4\right) = \sum_t \sum_m \langle t^{(A)}|m^{(A)}\rangle\langle m^{(A)}|t^{(A)}\rangle\ \sigma_m^4$$

$$= \sum_m \sigma_m^4 \sum_t \langle m^{(A)}|t^{(A)}\rangle\langle t^{(A)}|m^{(A)}\rangle$$

Using completeness, we can now contract the sum into the Identity

$$= \sum_m \sigma_m^4 \langle m^{(A)}|m^{(A)}\rangle$$

$$= \sum_m \sigma_m^4$$

If we repeat this for the reduced density matrix over B, $tr\left(\left(\hat{\rho}^{(/B)}\right)^2\right)$ we get ***THE SAME RESULT!!***

Although we made no constraint on the dimensions of subset A and B since

in general, $\boldsymbol{d_A \neq d_B}$

yet we still get that the purity of subparts are the same.

$$Purity^{(/A)} = Purity^{(/B)} = \sum_m \sigma_m^4 = \sum_m \lambda_m^2 \tag{121}$$

Making the Schmidt coefficients λ_m also the eigenvalues of $\hat{\rho}^{(/A \ or \ B)}$ .Thus, we conclude that in a Bipartite system ***purity***, and therefore ***mixedness***, are related to entanglement through Schmidt coefficients. We can therefore claim that the

more entangled a global system is, the more mixed the reduced density matrices of its substates.

We also know that purity comes in degrees, refer back to equation (42)

$$\frac{1}{d} \leq Purity \leq 1 \tag{122}$$

This leads to the conclusion that ***purity of reduced states*** can be a ***quantifier of entanglement***, and with that we can finally Identify a ***"maximally entangled state".*** With equation (42), the normalization $\sum_{m=1}^{r} |\sigma_m|^2 = 1$, and that the dimension of S is its rank r, we can deduce that a maximally entangled state can be defined by one whose decomposition has all non-zero singular values being equal to

$$\sigma_m = \frac{1}{\sqrt{r}} \tag{123}$$

when applying Schmidt decomposition, we represent the $|\psi^{(AB)}\rangle$ vector of the bipartite state, as a matrix $\hat{C}^{(AB)}$. This is achieved through assigning the eigenstate coefficients to matrix elements $\hat{C}_{ij}$. An example of how this works:

Consider a ***qubit*** for subsystem A (2 possible states $\{|0\rangle, |1\rangle\}$) and a ***qutrit*** for subsystem B (3 possible states $\{|0\rangle, |1\rangle, |2\rangle\}$), and the ***Global system AB***, so the global system has $(2 \times 3) = 6$ possible product states $\{|0,0\rangle, |0,1\rangle, |0,2\rangle, |1,0\rangle, |1,1\rangle, |1,2\rangle\}$. We can store the 6 states in a 2×3 matrix, and index it such that the states $\{|0,0\rangle, |0,1\rangle, |0,2\rangle, |1,0\rangle, |1,1\rangle, |1,2\rangle\}$ have corresponding coefficients stored in the matrix elements $\{C_{00}, C_{01}, C_{02}, C_{10}, C_{11}, C_{12}\}$. Given $|\psi^{(AB)}\rangle = \sum_{i,j} C_{ij}|i,j\rangle^{(AB)}$ we get the matrix

$$\hat{C}^{(AB)} = \begin{pmatrix} C_{00} & C_{01} & C_{02} \\ C_{10} & C_{11} & C_{12} \end{pmatrix} \tag{124}$$

So, a bipartite state $|\psi^{(AB)}\rangle = \alpha|00\rangle + \beta|10\rangle + \gamma|12\rangle$ could be mapped to a matrix

$$\hat{C}^{(AB)} = \begin{pmatrix} \alpha & 0 & 0 \\ \beta & 0 & \gamma \end{pmatrix}$$

For which we want to get the SVD

Now, one of the main goals of SVD is ***making low rank approximations of matrices***. Given d_A and d_B integers, we have one d_A-state-bit for substate A and one d_B-state-bit for substate B. This leads to a

$d_A \times d_B$ coefficient matrix $\hat{C}^{(AB)}$ with $d_A \times d_B$ entries that blow up for large integers that increase with the size of the system's Hilbert space. Now consider $d_A = d_B = N$ for simplicity of the following argument. From equations (117) and (118) let u_m be a column vector with entries U_{im} and v_m be another column vector with entries V_{jm} , we can therefore see that, owing to the S matrix being diagonal, we can rewrite a matrix $\hat{C}^{(AB)}$ whose elements are

$$\hat{C}_{ij}^{(AB)} = \sum_{m=1}^{r} U_{im}\sigma_m V_{jm}^* \tag{125}$$

As a matrix

$$\hat{C}^{(AB)} = \sum_{m} \sigma_m \, u_m v_m^\dagger \tag{126}$$

Where, $u_m v_m^\dagger$ is an $N \times N$ matrix with entries

$$\left(u_m v_m^\dagger\right)_{ij} = u_i v_j^* \tag{127}$$

We call matrices like $u_m v_m^\dagger$ that are tensor products of 2 vectors ***rank-1 matrices***. Equation (126) shows that the Schmidt decomposition has decomposed matrix $\hat{C}^{(AB)}$, into a ***sum of rank- 1 matrices weighted by the singular values***. There is significance to the decomposition being of rank- 1 matrices, and a closer look reveals that, since $\left(u_m v_m^\dagger\right)_{ij} = u_i v_j^*$ not all the C_{ij} matrix elements are independent, and that they depend on the components of the vectors u_m and $v_m^\dagger$, so that each individual rank-1 matrix has $N + N$ unique elements instead of the whole $N \times N$. Since singular values are always non-negative and are arranged in descending order, we can ***target the largest singular values instead of summing over the full Schmidt rank r***, to get an approximation of the matrix to a rank $r' < r$. ***This is called a rank-r' approximation for the matrix*** $\hat{\boldsymbol{C}}^{(\boldsymbol{AB})}$. If we only target the First (i.e. largest) σ_m, (a rank-1 approximation) then we have effectively reduced the $N \times N$ elements down to $N + N$, i.e. those of a single rank-1 matrix. We emphasized this point, since this concept will later be extended to tensor networks and matrix product states, a formulation that utilizes approximation of multipartite systems, by singling out specific states.

3.3.0 Quantizing entanglement in bipartite systems [1], [2], [5], [6], [10]

Much like purity, entropy can be used to communicate the mixedness of a density matrix. Additionally, entropy adds an "informational" aspect to this mixedness, and we eventually approach entropy and mixedness with questions like "how much information is associated with this system ", and we touch on

this question in a moment. Entropy seems to be a good quantifier of various quantum properties, including entanglement of pure multipartite states, which gives rise to quantities like ***entanglement entropy***. There are several ways to extend quantification to all types of ***correlations*** in general mixed states, one of which is ***mutual information***. In addition, we will introduce ***relative entropy***, which quantifies differences between two states by assigning ***distances between two density matrices***.

3.3.1 *Shannon entropy in quantification of information and compression*

The von Neumann entropy is not only related to informational entropy, as in quantifying the amount of information needed convey/ stored in a message, but its definition can also extend to quantifying entanglement. Since entanglement of a pure state is directly tied to the mixedness of its reduced density matrix, we can usually introduce a measure of entanglement using a measure of mixedness. One significant measure of mixedness is the ***von Neumann entropy*** defined for a ***density matrix*** as

$$S^{Neumann}(\rho) = -tr(\rho In\rho) = -\sum_k p_k \, In \, p_k \tag{128}$$

We can define the ***entanglement entropy*** of a pure state Ψ through the von Neumann entropy of its ***reduced density matrix***

$$E\left(\Psi^{(:\mathrm{AB})}\right) = S^{Neumann}\left(\rho^{(/A\ or\ /B)}\right) = -tr\left(\rho^{(/A)}\, In\rho^{(/A)}\right) = -tr\left(\rho^{(/B)}\, In\rho^{(/B)}\right) = \sum_m -\sigma_m^2 \, In \, \sigma_m^2$$

So, we can now use the Schmidt coefficients (non-zero eigenvalues of $\rho^{(/A\ or\ /B)}$)

$$E\left(\Psi^{(:\mathrm{AB})}\right) = -\sum_m \lambda_m \, In \, \lambda_m \tag{129}$$

Taking a step back, it's imperative that we take a look at ***Shannon entropy*** and its statistical thermodynamics and information theory roots. The Shannon entropy is actually the classical counterpart preceding the von Neumann entropy. It quantifies uncertainty for a classical variable and is a function of some probability distribution p. Given p, the Shannon entropy becomes

$$S^{Shannon}(p) = -tr(pInp) = -\sum_k p_k \, In \, p_k \tag{130}$$

With p_k being the probability of measuring some classical value k. The Shannon entropy originates from classical statistical thermodynamics with the Boltzmann entropy is defined as

$$S = k \, log \, W \tag{131}$$

, where k is the Boltzmann constant and W the number of available microstates in the system. In classical statistical thermodynamics, one key assumption was taking all the microstates to be equally probable, meaning we have a single $p = \frac{1}{W}$. Accordingly, The Boltzmann entropy becomes

$$S = k \log p^{-1} = -k \log p \tag{132}$$

Alternatively, by accounting for a probability distribution with the i^{th} microstate existing with probability p_i, the entropy becomes an average by summing over all the weighted microstates making the entropy

$$S = -k \sum_i p_i \log p_i \tag{133}$$

It is always preferable to express physical laws in dimensionless quantities if possible, so we usually drop the k, and will assume dimensionless entropy going forward. In information theory, the Shannon entropy is just an extension of this relation. To get a tighter grasp on what Shannon entropy is, consider a message coded for in N symbols, which were found from past messages to occur with probability p_i. To quantify the amount of information in an N symbol message, it is simply N times the average information per symbol transmitted, and the ***average information per symbol transmission is the Shannon entropy***

$$S^{shannon} = -\sum_i p_i \ln p_i \tag{134}$$

To understand how Shannon entropy quantifies the average information per symbol transmitted, and its relevance to quantum information, consider the following argument [5]. Consider a source producing a string of bits string $[X_1, X_2, \ldots]$ in the form of independent, identically distributed random variables. Shannon investigated the entropy of such a system, by asking what minimal physical resources are required to store the information being produced by the source. Let's say the source produces one of four symbols, $\{W, X, Y, Z\}$. Initially, 2 bits of storage space would be needed to code for the four possible outcomes, so 2 bits are consumed for each event from the source. Now if we add probabilities to different symbols, say W is produced with probability 0.5, X with probability 0.25, and Y and Z both with probability 0.125. We can take advantage of this uneven distribution to ***compress*** the source, meaning utilizing fewer bits for the more common symbols. The most efficient compression scheme here, turns out to be encoding W as the 1-bit string $\{0\}$, X as the 2-bit string $\{1,0\}$, y as the 3-bit string $\{1,1,0\}$, and Z as the 3-bit string $\{111\}$. Weighing all the probabilities against the lengths of the bit string code, we get

$$Length_{average} = (0.5)(1) + (0.25)(2) + (0.125)(3) + (0.125)(3) = \mathbf{1.75\ bits}$$

Therefore, ***on average we need just*** $\mathbf{1.75}$ ***bits to store the information per use of the source***. This is less than the 2 bits we required originally to store the equally distributed symbols. Here's the surprise, if we calculate the Shannon entropy of the source, we get

$$S^{Shannon} = -\sum_i p_i \log_2 p_i$$

$$= -(0.5 * \log_2 0.5) - (0.25 * \log_2 0.25) - (0.125 * \log_2 0.125) - (0.125 * \log_2 0.125) = \mathbf{1.75\ bits}$$

It turns out that the Shannon entropy quantifies the optimal compression state. This entropic quantity arose in response to an operational demand for the definition of entropy in terms of data compression. This highlights the following key philosophy of information theory that:

"fundamental measures of information arise as the answers to fundamental questions about the physical resources required to solve some information processing problem"[5].

Now that we've discussed the motivation behind the informational approach we took to entropy, there are different conventions for expressing the entropy. It is more physical to define entropy using the natural logarithm, however in information theory, some prefer to use $\log_2$, to account for the combinatorics of qubit microstates. Another important convention is how we define $0 \log_2 0$, since $log\ 0$ is undefined. Intuitively, such a state is irrelevant with its zero probability of occurring, so we conventionally drop its contribution to entropy, and agree that $0 \log_2 0 \equiv 0$. Also consider that $\lim_{x \to 0} 0 \log_2 0 = 0$, which further serves our case.

3.3.2 *Von Neumann Entropy as a measure of entanglement entropy in pure bipartite states* [2], [5], [10]

A peculiar adaptation arises when we apply equation (134) to ***probability distributions over quantum states***. The probabilities, as we discussed while introducing the density matrix formulation, are simply the ***eigenvalues of density matrices***, so we get that

$$S^{shannon\ for\ quantum\ probabilities} = -\sum_{quant} p_{quant}\ In\ p_{quant} = -\sum_{m} \lambda_m\ In\ \lambda_m \qquad (135)$$

, where λ_m are the quantum state probabilities, this means the eigenvalues of density matrices can be ***interpreted as a probability distribution*** and we proceed with the Shannon entropy derivation accordingly. A closer look at the last term in the equation (135), shows that it can be translated to a trace of a density matrix function, we can conclude that

$$-\sum_{m} \lambda_m\ In\ \lambda_m = -tr(\rho In \rho) = S^{von\ Neumann} \qquad (136)$$

This is what we define as the ***von Neumann entropy***, so mathematically, the von Neumann entropy is defined as the ***Shannon entropy of the eigenvalues of the density matrix***.

The behavior of this entropy function is shown in figure (3). A system exhibits zero entropy when $\lambda = 0$,1 and peaks at $p = 1/e$. Remember $\lambda \in [0,1]$, a domain over which ln(λ) is always negative, making $-\lambda ln\lambda$ an always positive term.

3.3.3 *Some properties of von Neumann entropy, its subadditivity and the defining difference between many-body quantum and conventional matter.*

1. A pure state in its diagonalizing basis, has a single diagonal entry of 1 in its $\hat{\rho}$ which is also its eigenvalue $\lambda = 1$. This translates to an entropy of zero, so the entropy is bounded from below by

$$S^{Neumann}(\rho_{pure}) = 0 \tag{137}$$

Let's give an example for the von Neumann entropy of a single qubit, and thus we refer to the density matrix of a mixed state (no reference to entanglement). Conversely for a bipartite entangled system we need to refer to the mixedness of their reduced density matrices and the sum of their individual entropies.

First, the density matrix of a single qubit can always be written as

$$\frac{1}{2}(\mathbb{I} + \bar{n} \cdot \bar{\sigma}) = \frac{1}{2}\begin{pmatrix} 1+n_z & n_x - in_y \\ n_x + in_y & 1-n_z \end{pmatrix}$$

as derived in equation (61). We can consider our 2 extreme states, pure and maximally mixed, so we can do so by defining n (the magnitude of the $\bar{n}$ state vector), and the eigenvalues $\frac{(1\pm n)}{2}$. The entropy here is

$$S = -\sum_m \lambda_m \, In \, \lambda_m = -\left(\frac{1+n}{2}\ln\frac{1+n}{2}\right) - \left(\frac{1-n}{2}\ln\frac{1-n}{2}\right)$$

$$n_{pure\ state} = 1\,, \quad \text{and} \quad n_{\text{maximally mixed state}} = 0\,, \quad \text{so}$$

$$S_{pure} = -(1\ln 1) - (0\ln 0) = 0, \qquad S_{maximally\ mixed} = -\left(\frac{1}{2}\ln\frac{1}{2}\right) - \left(\frac{1}{2}\ln\frac{1}{2}\right) = \ln 2$$

Notice that S_{pure} is the lowest value possible for S and $\ln 2$ is the upper limit defined by $\ln\left(d_{Hilbert\ space}\right)$ where the Hilbert space for a single qubit consists of 2 spin states, hence $\ln 2$.

Secondly, consider the Bipartite pure state

$$|\psi\rangle = \sqrt{p}\,(|0\rangle \otimes |1\rangle) + \sqrt{1-p}\,(|1\rangle \otimes |0\rangle) \tag{138}$$

This is completely separable for $p = 0{,}1$ and maximally entangled for $p = \frac{1}{2}$. We now need to emphasize that when calculating the entropy for a pure bipartite system AB, it is sometimes more important to consider the sum of the von Neumann entropy of the reduced density matrices of its subsystems A and B, rather than the entropy of the global system. Keep in mind that the entropy of pure global states will still be zero even if its parts have non vanishing entropies. How one might ask? Well, unlike the Shannon entropy where the ***entropy of substates of a classical bipartite system cannot exceed the Shannon entropy of the global system***, the Von Neumann entropy of an ***entangled bipartite pure quantum system*** is always zero regardless of the condition of its substates, since we know everything about the global system as a whole, and so our ignorance is minimal. However, we lack knowledge of the subsystems on their own, the ignorance on the subsystems A and B is quantified by $S\left(\rho^{(/A)}\right)$ and $S\left(\rho^{(/B)}\right)$. We therefore find that unlike classical systems, in ***quantum systems information can be encoded in the correlations among the parts of the system, which is info available when considering the system as a whole***, yet this info vanishes on considering subsystems one at a time. We touched on that whilst discussing the effect of tracing out a subsystem by taking its partial trace.

Also remember that bipartite entanglement is the quantum correlation between the two subsystems that ***might not be of the same dimension***, but will still have the same non zero Schmidt numbers for pure states, and we get that for ***pure bipartite states***

$$S\left(\rho^{(/A)}\right) = S\left(\rho^{(/B)}\right), \qquad S\left(\rho_{pure}^{(:AB)}\right) = 0 \tag{139}$$

To prove this, we first refer to equation (108) and derive the reduced density matrix using the spectral decomposition giving

$$\hat{\rho}_{pure}{}^{(/A)} = |a\rangle\langle a'|\ \delta_{b,b'} = p\,|0\rangle\langle 0| + (1-p)|1\rangle\langle 1|$$

So

$$S\left(\rho^{(/A)}\right) = S\left(\rho^{(/B)}\right) = -pln(p) - ((1-p)\ln(1-p))$$

and again, we can see that for maximally entangled states, we have $p = \frac{1}{2}$, so

$$S\left(\rho^{(/A)}\right) = S\left(\rho^{(/B)}\right) = \ln 2, \tag{140}$$

so, the entropy of a subsystem of maximally entangled bipartite states, is analogous to that of a maximally mixed state of a single qubit.

2. The von Neumann entropy is also bounded from above. To find the bound we use Lagrange multipliers to maximize the equation (134) using the normalization constraint $\sum_m \lambda_m = 1$.

$$\nabla_\lambda \left[-\sum_m^N \lambda_m \, In \, \lambda_m \right] = \Lambda \, \nabla_\lambda \left[\sum_m^N \lambda_m - 1 \right] \tag{141}$$

So

$$-\sum_m^N \left(\frac{\lambda_m}{\lambda_m} + In \, \lambda_m \right) = \Lambda \sum_m^N 1$$

$$ln(\lambda_m) = -\Lambda - 1 \ , \qquad \lambda_m = e^{-\Lambda - 1} \ldots\ldots (i)$$

Using

$$\sum_m \lambda_m = 1$$

So

$$\sum_m^N e^{-\Lambda-1} = 1 \ , \qquad e^{-\Lambda-1} \sum_m^N 1 = 1 \ , \qquad e^{-\Lambda-1} N = 1 \ , \qquad N = \frac{1}{e^{-\Lambda-1}} \ldots\ldots (ii)$$

From (i) and (ii) we get that

$$\lambda_m = \frac{1}{N} \tag{142}$$

This gives the maximum entropy

$$S^{Neumann}(\rho)_{MAX} = -\sum_m^N \frac{1}{N} \, In \frac{1}{N} = -\frac{1}{N} \, In \frac{1}{N} \sum_m^N 1 = -In \frac{1}{N} = lnN \tag{143}$$

For the density matrices the N is the dimension of our Hilbert space, which is the Schmidt rank for bipartite reduced density matrices, so

$$S^{Neumann}(\rho)_{MAX} = ln(d_{hilbert}) = ln(\min[d_A, d_B]), \qquad all\ \lambda_m = \frac{1}{N}, \qquad \rho_{MAX} = \frac{\mathbb{I}}{d_{hilbert}}$$

The ρ_{MAX} and λ_m results are exactly similar to the maximally mixed and entangled states we derived earlier, meaning that the maximum entropy occurs when the system is maximally mixed. The von Neumann entropy therefore varies between 0 for pure states and ln(d) for maximally mixed states, making

$$0 \leq S^{Neumann} \leq ln(\min[d_A, d_B]) \tag{144}$$

3. We establish the ***subadditivity*** property of the von Neumann entropy (a quantum generalization of subadditivity of Shannon entropy) in the relation

$$S\left(\rho^{(:AB)}\right) \leq S\left(\rho^{(/A)}\right) + S\left(\rho^{(/B)}\right) \tag{145}$$

equality is only achieved when the global state is a product state

$$\rho^{(:AB)} = \rho^{(/A)} \otimes \rho^{(/B)}$$

Which leads to

$$S\left(\rho_{product}^{(:AB)}\right) = S\left(\rho^{(/A)} \otimes \rho^{(/B)}\right) = S\left(\rho^{(/A)}\right) + S\left(\rho^{(/B)}\right)$$

We can derive this relation from the following

$$S\left(\rho^{(:AB)}\right) = -tr\left(\rho^{(:AB)} ln \rho^{(:AB)}\right) = -tr\left(\rho^{(:AB)} \ln\left(\rho^{(/A)} \otimes \rho^{(/B)}\right)\right) \tag{146}$$

Given $\quad ln\left(\rho^{(/A)} \otimes \rho^{(/B)}\right) = ln\left(\rho^{(/A)}\right) \otimes \mathbb{I}^{(B)} + \mathbb{I}^{(A)} \otimes ln\left(\rho^{(/b)}\right)$

so

$$S\left(\rho^{(:AB)}\right) = -tr\left(\rho^{(:AB)}\left(ln\left(\rho^{(/A)}\right) \otimes \mathbb{I}^{(B)} + \mathbb{I}^{(A)} \otimes ln\left(\rho^{(/b)}\right)\right)\right)$$

$$= -tr\left(\left(\rho^{(/A)} \otimes \rho^{(/B)}\right)\left(ln\left(\rho^{(/A)}\right) \otimes \mathbb{I}^{(B)} + \mathbb{I}^{(A)} \otimes ln\left(\rho^{(/b)}\right)\right)\right)$$

$$= -tr\left(\rho^{(/A)} ln\left(\rho^{(/A)}\right) + \rho^{(/B)} ln\left(\rho^{(/b)}\right)\right)$$

$$= -tr\left(\rho^{(/A)} ln\left(\rho^{(/A)}\right)\right) - tr\left(\rho^{(/B)} ln\left(\rho^{(/B)}\right)\right)$$

So, for product states

$$S(\rho^{(:AB)}) = S(\rho^{(/A)}) + S(\rho^{(/B)}) \tag{147}$$

We see that for a bipartite pure product state AB, we get pure substates A and B, making their entropies also zero so this is the only equality case, when

$$S(\rho^{(:AB)}) = S(\rho^{(/A)}) = S(\rho^{(/B)}) = 0 \tag{148}$$

Using the Klein inequality (explained with relative entropy at the end of the property list) the proof of subadditivity becomes a simple application of Klein's inequality, $S(\hat{\rho}||\hat{\sigma}) \geq 0$, where $S(\hat{\rho}||\hat{\sigma}) = tr(\hat{\rho}\log_2(\hat{\rho}) - \hat{\rho}\log_2(\hat{\sigma}))$, therefore by letting $\hat{\rho} = \rho^{(:AB)}$, and $\hat{\sigma} = \rho^{(/A)} \otimes \rho^{(/B)}$, we get

$$S(\rho^{(:AB)}||\rho^{(/A)} \otimes \rho^{(/B)}) \geq 0$$

so

$$tr(\rho^{(:AB)}\log_2(\rho^{(:AB)}) - \rho^{(:AB)}\log_2(\rho^{(/A)} \otimes \rho^{(/B)})) \geq 0$$

$$tr(\rho^{(:AB)}\log_2(\rho^{(:AB)})) \leq -tr(\rho^{(:AB)}\log_2(\rho^{(/A)} \otimes \rho^{(/B)}))$$

$$S(\rho^{(:AB)}) \leq -tr(\rho^{(:AB)}\log_2(\rho^{(/A)} \otimes \rho^{(/B)}))$$

We have already showed in equation (146) that the right-hand term reduces to $S(\rho^{(/A)}) + S(\rho^{(/B)})$

making

$$S(\rho^{(:AB)}) \leq S(\rho^{(/A)}) + S(\rho^{(/B)}) \tag{149}$$

This ***subadditivity property*** of the von-Neumann entropy is vital to understanding one of the tenants of quantum information, that by taking the partial trace, entropy increases, this corresponds to one losing information. This aligns with what we deduced before regarding the ambiguity in transitioning from density matrices of bipartite states to their constituents' reduced density matrices, which is a partial trace operation. There is a more subtle implication of this, and it lies at the defining border between ***quantum and conventional matter***. In conventional matter, we use Shannon entropy, which, as we discussed in property 1, the Shannon entropy of substates of a classical bipartite system cannot exceed the Shannon entropy of the global system, in contrast to the Von Neumann entropy of an entangled bipartite pure quantum system. In a conventional system, looking at enough local sites can give one enough information to deduce the behavior of an entire system, and thus the ambiguity in the parts never exceeds the whole. The same is not true for a quantum system, and the subadditivity of the von Neumann entropy shows this. Without considering the system as a whole, we get an increase in entropy, i.e. in the ambiguity of the state of a system. Therefore, in quantum matter, looking at the local substates, will always leave the observer with missing information about the behavior of the system as a whole. This information that is strictly quantum in nature, it is actually the quantum correlations between the substates, i.e. the entanglement between substates, which results from the interactions between the system's subparts. We will discuss the nature of these interactions and how local hamiltonians depict them in the next section.

4. Since calculating the von Neumann entropy is a trace operation, and traces are cyclic, so the von Neumann entropy of a state density matrix is ***invariant under unitary transformation***,

$$S\left(U\rho U^{\dagger}\right) = S(\rho) \tag{150}$$

Since $U^{\dagger}U = \mathbb{I} \quad tr\left(U\rho U^{\dagger}\right) = tr\left(U^{\dagger}U\rho\right) = tr(\rho)$,

This time evolving state takes the form we derived for the time evolution of density matrices, the integral form of the von Neumann equation. We therefore know that for time independent hamiltonians***, time evolution*** is implemented by unitary transformations, which means that the entropy, and similarly the purity, are ***constants of motion***, and we conclude that ***unitary evolutions do not change the mixedness of a state***, which if illustrated on a Bloch sphere, equates to the preservation of the length of the state vector $\bar{n}$. Remember that this applies to closed quantum systems, for open quantum systems this will no longer be the case, as we have shown with decoherence.

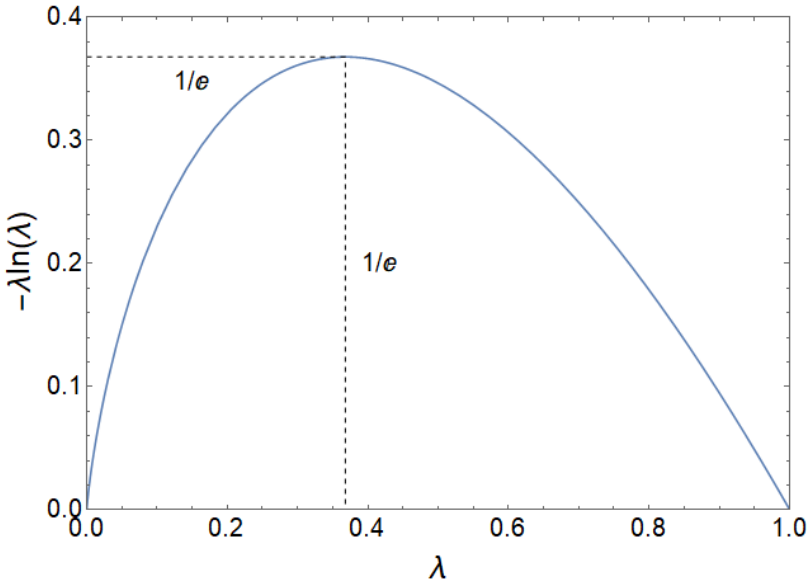


Figure (3): Behavior of the von Neumann entropy (ln base) as a function of the Schmidt coefficients

Another relevant quantity we should mention is the ***conditional entropy*** $S(X|Y)$***.*** The conditional entropy quantifies the remaining ignorance about substate B after substate A is measured. If knowing B makes one certain of the state of A, we have zero conditional entropy. Using X, Y for classical states and A, B for quantum states, we find that the ***Shannon conditional entropy is always positive***

$$S^{Shannon}(X|Y) \;=\; S^{Shannon}(XY) \;-\; S^{Shannon}(Y) \;\geq\; 0. \tag{151}$$

Compare this to the von Neumann conditional entropy where $S^{Von\ Neumann}(AB)$ can be

$$S^{Von\ Neumann}(A|B) \;=\; S^{Von\ Neumann}(AB) \;-\; S^{Von\ Neumann}(B)$$

$$which\ \boldsymbol{is\ negative\ for\ entangled\ pure\ states}$$

So, for a bipartite pure state, if the global system is entangled, then $S^{Von\ Neumann}(A|B) < 0$

3.3.4 *Relative entropy, the distance between density matrices*

Another quantity to consider is the ***relative entropy.*** For two density matrices $\hat{\rho}_A$ and $\hat{\rho}_B$, the quantum relative entropy of $\hat{\rho}_A$ with respect to $\hat{\rho}_B$, $S(\hat{\rho}_A||\hat{\rho}_B)$ is given by

$$S(\hat{\rho}_A||\hat{\rho}_B) = tr(\rho_A \log_2(\rho_A) - \rho_A \log_2(\rho_B)) \tag{152}$$

The relative entropy is a measure of the closeness of the two probability distributions represented by the $\lambda_i^{(A)}$ and $\lambda_i^{(B)}$ eigenvalues of the density matrices $\hat{\rho}_A$ and $\hat{\rho}_B$, and we will give the motivation as to why it is regarded as a ***distance-like measure***. An important property of the relative entropy is its ***non-negativity.***

$$S(\hat{\rho}_A||\hat{\rho}_B) \geq 0\ , \quad and\ only\ equal\ iff \qquad \hat{\rho}_A = \hat{\rho}_B \tag{153}$$

Which makes sense, since if distributions are identical, we expect the "distance" between the two states to be zero, so this result is ensuring. The derivation adds nothing to the conversation, so one can look up the derivation in many sources, including reference [5], page 513. This vital relation is called ***Klein's inequality.*** We've used this inequality in property 3, of the von Neumann entropy to prove subadditivity.

We can use this relation to derive our original statement that $S^{Neumann} \leq \ln or \log_2(d_{Hilbert})$. Looking at equations (152,153)

$$tr(\rho_A \log_2(\rho_A) - \rho_A \log_2(\rho_B)) \geq 0\ , With\ equality\ iff\ distributions\ are\ identical$$

Identical distributions entails that

$$\text{all the probabilities } \lambda_i^{(A)} \text{ and } \lambda_i^{(B)} \text{are } \frac{1}{d_{Hilbert}}\ , meaning\ \hat{\rho}_A = \hat{\rho}_B = \frac{\mathbb{I}}{d_{Hilbert}},$$

making

$$tr(\rho_A \log_2(\rho_A) - \rho_A \log_2(\rho_B)) = 0$$

So

$$tr(\rho_A \log_2(\rho_A)) = tr(\rho_A \log_2(\rho_B)) \tag{154}$$

$$Knowing\ S^{von\ Neumann}(\rho_A) = -tr(\rho_A\ log_2(\rho_A)), \quad and\ \lambda_i^{(A)} = \lambda_i^{(B)} = \frac{1}{d_{Hilbert}},\ \ so$$

$$S^{von\ Neumann}(\rho_A) = -\sum_{i=1}^{d_{Hilbert}} \frac{1}{d_{Hilbert}} log_2\left(\frac{1}{d_{Hilbert}}\right) = -d_{Hilbert} * \frac{1}{d_{Hilbert}} log_2\left(\frac{1}{d_{Hilbert}}\right)$$

$$= -log_2\left(\frac{1}{d_{Hilbert}}\right) = log_2(d_{Hilbert})$$

So

$$S^{von\ Neumann}(\rho_A) = log_2(d_{Hilbert}) \tag{155}$$

We just showed that the maximum value of $S^{von\ Neumann}$ is $log_2(d_{Hilbert})$, and given the minimum is zero we have that

$$0 \leq S^{von\ Neumann}(\rho_A) \leq log_2(d_{Hilbert}) \tag{156}$$

which is when we have a maximally distributed state. We have showed before that this corresponds to a maximally mixed substate $A\ or\ B$, and in turn, these substates together constitute a maximally entangled pure bipartite global state.

Using the relative entropy, we can attempt ***to define the information we have about a quantum state*** as $I\left(A; \frac{\mathbb{I}}{d_{hilbert}}\right)$, so the more you know about a state, the less the ambiguity, the less the entropy, the less the number of bits needed to store the information and the larger the relative entropy,

$$I\left(A; \frac{\mathbb{I}}{d_{hilbert}}\right) = S\left(\rho^{(:AB)} || \frac{\mathbb{I}}{d_{hilbert}}\right) = S\left(\frac{\mathbb{I}}{d_{hilbert}}\right) - S\left(\rho^{(:AB)}\right)$$

So

$$I\left(A; \frac{\mathbb{I}}{d_{hilbert}}\right) = log_2(d_{Hilbert}) - S\left(\rho^{(:AB)}\right) \tag{157}$$

Such that if we wish to ***quantify information***, we can consider the distance between our state and a reference state, that reference being naturally the maximally mixed state $\frac{\mathbb{I}}{d_{hilbert}}$, which is the state for which we have zero information, and thus zero relative entropy, and thus the state with the highest entropy.

3.3.5 Mutual Information and quantifying correlations

Finally, if we look at the relative entropy and the subadditivity, we can identify a very important, strictly positive quantity, called the ***mutual information*** $I(A;B)$. We define the mutual information as

$$I(A;B) = S\left(\rho^{(:AB)} || \rho^{(/A)} \otimes \rho^{(/B)}\right) \tag{158}$$

so

$$I(A;B) = tr\left(\rho^{(:AB)} \log_2\left(\rho^{(:AB)}\right) - \left\{\rho^{(/A)} \log_2\left(\rho^{(/A)} \otimes \rho^{(/B)}\right)\right\}\right)$$

$$= -S\left(\rho^{(:AB)}\right) + \left(\left\{\rho^{(/A)} \log_2\left(\rho^{(/A)} \otimes \rho^{(/B)}\right)\right\}\right) = -S\left(\rho^{(:AB)}\right) + \rho^{(/A)} \log_2\left(\rho^{(/A)}\right) + \rho^{(/A)} \log_2\left(\rho^{(/B)}\right)$$

$$= -S\left(\rho^{(:AB)}\right) + S\left(\rho^{(/A)}\right) + S\left(\rho^{(/B)}\right)$$

And since $S\left(\rho^{(:AB)}\right) \leq S\left(\rho^{(/A)}\right) + S\left(\rho^{(/B)}\right)$

So

$$I(A;B) = S\left(\rho^{(/A)}\right) + S\left(\rho^{(/B)}\right) - S\left(\rho^{(:AB)}\right) \geq 0 \tag{159}$$

Since mutual information is always positive, and only zero when the two systems are uncorrelated i.e. $\rho^{(:AB)} = \rho^{(/A)} \otimes \rho^{(/B)}$, this makes it a great candidate for a ***quantifier of the total degree of correlations***. It represents the amount of information stored in AB which is not stored in A and B separately. Note that It measures correlation in general and not just quantum correlations (entanglement), this suggests it will have a role to play in entangled mixed multipartite states and other quantum-classical hybrid systems.

Consider the following example [10], given a typical mixed substate $\rho^{(/A)} = \rho^{(/B)} = \begin{pmatrix} p & 0 \\ 0 & 1-p \end{pmatrix}$ and how the correlations in $\hat{\rho}_{AB}$ are related to the populations through the inequality $|\alpha| < p(1-p)$

$$\hat{\rho}_{AB} = \begin{pmatrix} p^2 & 0 & 0 & 0 \\ 0 & p(1-p) & \alpha & 0 \\ 0 & \alpha & p(1-p) & 0 \\ 0 & 0 & 0 & (1-p)^2 \end{pmatrix} \tag{160}$$

Plugging in these matrices into Mathematica we immediately get the mutual information

$$I(A;B) = p(1-p) \log_2\left(\frac{p^2(1-p)^2 - \alpha^2}{p^2(1-p^2)}\right) + \alpha \log_2\left(\frac{p(1-p)+\alpha}{p(1-p)-\alpha}\right) \tag{161}$$

The larger the α the greater the mutual information, where due to the upper limit provided by the population terms, the maximum mutual information reaches $2p(1-p)\,ln(2)$. If we take into consideration that $\rho^{(:AB)}$ is a pure state, we know that it follows that $\rho^{(/A)} = \rho^{(/B)}$ and $S\left(\rho^{(:AB)}\right) = 0$. Hence, the mutual information becomes

$$I(A;B) = 2\,S\left(\rho^{(/A)}\right) = 2S\left(\rho^{(/B)}\right) \tag{162}$$

This means that for pure states the maximum amount of information coded into correlations is twice the information of each of the subsystems. With this more general quantifier of quantum correlations, we now have means of addressing both the entanglement in pure multipartite systems and mixed states.

4.0.0 Section 4: Tensor networks and the Matrix Product State (MPS) formulation

4.1.0 Many-Body wavefunction of a pure multipartite system

As we have elaborated many times by now, all the info of a pure system of quantum matter can be found in its Many-body wavefunction, and we have specified how quantum matter involves entanglement. To illustrate further with an example, let's take a 3-qubit system, it will have 2^3, so 8, unique eigenstates, making up our Global Hilbert space. The most general many-body wavefunction will be a weighted sum of those 8 states. The $\psi^{\uparrow\uparrow\downarrow}$ (whose superscripts reference the spin combination of that unique state), is the coefficient of its corresponding state $|\uparrow\uparrow\downarrow\rangle$, which is also shorthand for $|\uparrow\rangle\,|\uparrow\rangle\,|\downarrow\rangle$ and $|\uparrow\rangle \otimes |\uparrow\rangle \otimes |\downarrow\rangle$. Accordingly, we get the following most general 3 qubit wavefunction:

$$|\psi\rangle \equiv \psi^{\uparrow\uparrow\uparrow}|\uparrow\uparrow\uparrow\rangle + \psi^{\uparrow\uparrow\downarrow}|\uparrow\uparrow\downarrow\rangle + \psi^{\uparrow\downarrow\uparrow}|\uparrow\downarrow\uparrow\rangle + \psi^{|\downarrow\uparrow\uparrow}|\downarrow\uparrow\uparrow\rangle + \psi^{\uparrow\downarrow\downarrow}|\uparrow\downarrow\downarrow\rangle + \psi^{\downarrow\uparrow\downarrow}|\downarrow\uparrow\downarrow\rangle + \psi^{\downarrow\downarrow\uparrow}|\downarrow\downarrow\uparrow\rangle + \psi^{\downarrow\downarrow\downarrow}|\downarrow\downarrow\downarrow\rangle$$

We can also write more compactly for any arbitrary number N of bodies:

$$\sum_{s_1 s_2 s_3 \dots s_N} \psi^{s_1 s_2 s_3 \dots s_N} |s_1 s_2 s_3 \dots s_N\rangle \tag{163}$$

Where $\psi^{s_1 s_2 s_3 \dots s_N}$ is the "***Amplitude Tensor***", which is a legitimate ***tensor*** in the sense that it's a multidimensional array with 2^N unique entries, acting as coefficients to corresponding 2^N unique states, and for the first time the physics starts hinting at some potentially significant overlooked geometry. The states corresponding to the amplitude tensor components could in turn be extended to an equal number of 2^N-component ***state vectors***. Now, with everything encoded in the Amplitude tensor, the question becomes, how efficiently can one extract the physics from the given state, without representing the quantum state by giving the coefficients of the wavefunctions in some huge basis.

Although the many-body Hilbert space is too large to consider in its entirety like we just argued, it turns out that not that all of it is equally important. Our goal becomes ***figuring out what parts of the many-body Hilbert space are relevant***. By addressing the ground states of multipartite systems with ***short-range interaction Hamiltonians***, one uncovers a behavior unique to these groups called ***Area Laws of entanglement entropy***[6], [9], [17]–[19]**.** One currently novel approach is through implementing "***Tensor Networks***". Tensor network states are the subject of this section, where we address their relation to the

area law of entanglement entropy, Schmidt decomposition and graph theory. Armed with these tools, one could devise novel approaches to many -body condensed matter physics, including ***tensor networks***[6], [7], [12], [17], [19], ***Matrix Product States (MPS)*** and ***DMRG***[12], [13], [17].

4.2.0 *Area laws for entanglement entropy* [6], [9], [18], [19]

So far, we have not specified any interactions between lattice points in a many-body system, nor given any reference to temperature. The models used thus far have illustrated simple cases of entanglement for 2 to 3-body states, specifically qubits, with no reference to energy states, ground levels and their degeneracy, equilibrium conditions or the nature of the interactions governing entanglement. To address useful cases, we need to examine how many-body systems interact, say in a lattice, given how prominent this arrangement is in nature, and how well multidimensional lattice structures model the physics of quantum matter. It seems however, that some of the most interesting and representative properties of quantum matter in many-body systems, are studied when the system is in ***thermal equilibrium*** with short range interactions, most commonly, ***nearest neighbor interactions***. Our preferable Hamiltonian to study is therefore one that is gapped, local Hamiltonian. We will focus here on one key property of such Hamiltonians, one that underlies the behavior of systems in equilibrium, and thus one that strongly characterizes many-body quantum states as they appear in Nature. This one property is conveniently termed the ***Area Law***. The concept of area laws exists for both zero and finite temperature systems. The first zero temperature case is the one we are interested in, given how close it models quantum matter within the applications of quantum information. Consequently, we confine our analysis to pure states and can thus use ***entanglement entropy and a bipartition of our system***. In the latter case, a finite temperature induces a randomness that would warrant more general quantifiers of entanglement, forcing a switch from entanglement entropy to ***quantum mutual information***, since quantum mutual information is a more general quantifier of correlations.

Our model is a spin system with sites represented by a d-dimensional lattice, where we will be focusing primarily on 1-D, and sometimes 2-D lattices for simplicity. Since spins interact according to a Hamiltonian $\widehat{H}$, we need to consider many-body system properties such as entanglement and entropy under the influence of their respective Hamiltonian. Hamiltonians will therefore, through factors like correlation length, govern the nature of the interactions, and in turn entanglement, in a lattice. As discussed previously we can extend the bipartite entanglement model to multipartite systems, so most of our discussion will reference bipartite systems. In turn these 2-part models may be extended, if need be, to more partitions. In doing so, we will mostly reference finite-range interactions down to, in many cases, nearest-body interactions.

4.2.1 *Gapped local hamiltonians and the Area law at zero temperature*

To introduce area laws, we look at a gapped pure system AB at temperature $T = 0$. Zero temperature renders the system in the ground state of the gapped Hamiltonian $\widehat{H}$. Having a gap in the energy levels implies that for the thermodynamic limit, i.e. as the system size $N \to \infty$ there is a persistent energy gap Δ $(\text{which is usually a function of the size of the system})$, between the ground state energy E_0, and the first excited state energy. This ensures that no matter how large the state gets; its ground state is stable under minor temperature changes or any perturbation, and thus presents stable physics, so

$$\Delta_N \geq \gamma > 0 \quad \forall N \,, \tag{164}$$

Where γ, is some constant non zero value. The eigenvectors corresponding to the ground state energy of the Hamiltonian, form a ***Hilbert space*** $\mathcal{G}$ called the ***ground space***, a ***subspace of the global Hilbert space*** $\mathcal{H}$. We focus on ground states since they can often capture most of the low temperature physics. The gapped case ensures that interactions decay exponentially at distances beyond the nearest or $2^{nd} to\ 3^{rd}$ nearest neighbors, which is ideal for our finite interaction model. These interactions translate into correlations between regions. Given a gapped system at $T = 0$, with a non-degenerate ground state, all correlation functions $C(r)$ will exponentially decay with r, the distance between any 2 regions, giving

$$C(r) \sim e^{-r/\xi} \tag{165}$$

where, ξ is the ***correlation length*** of the system, so the smaller the correlation length, the faster these interactions decay and the more local their influence. The correlation function nature as $r \to \infty$ is a key attribute of many body systems, since it reflects many of the physical properties of the global multipartite system. It is important to mention that if the system is gapless, correlation functions in the ground state decay polynomially with r, giving rise to different properties of matter.

4.2.2 Corner of Hilbert space carrying low energy eigenstates of gapped local Hamiltonians

The fact that degrees of freedom can interact only locally with one-another in many condensed matter systems, puts a strong constraint on the amount and form of entanglement that can be present in many-body systems. To be more specific, ***low-energy eigenstates*** of most realistic Hamiltonians ***(gapped Hamiltonians with local interactions)*** are certain states in the Hilbert space, ***ones that are forced by locality to obey the area-law of entanglement entropy***. The ***states that follow the area law form a very small manifold of quantum states*** occupying a corner of the total Hilbert space. Accordingly, if a quantum state is arbitrarily selected from the many-body Hilbert space, it will most likely exhibit an entanglement entropy between its subregions, which breaks the area law and scales with the number of particles, i.e. volume. Stating this finding differently, we can say that ***not all quantum states in the Hilbert space can be a low-energy state of a gapped, local Hamiltonian, and only those satisfying an area-law are valid candidates***. If we are to study the low temperature physics of many-body quantum matter, it would be

most efficient to ***target this corner***. It turns out that, by construction, TN states can target the states abiding by the area law of entanglement entropy, and we will elaborate on this extensively in this section.

4.2.3 1-D, 2-D and 3-D area laws and the upper limit they place on entanglement entropy

To introduce area laws, consider a connected smooth region A of on a 2-D lattice and the complementary region B as shown in figure (4). In the previous section, we pitched the entanglement entropy $E(\rho^{(:AB)})$ between the two regions of AB, as being the von Neumann entropy of one of its reduced density matrices, say $\rho^{(/A)}$. Being an extensive property, the entanglement entropy grows as the regions grow. One would expect that it scales up with the total number of spins inside region A, which is in fact the case for random states beyond the ground state. However, for ground states of the Hamiltonians, a peculiar relation arises, owing to the locality of finite interactions (reflected in the correlation function) and the presence of the energy gap, the entanglement between region A and the bordering region B, instead scales with the number of ***particles bordering*** the region A, $N_{\delta A}$, which is in a sense the "perimeter" or boundary, surrounding a region. Consequently, the entanglement scales with the boundary of a region, so the perimeter of a given area, or the area of a given volume, hence this scaling behavior, takes the poorly representative name of ***area law***. To better illustrate this, consider how the entanglement entropy of the subregion A would be proportional to L^2 for generic many-body entangled states, however, for ***gapped ground states*** the entanglement entropy would be proportional to the length of the boundary L, so

$$S_{A(2D)} \sim \alpha L \quad , \quad S_{A(3D)} \sim \alpha L^2 \tag{166}$$

The interaction shifting to the boundaries is illustrated in the shaded area shifting to a shaded perimeter in figure (4).

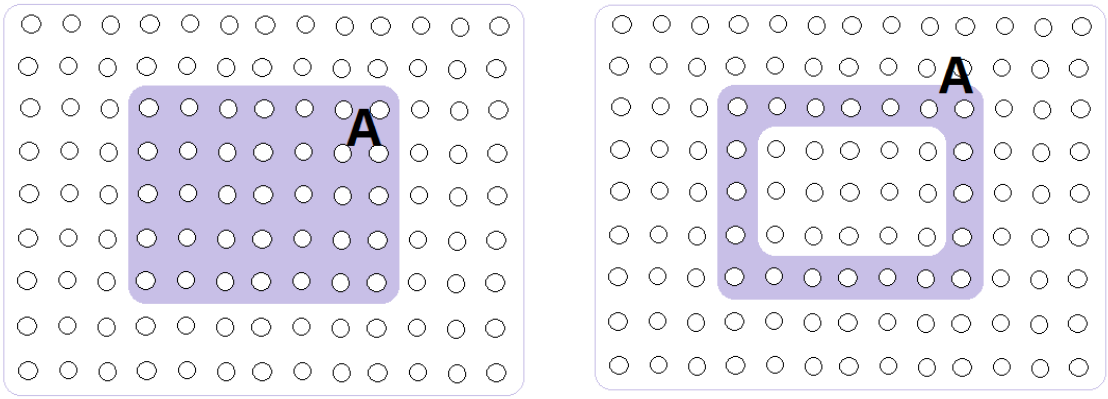


Figure (4): an illustration of how short interactions between two regions can lead to a situation where only bodies at the border effectively interact giving rise to an area law behavior.

We can conclude that a ***gapped ground state*** in a ***locally interacting system*** always exhibits much less entanglement than a generic quantum many-body entangled state. This minimalism is why we are interested in such states, as it seems the entanglement entropy can be concentrated at the borders of massive Hilbert spaces which get exponentially large the larger the involved subregion. A particularly interesting case is when, a one-dimensional system has 2 points, one at each end, as its boundary. This 2-point boundary is independent of the size of either region; therefore, entanglement entropy of a one-dimensional gapped system in the ground state, is bounded by a constant maximum vale

$$S_{A(1D)} \leq Const \tag{167}$$

For the limit of no interactions, and thus no correlations, between adjacent subregions, the ground state becomes a product state and in turn entanglement vanishes. It is important to emphasize that if the locality behavior of interactions is removed, the area law is violated. Other violations of the area law occur if the system is not gapped as well, however this is beyond our scope here, and we only focus on gapped ground states pure systems with finite local interactions for their usefulness, as we will discuss in tensor networks, and their utility in approximating complex systems.

Finally, we emphasize that the ***coefficient of the area law scaling,*** α, in 2D and 1D, as well as the constant bound in 1D, strongly reflect the nature of the interactions in the system. For a closer look, consider the ***1D transverse Ising model*** with nearest neighbor interactions, described by the Hamiltonian

$$\widehat{H}^{Ising} = -J\sum_{i} S_i^Z S_{i+1}^Z - B\sum_{i} S_i^X \tag{168}$$

The model describes spins of adjacent sites that interact according to J and tend to align to the Z-direction to lower the energy of the spin chain. Interaction only occurs between sites i and $i+1$,(nearest neighbors). The second term expresses the competing reaction of the spin to an external field interacting with the lattice at site i which tends to align the spins along X (transverse field). This transverse effect is proportional to B and is not a correlation. The model can exhibit several cases, the first being a gapped Hamiltonian as long as $J \neq B$. Secondly if $|J| > |B|$, the ground state is degenerate. Thirdly if $|J| < |B|$, the ground state is unique. Lastly if there is ***no interaction, i.e. $J = 0$, the ground state is a product state,*** so entropy vanishes for both global states and any subregion.

4.3.0 Tensor networks in many-body quantum matter and their relation to area laws

Having identified area laws and the existence of unique attributes of ground state gapped systems in thermal equilibrium, we can study this group of many-body quantum states that readily appear in Nature. As you might have guessed, we can utilize this behavior to target the exponentially small corner of relevant states in the gigantic Hilbert space. One of the most prominent tools that utilize these local properties are the tensor network states (TNS), which target this corner directly. A particular TNS technique that was shown to be very effective in modeling 1D gapped latices, is the matrix product states (MPS). MPS are the cornerstone of some of the most widely used numerical methods in condensed matter physics, including

density matrix renormalization group (DMRG.) Renormalization Group (RG) methods for many-body systems aim to tackle relevant degrees of freedom to describe a system, and are perfect candidates for targeting this relevant corner of quantum states by relying on TN states.

4.3.1 ***Tensor network theory, MPS and Tensor diagrams*** [6], [7], [12], [17], [19]

There are many ways to introduce TNS and MPS, we will try and approach this from a practical perspective first, addressing the utility of the tensorial representation of many-body global quantum states, and their diagrammatic equivalence in tensor diagrams, then later delve into the mathematical rigor of this representation. We will focus on 1-D lattices for simplicity, and given they are the best established in literature, specifically MPS. We have already seen, in section 4.1, how we get an amplitude tensor $\psi^{s_1 s_2 s_3 \dots s_N}$ in the representation

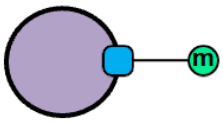
$$|\psi\rangle = \sum_{s_1 s_2 s_3 \dots s_N} \psi^{s_1 s_2 s_3 \dots s_N} |s_1 s_2 s_3 \dots s_N\rangle \tag{169}$$

This representation hints at the tensorial nature of multipartite states, and we will see in a moment how this transition will be made. First, we define a Tensor Network (TN), which is a set of tensors where some, or all, of its indices are contracted following some pattern. These tensors abide by the basic operations of tensor algebra, for example, summing over a repeated index is analogous to contracting the TN. Consider these two examples of tensor networks,

$$M_{ij} = \sum_{k=1}^{\dim_k} A_{ik} B_{kj} \; ; \; S = \sum_{k=1}^{\dim_k} C_k D_k \tag{170}$$

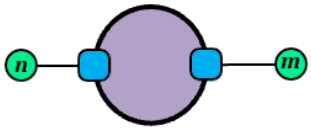
The tensor network depicted by $M_{ij} = \sum_{k=1}^{\dim_k} A_{ik} B_{kj}$ is equivalent to a matrix product between matrices A and B, which contract to the matrix M. The components of M are M_{ij}, where i and j are the ***open indices*** remaining after the contraction over the dummy index k. The second tensor network $S = \sum_{k=1}^{\dim_k} A_k B_k$, results in the contraction of two vectors into a scalar, and is equivalent to a tensor product of vectors C and D, leaving S with no open indices. Keeping track of indices becomes exponentially harder the larger the Hilbert space of the constituent tensors, we therefore look to ***tensor diagrams***, which are pictorial illustrations of these networks. The tensor diagram representation and will play an irreplaceable role later on when utilizing tensor networks to model lattices in many-body quantum matter. Tensor network diagrams offer novel visualization and utility by introducing a new language to quantum physics, one that makes everything much more visual and brings new intuitions into play. To restate this for emphasis before we move on, a tensor network is a diagram which guides us into combining several tensors (contraction) into a single composite tensor, whose rank is equal to the number of free legs (open indices) in the diagram.

The following are tensor diagram representations of different tensor networks, an application called ***Tensor Trace*** ♦ was used here to help form these diagrams [20]:



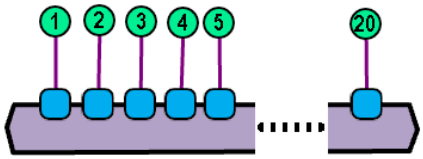
T_m

Figure (5): Vector



$T_{n,m}$

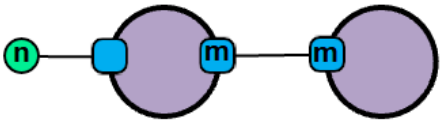
Figure (6): Matrix



$T_{m_1,m_2,m_3\ldots..m_{20}}$

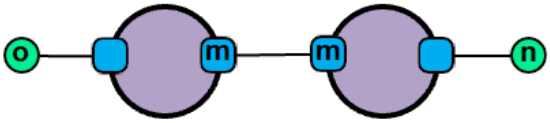
Figure (7): Rank 20 Tensor

Each partition is a tensor, depicted by the circular nodes. Tensor indices are denoted by the arms on the diagram. The tensor diagram in figure (5) has a single open index, which means it does not share in any interactions, it is simply a vector T_m, a rank-1 tensor. The tensor diagram in figure (6), has 2 open indices, and is therefore a rank-2 tensor $T_{n,m}$. The tensor diagram in figure (7) is a larger tensor with 20 open indices, it's a $T_{m_1,m_2,m_3\ldots..m_{20}}$ rank-20 tensor with 20 open indices.



$$T_n = \sum_m V_{n,m} U_m$$

Figure (8): Matrix acting on a Tensor



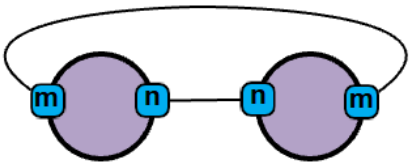
$$T_{o,n} = \sum_m V_{o,m} U_{m,n}$$

Figure (9): Matrix Multiplication

Joined arms or ***"bonds"*** between two partitions denote interactions and are the equivalent of contractions in tensor operations. Therefore, the tensor diagram in figure (8) is a contraction between V and U over the index m, as shown by the $m - m$ bond between the 2 tensors in the diagram. The resulting tensor T_n is a rank-1 tensor, as is depicted by the single free arm, the sole open index n. This diagram shows that the result of the product of a matrix (the partition with 2 indices n and m) and a vector (the partition with the index m), is another vector (the combined state has a single open index n). The Tensor diagram in figure (9) is also a contraction over the index m, but this time the resulting tensor is a rank-2 tensor $T_{o,n}$ as depicted by the 2 free arms o and n. This is actually a matrix multiplication, and the diagram shows

♦ Tensor Trace was used here to inspire our own notations for these diagrams [19] we used additional customized labeling and added some modifications using Microsoft PowerPoint

that the product of 2 matrices, one with indices o, m and the other with indices m, n, is another matrix as indicated by the resulting tensor's 2 open indices o and n.



$$T = \sum_{m,n} V_{m,n} U_{n,m} = Tr(VU)$$

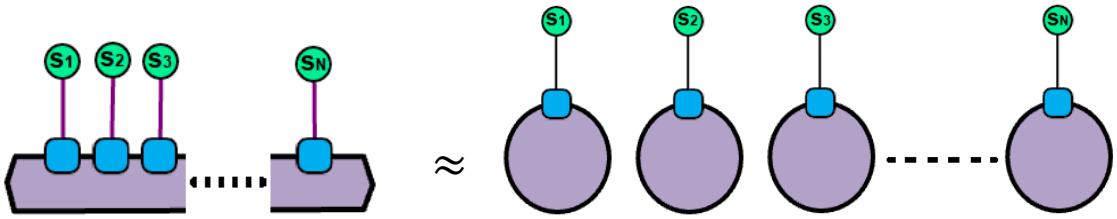
Figure (10): Trace operation over a matrix product

The final tensor diagram in figure (10) is a double contraction over the indices n and m. This is actually a trace operation, and the resulting Tensor is a rank-0 tensor, i.e. a scalar T, as depicted by the lack of any free arms, and in turn missing open indices. This is actually a trace operation, and will come in handy when depicting reduced density matrices and SVD operations in tensor diagrams later on.

4.3.2 Breaking down wavefunctions into tensor networks

Now, given $|\psi\rangle = \sum_{s_1 s_2 s_3 \dots s_N} \psi^{s_1 s_2 s_3 \dots s_N} |s_1 s_2 s_3 \dots s_N\rangle$, is there a way to break this wavefunction down into a tensor network? The answer is yes, so instead of the computationally-inefficient approach of specifying the values of each of the 2^N $\psi^{s_1 s_2 s_3 \dots s_N}$ coefficients of the amplitude tensor, an alternative representation utilizing TN states is achieved by replacing the inefficiently large amplitude tensor, by an equivalent TN of smaller ranked tensors. We will tackle this from a somewhat qualitative approach and later introduce the rigorous process of this breakdown using MPS in the next section 4.4.

There are different approaches to breaking down the wavefunction into smaller tensor partitions one can handle more efficiently, but not all methods will effectively reflect rapidly decaying interactions (where rapidly decaying interactions imply rapidly decaying correlation function $C(r)$ and thus a small correlation length ξ) and entanglement in gapped local Hamiltonians. One method, uses ***mean field theory*** to achieve this, by negating all interactions and simply breaking down the wavefunction into an approximate tensor product as follows:



$$\Psi^{s_1 s_2 s_3 \dots s_N} \approx \psi^{s_1}\psi^{s_2}\psi^{s_3} \dots \psi^{s_N}$$

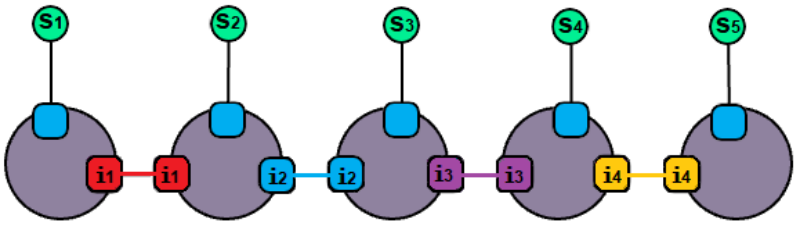
Figure (11): mean field theory breakdown of a many-body quantum state into individual non interacting partitions

The approach depicted in figure (11), is one type of mean field theory approximation, one that preserves the expectation values of individual spins, but totally negates the correlations, and thus misses all the entanglement. Since the wavefunction has essentially factorized, and we consider each site separately, so instead of the exponentially large amplitude tensor $\Psi^{s_1 s_2 s_3 \dots s_N}$, each site is considered individually, so we bring it down to one factor per site, where the value of each factor depends on the local state on each site, here we are dealing with spin states, so that's 2, over N many sites. Therefore, the simplification is cutting the coefficients down from

$$2^N \to 2N\,, \tag{171}$$

a significant improvement, but for the huge price of abandoning quantum correlations/entanglement. Some possible corrections to this model, to better approximate moderate correlations between spin particles, is to include several other Slater determinants with the product above, but that breaks down in different ways, that's not our main focus however so we'll leave it at that.

Now comes the novel way of treating this partitioning, we can promote those single tensors in the product in figure (11) by adding extra indices, such that each index resembles a possible correlation/interaction, a figurative "bond" between neighboring partitions/tensors. We will refer to these types of partitioning resulting from breaking larger tensors into smaller ones, as ***t-partitions***, to avoid mistaking them for the very closely related partitioning of global systems into bipartites, even though they are basically the same idea, it would be beneficial to make the distinction since we later aim to partition tensor networks, with several t-partitions themselves, into a bipartite system. That way we no longer ignore correlations, as we did with the mean field theory example, when we approximated an N-index tensor with N smaller 1-index tensors. Instead of leaving the t-partitions as uncorrelated products, we create an even larger (more degrees of freedom) but more ***exact*** tensor, by including extra dummy indices between neighboring t-partitions/tensors, that way we get $2 + 2(N-2)$ extra indices, 1 extra for each end tensor, given tensors on the end of a chain share an index with one neighboring tensor further inside, and 2 for the remaining $N-2$ inner tensors. The following diagram is an example of a rank-5 amplitude tensor of a wavefunction, broken down into 5 individual tensors with correlations intact:



$$\Psi^{s_1 s_2 s_3 s_4 s_5} \rightarrow \psi^{s_1}\psi^{s_2}\psi^{s_3}\psi^{s_4}\psi^{s_5} \rightarrow \psi^{s_1}_{i_1}\psi^{s_2}_{i_1,i_2}\psi^{s_3}_{i_2,i_3}\psi^{s_4}_{i_3,i_4}\psi^{s_5}_{i_4}$$

Figure (12): Maintaining correlations between the partitions resulting from a rank 5 tensor partitioning, this is an example of an MPS.

These 5 tensors are arranged in a way to depict correlations using dummy indices between adjacent t-partitions to represent correlations. Therefore, even though we increased the number of indices from N to $N + (2 + 2(N - 2))$, which is in this case from $5 \rightarrow 13$, their contraction leaves the overall tensor with the original 5 open indices as clearly shown in the tensor diagram, or from tensor operations where each contraction deducts 2 indices, leaving $13 - (2 \times 4) = 5$ indices. Notice how we have tensors at each end with only 2 indices, one of which is an open index, and the other a correlation to be contracted, since they have a single neighboring tensor. In contrast, the middle tensors each have 3 indices, one of which is also an open index, and the other 2 correlations to be contracted with their left and right neighboring tensors respectively.

4.3.3 *How TNS inherently represent the corner of Hilbert space abiding by area laws, and how they achieve compression.*

This pattern chosen for our previous tensor diagram example works extremely well with 1-D systems, and preserves both local expectation values while preserving the exponentially decaying correlations between the t-partitions (entanglement). This is a trivial example of a matrix product state (MPS). To assess how well this tensor network might have compressed the original amplitude tensor, we need to realize that the exponential 2^N parameters needed for an amplitude tensor are not independent of one another. Why? Because of the contractions we've made, they can be related to smaller tensor partitions, which are, $\psi^{s_1}, \psi^{s_2}, \psi^{s_3}, \psi^{s_4}, \psi^{s_5}$, in our rank 5 tensor example. These t-partitions clearly have way fewer degrees of freedom, than the exponential number of amplitude tensor components encompassing the entire Hilbert space indiscriminately. In contrast, these t-partitions cover a ***small corner*** of the entire Hilbert space. Being able to contract these individual t-partitions back into a 2^N component tensor, means these components are heavily dependent, and depend indirectly on the much fewer components of the individual t-partitions. This again is analogous to what we did with pure states and their Schmidt decomposition, where the $N \times N$ parameters in the rank-1 U and V matrices were traced back to components of two N-component vectors u and v, compressing the parameters from $O(N^2) \rightarrow O(2N)$.

We can also choose a more descriptive way of presenting the concept of MPS, without relying as much on tensor algebra. One can consider an amplitude tensor as a map which takes in spin configurations, for example $\Psi^{s_1 s_2 s_3 s_4 s_5}$, where we input a spin configuration (either spin up or down) in every one of the 5 S_i sites. The amplitude tensor maps these input configurations to a complex number, the coefficient in the state vector corresponding to this basis. Now the trick here is to adjust this mapping, such that it takes in a matrix, Instead of a scalar spin configuration, by giving every spin configuration its own matrix, so for our purpose here, every i^{th} site has 2 possible matrices, one matrix $M_i^{\uparrow}$, representing the up and the other, $M_i^{\downarrow}$ representing the down spins. Remember these matrices are totally arbitrary, the only constraint being that on a given site the dimensions of the different state matrices are equal. Next, we want to map to a number by multiplying these matrices. This is possible if we let the matrix on the first site have a row size of 1, and the matrix on the last site have a column size of 1, this matrix multiplication regardless of the inner matrices' dimensions would give a final 1×1 matrix as follows

Given $\Psi^{s_1 s_2 s_3 \ldots . s_N} \approx M_1^{s_1} M_2^{s_2} M_3^{s_3} \ldots M_N^{s_N}$ let $M_1^{s_1}$ be a row vector, and $M_N^{s_N}$ be a column vector, so

$$M_1^{s_1} M_2^{s_2} M_3^{s_3} \ldots M_N^{s_N} = \left(a, b, c \ldots \quad d_{M_1^{s_1}}\right) M_2^{s_2} M_3^{s_3} \ldots M_{N-1}^{s_{N-1}} \begin{pmatrix} a \\ b \\ \vdots \\ d_{M_N^{s_N}} \end{pmatrix} = (Number) \qquad (172)$$

So, the result is a number, as we intended. In our rank-5 amplitude tensor example, it would have the following form:

Given $\Psi^{s_1 s_2 s_3 s_4 s_5}$ with s_1 (the spin at site 1), being $\uparrow$, $s_2 = \downarrow$, $s_3 = \downarrow$, $s_4 = \uparrow$ $s_5 = \uparrow$ we get

$$\Psi^{\uparrow\downarrow\downarrow\uparrow\uparrow} \approx M_1^{\uparrow} M_2^{\downarrow} M_3^{\downarrow} M_4^{\uparrow} M_5^{\uparrow} \qquad (173)$$

We immediately see why this was called a "***Matrix Product State***", since it is literally a product of matrices. This representation is equivalent to the tensorial approach we did earlier.

To appreciate the compression this tensor network achieves, remember that the full wavefunction has 2^N parameters, then consider the following. Typically, the size of the matrices in the middle will be $m \times m$, by dividing into N t-partitions we get $2N$ matrices, since every site has 2 matrices, one for each spin state, then every matrix in turn has m^2 parameters. Therefore, after ignoring the small contribution from the outer $1 \times m$ and $m \times 1$ matrices at the ends of the chain we get a compression of

$$O(2^N) \rightarrow O(2Nm^2). \qquad (174)$$

To make this more general, consider we are given N sites, and thus obtained N t-partitions initially, where each site-t tensor has an $m(t)$ number of parameters. This makes the total number of parameters m_{total} for the tensor network:

$$m_{total} = \sum_{t=1}^{N} m(t) \tag{175}$$

For each partition-t, the tensor for that t^{th} site has its own rank, $rank(t)$, so rank is a function of site, and so each tensor has a number of indices equal to $rank(t)$, with each i^{th} index running over $\dim(i)$ dimensions, and we multiply the number of parameters for each index to get the total. This makes $m(t)$:

$$m(t) = O\left(\prod_{i=1}^{rank(t)} \dim(i)\right) \tag{176}$$

If we call the maximum dimension among the different indices of the tensor-t, "$\dim_t$ ", so a maximum value estimate for $m(t)$ would be that

$$m(t) = O\left(\prod_{i=1}^{rank(t)} \dim_{\mathrm{t}}\right) = O\left(\dim_t^{rank(t)}\right) \tag{177}$$

If we then take the maximum $\dim_t^{rank(t)}$, among all the t-partitions and call it χ, then the maximum estimate for $m(t)$ is some polynomial in χ, $O(pol(\chi))$.

This estimate makes the total number of parameters of the tensor network

$$m_{total} = \sum_{t=1}^{N} O\left(\dim_t^{rank(t)}\right) = \sum_{t=1}^{N} O(pol(\chi)) \tag{178}$$

This sum gives us some polynomial in N setting the final estimate to

$$m_{total} = O(pol(N)pol(\chi))$$

If we apply this to our $spin\ half\ 2$ state case with MPS, with the inner tensors all being of rank-2, since they are matrices, which makes the polynomial of $\chi = O(\chi^2)$. Each matrix was $(m \times m)$, which makes $\chi = m$. Open indices of the MPS could take up to 2 values, i.e. 2 spins, so the polynomial of N is just $2N$, we therefore get

$$m_{total} = O(2Nm^2)$$

This is consistent with our original assessment, and the compression goes like

$$O(2^N) \to O(pol(N)pol(\chi)) = O(2Nm^2).$$

That's a huge compression, we've taken the N from being an exponent to it being a multiple, which reduces exponential behavior to polynomial behavior. We can therefore represent any exponentially large number of tensor components if m is made large enough.

4.3.4 *Significance of connecting/bond indices in tensor networks and their Influence on entropy area laws in TNS.*

We've already implied that the connecting indices between t-partitions, depict interactions between them; it's time we illustrate their important physical meaning. First, dimension $\dim(i)$ of the i^{th} index in tensor-t, is a measure of the amount of quantum correlations (entanglement) that t-partition has within the full system, accordingly, these dummy indices are usually called ***bond indices***, and their dimensions are referred to as ***bond dimensions***. The largest of these dimensions, which we called D, is in turn, referred to as the ***bond dimension of the tensor network***.

Let's give an example to emphasize the importance of bond indices in setting entanglement. Imagine we have a TN with maximum bond dimension χ for all the indices, so all t-partitions are $\chi^{rank(t)}$. One possible TN is shown in figure (13), this is a TNS called ***Projected Entangled Pair State (PEPS)***, which is a 2-D equivalent of MPS. This one is a (6×6) PEPS, so has a total of 36 nodes. If we turn the PEPS into a bipartition, we can divide the Global system into 2 subsets, subsystem A (inner 4×4 block) and subsystem B(outer region). Subsystem A is a square region with 4 sides each of length L. We want to estimate the ***entanglement entropy*** of subsystem A. The boundary between our 2 regions includes 4 arms, each 4-sites, or rather 4-t-partitions, long, so $L = 4$.

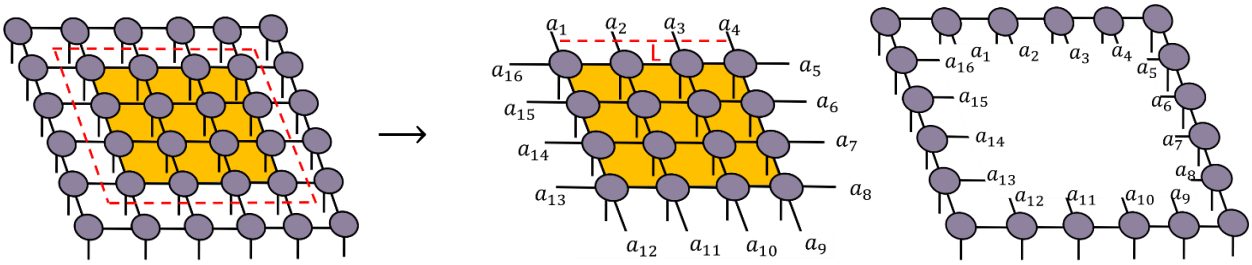


Figure (13): Dividing a 6×6 PEPS into a bipartition AB, where subsystem A is the inner 4×4 block, and subsystem B is the outer region.

This TN pattern and subsequent partitioning into a bipartite state, allows us to identify ***combined index*** $\bar{a}$, where $\bar{a}$ is a combined index from the separate TN indices connecting the tensor sites bordering the two regions at the cut. With each side being 4-t-partitions long, and with 4 sides we expect $4L = 16$ indices across the boundary. We identify

$$\bar{a} = \{a_1 a_2 a_3 a_4 a_5 a_6 a_7 a_8 a_9 a_{10} \dots a_{4L}\} \tag{179}$$

Given each index can take up to χ values itself, then index vector $\bar{a}$ itself actually has χ^{4L} different values, and Keep in mind, that this is still less than the wavefunction amplitude tensor, which given the 36 nodes

in this case, would have had χ^{36} components. We can now write the state as a tensor product of the two substate kets, the inner substate A, given by $|A(\bar{a})\rangle$, and outer substate B, given by $|B(\bar{a})\rangle$. Since there are χ^{4L} different $\bar{a}_i$ index tensors, each one gives a different pattern of bond indices at the border, so to get the full state we need to sum over all $\bar{a}$'s, so we have a sum of χ^{4L} terms to cover all possible arrangements between subsection A and B. Each one of these arrangements depicts a different index configuration at the boundary between the 2 cuts, and thus codes for the interactions between the two substates. by ignoring normalization this gives

$$|\Psi\rangle = \sum_{\bar{a}=1}^{\chi^{4L}} |A(\bar{a})\rangle \otimes |B(\bar{a})\rangle \tag{180}$$

And since they are pure states, we know from equation (108), that this is a special case of an operator $\hat{\rho}_{pure}{}^{(:AB)} = |a, b\rangle\langle a', b'|$, and that the reduced density matrix of A, which will be equal to that of B, can be obtained using

$$\rho^{(/A)} = \rho^{(/B)} = \sum_{\bar{a},\bar{a}'} |A(\bar{a})\rangle\langle A(\bar{a}')| \ \ \delta_{B(\bar{a}),B(\bar{a}')}$$

$\rho^{(/A)}$ will, at most, have at most a χ^{4L} rank. Now that we have $\rho^{(/A)}$, we can get the entanglement entropy through calculating the von Neumann entropy

$$S\left(\rho^{(/A)}\right) = -tr\left(\rho^{(/A)} \log_2\left(\rho^{(/A)}\right)\right)$$

And we also know from property 3 of the von Neumann entropy, that it has an upper bound of

$$S^{Neumann} \leq \log_2(\min\,[d_A, d_B])$$

Applying this here gives the upper bound

$$S^{Neumann} \leq \log_2(\chi^{4L}) = 4L\ log_2(\chi) \tag{181}$$

This is a spectacular result, since if we look back at the area laws for $2D$ lattices we derived that

$$S_{A(2D)} \sim \alpha L$$

Which is consistent with what we just derived using tensor network states, for this specific PEPS TNS and this square-cut partitioning, the TNS turn out to be the states which obey the area law with

$$S_{A(2D)} \sim \alpha L \ \ ; \ \ and\ \alpha = cosntant = 4\log_2(\chi)$$

In addition, this result adds further intuition to the meaning of ***bond indices***, as we can interpret us having $4L$ indices, and $S^{Neumann} \leq 4L\ log_2(\chi)$, that along the boundary between our A and B partitions,

every bond index gives an entropy contribution of at most $\boldsymbol{log_2(\chi)}$.

Before we move on, let's check the entanglement entropy result for different values of χ. First, if $\chi = 1$, then the upper bound says that $S^{Neumann} \leq \log_2(1^{4L}) = 0$, regardless of L, which means no correlations and in turn no entanglement. This makes sense since if we have $\chi = 1$, we get a TNS that is just a product state, the same pattern we resorted to during our mean field theory example, illustrated in figure (11) where the t-partitions had no bond indices, and every node just had its own open index. Second, for any $\chi > 1$, the ***TN states are inherently targeting the states abiding by the area law for entanglement entropy***, and thus ***operate within the vital corner of Hilbert space of the gapped Local Hamiltonians*** we discussed earlier. Therefore, TN states are able to represent quantum many-body states in a way which is both computationally efficient (due to heavy compression), and quantumly correlated, i.e. entangled. This is the goal we were trying to achieve in order to tackle the exponentially large amplitude tensor of a wavefunction. Lastly, it is worthwhile to notice that given $S^{Neumann} \leq \log_2(\chi^{4L})$, changing the bond dimension χ alters a multiplicative factor in front of L, and is therefore still abiding by the area law. However, in order to modify the scaling with L we would need to rearrange the geometric pattern of the TN. The TNs we discussed thus far, were MPS and PEPS, which both are perfect for representing gapped local Hamiltonians of quantum many body systems at low temperatures/in ground state; the key reason is their inherent tie to area laws of entanglement entropy. Other patterns of TNS however, will change the entanglement through altering the contraction geometry. That said, we have shown that entanglement in TNS is governed by both the size of bond indices χ, and the geometric pattern, where although MPS and PEPS following area laws uneventfully, other different families of TNS will have very different entanglement behavior.

4.4.0 Deeper look into building MPS, the most common 1-D tensor networks [7], [9], [12], [17], [19]

Now that we have shown the advantages of writing the wavefunction as a TN state, and how TNS target the corner of Hilbert space with low energy eigenstates, and that these states will exhibit area law abiding entanglement entropy between their partitions, the question that remains; how can the specific t-partitions of a TN pattern be determined given an arbitrary wavefunction? We have briefly discussed two types of TNS, MPS and PEPS. We will focus on MPS for the rest of our work, since it is arguably the most established in literature.

MPS are a 1-D array of tensors and are among the most famous examples of TN states for treating 1-D quantum many-body systems. MPS also underly several of the leading methods of numerically simulating and solving many-body quantum systems, most prominently the Density Matrix Renormalization Group (DMRG) algorithm, so let's investigate them more closely, and give specific examples of MPS t-partitioning of some common wavefunctions.

4.4.1 The MPS as a 1-D tensor network

In a matrix product state, we have one tensor per site t, and we have N-many sites generating a TN with N nodes. Nodes have both bond indices and open indices as discussed earlier. The ***bond indices*** glue the

matrices together, and can take up to χ values, the bond dimension of the TN. Open indices however, correspond to the ***physical degrees of freedom*** of the local Hilbert spaces of the state function. The MPS can either be with open boundary conditions (OBC)/ infinite boundaries, which is more favorable model, especially with DMRG, or it could have periodic boundary conditions (PBC). PBC's would require an extra bond index between the first and last tensor in the network, this poses some problems with entanglement entropy due to this interaction. We won't discuss this further, but just emphasize that we are sticking with OBC in our following examples.

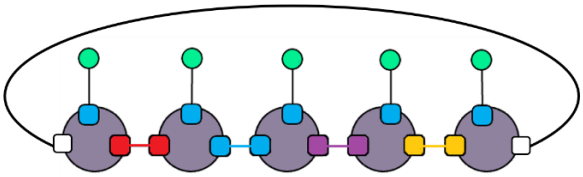


Figure (14): an MPS with periodic boundaries

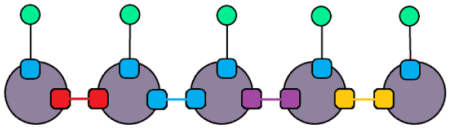


Figure (15): an MPS with open boundaries

The main advantage of MPS, is that ***MPS with finite χ reproduce the properties of the corner of the Hilbert space satisfying the 1-D area-law for the entanglement entropy***. However, we can make the MPS cover larger parts of the many-body Hilbert space, up to engulfing the entire space, just by making the χ large enough. To engulf the entire Hilbert space, χ will have to be of order exponential in N. An exact breakdown of how the bond index controls MPS chains is covered in section (4.4). This divergence of achieved in the number of parameters by increasing χ, blows up a partitions' entanglement entropy, and is therefore counterproductive, as we want the behavior of the system to abide by the area law in order to reflect the low energy corner states of gapped local Hamiltonians.

.

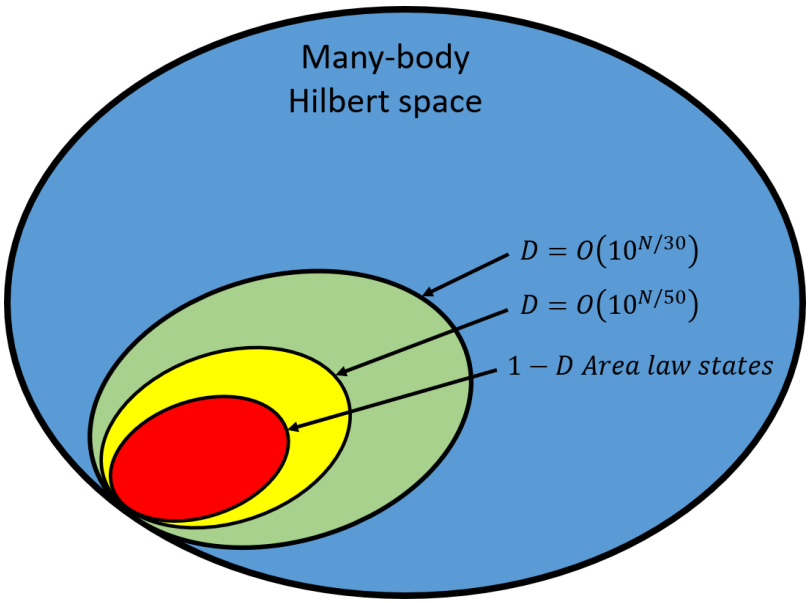


Figure (16): The many body Hilbert space of a 1-D lattice, and the effect of varying the order of D, the red region comprises the corner of Hilbert space abiding by the 1-D area law

Since we showed how PEPS satisfy the 2-D area-law of entanglement entropy, showing that t-partitioning left them with an upper limit of $S_{A(2D)} \leq \alpha L$ shown in equation (181); by virtue of the same argument we expect MPS to have a limit $S_{A(1D)} \leq \alpha$, a 1-D area law. This simply means that the entanglement entropy of some bipartition block of sites, is bounded from above by a constant, making

$$S(L) \; = \; O(\log_2 \chi), \tag{182}$$

This is in accordance with the behavior usually observed in the ground states (low energy states) of gapped 1-D local Hamiltonians.

As you might've guessed, there are several ways to obtain satisfactory MPS, MPS are not unique, but one very convenient way has been identified when given a quantum state with open boundary conditions, for which there is a choice of tensors called canonical form of the MPS. An MPS, is said to be in canonical form, if for each i^{th} bond index, there is a corresponding i^{th} Schmidt coefficient and Schmidt vectors of the Schmidt decomposition of the wavefunction across the corresponding index, we explain this in the following section, section (4.4.2).

4.4.2 *Left-, Right-, and Mixed-Canonical MPS decompositions of arbitrary pure quantum states* [7], [9], [12], [17], [21]

In section (4.3) we had a qualitative introduction to the motivation and reasoning behind dividing an arbitrary state amplitude tensor into a tensor network. Here we solidify the concept through a rigorous introduction of the 3 most common formulations of MPS, the ***left, right and mixed -Canonical MPS decomposition***, an algorithm used to write any arbitrary pure state as an MPS. Remember, an MPS in simplest terms is just a decomposition of a 1-D system's state amplitude tensor rank N, into N smaller ranked tensors.

We again consider a lattice of N sites, where each i^{th} site has its own local state S_i, where each S_i is d-dimensional. Given we are dealing with spin states, S_i has an extra index besides i which runs over 2 states, so $d = 2$. The most general ***pure quantum state*** on the lattice reads

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

We can restructure the ***state vector of a multipartite system*** into a bipartite system like we did in section 2.2.5, by utilizing ***collective indices*,** which results in the partitioning of $\{s_1, s_2, s_3, \dots s_N\}$ into 2 subsets A and B, where if we choose the A partition to be the single state S_1 , we get

$$A = \{s_1\} \ , \quad B = \{s_3, \dots s_N\}$$

This reduces the global system's $C^{s_1 s_2 s_3 \dots s_N}$ state vector into

$$C^{s_1 s_2 s_3 \dots s_N} = \psi^{\{s_1\},\{s_2, s_3, \dots s_N\}} = \psi^{AB}$$

So now ψ^{AB} is our restructured bipartition, generated by applying collective indices to the multipartition amplitude tensor $C^{s_1 s_2 s_3 \dots s_N}$. The trick here is that ψ^{AB} is also mapped to the components of a matrix, making the transformation

$$|\psi\rangle = \sum_{s_1 s_2 s_3 \dots s_N} C^{s_1 s_2 s_3 \dots s_N} |s_1 s_2 s_3 \dots s_N\rangle \ \rightarrow |\psi^{(AB)}\rangle = \sum_{i,j} \psi^{ij} |i, j\rangle^{AB}$$

Now the coefficients ψ^{ij} can be mapped to a matrix of dimension $(d \times d^{N-1})$, since index i runs through 1,2 since it's just the local state S_1 (where $d = 2$ for spin states), while the index j runs through all the iterations of $(S_2, S_3 \dots\dots S_N)$, which are $(N-1)$ many local states giving d^{N-1} permutations. Now comes the key part of the algorithm, given ψ^{ij} is a matrix, we refer back to section 3.1, we can apply SVD to it as shown in equation (115), which was

$$\hat{C}_{ij}^{\ (AB)} = \sum_{m=1}^{r} U_{im} \sigma_m V_{jm}^*$$

Let's apply SVD to the matrix coefficients ψ^{ij} of the Global **Bipartite Pure State**, we will index individual SVD using m_i, and r_i since we will apply several of them, so after the first SVD we get

$$\psi^{ij} = \psi^{S_1,(S_2 \dots S_N)} = \sum_{m_1=1}^{r_1} U_{S_1,m_1} \, \sigma_{m_1} V^*{}_{m_1,(S_2,\dots,S_N)} \equiv \sum_{m_1=1}^{r_1} U_{S_1,m_1} \, c_{m_1,S_2 \dots S_N}$$

Where $\sigma_{m_1} V^*{}_{m_1,(S_2,\dots,S_N)}$ were multiplied, and the resulting matrix was remapped back into a state vector. The matrix U_{S_1,m_1} will also be decomposed into a collection of d row vectors $M_{m_1}^{S_1}$ is mapped to the entries

$$M_{m_1}^{S_1} = U_{S_1,m_1} \tag{183}$$

This translates to every spin state in S_1, having its own row vector $M_{m_1}^{S_1}$, and in the sum we can include as many r_1 values as we need up till d, the same way we exclude as many Schmidt coefficients we need to in the Schmidt decomposition, so $r_1 \leq d$.

We now repeat the process by restructuring $c_{m_1,S_2 \dots S_N}$, into a matrix $\psi^{(m_1,S_2),(S_3 \dots S_N)}$ such that

$$c_{m_1,S_2,S_3 \dots S_N} \rightarrow \psi^{(m_1,S_2),(S_3 \dots S_N)}$$

Where the $\psi^{(m_1,S_2),(S_3 \dots S_N)}$ matrix is of dimension $(r_1 d \times d^{N-2})$, we therefore get

$$\sum_{m_1=1}^{r_1} M_{m_1}^{S_1} \, \psi^{(m_1,S_2),(S_3 \dots S_N)}$$

Reapplying the SVD to the $\psi^{(m_1,S_2),(S_3\ldots S_N)}$ we get

$$C^{s_1s_2s_3\ldots s_N} = \sum_{m_1=1}^{r_1} \sum_{m_2=1}^{r_2} M^{S_1}_{m_1} U_{(m_1,S_2),m_2}\, \sigma_{m_2} V^*_{m_2,(S_3,\ldots,S_N)}$$

We can now replace $U_{(m_1,S_2),m_2}$ by a set of d matrices $M^{S_2}_{m_1,m_2}$ (since S_2 takes d-values) where the matrices are of dimension $(r_1 \times r_2)$, since the index m_1 takes r_1 values, while index m_2 takes r_2 values, where r_2 runs over $r_1 d$ values making $r_2 \leq r_1 d \leq d^2$. Recall that we did basically the same thing with the first SVD but we had a single index m_1, so it was mapped to d many column vectors instead of matrices. Finally, $\sigma_{m_2} V^*_{m_2,(S_3,\ldots,S_N)}$, and reshape them into a matrix $\psi^{(m_2,S_3),(S_4\ldots S_N)}$, of dimension $(r_2 d \times d^{N-3})$, this gives us

$$C^{s_1s_2s_3\ldots s_N} = \sum_{m_1=1}^{r_1} \sum_{m_2=1}^{r_2} M^{S_1}_{m_1} M^{S_2}_{m_1,m_2}\, \psi^{(m_2,S_3),(S_4\ldots S_N)}$$

We can reapply the SVD till we deplete the complementary state vector on the right-hand side to get

$$C^{s_1s_2s_3\ldots s_N} = \sum_{m_1,m_2,m_3\ldots,m_{N-1}}^{r_1,r_2,r_3\ldots r_{N-1}} M^{S_1}_{m_1}\, M^{S_2}_{m_1,m_2}\, M^{S_3}_{m_2,m_3} \ldots\ldots M^{S_{N-1}}_{m_{N-2},m_{N-1}} M^{S_N}_{m_{N-1}} \qquad (184)$$

Where the last set of matrices can be restructured to d-many column vectors, and d is the dimension of the Hilbert space. It might have become obvious to the reader that equation (184) is starting to take the form we introduced qualitatively in section (4.3.2), specifically equation (172) that is nicely illustrated in figure (12). We show once more that by simply reshaping the state amplitude vector, we achieved a compression of the order

$$O(2^N) \rightarrow O(2Nm^2)$$

as we derived in equation (174). However there was still a catch, the values taken by the indices, i.e. the bond dimensions, are not necessarily independent of the system size, and the indices here might be forced to match exponentially large values due to the divergence of r_i's making matrices of dimensions $r_i \times r_{i+1}$ where $r_{i+1} = r_i d$. This exponential growth was clearly illustrated in equations (185, 186). This unmodified raw MPS can actually give an ***EXACT*** representation of the exponentially large state vector $C^{s_1s_2s_3\ldots s_N}$. In general, each i^{th} site on the lattice has d many $r_i \times r_{i+1}$ matrices. It is important to re-emphasize that the two indices between adjacent matrices, indicate tensor contractions and are therefore referred to as "virtual", "dummy" or "auxiliary" indices, to distinguish them from the "physical" indices, which are in this case spin state indices S_i. This is exactly similar to the bonding indices we presented back in section (4.3.2), and we will show, in the following section, that these indices do run over singular values of the bipartition's reduced density matrices and therefore do reflect entanglement entropy between two partitions of the chain. Accordingly, we were right to refer to these indices as ***"bond dimensions"***.

One might start to ask, if there is still exponential growth, then why are we using this representation? Well, we have already discussed local gapped Hamiltonians and their exponentially decaying correlation length, and we described how that instills the are laws of entanglement entropy on the ground states, and that considering a cut generating a bipartite system, the entanglement entropy of low-energy eigenstates will accordingly mostly depend on the nearest interactions only, rendering most indices near zero, and this will be reflected in the spectrum of the singular values. This will be discussed in detail in the decimation of MPS at the end of this section and in section (4.4.3)

Now back to the topic in hand, given each site in the MPS now has three indices, we usually group them together into a $rank-3$ tensor. In addition, sometimes the first and last matrices had a constant index (one which is always 1) added to them for convenience and consistency with the rest of the matrices in the chain, which turns them from d-many vectors, into d-many $(1\times r)$ and $(r\times 1)$ matrices respectively.

A closer look at our MPS reveals that the maximum dimension of M matrices is achievable when for each SVD the non-zero singular values are also equal to the $min[dim\,A\,, dim\,B]$ of our restructured bipartitions in every successive SVD . Our matrix dimensions will maximally be

$$(1\times r_1),(r_1\times r_2),\dots,\left(r_{\left(\frac{N}{2}\right)-1}\times r_{\left(\frac{N}{2}\right)}\right),\left(r_{\left(\frac{N}{2}\right)}\times r_{\left(\frac{N}{2}\right)-1}\right),\dots,(r_{N-1}\times r_N),(r_N\times 1) \tag{185}$$

These will take the values

$$(1\times d),(d\times d^2),\dots,\left(d^{\left(\frac{N}{2}\right)-1}\times d^{\left(\frac{N}{2}\right)}\right),\left(d^{\left(\frac{N}{2}\right)}\times d^{\left(\frac{N}{2}\right)-1}\right),\dots,(d^2\times d),(d\times 1) \tag{186}$$

We observe such a pattern since as we increase the values taken by the m_i indices available to the left partition A, by a multiple of d in every i^{th} increment, we also lose an entire index from the $(S_i\dots S_N)$ term corresponding to the left partition B. Therefore, the transition will go like

$$\psi^{S_1,(S_2\dots S_N)}\rightarrow\psi^{(m_1,S_2),(S_3\dots S_N)}\rightarrow\psi^{(m_2,S_3),(S_4\dots S_N)}\rightarrow\dots\dots\psi^{(m_{N-1}),(S_N)}$$

Another look at the patter in equation (186), shows that the matrices placed in corresponding positions equally away from the center of the chain are actually transposes of one another given we are dealing with indistinguishable local states. We will not go into full details here, but as can be shown in reference [12], since in SVD $U^\dagger U=\mathbb{I}$, and that at SVD iteration we replace U with an M matrix, we can establish a sort of orthonormality between these matrices in the relation

$$\sum_{S_i} M^{S_i^\dagger} M^{S_i} = I \tag{187}$$

Matrices abiding by equation (187), in that specific order where the left side matrix is transposed meaning our chains always start with a row vector and end with a column vector, we refer to the constituting matrices as ***left-normalized***, making our MPS ***left-canonical***.

With respect to right-canonical MPS, as the name implies, entails the exact same process but starts the partitioning from the right end, giving

$$\psi^{ij} = \psi^{(S_1 \dots S_{N-1}),S_N} = \sum_{m_{N-1}=1}^{r_1} U_{(S_1 \dots S_{N-1}),m_{N-1}} \, \sigma_{m_{N-1}} V^*{}_{m_{N-1},(S_N)}$$

and obeys the ***right-normalization***

$$\sum_{\sigma_i} R^{S_i^\dagger} R^{S_i} = I \tag{188}$$

An important behavior of left or right Matrix Product States, is that if we want to breakdown their MPS chain into a bipartition we would get a very debilitating inconvenience. To illustrate this inconvenience, consider we split an N-site lattice into subparts A, B, where A encompasses sites $1 \to \ell$ and B sites $\ell + 1 \to N$. If we decompose parts A and B using left or right canonical MPS we would get

$$|m_\ell^{(A)}\rangle = \sum_{S_1,\dots,S_\ell} (M^{S_1} M^{S_2} \dots M^{S_\ell})_{1,m_\ell} |S_1, \dots, S_\ell\rangle \tag{189}$$

$$|m_\ell^{(B)}\rangle = \sum_{S_{(\ell+1)},\dots,S_L} (M^{S_{(\ell+1)}} M^{S_{\ell+2}} \dots M^{S_N})_{m_{\ell,1}} |S_{\ell+1}, \dots, S_N\rangle \tag{190}$$

Such that the MPS becomes

$$|\psi\rangle = \sum_{m_\ell} |m_\ell^{(A)}\rangle |m_\ell^{(B)}\rangle$$

The inconvenience here, is that this is ***almost*** a Schmidt decomposition, ***but not quite***. The reason it is not a legitimate Schmidt decomposition, is that if used left canonical MPS ***the*** $\{|\boldsymbol{m_\ell}\rangle_{\boldsymbol{A}}\}$ ***basis forms an orthonormal set, while*** $\{|\boldsymbol{m_\ell}\rangle_{\boldsymbol{B}}\}$ ***do not***, and if we go back to the Schmidt decomposition, section (3.1.0), we can see that a key property of the Schmidt decomposition, the one that renders it an indispensable tool for approximating large Hilbert spaces, is that its two basis sets $\{|m_\ell\rangle_A\}$ and $\{|m_\ell\rangle_B\}$ MUST be orthonormal. This inconsistency is due to the MPS chain being of left normality, and the opposite goes if we had used right canonical MPS. The inconsistency can be easily illustrated given

$$\langle m'_\ell | m_\ell \rangle_A = \sum_{S_1,\dots,S_\ell} (M^{S_1} \dots M^{S_\ell})^*_{1,m'_\ell} (M^{S_1} \dots M^{S_\ell})_{1,m_\ell}$$

$$= \sum_{S_1,\dots,S_\ell} (M^{S_1} \dots M^{S_\ell})^\dagger_{m'_\ell,1} (M^{S_1} \dots M^{S_\ell})_{1,m_\ell}$$

$$= \sum_{S_1,\dots,S_\ell} \left(M^{S_\ell \dagger} \dots M^{S_1^\dagger} M^{S_1} \dots M^{S_\ell} \right)_{m_l', m_\ell}$$

So, using equation (188)

$$\langle m_\ell' | m_\ell \rangle_A = \delta_{m_\ell', m_\ell} \tag{191}$$

In contrast for B

$$\langle m_\ell' | m_\ell \rangle_B = \sum_{S_{\ell+1},\dots,S_N} (M^{S_{\ell+1}} \dots M^{S_N})^*_{m_\ell',1} (M^{S_{\ell+1}} \dots M^{S_N})_{m_\ell,1}$$

$$= \sum_{S_1,\dots,S_N} \left(M^{S_N \dagger} \dots M^{S_{\ell+1}\dagger} \right)_{1,m_\ell'} \left(M^{S_{\ell+1}} \quad \dots M^{S_N} \right)_{m_\ell,1}$$

$$= \sum_{S_1,\dots,\mathrm{N}} \left(M^{S_{\ell+1}} \dots M^{S_N} \, M^{S_N^\dagger} \dots M^{S_\ell+1\dagger} \right)_{m_l', m_\ell}$$

But in general

$$\sum_S M^S M^{S\dagger} \neq I$$

Since these are all left normalized matrices. The same arguments extend to right canonical MPS and right normalized matrices.

The solution to this problem, is using a ***Mixed Canonical MPS***. This entails that we form a left canonical MPS of subset A with its left normalized M matrices, up till site $\ell + 2$, this leaves the last SVD on subset A with the right-most term

$$U_{(m_\ell S_{\ell+1}), m_{\ell+1}} \sigma_{\ell+1}$$

We reshape $U_{(m_\ell S_{\ell+1}), m_{\ell+1}} \sigma_{\ell+1}$ into $R^{S_{\ell+1}}_{m_\ell, m_{l+1}}$, then start a right canonical MPS of subset B with its right normalized R matrices. Notice that this leaves us with a coefficient $\sigma_{\ell+1}$. This mixing gives

$$|m_\ell^{(A)}\rangle = \sum_{\mathrm{S}_1,\dots,S_\ell} (M^{S_1} M^{S_2} \dots M^{S_l})_{1,m_\ell} |S_1, \dots, S_\ell\rangle \tag{192}$$

$$|m_\ell^{(B)}\rangle = \sum_{S_{(\ell+1)},\dots,S_N} (R^{S_{(\ell+1)}} R^{S_{\ell+2}} \dots R^{\sigma_N})_{m_{\ell,1}} |S_{\ell+1}, \dots, S_N\rangle \tag{193}$$

Through this minor modification, we have restored the orthonormality of the right-hand partition B basis where

$$\delta_{m'_\ell,m_\ell} = \sum_{\sigma_{\ell+1}} \left(R^{S_{\ell+1}} R^{S_{\ell+1}\dagger}\right)_{a_\ell,a'_\ell}$$

Which ensures

$$\langle m'_\ell | m_\ell \rangle_B = \delta_{m'_\ell,m_\ell} \tag{194}$$

By orthonormalizing both left and right pats of a partitioned MPS chain, we can finally overcome the drawback of using either left or right canonical MPS individually, by resorting to mixed canonical MPS. Mixed canonical MPS gives us a decomposition of the state vector that looks like

$$C^{s_1 s_2 s_3 \dots s_N} = M^{S_1} \dots M^{S_\ell}\, O_{\ell+1}\, R^{S_{l+1}} \dots R^{S_N}$$

Notice that this formulation leaves us with the singular value matrix $O_{\ell+1}$, holding all the singular values $\sigma_m^{(\ell+1)}$ on the bond $(\ell, \ell+1)$ on its diagonal, which is the bond between the two subregions A, B. The mixed decomposition can be shown in figure (17)

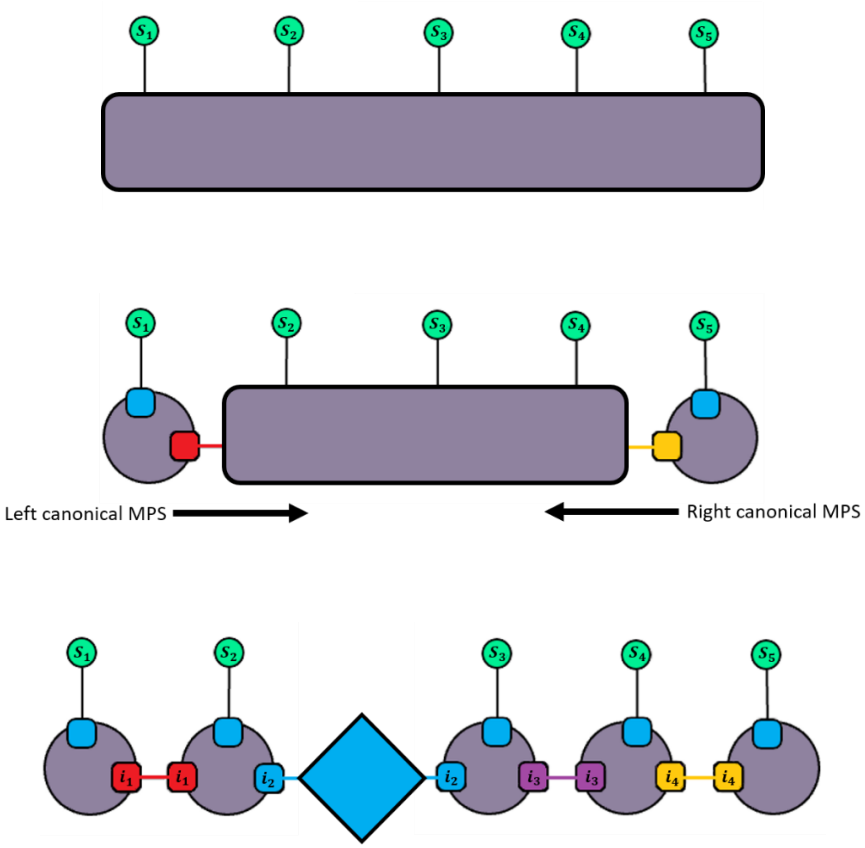


Figure (17): A series of tensor diagrams of a mixed canonical MPS process obtained by a sequence of singular value decompositions, from both the left and right. The diamond unit represents the diagonal singular value matrix with entries σ_m. The singular value matrix bonds the right and left partitions together.

The mixed canonical MPS formulation allows us to divide an MPS chain into a bipartite system using Schmidt decomposition resulting in an ***MPS based Schmidt decomposition*** in equation (195) where

$$|\psi\rangle = \sum_{m_\ell}^{r} \sigma_{m_\ell} |m_\ell\rangle_A |m_\ell\rangle_B \tag{195}$$

We need to emphasize that, to study the entanglement entropy of the system we have to consider a bipartition, so let us consider partitioning an MPS chain like we did using the Schmidt decomposition, with the first site to the left as A and the rest of the chain B. The sum we have here will not necessarily iterate up to d; like we discussed earlier, the number of non-zero singular values reflects the amount of entanglement of site 1 with the rest of the system, so when dealing with eigenstates of local gapped hamiltonians, we know the limited correlation length limits the entanglement to area laws. So, we can get compression from forming an MPS when the global Hamiltonian is a sum of local Hamiltonians. For simple cases such as the Ising model whose Hamiltonian reflects only nearest neighbor interactions, many singular values will present themselves as zero and can be discarded. In complicated Hamiltonians however, we might have to select an arbitrary cutoff within acceptable error. In any case, it is key that this decimation process revolves around SVD, and the mixed canonical MPS formulation is the one allowing best utilization of Schmidt decomposition. We discuss the most general decimation process in the following section.

4.4.3 *MPS and single site decimation to target the low energy eigenstates of gapped local Hamiltonians* [7], [9], [12], [17]

Since The canonical MPS formulation is an ***exact*** representation of the state vector, so if we were to carry out this exact MPS decomposition we would be faced with the same exponential complexity that renders state vectors inefficient tools in treating many-body quantum systems. Accordingly, we must introduce a way to focus on certain parts of the large M^{S_i} to isolate states in the corner of Hilbert space of gapped local Hamiltonians.

One way to do this is to consider mixed-canonical MPS, we have identified a Schmidt decomposition of the state, and we know from section (3.1.0), that singular values are arranged in decaying order from highest to lowest. We argued that a cutoff of the sizes of the matrices involved is viable and by setting an $r' < r$ we are able to make the best low rank approximation of the original matrix, which we have mapped the state vector to, following its partitioning into subparts A and B. Using equation (120)

$$\left|\tilde{\psi}^{(AB)}\right\rangle = \sum_{m}^{r'} \sigma_m \left|m^{(A)}\right\rangle \left|m^{(B)}\right\rangle$$

we can choose an appropriate cutoff given we have an exponentially decaying eigenvalue spectra of reduced density operators, and in turn, exponentially decaying singular values. This is exactly the case

with gapped local Hamiltonians which exhibit an exponentially decaying correlation factor as we discussed with area laws in section (4.2), specifically equation (165). We have therefore shown that given a mixed canonical MPS chain with

$$C^{s_1 s_2 s_3 \dots s_N} = M^{S_1} \dots M^{S_\ell}\, O_{\ell+1}\, R^{S_{l+1}} \dots R^{S_N}$$

we can decimate the matrices involved this Schmidt decomposition by setting an appropriate r' value, which leads to the retention of the first r' columns of M^{S_ℓ}, the first r' rows of $R^{S_{l+1}}$, and the first r'rows and columns of $O_{\ell+1}$. We should also renormalize the cut off states by rescaling the retained singular values. It is important to note that this decimation occurs with respect to a single bond $(\ell, \ell+1)$, i.e. the one between sites ℓ and $(\ell+1)$, it is only possible since we have a right normalized chain to the left, and a right normalized chain to the right, and this procedure rests on the states $|m^{(A)}\rangle$, $|m^{(B)}\rangle$ of subsets A and B respectively, being orthonormal. This condition is only applicable to this single bond in-between the 2 canonizations. We usually need to shrink all the other matrices, and this requires we repeat this procedure with all the different available mixed canonical representations to cover bonds on all sites in the original chain.

To demonstrate this idea, consider we start with all left canonized matrices, and we start from the right using SVD to turn the far-right end matrix right canonized, so after the first canonization we get the truncated form

$$|\psi\rangle = \sum_{S_1 \dots S_N}^{r} M^{S_1} \dots M^{S_{N-1}}\, U O_{N-1} R^{S_N}\, |S_1 S_2 \dots S_N\rangle,$$

This leaves us with

$$|m_\ell^{(A)}\rangle = \sum_{S_1,\dots,S_{N-1}} (M^{S_1} M^{S_2} \dots M^{S_{L-1}} U)_{1,m_{N-1}} |S_1, \dots, S_{N-1}\rangle$$

$$|m_{N-1}^{(B)}\rangle = \sum_{S_N} R^{S_N}{}_{m_{N-1},1} |S_N\rangle$$

With each basis sets orthonormal. Accordingly, the Schmidt decomposition for that becomes

$$|\psi^{(N-1)}\rangle = \sum_{m_{N-1}}^{r} \sigma_{m_{N-1}} \left|m_{N-1}^{(A)}\right\rangle \left|m_{N-1}^{(B)}\right\rangle$$

We then truncate the matrices U, O and R^{S_N} into $\tilde{U}, \tilde{O}$ and $\tilde{R}^{S_N}$ by omitting all but the largest singular values in σ_m forming the O matrix. The process is repeated till we truncate all the matrices, this goes like

$$|\psi\rangle = \sum_{S_1 \dots S_N}^{r} M^{S_1} \dots M^{S_{N-2}} (M^{S_{N-1}} U O_{N-1})\, R^{S_N} |S_1 S_2 \dots S_N\rangle$$

↓

$$|\tilde{\psi}\rangle = \sum_{S_1 \dots S_N}^{r} M^{S_1} \dots M^{S_{N-2}} \left(M^{S_{N-1}} \tilde{U} \tilde{O}_{N-1}\right) \tilde{R}^{S_N} |S_1 S_2 \dots S_N\rangle$$

We apply SVD to the product in the bracket, which results in $M^{S_{N-1}} \tilde{U} \tilde{O}_{N-1} \rightarrow U \tilde{O}_{N-1} \tilde{R}^{S_{N-1}}$

$$\downarrow$$

$$|\tilde{\psi}\rangle = \sum_{S_1 \dots S_N}^{r} M^{S_1} \dots M^{S_{N-3}} \left(M^{S_{N-2}} U O_{N-2}\right) R^{S_{N-1}} \tilde{R}^{S_N} |S_1 S_2 \dots S_N\rangle$$

$$\downarrow$$

$$|\tilde{\psi}\rangle = \sum_{S_1 \dots S_N}^{r} M^{S_1} \dots M^{S_{N-3}} \left(M^{S_{N-2}} \tilde{U} \tilde{O}_{N-2}\right) \tilde{R}^{S_{N-1}} \tilde{R}^{S_N} |S_1 S_2 \dots S_N\rangle$$

$$\downarrow$$

$$\vdots$$

$$|\tilde{\psi}\rangle = \sum_{S_1 \dots S_N}^{r} \tilde{R}^{S_1} \tilde{R}^{S_2} \dots \tilde{R}^{S_{N-2}} \tilde{R}^{S_{N-1}} \tilde{R}^{S_N} |S_1 S_2 \dots S_N\rangle \quad (196)$$

This eventually generates a decimated right canonical MPS. Finally, it turns out that this is pretty much the decimation algorithm used by numerical methods like DMRG to approximate the low energy eigenstates in the corner of Hilbert space of gapped local Hamiltonians. It is important to emphasize that MPS states are not unique, and as we have shown there are many variables one could modify to alter the shape and orientation of the chain, not to mention the degree to which the MPS approximates the state vector, including the limit of being an exact representation.

4.5.0 *Explicit examples of quantum states in exact and compressed MPS form*

[7], [9], [12], [17], [21]

To better visualize MPS formulations let us give a few examples starting with some exact MPS 's. Starting with the easiest example, product states, which exhibit no entanglement to more non trivial cases with many bodies.

4.5.1 *MPS of the Transverse Ising model*

given we have a product state, we can simply divide the state vector into a product of matrices, each with both their bond dimensions of 1, implying there is no entanglement between the sites. As expected, this gives us a product of (1×1) matrices $\phi^{(i)}$ at every i^{th} site, yielding

$$|\psi\rangle = |\phi^{(1)}\rangle \otimes |\phi^{(2)}\rangle \otimes |\phi^{(3)}\rangle \ldots\ldots \otimes |\phi^{(N)}\rangle \tag{197}$$

A perfect example of this kind of state is the ground state of the transverse Ising model we discussed in equation (168), where if $J = 0$ we have no interactions, and we get the simplification

$$\hat{H}^{Ising} = -J\sum_i S_i^Z S_{i+1}^Z - B\sum_i X_i$$

$$\downarrow$$

$$\hat{H}^{Ising} = -B\sum_i S_i^X \tag{198}$$

The ground state in this case is the symmetric product state, in which all spins are polarized by the transverse field in x-direction, and so can be decomposed into the trivial form in equation (197). We already know the eigenstates of states with spin along $\pm x$ in the Z-basis, so explicitly the product state is a tensor product of N-many spin$-x$ eigenstates:

$$|\leftarrow,\leftarrow,\cdots\leftarrow,\leftarrow\rangle \equiv \left(\frac{1}{\sqrt{2}}|\uparrow\rangle \pm \frac{1}{\sqrt{2}}|\downarrow\rangle\right) \otimes \cdots \otimes \left(\frac{1}{\sqrt{2}}|\uparrow\rangle \pm \frac{1}{\sqrt{2}}|\downarrow\rangle\right)$$

Which we can reproduce in MPS by substituting (1×1) matrices $M^{[i]\uparrow} = \left(\frac{1}{\sqrt{2}}\right)$ and $M^{[i]\downarrow} = \left(-\frac{1}{\sqrt{2}}\right)$

The other observed limit for the transverse Ising model Hamiltonian, is when $B = 0$ and we are in the thermodynamic limit, i.e. $N \to \infty$ and at $T = 0$, we therefore get a special ground state in this case, which is the twofold degenerate ***ferromagnetic product state,*** which are antisymmetric states

$$|\uparrow,\uparrow,\uparrow,\uparrow,\ldots\rangle \text{ or } |\downarrow,\downarrow,\downarrow,\downarrow,\ldots\rangle \text{ where } |\uparrow,\uparrow,\uparrow,\uparrow,\ldots\rangle = |\uparrow\rangle \otimes |\uparrow\rangle \otimes |\uparrow\rangle \cdots \otimes |\uparrow\rangle$$

The MPS is therefore eigenstates of the spin-Z elements which are the (1×1) matrices (1) and (0), so

$$M^{[2i-1]\uparrow} = M^{[2i]\downarrow} = (1) \quad ; \quad M^{[2i-1]\downarrow} = M^{[2i]\uparrow} = (0)$$

Finally, for intermediate values of B, we get interactions, so bond indices are no longer equal to one, and the matrices grow in dimension. However, ground states for the transverse Ising Hamiltonian with

intermediate transverse field strength, are area law abiding entangled states, and we can get approximations if the state tensor is too large by decimating our matrices through excluding the lowest Singular values.

4.5.2 *MPS of a singlet state*

The second example is one with entanglement, is the singlet state $|\Psi\rangle = \frac{1}{\sqrt{2}}|1\rangle|0\rangle - \frac{1}{\sqrt{2}}|0\rangle|1\rangle$, we have just 2 bodies here, so the MPS chain will consist of just the two ends, which we showed in equations (185) and (186), given $d = 2$ here for half spin states, must be (1×2) and (2×1) matrices:

$$|\Psi\rangle = \frac{1}{\sqrt{2}}\begin{bmatrix} |1\rangle & |0\rangle \end{bmatrix}\begin{bmatrix} |0\rangle \\ -|1\rangle \end{bmatrix} \tag{199}$$

We can also clarify that the state tensor of coefficients can also be written more familiarly in the tensorial form we've been introducing as

$$\Psi^{S_1,S_2} = M^{S_1}M^{S_2} \tag{200}$$

Where M^{S_i} are d-many matrices, so we have two matrices at each site i, one for each spin state, making

$$M^{\uparrow_1} = \begin{bmatrix} 1 & 0 \end{bmatrix} \quad ; \quad M^{\downarrow_1} = \begin{bmatrix} 0 & 1 \end{bmatrix} \quad ; \; M^{\uparrow_2} = \begin{bmatrix} 0 \\ -1 \end{bmatrix} \quad ; \; M^{\downarrow_2} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \tag{201}$$

So in the MPS chain, M^{S_1} points to two matrices, $M^{\uparrow_1}$ and $M^{\downarrow_1}$, and M^{S_2} points to $M^{\uparrow_2}$ and $M^{\downarrow_2}$

4.5.3 *W and GHZ states as MPS's*

Let us now consider matrix product states that exhibit entanglement. We will cover two states which are of huge significance in many quantum information applications, the $|W\rangle$ and $|GHZ\rangle$ states. The property that makes them so unique is that they are two common examples of 3-qubit systems that are maximally entangled. They can also be generalized to higher dimensions.

The previous examples were for exact representations of states, let's consider an example of a large state vector where MPS compresses a much bigger wave function as a matrix product state. Consider the wave function of a single particle state, which has access to N unique states, making the wave function

$$|\Psi\rangle = \phi_1|100\dots0\rangle + \phi_2|010\dots0\rangle + \phi_3|0010\dots0\rangle + \dots \phi_{N-1}|00\dots10\rangle + \phi_N|00\dots01\rangle \qquad (202)$$

this is a generalization of the W-state. We know that for N sites we get a state which exists in 2^N dimensional space, however since it is obvious that there are mostly zeros in the matrices comprising an MPS of this state, we can remove the zero singular values and actually decimate it down to an MPS of just (2×2) matrices. This compression gives us the MPS

$$\begin{bmatrix} \phi_1|1\rangle & |0\rangle \end{bmatrix} \begin{bmatrix} |0\rangle & 0 \\ \phi_2|1\rangle & |0\rangle \end{bmatrix} \begin{bmatrix} |0\rangle & 0 \\ \phi_3|1\rangle & |0\rangle \end{bmatrix} \begin{bmatrix} |0\rangle & 0 \\ \phi_4|1\rangle & |0\rangle \end{bmatrix} \dots \begin{bmatrix} |0\rangle & 0 \\ \phi_{N-1}|1\rangle & |0\rangle \end{bmatrix} \begin{bmatrix} |0\rangle \\ \phi_N|1\rangle \end{bmatrix} \qquad (203)$$

Notice that the zeros in red are empty vector space, it's not the state ket zero

Now to better illustrate how this form turns into the original state vector lets multiply them out. After the first multiplication we get

$$|\Psi\rangle = [\,\phi_1|10\rangle + \phi_2|01\rangle \quad |00\rangle] \begin{bmatrix} |0\rangle & 0 \\ \phi_3|1\rangle & |0\rangle \end{bmatrix} \begin{bmatrix} |0\rangle & 0 \\ \phi_4|1\rangle & |0\rangle \end{bmatrix} \dots \begin{bmatrix} |0\rangle & 0 \\ \phi_{N-1}|1\rangle & |0\rangle \end{bmatrix} \begin{bmatrix} |0\rangle \\ \phi_N|1\rangle \end{bmatrix}$$

$$= [\,\phi_1|1000\rangle + \phi_2|0100\rangle + \phi_3|0010\rangle + \phi_4|0001\rangle \quad |0000\rangle] \dots \begin{bmatrix} |0\rangle & 0 \\ \phi_{N-1}|1\rangle & |0\rangle \end{bmatrix} \begin{bmatrix} |0\rangle \\ \phi_N|1\rangle \end{bmatrix}$$

$$= \phi_1|100\dots0\rangle + \phi_2|010\dots0\rangle + \phi_3|0010\dots0\rangle + \dots \phi_{N-1}|00\dots10\rangle + \phi_N|00\dots01\rangle$$

Which is the original state.

Our final example will be one for an N-body GHZ state. Its generalization takes the form

$$|GHZ\rangle = \frac{1}{\sqrt{2}}(|1,1,\dots,1\rangle + |0,0,\dots,0\rangle) \tag{204}$$

The MPS here is pretty straight forward and would look like

$$|GHZ\rangle = \frac{1}{\sqrt{2}}\begin{bmatrix} |1\rangle & |0\rangle \end{bmatrix}\begin{bmatrix} |1\rangle & 0 \\ 0 & |0\rangle \end{bmatrix}\begin{bmatrix} |1\rangle & 0 \\ 0 & |0\rangle \end{bmatrix}\begin{bmatrix} |1\rangle & 0 \\ 0 & |0\rangle \end{bmatrix}\dots\begin{bmatrix} |1\rangle & 0 \\ 0 & |0\rangle \end{bmatrix}\begin{bmatrix} |1\rangle \\ |0\rangle \end{bmatrix} \tag{205}$$

It is obvious from the (2×2) matrices that the dimension of bond indices is two

We could also express the MPS in the same form we did with the singlet state, where in equation (199), we expressed the MPS as a product of matrices whose matrix coefficients are themselves states, while in equations (200),(201), we expanded it into a more tensorial form to dissect it into the d many matrices at every site, which is two given we are dealing with spin half states. This gives

$$\Psi^{S_1,S_2,S_3\dots S_N} = M^{S_1}M^{S_2}\dots M^{S_N} \tag{206}$$

So, the start and end (1×2) and (2×1) matrices are

$$M^{\uparrow_1} = \begin{bmatrix} \frac{1}{\sqrt{2}} & 0 \end{bmatrix} \quad ; \quad M^{\downarrow_1} = \begin{bmatrix} 0 & \frac{1}{\sqrt{2}} \end{bmatrix} \quad ; M^{\uparrow_N} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \quad ; M^{\downarrow_N} = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \tag{207}$$

While the intermediate matrices in the chain are

$$M^{\uparrow_i} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \quad ; \quad M^{\downarrow_i} = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \tag{208}$$

5.0.0 *Section 5: Running Operations on Tensor Network Diagrams* [9], [14], [22]

Now that we've established the tensor diagram and MPS formalisms, we should discuss briefly how convenient it is to run different operations, including extracting reduced density matrices from tensor diagrams, calculating expectation values and taking partial traces, both from a density matrix formalism and an MPS formalism. The goal is to systematically lead the reader through utilizing the functionality of general tensor diagrams with the conventional density matrix techniques introduced earlier for running algorithms like extracting entanglement entropy; the follow with the even smarter utility of MPS decompositions to very efficiently calculate, or approximate, entanglement entropy of an exponentially large multipartite system given a certain bipartition cut in the MPS chain. This is where all the work done in this study converges. We will start by illustrating some basic operations using tensor diagrams and conventional density matrix formalism, then extend to calculating the entanglement entropy whilst utilizing general tensor networks with density matrices, then upgrade specifically to MPS operations.

5.1.0 *Some basic operations in tensor diagram notation*

We've already discusses trivial operations such as traces and matrix multiplications in section (4.3.1), it's time to utilize tensor diagrams to visualize more complicated operations such as partial traces and SVD, which are the keys behind deriving entanglement entropy. To define entanglement of a tensor diagram, we will follow the same steps we introduced for the generic many-body states need. So first we specify a bond index at which we separate the chain into a bipartition with subsets A and B. Referring back to section (3.3.0) , equation (136) gives the von Neumann entropy for a bipartite system is

$$S^{von\ Neumann} = -\sum_m \lambda_m\, In\, \lambda_m = -tr\left(\rho^{(/A\ or/\ B)} In \rho^{(/A\ or/\ B)}\right)$$

Where we use reduced density matrices of either subsystem. This involves taking the partial trace, so we need to define these operations in tensor diagram notation. To do so, we refer back to the criteria we introduced in section (4.3.1) with a small modification of writing indices as plain lines for simplicity. Given a partitioning as in figure (18) we can derive the reduced density matrix of sub-part A, $\rho^{(/A)}$, by tracing over region B from the density matrix of the pure state AB, which has a density matrix $\left(\rho_{pure}{}^{(:AB)}\right) = (|a,b\rangle\langle a',b'|)$, since as we derived in equation $\hat{\rho}_{pure}{}^{(/A)} = tr^{(/B)}(\,|a,b\rangle\langle a',b'|\,)$

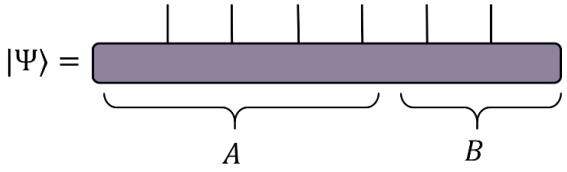


Figure (18): Tensor diagram of a bipartite system's Wave function

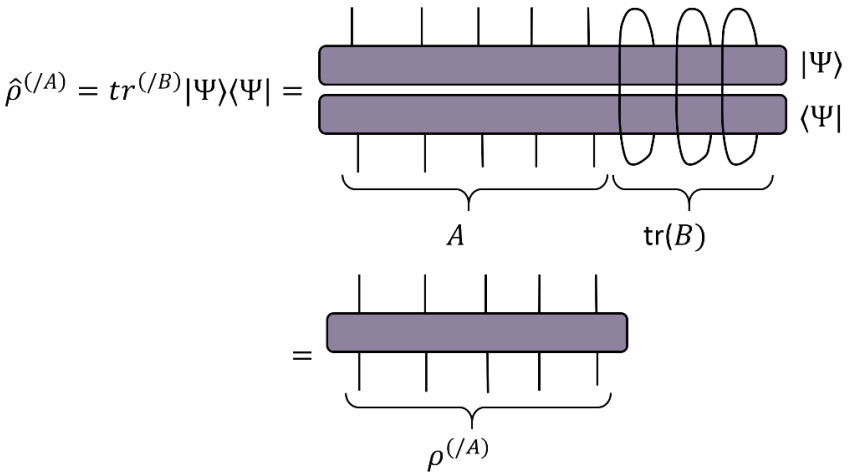
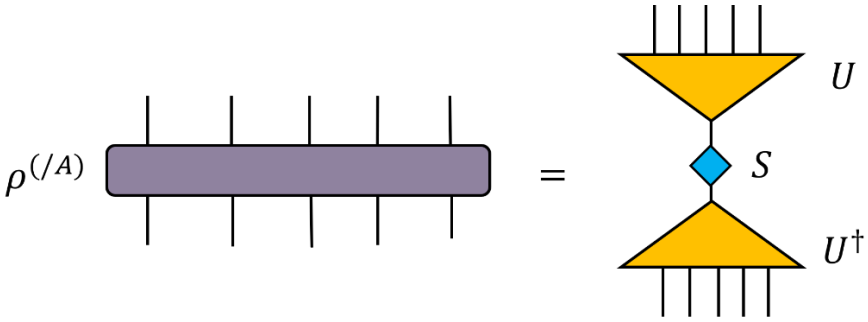


Figure (19): Tensor diagrams illustrating the Derivation of a reduced density matrix of a bipartite state

Now, given $\rho^{(/A)}$ shown in figure (19), we can extract the entanglement entropy using

$$S^{von\ Neumann} = -\sum_{m} \lambda_m\ In\ \lambda_m$$

by diagonalizing $\rho^{(/A)}$ to get its eigenvalues λ_m, achievable by a similarity transformation $\rho = USU^{\dagger}$, which extracts the eigenvalues into a central diagonal matrix S. In tensor diagram notation this is simplified to



$$S = \begin{pmatrix} \lambda_1 & 0 & 0 & 0 & 0 \\ 0 & \lambda_2 & 0 & 0 & 0 \\ 0 & 0 & \lambda_3 & 0 & 0 \\ 0 & 0 & 0 & \lambda_4 & 0 \\ 0 & 0 & 0 & 0 & \lambda_5 \end{pmatrix}$$

Figure (20): Tensor diagram of a unitary transformation of the reduced density matrix

We've already discussed as we showed in equation (142) that if we have maximum entanglement for spin states, we get all equal eigenvalues λ_m so all contribute equally to the entanglement entropy, where $\lambda_m = \frac{1}{N}$, so if $\rho^{(/A)}$ has five bodies like in this example $N = 5$ and the sum is just a multiple of N due to equal contributions in this case and we get

$$S^{von\ Neumann} = -\sum_{m=1}^{5} \lambda_m\ In\ \lambda_m = -5 * \left[\ \frac{1}{5} In\ 5^{-1}\ \right] =\ In\ 5 = In\ N$$

Which as we can see is not a constant in N, so this is an example of entropy scaling with the size of the system, an ***entropic volume law***, given all particles interact with the full length of the system because of maximum entanglement.

Next, we introduce operators. An operator will receive a wavefunction's amplitude tensor, and convert it into another amplitude tensor, so it has two copies of physical indices, one set to receive a wavefunction and the other to output the other with the same number of free indices. An operator in tensor diagram notation will look like

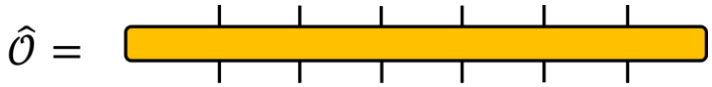


Figure (21): An operator in tensor diagram notation acting on a six body wavefunction

An operator can also be used to derive ***expectation values of operators***, so basically ***observables***. Here the operator connects a ket and a bra version of the wavefunction, and outputs a scale with no open indices, a scalar looking like

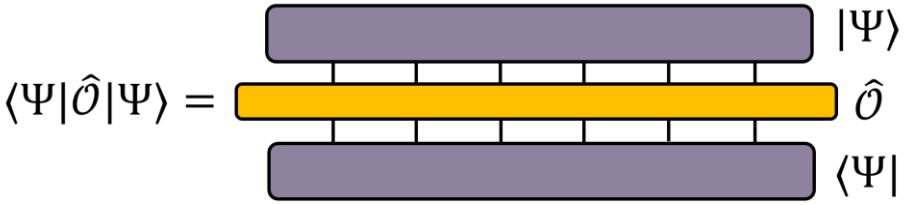


Figure (22): Tensor diagram calculating the expectation value of an operator acting on the entire wavefunction

An operator can also act on a specific part of the wavefunction, so it might only affect a region of the Hilbert space. For example, if we return to our example of a bipartite system in figure (19), we can have an operator which acts only on subsystem A. We've already derived how to simplify these types of

operations in section (2.2), specifically equations 100 and 106, where the expectation value of an operator acting in a global state AB, but targets the subspace A will be given by

$$\langle \mathcal{O}^{(A)} \rangle^{(:AB)} = tr^{(/A)}(\rho^{(/A)}\, \mathcal{O}^{(A)})$$

This process in tensor diagram notation simplifies to

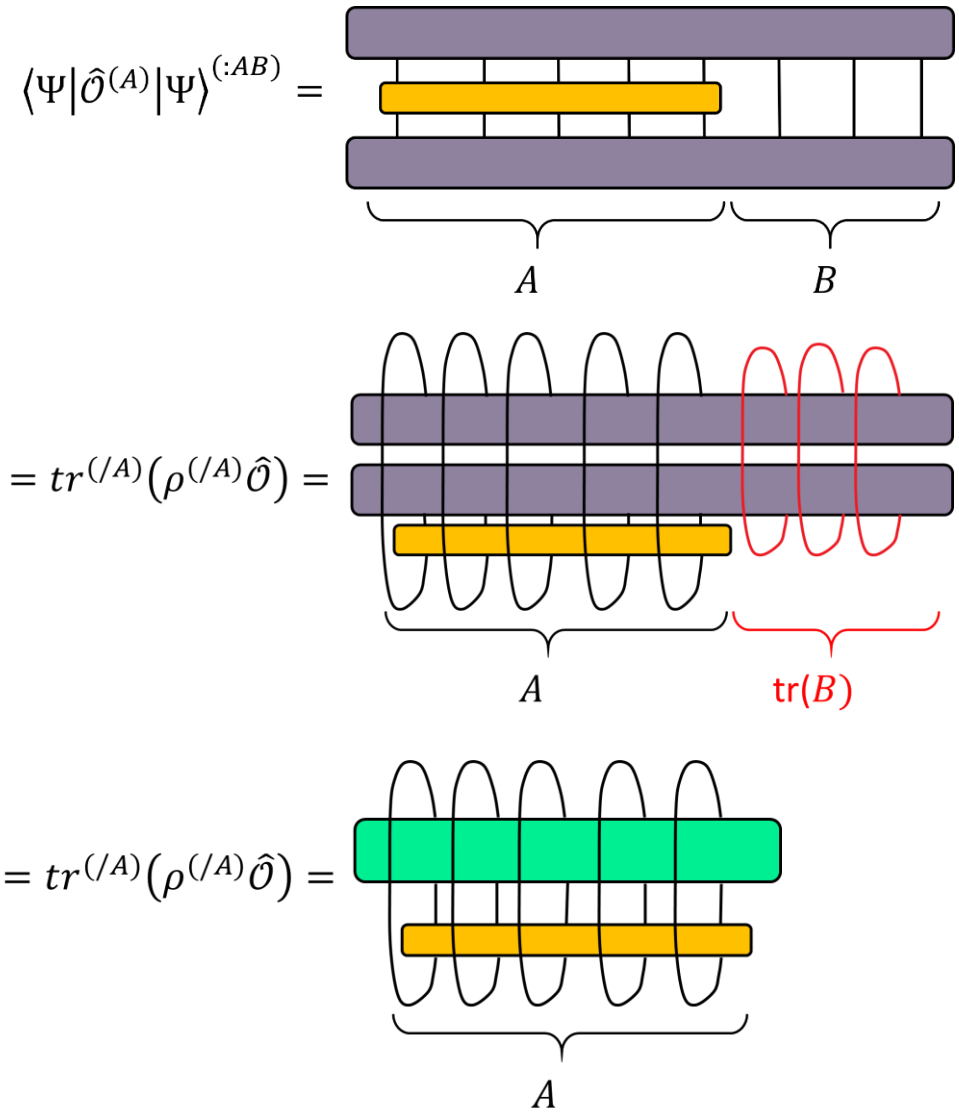


Figure (23): Tensor diagram for the expectation value of partial operator acting on a subset of the Hilbert space containing the wavefunction

Instead of using density matrices, we could instead utilize Singular value decomposition of the state amplitude tensor to run very similar operations and derive the entanglement entropy that way. Like we discussed in section (4.4.2) , we can use collective indices to partition the state vector into two parts and apply SVD to it, analogous to what we did with similarity transformations on reduced density matrices.

Through SVD we can then extract the singular values from the diagonal matrix. To refresh one's memory, refer back to section (3.1.0), where we go over the SVD and Schmidt decompositions. Now, given an 8-body state $\Psi^{(s_1 s_2 s_3 s_4 s_5 s_6 s_7 s_8)}$, we can decompose it into

$$\Psi^{(s_1 s_2 s_3 s_4 s_5 s_6 s_7 s_8)} = \sum_m U_m^{(s_1 s_2 s_3 s_4)} O_{m,m} V_m^{(s_5 s_6 s_7 s_8)}$$

Which in turn has the tensor diagram representation of

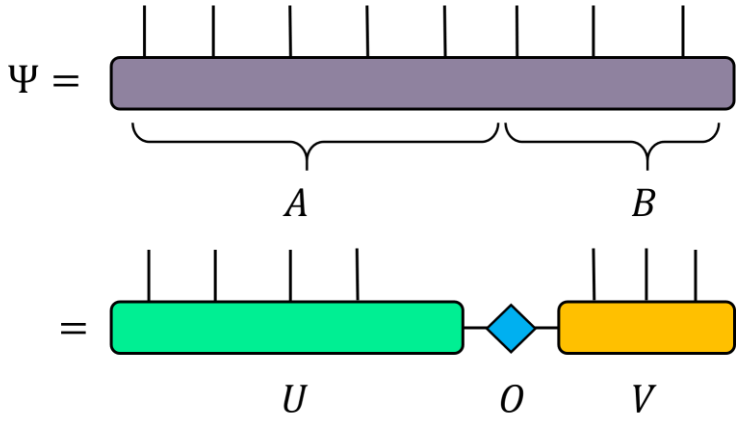


Figure (24): Tensor diagram of an SVD

Where the diamond matrix here is the diagonal central matrix representing the bond between the two subsystems. We can see that this single decomposition around a single site in along the wave vector has a significant effect on the number of parameters. Going back to section (4.4), we can see how the benefits from a single decomposition at one bond site, can be extended to the entire wave vector, by iterating this process over all bonds. This was the motivation behind the MPS formulation. It is now time to consider what MPS diagrams in particular have to offer when it comes to running operations.

5.2.0 *Operations and scaling of calculations with Matrix Product States and Matrix Product Operators*

The easiest operation we could use to demonstrate the kind of utility introduced by the MPS formalism is calculating the norm of a state. Given some large six-body state, we define the norm as

$$\langle \Psi | \Psi \rangle = \sum_{s_1, s_2 \ldots s_9} \Psi^{s_1 s_2 s_3 s_4 s_5 s_6} \overline{\Psi}_{s_1 s_2 s_3 s_4 s_5 s_6}$$

This is a scalar resulting from a contraction across all five indices, which would give the diagram

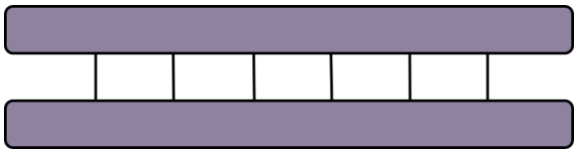


Figure (25): A tensor diagram of taking the norm of two state amplitude tensors

Now this operation is basically summing 2^6 terms, which is not very efficient. However, if we break it down to an MPS we discover a very clever way to contract the chain incrementally. The MPS version of figure (25) is

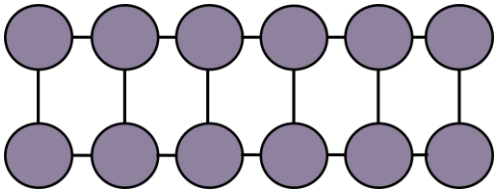


Figure (26): The MPS version of taking the norm of two state amplitude tensors

We can now seek to contract these matrices starting from the two left most matrices then incrementing the other matrices one at a time. The process looks something like

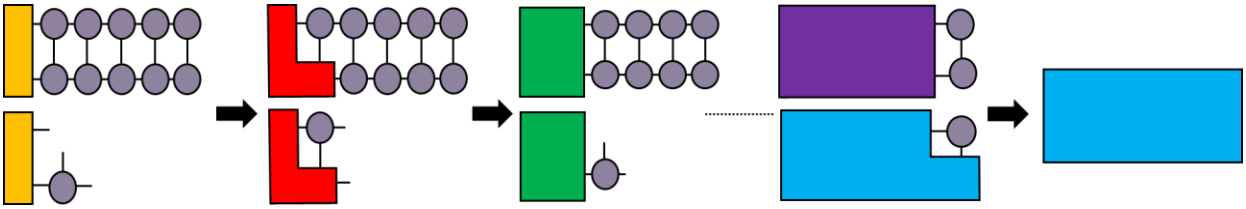


Figure (27): Contracting an MPS incrementally

This algorithm reduces the complexity of the contraction considerably. Consider the contraction in figure (28-b), If the physical legs (i.e. the site indices s_i) have dimension d and the bond indices have maximum dimension χ as shown in the figure (28-a) tensor diagram; The scaling on this contraction is $m^3 d$, since we contract the indices in gold, where they contribute a factor of $(\chi \times d)$, while we have to repeat this process for every unique combination of the open indices in black. This is analogous to taking a cross section in the tensor at a given value of each open index, which means we have $(\chi \times \chi)$ repetitions, so overall the cost is

$$(\chi \times d) \times (\chi \times \chi) = \chi^3 d \tag{209}$$

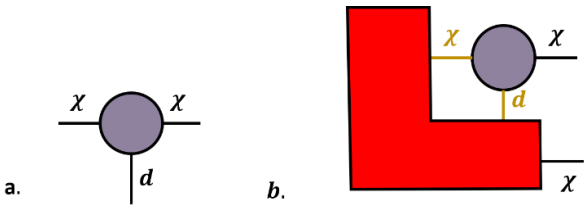


Figure (28): a) a single rank three tensor in an MPS, with its bond indices, χ, and its physical index d. b) A contraction step in deriving the norm depicted in figure (27), with two contracting indices and two open indices.

If we run the same incremental contraction, we get a huge compression of the norm operation from an exponentially growing order to a polynomial. This same cost applies to all the other MPS contractions at all the available bonds along the chain, we can therefore find that we can get operational costs down to some order $O(\chi^3)$, and we know that usually the MPS bond index would diverge polynomially with size in case of moderate entanglement in a system.

An important example of an operator whose expectation value we regularly calculate is the Hamiltonian, whose expectation value is simply the energy of a system, which again we usually seek to minimize to reach ground states. The tensor diagram for this operation would be

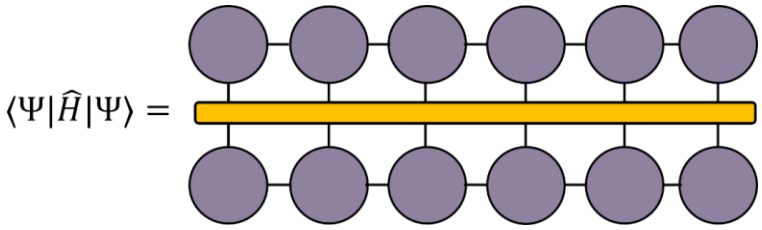


Figure (29): The expectation value of the Global Hamiltonian acting on an MPS

In figure (29) the Hamiltonian is acting on the entire wavefunction, so it is acting on all the sites simultaneously, and we can clearly see that it has no open indices and the result is a scalar, the expectation value of the operator, in this case, the expectation value of the Hamiltonian, which is the energy of the MPS. To emphasize how instructive this representation is, it is worthwhile to mention that DMRG algorithm, very crudely put, is basically an algorithm that would update the indices in figure (29), in order to minimize the expectation value of the Hamiltonian, thereby obtaining the ground state energy.

One important thing to mention here is that if we were dealing with an operator that, as we mentioned earlier acts on the entire tensor indices simultaneously, we wouldn't really be efficient. Therefore, we discussed in great detail how MPS's were mostly efficient when dealing with Gapped Local hamiltonians, with decaying correlation length. The Hamiltonian of such systems is the sum of individual local Hamiltonians, an example of which would be the Ising model negligible transverse field

$$\hat{H}^{Ising} = -J \sum_i S_i^Z S_{i+1}^Z$$

The $S_i^Z S_{i+1}^Z$ interaction term runs over nearest neighbor interactions, so it is by construction a local Hamiltonian. This can be elegantly expressed in terms of tensor diagrams by

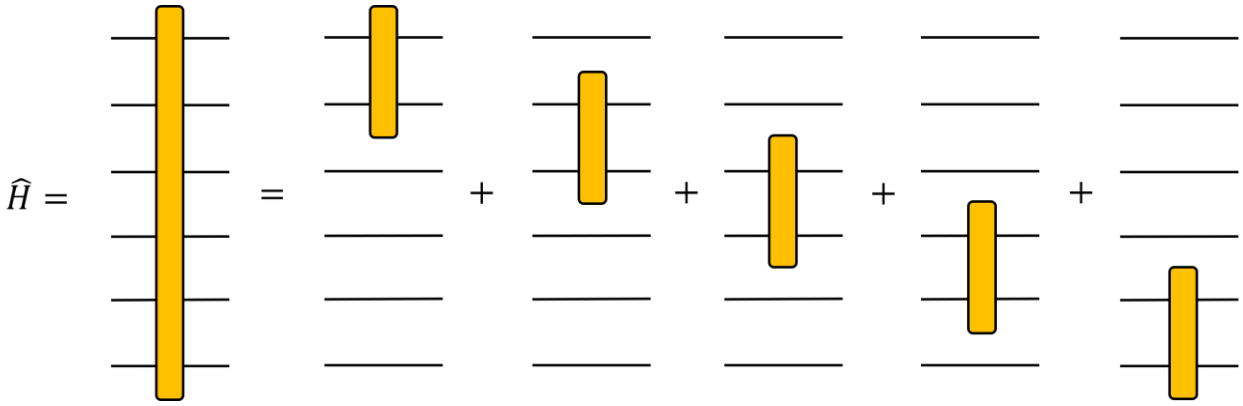


Figure (30): Global hamiltonian as a sum of local Hamiltonians in tensor diagram notation

Figure (30) shows a local Hamiltonian with nearest neighbor interactions, where each local hamiltonian term interacts with just two adjacent indices leaving the rest untouched. The untouched indices are simply tensor products of local identity operators. This allows us to deal with reduced parts of our wavefunction, which are effectively reduced density matrices of our global state. Now given we can decide an operator into multiple local operators, if we connect all the local operators in a certain manner, we get a peculiar structure called Matrix product operators. It's outside our scope to go into rigorous details as we did with MPS's, but the idea is generally pretty simple to follow. If we connect the local structures shown in figure (30) at their corresponding adjacent indices we could get the following structure

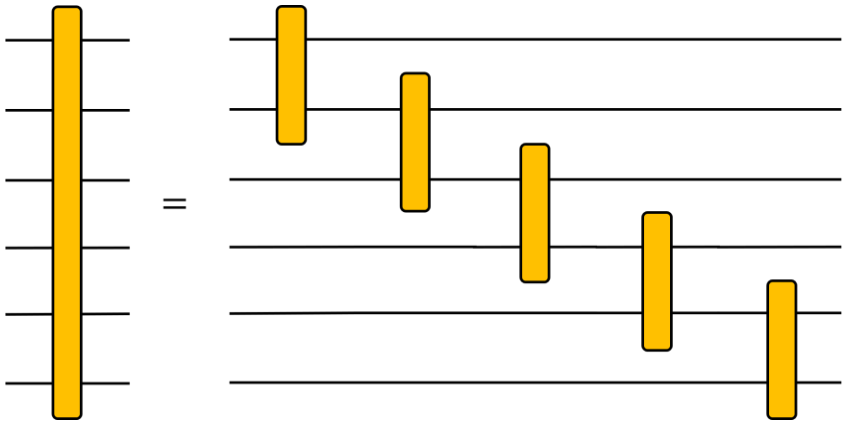


Figure (31): Partitioning a global operator into a Matrix product Operator

The structure shown in figure (31) is a simple example of a Matrix Product Operator (MPO). If figure (31) looks familiar to you, then it probably is, this looks like a quantum circuit of entangled qubits, and in fact, this was not intentional, it turns out that this similarity was the key behind realizing that tensor diagrams could be reproduced faithfully into quantum circuits, and this is the subject of section (6.1.2). Anyway, like any operator it acts on a collection of local states to output another state, or it can sandwich between two states and output a scalar, for example, energy, or any other observable. A quick example is the nearest neighbor Hamiltonian analogous to that of the Ising model, let us say

$$\hat{H} = \sum_i S_i^z S_{i+1}^z$$

For a 3-qubit system we get 2 sums giving us the local Hamiltonian term

$$\hat{H} = S_1^z S_2^z + S_2^z S_3^z$$

To deal with the full Hilbert space we expand the operators by an extra tensor product with the Identity matrix to indicate inaction towards the Hilbert space basis they do not interact with, which is the third state for the first term and the first state for the second term, giving

$$\hat{H} = S_1^z S_2^z + S_2^z S_3^z = S_1^z S_2^z \mathbb{I}_3 + \mathbb{I}_1 S_2^z S_3^z \tag{210}$$

This local Hamiltonian can be broken down into the matrix product

$$\begin{bmatrix} 0 & S_1^z & \mathbb{I}_1 \end{bmatrix} \begin{bmatrix} \mathbb{I}_2 & 0 & 0 \\ S_2^z & 0 & 0 \\ 0 & S_2^z & \mathbb{I}_2 \end{bmatrix} \begin{bmatrix} \mathbb{I}_2 \\ S_3^z \\ 0 \end{bmatrix} \tag{211}$$

This is a matrix product operator for the given Hamiltonian. Another common Hamiltonian is that of the Heisenberg model, with the general N-body Hamiltonian is given by

$$\hat{H} = \sum_{i=1}^{N} J S_i^z S_{i+1}^z + \frac{J}{2}(S_i^+ S_{i+1}^- + S_i^- S_{i+1}^+) \tag{212}$$

This will have an MPO of

$$\begin{bmatrix} 0 & \frac{J}{2}S_1^- & \frac{J}{2}S_1^+ & JS_1^z & \mathbb{I}_N \end{bmatrix} \begin{bmatrix} \mathbb{I}_N & 0 & 0 & 0 & 0 \\ S_i^+ & 0 & 0 & 0 & 0 \\ S_i^- & 0 & 0 & 0 & 0 \\ S_i^z & 0 & 0 & 0 & 0 \\ 0 & \frac{J}{2}S_i^- & \frac{J}{2}S_i^+ & JS_i^z & \mathbb{I}_N \end{bmatrix}^{(N)} \begin{bmatrix} \mathbb{I}_N \\ S_N^+ \\ S_N^- \\ S_N^z \\ 0 \end{bmatrix} \tag{213}$$

The importance of the MPS formulation here, is that we can take the ability to compute local terms efficiently, then we can operationalize that into a very effective algorithm by extending this to many-body Hamiltonians. In the next section, we will show how one can use the ITensor library to build any MPO given a sum of local or even a global Hamiltonian.

5.3.0 *Running numerical calculations on tensor networks using the ITensor C++ Library* [14]

There are several powerful libraries out there that were specifically designed to deal with numerical calculations on tensor networks and MPS's. One such library is the ITensor library [14] available in both C++ and Julia programming lan. We will go over some of its main functionalities, including a tensor network operation on a singlet state, and afterwards a DMRG numerical calculation on an MPS generated by generating and running an MPO given some sum of local Hamiltonians. We finally use SVD on the resultant MPS to extract von Neumann entanglement entropy around a specific MPS bond.

5.3.1 *Building the singlet state and calculating its energy under the Heisenberg Hamiltonian*

The full code corresponding to following steps illustrated in the figures can be found in Appendix (2). First, we define the operators as tensors using the Heisenberg hamiltonian terms given in equation (212), and use ITensor to create these operators individually. Afterwards, we can multiply and sum them up accordingly into the full Hamiltonian for $N = 2$, since we will be acting on a Singlet. This whole process in the algorithm is fully followed in the code and the corresponding tensor diagram notation is given in figures (32).

First consider the singlet, which as we already know, will be a superposition as shown in figure (32), analogous to two slices in the tensor. This is a tensor with two indices, illustrated in figure (32a), so it represents the most general two spin state. If we take elements $(1,2)$ and $-(2,1)$ each weighted at $\frac{1}{\sqrt{2}}$, we get the singlet state, as illustrated in figure (32b) and it can be given by the tensor

$$\begin{pmatrix} 0 & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & 0 \end{pmatrix} \tag{214}$$

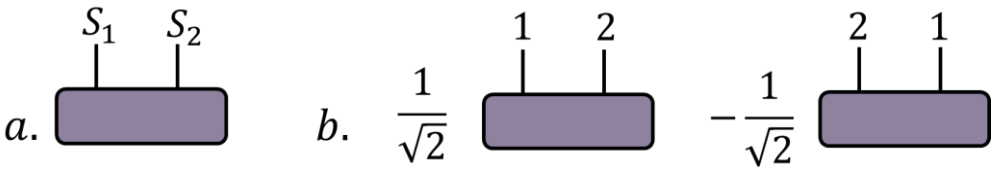


Figure (32): a) The most general two-half spin particle state. b) The singlet state

The Hamiltonian here is a sum of local Hamiltonians and is illustrated in figure (33). The lower-side indices Hamiltonian is now ready to receive the S_1 and S_2 indices of the singlet and generate a new state.

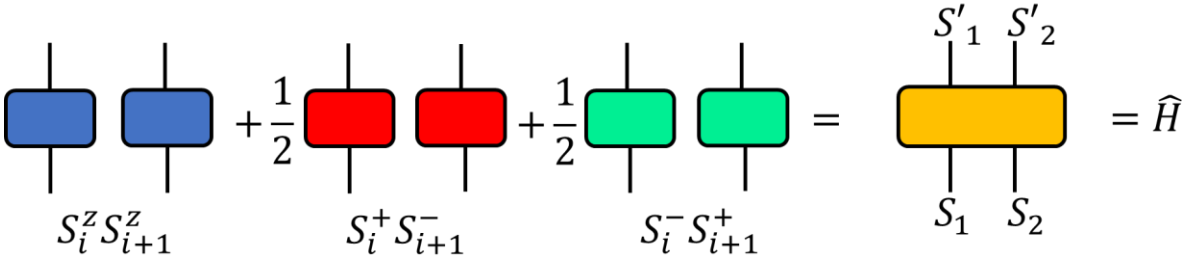


Figure (33): Forming the general Heisenberg model Hamiltonian from a sum of local operators

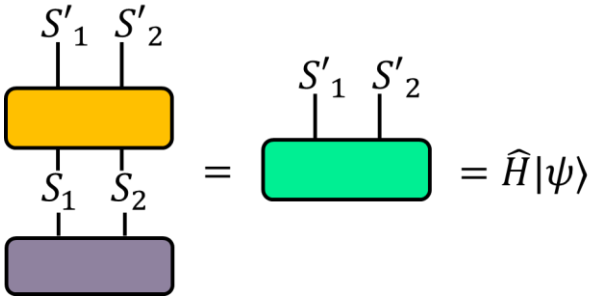


Figure (34): The new state from applying the Heisenberg Hamiltonian on the singlet

All That remains is calculating the energy, which is the expectation value $\langle\psi|\hat{H}|\psi\rangle$. That would mean contracting the $\hat{H}|\psi\rangle$ result with $|\psi\rangle^{\dagger} = \langle\psi|$ to get the energy, so $E = \langle\psi|\hat{H}|\psi\rangle$, and the corresponding tensor diagram is

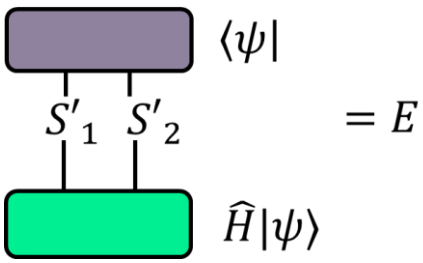


Figure (35): Expectation value of the Heisenberg Hamiltonian acting on the singlet

After running the ITensor code in Appendix (2), we got the energy as the output, which read

Initial energy = -0.75

Now we wish to utilize the Library to its full potential in the next section, we will run some DMRG code and extract the entanglement entropy from a 20-body state on which we apply an accordingly much larger Heisenberg Hamiltonian than the one we used for the singlet.

5.3.2 *Constructing a 20-body MPO for the Heisenberg Hamiltonian and using DMRG and SVD to extract von Neumann entanglement entropy from the output MPS*

Here it would be very difficult with larger chains to define the Hamiltonian by writing-in local contributions, term by term, like we did with the Heisenberg Hamiltonian on a singlet. We instead utilize ITensor's functionality to construct the Heisenberg Hamiltonian as a summation term to be iterated N times over, as we conventionally do with pencil-and-paper notation, similar to the sum in equation (212). The full code can be found in Appendix (2). After we write down the general terms in the Hamiltonian, we define an MPO using a special function in the library and initiate the Hilbert space, which in this case translates into generating a 20 site chain, creating a Hilbert space of 20 spin-half sites. Next the initial state for the sites can be set, here we will set each even site to "spin up" and odd site to "spin down" product states, and use this as our starting MPS $\psi_{initial\ MPS}$ defined in the same 20-unit Hilbert space. Finally, we call the DMRG function, which approximates the ground state energy by decimating the least involved bond indices. DMRG overrides the ψ_{MPS} with an ***optimized ground state wavefunction*** $\psi_{DMRG\ MPS}$. We can then gauge this MPS to a certain bond, in this case the 4^{th}. The MPS gauge renders the matrices to its right as right canonical, and those to its left as left canonical. This was discussed in section (4.4.2) and (4.4.3), where we explained how a mixed canonical MPS allows us to use an Schmidt decomposition by introducing orthonormality to both sides of the bond site separating the bipartitions. This gauging identifies which bond we will measure the von Neumann entropy across. Now given the MPS gauged at bond 4, we can run a Schmidt decomposition, and extract the remaining relevant singular values

between the two partitions, then directly use the von Neumann entanglement entropy equation to calculate the entanglement entropy. Translating this algorithm to tensor diagram notation gives the following flowchart in figure (36).

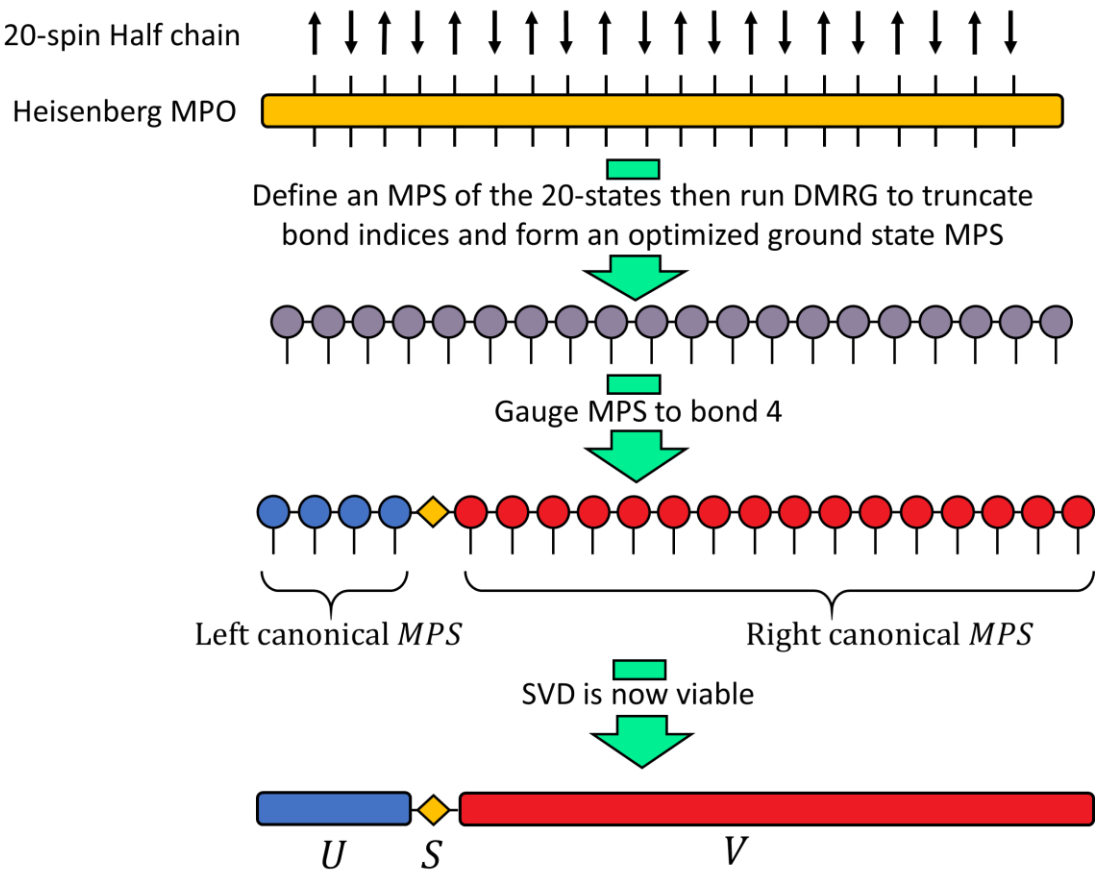


Figure (36): algorithm followed to run DMRG and extract singular values around a given MPS gauge

The final step was using the diagonal S matrix, to extract the relevant singular values, after running our code the output showed that the square of the diagonal matrix was

$$\text{◆}^{2} = \begin{pmatrix} 0.0491998 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.0000595 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.879603 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.0391701 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.00002186 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.0319458 \end{pmatrix} \tag{215}$$

The final line of code then uses the squares of the singular values to calculate the entanglement entropy as per equation (129) giving the von Neumann entropy SvN

$$\text{Across bond b=4} \quad \begin{matrix} SvN_{base\ ln} = 0.4987535674 \\ SvN_{base\ 2} = 0.7195492005 \end{matrix} \tag{216}$$

6.0.0 *Section 6: Matrix Product States as faithful representations of quantum circuits*

In section (5.2.0) we showed how using MPS's reduces the computational cost of contracting over bond indices χ, and by extension, running numerical operations for computing different expectation values in general. Given χ is bounded, many tensor network factorizations can be efficiently evaluated to a polynomial order computational cost for many-body entangled structures, with emphasis on entanglement dictating the behavior of the bond indices. We also saw how the individual design and bond placements, affect the efficiency of calculations on a network, so a poorly designed tensor network will experience rapidly increasing χ's to keep within acceptable decimation error.

Now that we've determined the importance of the entanglement structure in a many-body wavefunction, we look at a state of multiple qubits on a quantum computer for reasons that will become clear to the reader in a moment. Entanglement in multi-qubit systems is dependent on its quantum circuit, and these circuits are very case sensitive. In practice, the choice of quantum circuit design will often be motivated by numerical convenience and operation, and no single design can fit all. It turns out that the same utility we demonstrated with MPS's extends to quantum circuit design as well. The reason for this relation is that quantum circuits are nothing more than networks in a quantum computer, which are set up to ***manipulate multi-qubit quantum states*** [23]–[27].

6.1.0 *Quantum computer circuits as tensor networks*

Using quantum circuits to run computational operations is one of the pillars of contemporary quantum information theory and computing. Since quantum computers, like their conventional counterparts, have memory requirements and computational runtimes, the cost of operations on the states becomes a determining factor in the success of a circuit. Considering that multi-qubit states are basically many-body quantum systems, and accordingly have entanglement entropies governed by the same physics we introduced thus far in the study, we expect to be able to represent the state vector of multiple qubits as tensor amplitudes and work our way through. One can already see where we're going with this; we will find that there is a direct ***one-to-one mapping of quantum circuits to tensor networks***.

6.1.1 *Qubit state initialization, single and Multiple qubit Quantum gates and operators* [5], [26], [28]

Quantum circuits are comprised of three main elements; quantum gates; wires; input states. The input states are single qubits. We've already described the qubit as being the simplest quantum state and have mapped its Hilbert space to Bloch spheres in section (1.1.0). A qubit is the building block of the global state on which a quantum circuit will run operations governed by quantum gates. Referring back to section (5), we express operations as contractions with matrices, hence the quantum gates are basically matrices acting on qubit states. The matrices here either act on a single qubit, or multiple qubits.

Single qubit gates must be must be 2×2 , since they act on a single 2 state system, and must be unitaries, because the gate is meant to change qubit states, as in evolve the quantum system by applying a transformation. Unitary transformations, as we have mentioned multiple times, do not change the eigenvalues, and simply alter the eigenbasis, and are therefore reversible and preserve the norm of the state vector. ***This transformation on the Block sphere is therefore a rotation of Hilbert space***, and the pure state remains on the surface of the Bloch sphere. Looking back at equations (1) and (3), we presented the general qubit state as

$$|\psi\rangle = \cos(\theta/2)|0\rangle + e^{i\varphi}\sin(\theta/2)|1\rangle$$

Referring to IBM's quantum circuit builder [26] and open cloud quantum computing services, IBM Q Experience ®, we'll identify the most basic gates available in the builder. First we start with single qubit gates; there's the universal gate, which is the most general transformation gate, which takes an input of three angles θ, φ, λ as shown in equation (217), and so rotates the computational basis, which is conventionally the Z-states, towards an arbitrary new state on the surface of the sphere. Therefore, a general single qubit quantum gate $U_3(\theta, \varphi, \lambda)$ will be given by

$$U_3(\theta, \varphi, \lambda) = \begin{pmatrix} \cos(\theta/2) & -e^{i\lambda}\sin(\theta/2) \\ e^{i\varphi}\sin(\theta/2) & e^{i(\lambda+\varphi)}\cos(\theta/2) \end{pmatrix} \tag{217}$$

Second, we have the X, Y, Z ***Pauli gates***, which are common specific cases of the U_3. From the name we can immediately guess what they do and that they are represented by the corresponding Pauli matrices

$$X = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \qquad Y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \qquad Z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \tag{218}$$

These are used to perform a ***half rotation of the Bloch sphere around the x, y and z axes*** respectively. For example, the Z gate will interchange $|X_+\rangle$ and $|X_-\rangle$ by rotating the sphere around the z-axis, as illustrated in figure (37). The X gate similarly, as is evident in its off-diagonal matrix, reverses the z-basis.

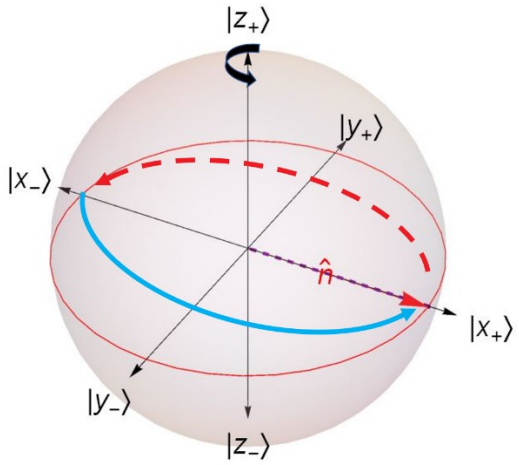


Figure (37): The effect of the Z-pauli gate on the X-states

Thirdly, we have the Hadamard gate H. This gate is key, since it generates superposition states by having the matrix

$$H = U_3(\pi/2,0,\pi) = \frac{1}{\sqrt{2}}\begin{pmatrix}1 & 1\\ 1 & -1\end{pmatrix} \tag{219}$$

It is clear from equation (219) that Superposition is achieved by rotating the z-basis into the x-basis, so

$$|0\rangle \rightarrow |X_+\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$$

$$|1\rangle \rightarrow |X_-\rangle = \frac{1}{\sqrt{2}}(|0\rangle - |1\rangle) \tag{220}$$

Since the x-basis is just a superposition of the z-basis

This is simply a twostep rotation as is indicated by the equivalence of the H-gate to the general rotation $U_3(\pi/2,0,\pi)$, as shown in equation (219), which receives input angles $\theta = \frac{\pi}{2}$, and $\lambda = \pi$, so on the Bloch sphere, it would rotate the sphere about the y-axis by $\frac{\pi}{2}$, then about the x axis by π.

Fourthly we have a single qubit operator, the ***irreversible measurement operator***, which collapses the qubit state by measuring it in the z-basis.

The final gate is the only two-qubit gate we will be dealing with, the Controlled Not gate $(CNOT)$, also known as the Controlled-X gate CX-gate. This gate is the ***entanglement gate***, which makes sense that it be a multiqubit gate. Multi-qubit gates will therefore act on all the available m-qubit states, which as we know are 2^N, so exponential in the N of qubits. Of interest is the two-qubit gate, so the matrix is of dimension $2^2 \times 2^2 = 4 \times 4$. The $CNOT$ gate will relate a ***control qubit***, which does not change, to a ***target qubit,*** which will flip z-states as the CX performs an X transformation on the target iff the control qubit is $|1\rangle$, which means ***if the control qubit is in a superposition***, which might be brought on by the H-gate on the control beforehand, then the CX-gate will result in entanglement.

$$CNOT = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix} \tag{221}$$

We can clearly identify from equation (221), that the $CNOT$ is block diagonal, where the first block acts on the first qubit, and is simply a (2×2) identity block, and the second block is just a Pauli X.

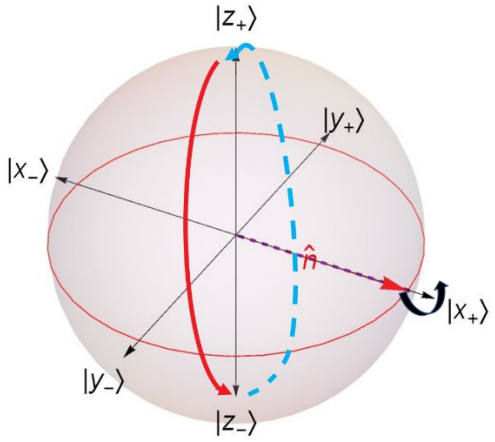


Figure (38): Bloch sphere illustration of the action of a $CNOT$ gate

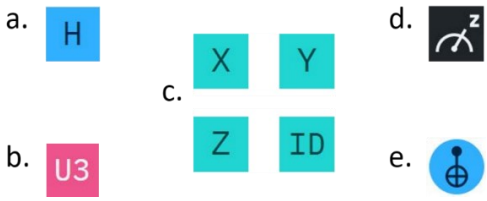


Figure (39): a) The Hadamard gate in the circuit builder interface of IBM's Q-Experience [26]. b) the general transformation gates. c) Pauli X, Y, Z gates and the Identity gate. d) The irreversible measuring operator. e) The 2-Qubit entanglement gate, $CNOT$,or CX gate.

Since the available multi-qubit states grow exponentially with the system size, as is expected from a many-body quantum system, these circuits can become increasingly inefficient with the size of the system. That's where our work with the functionality and efficiency of tensor networks, in particular MPS's comes in.

6.2.0 *Running an MPS quantum circuit for a 4-qubit GHZ state on an IBM quantum computer with Qiskit®* [21], [23]–[28]

To demonstrate actual examples of this mapping, we will run a quantum circuit on one of IBM's quantum computers to generate a GHZ state, but first, we analyze the theoretical case

6.2.1 *The 4-qubit GHZ state*

first. Using equation (107) we can analytically derive the density matrix of this GHZ state. With a 4-qubit GHZ state is given by

$$GHZ^{(4\otimes)} = \frac{1}{\sqrt{2}}(|0,0,0,0\rangle + |1,1,1,1\rangle) \tag{222}$$

We get that

$$\hat{\rho}^{(GHZ)} = |\psi_{GHZ}\rangle\langle\psi_{GHZ}| = \frac{1}{2}\{(|0,0,0,0\rangle + |1,1,1,1\rangle) \otimes (\langle 0,0,0,0| + \langle 1,1,1,1|)\}$$

so

$$\hat{\rho}^{(GHZ)} = \frac{1}{2}\{|0,0,0,0\rangle\langle 0,0,0,0| + |0,0,0,0\rangle\langle 1,1,1,1| + |1,1,1,1\rangle\langle 0,0,0,0| + |1,1,1,1\rangle\langle 1,1,1,1|\} \tag{223}$$

Remember that given an arbitrary 4-qubits, we have $2^4 = 16$ unique states in the set

$$\{|0,0,0,0\rangle, |1,0,0,0\rangle, |0,1,0,0\rangle, |0,0,1,0\rangle, |0,0,0,1\rangle, |1,1,0,0\rangle, |1,0,1,0\rangle, |1,0,0,1\rangle, |0,1,0,1\rangle, |0,1,1,0\rangle, |0,0,1,1\rangle, |1,1,1,0\rangle, |0,1,1,1\rangle, |1,1,0,1\rangle, |1,0,1,1\rangle, |1,1,1,1\rangle\}$$

These are a basis vector set for a general 4-qubit system. We could conveniently map them to a standard basis of 16-component state vectors. The standard basis would therefore be

$$\left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} \right\}$$

From equation (222), we see that the only two basis vectors we need to represent a GHZ state are the ones in red and blue, which map to the $|0,0,0,0\rangle$ and $|1,1,1,1\rangle$ states. Using the standard basis, we can build the density matrix of the GHZ state, given equation (223), we get

$$\hat{\rho}^{(GHZ)} = \frac{1}{2}\{|0,0,0,0\rangle\langle 0,0,0,0| + |0,0,0,0\rangle\langle 1,1,1,1| + |1,1,1,1\rangle\langle 0,0,0,0| + |1,1,1,1\rangle\langle 1,1,1,1|\}$$

$$= \frac{1}{2}\left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \right.$$

$$
+\begin{bmatrix}0\\0\\0\\0\\0\\0\\0\\0\\0\\0\\0\\0\\0\\0\\0\\1\end{bmatrix}\begin{bmatrix}1&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0\end{bmatrix}+\begin{bmatrix}0\\0\\0\\0\\0\\0\\0\\0\\0\\0\\0\\0\\0\\0\\0\\1\end{bmatrix}\begin{bmatrix}0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&1\end{bmatrix}\left.\vphantom{\begin{bmatrix}0\\0\\0\\0\\0\\0\\0\\0\\0\\0\\0\\0\\0\\0\\0\\1\end{bmatrix}}\right\}
$$

So, we get the density matrix

$$
\hat{\rho}^{(GHZ)} = \begin{bmatrix}
0.5&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0.5\\
0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0\\
0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0\\
0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0\\
0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0\\
0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0\\
0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0\\
0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0\\
0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0\\
0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0\\
0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0\\
0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0\\
0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0\\
0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0\\
0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0\\
0.5&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0.5
\end{bmatrix} \tag{224}
$$

Since the GHZ state is a superposition of two states we expected off diagonal terms. Now, from inequality (42), we conclude that

$$
tr\left({\hat{\rho}^{(GHZ)}}^2\right) = 1
$$

solidifying that the GHZ state is a pure state.

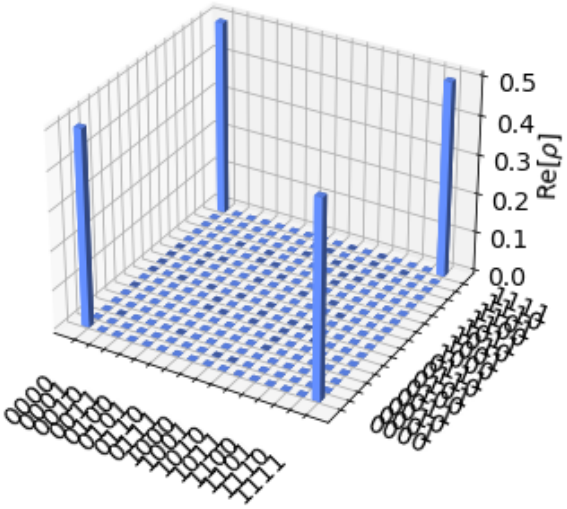


Figure (40): Density matrix (real part) for the GHZ state, the graphical equivalent of the density matrix in equation (224)

6.2.2 *Designing an MPS circuit to realize a GHZ state and the one-to-one mapping of quantum circuits to tensor networks and MPS* [23]–[27], [29]

We've already hinted at there being a direct connection between quantum circuits and tensor networks. Let us try and encode a left orthogonal MPS to a circuit design which will therefore be consisting of only one and two-qubit gates. The simplest mapping will be of the simplest nontrivial MPS's, meaning the maximum bond index $\chi = 2$. The model can also be extended to larger bond dimensions. First of all, generative quantum computing starts with multiple qubits all prepared in arbitrary states, usually all $|0\rangle$'s, we then act on them with unitary gates as specified in the previous section. The goal is to construct MPO's that link an MPS of multiple-qubits to disentangled qubit product states [23]–[27]. If we look back at figure (31), we can construct a similar MPO but one which acts on 4 qubit product states. The unitary gates would act locally through single and two qubit gates. The resulting network would consist of the initial, all $|0\rangle$'s, multi-qubit product states, the MPO unitary that evolves the product state, and the single qubit measurement operators; This is the basic architecture of a quantum circuit. Finally, concerning measurements, the measurement operation measures all the qubits to get a single set of outputs, however since quantum states are probabilistic, the circuit needs to iterate enough number of times to get useful results. Accordingly, we will require several runs to study the distribution of output values and interpret the results[5], [26], [28].

If it is hard to visualize the stunning similarity between tensor networks and quantum circuits, we will create a 4-qubit GHZ state in section 6.2.0 using an MPS and MPO bases circuit. Figures 35 shows the MPO used to model the unitary gates in the circuit, side-by-side with the corresponding tensor network, which, if you look back at figure (31), has the exact same architecture but receives 4 qubits into its physical indices.

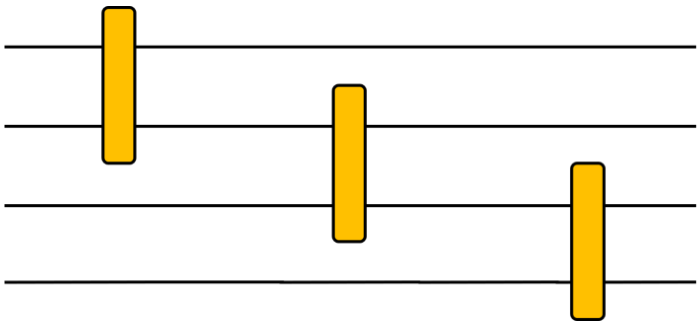


Figure (41): MPO with 8 physical indices (open indices not connected to any tensor), 4 on each side, meaning they act on a 4-qubit product state and output another 4-qubit state

Now that's we've covered the theory underlying the GHZ states, let us design an appropriate quantum circuit using IBM's circuit builder and Qiskit [26]. The circuit design we picked for GHZ was

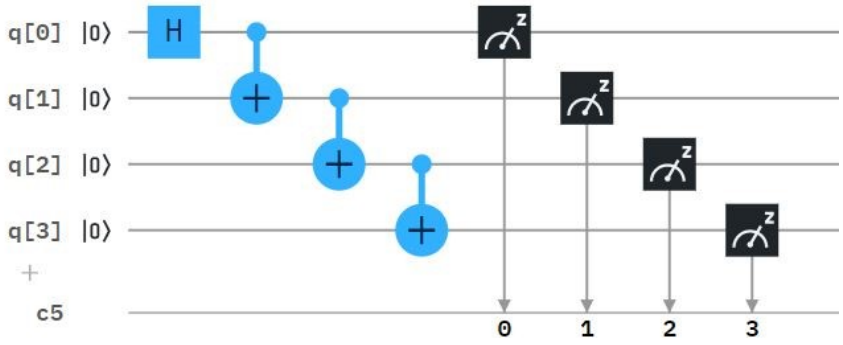


Figure (42): an MPS analogous quantum circuit for preparing a 4-qubit GHZ state

The qubits on the left are prepared as a product state each in the $|0\rangle$ state. It's important to notice that the circuit also involves a time factor, and time runs from left to right along the quantum wires starting at the product state. The first qubit is acted upon by a Hadamard gate, which puts it into a superposition by applying the operator $U_3(\pi/2,0,\pi)$ as dictated by equations (219,220). Following this superposition, the second qubit is entangled to the first through another unitary, the $CNOT$ gate, given the control qubit is in a superposition. The entanglement is then extended between the second and third, and finally, the third and fourth qubit in increments of 2-qubit gates unitary gates. At the end as indicated by the right most operators, the measurement gates measure the state of all the qubits.

If figure (42) looks familiar, well it's because it's a one-to-one map to the 4-qubit MPO we formed in figure (42). For further comparison consider the two where the MPO is made to act similarly on 4-qubit product state

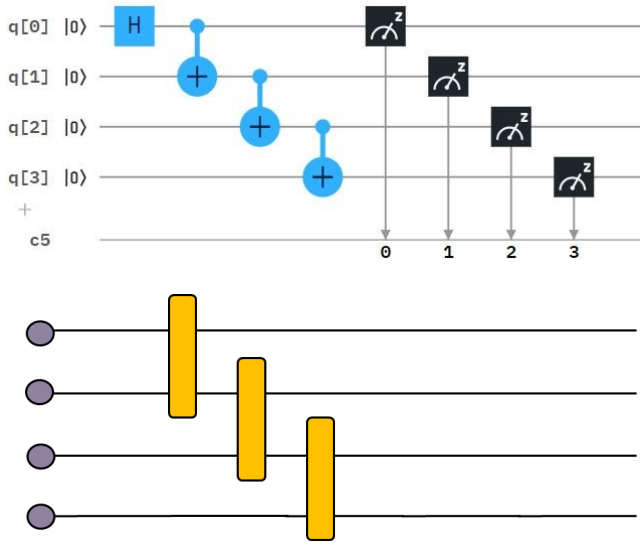


Figure (43): Analogy between quantum circuits and Tensor networks

To break this down even further, let us look at figure (43,44), where we can rearrange the MPO in a familiar way such that it is clear that it is actually acting on 4-qubit product states to contract into an MPS of multiple-qubits. From one perspective, the MPS here can be considered as acting to disentangles the MPS into a qubit product state.

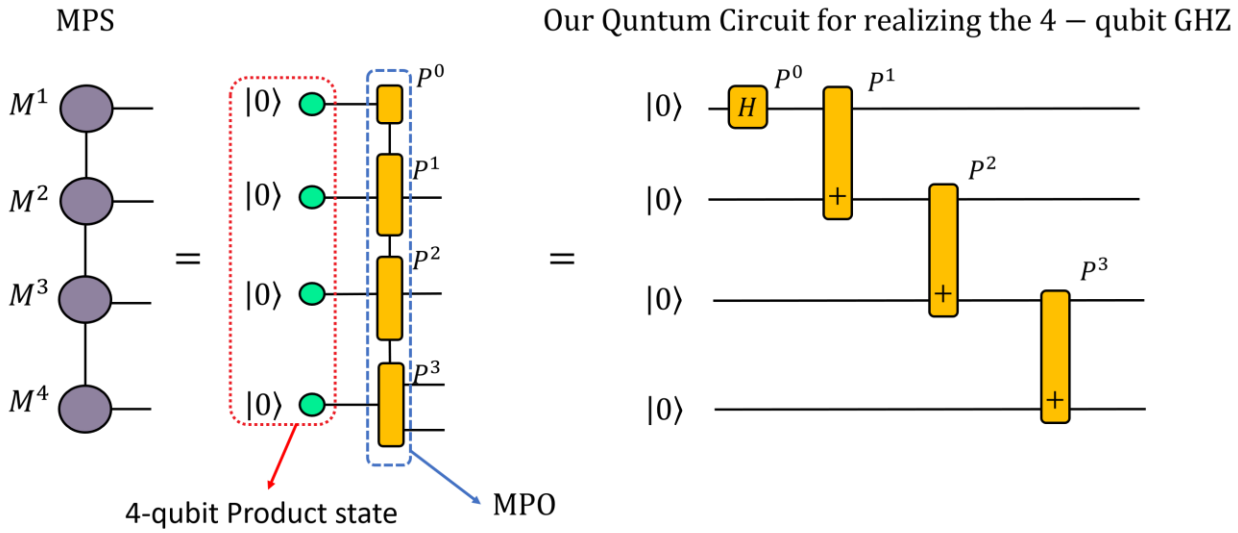


Figure (44): The one-to-one mapping of a quantum circuit to an MPS tensor network

In figure (44), we can see that an MPO acts on a 4-qubit product state, it has four indices to interact with them on its left, and gives rise to four indices on the right which generates a legitimate entangled MPS. This is equivalent to running the 4-qubit quantum circuit in figure (42), which acts on four disentangled qubits in the $|0\rangle$ state, and runs unitary gates on them to generate the GHZ state. Recall that we were able to express as an MPS in section (4.5.3), specifically equation (205).

With this realization, that quantum circuits can in fact operate and act like tensor networks, one can use either to inspire the construction of the other, and not only is it possible, but recommended, that one consider this faithful representation during their design. One rising field where this equivalence was found to be most useful, is quantum machine learning, where tree networks for machine learning could be translated into circuits and run on quantum computers. In fact, by seeking motivation from tensor networks in classical machine learning, researchers have run quantum circuits based on tree and MPS tensor networks, and were able to unify a framework where machine learning models can be trained classically then transferred to a tensor network-based quantum setting for further optimization. This has even been demonstrated on a training model to perform handwriting recognition which was optimized on quantum hardware [29].

Inspired by the prospect of running simulations and training models through quantum machine learning, we ran our own trivial MPS inspired circuit on a quantum computer to assess how well it would reproduce theoretical results, specifically through entanglement entropy.

6.2.3 Obtaining the density matrices from running an MPS quantum circuit for realizing the 4-qubit GHZ state using Quantum State Tomography

Going back to our analytical results, it would be informative to compare how well the quantum computer could reproduce the idealized results. Accordingly, we ran this circuit on the IBM's "***ibmq_london"*** quantum computer and the classical simulator "***ibmq_qasm_simulator***", [26], over exactly 8192 iterations. The output is the Probability distribution among the 16 different 4-qubit states. Figure (45) shows a histogram of the results for the quantum computer.

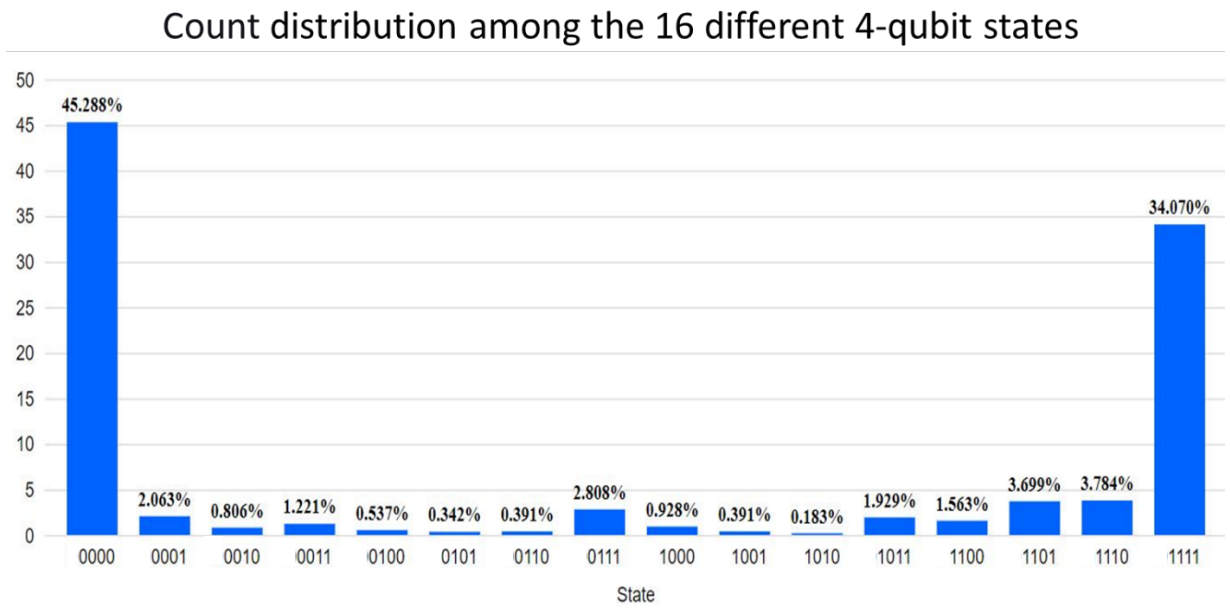


Figure (45): measured probability distribution after running our circuit in figure (42) 8192 times

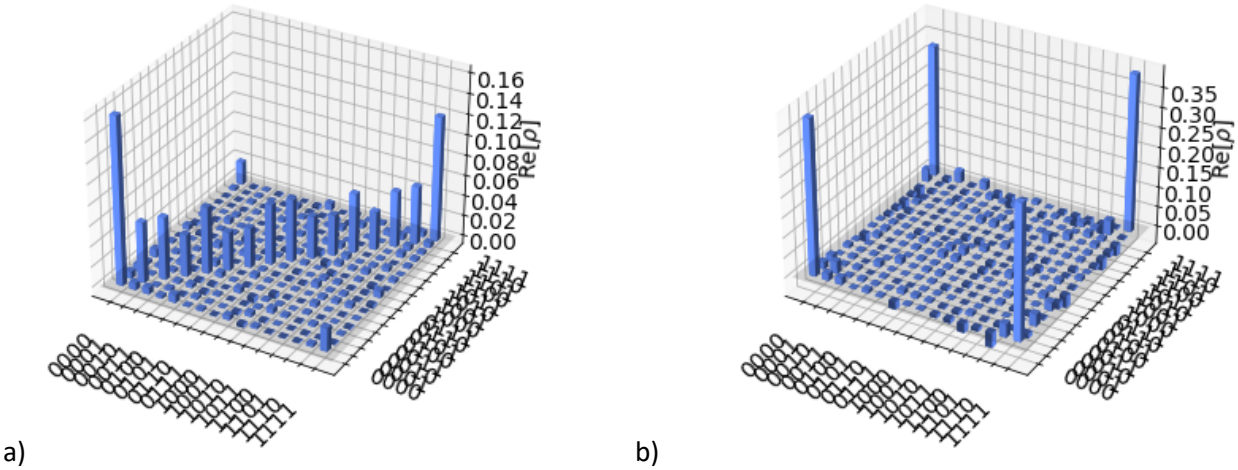
The simulator perfectly matched the GHZ state, yielding a distribution within 0.1% of the exact state given in equation (222). The quantum computer circuit results were quite close to, but not quite the same as, the exactly equal distribution between the $|0,0,0,0\rangle$ and $|1,1,1,1\rangle$ eigenstates we expected to see in an ideal operating quantum computer. Instead, more states besides the $|0,0,0,0\rangle$ and $|1,1,1,1\rangle$ were produced. We could infer a pre-measurement density matrix for such a measurement distribution of classical bits through quantum state tomography, then use our matrix to quantify the decoherence away from the ideal state indirectly. Since applying tomography circuits on a real quantum computer is limited by the number of available circuits in IBM's cloud service, we chose to repeat the analysis using the "ibmq***_qasm_simulator***", however to make it as realistic as possible, we applied a measurement noise model, and ran the circuit, once with, and another time without error correction methods. With error correction, the fidelity was ~ 0.166, while with error correction circuits, the fidelity rose to ~ 0.725. resulting in the following density matrices shown in equation (225), set to the fourth decimal place. This was all done using Qiskit and Mathematica.

$$\hat{\rho}^{GHZ}_{IBM} =$$

$$\begin{bmatrix}
0.3884 & 0.0175+0.0356i & 0.0175+0.0113i & 0.0006-0.0331i & 0.0082+0.0005i & 0.0032+0.0219i & -0.0189+0.0122i & 0.0024+0.017i \\
0.0175-0.0356i & 0.0294 & 0.0034-0.0108i & 0.0054+0.0047i & 0.0031-0.0032i & 0.002+0.0097i & 0.0105+0.0087i & 0.0078+0.0054i \\
0.0175-0.0113i & 0.0034+0.0108i & 0.0065 & -0.0004+0.0025i & 0.0036+0.0012i & -0.0023+0.0008i & -0.001+0.0043i & 0.0026+0.0008i \\
0.0006+0.0331i & 0.0054-0.0047i & -0.0004-0.0025i & 0.0131 & 0.0018-0.0053i & 0.0009+0.0012i & 0.0024-0.0051i & 0.0044-0.0057i \\
0.0082-0.0005i & 0.0031+0.0032i & 0.0036-0.0012i & 0.0018+0.0053i & 0.015 & 0.0018+0.0018i & 0.0062-0.0061i & 0.0082-0.0064i \\
0.0032-0.0219i & 0.002-0.0097i & -0.0023-0.0008i & 0.0009-0.0012i & 0.0018-0.0018i & 0.0054 & 0.0037-0.004i & 0.0048-0.0026i \\
-0.0189-0.0122i & 0.0105-0.0087i & -0.001-0.0043i & 0.0024+0.0051i & 0.0062+0.0061i & 0.0037+0.004i & 0.0143 & 0.0125+0.0006i \\
0.0024-0.017i & 0.0078-0.0054i & 0.0026-0.0008i & 0.0044+0.0057i & 0.0082+0.0064i & 0.0048+0.0026i & 0.0125-0.0006i & 0.0175 \\
0.0021-0.0152i & 0.0048+0.0089i & 0.0025+0.0013i & -0.0075+0.0077i & 0.0003+0.0062i & -0.0017+0.0032i & 0.0022+0.009i & 0.0017+0.0121i \\
-0.0076+0.001i & -0.0039+0.0071i & 0.0009+0.0005i & -0.0014-0.0013i & -0.0076-0.0067i & -0.0042-0.0033i & -0.0109+0.003i & -0.0114+0.0006i \\
-0.0096-0.0076i & -0.0075-0.0037i & -0.0011+0.0023i & -0.0013-0.0013i & -0.0003-0.0007i & 0.0016-0.0038i & -0.0018-0.0039i & 0.-0.0038i \\
-0.0224-0.0276i & 0.0077-0.0035i & 0.0016-0.0014i & -0.0035+0.0064i & 0.0014+0.0144i & 0.0016+0.003i & 0.014+0.0081i & 0.014+0.008i \\
-0.0077+0.0017i & -0.0031-0.0122i & -0.0015-0.0019i & 0.0072-0.0044i & 0.0074-0.0024i & 0.0046-0.0026i & 0.0053-0.0119i & 0.007-0.0148i \\
-0.0369-0.0381i & -0.0069-0.0019i & 0.0008+0.0008i & -0.0058+0.004i & 0.0031-0.0011i & 0.0019-0.0057i & -0.0001-0.0033i & 0.0011-0.0053i \\
0.0327-0.0782i & -0.0031-0.0169i & -0.0021+0.0014i & -0.0064-0.0094i & 0.0022+0.0016i & 0.0112-0.0024i & 0.0043-0.0055i & 0.0091-0.0059i \\
0.3341+0.0926i & -0.0106+0.0742i & 0.0295+0.0224i & -0.0038-0.0164i & 0.0212+0.0097i & -0.0145+0.0132i & -0.0298+0.0111i & -0.0043+0.0133i
\end{bmatrix}$$

$$\begin{bmatrix}
0.0021+0.0152i & -0.0076-0.001i & -0.0096+0.0076i & -0.0224+0.0276i & -0.0077-0.0017i & -0.0369+0.0381i & 0.0327+0.0782i & 0.3341-0.0926i \\
0.0048-0.0089i & -0.0039-0.0071i & -0.0075+0.0037i & 0.0077+0.0035i & -0.0031+0.0122i & -0.0069+0.0019i & -0.0031+0.0169i & -0.0106-0.0742i \\
0.0025-0.0013i & 0.0009-0.0005i & -0.0011-0.0023i & 0.0016+0.0014i & -0.0015+0.0019i & 0.0008-0.0008i & -0.0021-0.0014i & 0.0295-0.0224i \\
-0.0075-0.0077i & -0.0014+0.0013i & -0.0013+0.0013i & -0.0035-0.0064i & 0.0072+0.0044i & -0.0058-0.004i & -0.0064+0.0094i & -0.0038+0.0164i \\
0.0003-0.0062i & -0.0076+0.0067i & -0.0003+0.0007i & 0.0014-0.0144i & 0.0074+0.0024i & 0.0031+0.0011i & 0.0022-0.0016i & 0.0212-0.0097i \\
-0.0017-0.0032i & -0.0042+0.0033i & 0.0016+0.0038i & 0.0016-0.003i & 0.0046+0.0026i & 0.0019+0.0057i & 0.0112+0.0024i & -0.0145-0.0132i \\
0.0022-0.009i & -0.0109-0.003i & -0.0018+0.0039i & 0.014-0.0081i & 0.0053+0.0119i & -0.0001+0.0033i & 0.0043+0.0055i & -0.0298-0.0111i \\
0.0017-0.0121i & -0.0114-0.0006i & 0.+0.0038i & 0.014-0.008i & 0.007+0.0148i & 0.0011+0.0053i & 0.0091+0.0059i & -0.0043-0.0133i \\
0.0121 & -0.0006-0.0061i & -0.0025-0.0008i & 0.0078+0.0069i & -0.0112+0.0052i & 0.0009+0.0005i & -0.0036-0.0003i & 0.0079-0.0186i \\
-0.0006+0.0061i & 0.0117 & 0.0004-0.004i & -0.0099+0.009i & -0.0069-0.0085i & 0.001-0.0063i & -0.0087-0.0076i & 0.0005-0.0034i \\
-0.0025+0.0008i & 0.0004+0.004i & 0.0034 & -0.0013-0.0021i & 0.0033-0.0026i & 0.0046+0.0012i & 0.0061-0.0069i & -0.0074+0.0089i \\
0.0078-0.0069i & -0.0099-0.009i & -0.0013+0.0021i & 0.0226 & -0.0006+0.0154i & 0.0018+0.0045i & 0.0057+0.0017i & -0.0271-0.0137i \\
-0.0112-0.0052i & -0.0069+0.0085i & 0.0033+0.0026i & -0.0006-0.0154i & 0.017 & 0.0026+0.0029i & 0.0105-0.0012i & -0.0122+0.0119i \\
0.0009-0.0005i & 0.001+0.0063i & 0.0046-0.0012i & 0.0018-0.0045i & 0.0026-0.0029i & 0.0124 & 0.0076-0.0192i & -0.0264-0.0138i \\
-0.0036+0.0003i & -0.0087+0.0076i & 0.0061+0.0069i & 0.0057-0.0017i & 0.0105+0.0012i & 0.0076+0.0192i & 0.0368 & -0.0035-0.0428i \\
0.0079+0.0186i & 0.0005+0.0034i & -0.0074-0.0089i & -0.0271+0.0137i & -0.0122-0.0119i & -0.0264+0.0138i & -0.0035+0.0428i & 0.3944
\end{bmatrix}$$

(225)



a) b)

Figure (46): State Tomography generated density matrix (real part) visualization for the a) noise simulated run of the GHZ state circuit, decoherence forced by non-ideal conditions forces other states to emerge along the diagonal whilst lowering the distributions amongst the off-diagonal elements; and b) the same circuit but accompanied by an error correction circuit, one can see the off-diagonal elements regain their values in favor of decreasing diagonal expression.

6.2.4 *Comparing the Von Neumann entropy of the ideally GHZ state to the decoherence and mutual Information of the quantum circuit GHZ with noise model and error correction circuits*

Another important quantity to calculate is the entanglement entropy of the GHZ state. The simulation and exact GHZ state have a density matrix of $\hat{\rho}^{(GHZ)}$, we need to define a bipartition, and therefore a reduced density matrix. Let us say we calculate the entanglement between the first qubit and the rest of the system, designating the first qubit as subsystem A, and the rest as subsystem B. There are several ways to trace out parts of the system and obtain the reduced density matrix, one way is to use readymade packages. We used Qiskit's quantum information package (quantum_info), which has built in functions, that when given a density matrix of an n-qubit system, and told which qubits to trace out, can give the reduced density matrix of the resulting subsystem. Accordingly, we used the package to obtain $\rho^{(/A)}$, by tracing out the second third and fourth qubits (code in appendix 2- a Jupyter notebook excerpt), this gave

$$\rho^{(/A)} = \begin{bmatrix} 0.5 & 0 \\ 0 & 0.5 \end{bmatrix} \tag{226}$$

Using equation (136) we find that the entanglement entropy between the first qubit and the other three

$$S^{von\ Neumann} = -\sum_m \lambda_m \log_2 \lambda_m = \log_2(2) = 1$$

Meaning that the system is maximally entangled as was expected

On the other hand, we could have calculated the entanglement entropy using subsystem B, by tracing out the first qubit from $\hat{\rho}^{(GHZ)}$, instead, so again using the Mathematica package we find it to be

$$\rho^{(/B)} = \begin{bmatrix} 0.5 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.5 \end{bmatrix} \tag{227}$$

Where $\rho^{(/B)}$has different eigenvectors from $\rho^{(/A)}$, but shares the same non zero eigenvalues, exactly as expected for a pure state, and we get

$$S^{von\ Neumann} = -\sum_m \lambda_m \log_2 \lambda_m = \log_2(2) = 1$$

Which again verifies our finding in equation (139) that for a bipartite pure state

$$S^{von\ Neumann} = S\left(\rho^{(/A)}\right) = S\left(\rho^{(/B)}\right)$$

Repeating the same analysis with the quantum computer results, we showed in the previous section how decoherence spreads the distribution of density matrix elements across its diagonals, and so we will inevitably lose information once stored in them from being entangled. We've already touched on the problems in calculating von Neumann entropy for mixed states in section (3.3.5), where instead of the von Neumann entropy we suggested using the ***mutual information*** $I(A;B)$, a more general quantifier of correlations. Here, we will apply the aforementioned method to derive two density matrices and compare loss of entanglement by referencing their mutual information. First, Using Mathematica we found for the error corrected tomography results the reduced density matrices of the subsystems were:

$$\rho^{(/A)} = \begin{bmatrix} 0.5753 & 0.0754 \\ 0.0754 & 0.4246 \end{bmatrix} \tag{228}$$

$$\rho^{(/\mathrm{B})} =$$

$$\begin{bmatrix} 0.4005 & 0.0169+0.0295i & 0.0150+0.0104i & 0.0084-0.0261i & -0.0030+0.0057i & 0.0041+0.0224i & -0.0225+0.0119i & 0.0103-0.0016i \\ 0.0169-0.0295i & 0.0411 & 0.0038-0.0148i & -0.0045+0.0137i & -0.0038-0.0117i & 0.003+0.0034i & 0.0018+0.0011i & 0.0083+0.0020i \\ 0.0150-0.0104i & 0.0038+0.0148i & 0.0098 & -0.0017+0.0004i & 0.0069-0.0014i & 0.0023+0.002i & 0.0051-0.0026i & -0.0048+0.0097i \\ 0.0084+0.0261i & -0.0045-0.0137i & -0.0017-0.0004i & 0.0357 & 0.0012+0.0101i & 0.0027+0.0057i & 0.0081-0.0034i & -0.0227-0.0194i \\ -0.0029-0.0057i & -0.0038+0.0117i & 0.0069+0.0014i & 0.0012-0.0101i & 0.032 & 0.0044+0.0047i & 0.0167-0.0073i & -0.004+0.0055i \\ 0.0041-0.0224i & 0.003-0.0034i & 0.0023-0.002i & 0.0027-0.0057i & 0.0044-0.0047i & 0.0178 & 0.0113-0.0232i & -0.0216-0.0164i \\ -0.0225-0.0119i & 0.0018-0.0010i & 0.0051+0.0026i & 0.0081+0.0034i & 0.0167+0.0073i & 0.0113+0.0232i & 0.0511 & 0.0090-0.0422i \\ 0.0103+0.0015i & 0.0083-0.0020i & -0.0048-0.0097i & -0.0227+0.0194i & -0.004-0.0055i & -0.0216+0.0164i & 0.0090+0.0422i & 0.4119 \end{bmatrix}$$

(229)

Now Using equation (159), we get that

$$I(A;B) = S\left(\rho^{(/A)}\right) + S\left(\rho^{(/B)}\right) - S\left(\rho^{(:AB)}\right)$$

Using Mathematica and Starting with

$$S\left(\rho^{(/A)}\right) = -tr\left(\rho^{(/A)} \log_2 \rho^{(/A)}\right)$$

Whilst ignoring the eigenvalues of the order 0.001, We get

$$S\left(\rho^{(/A)}\right) = -\sum_{m_A} \lambda_{m_A} \log_2 \lambda_{m_A} \sim 1$$

$$S\left(\rho^{(/B)}\right) = -\sum_{m_B} \lambda_{m_B} \log_2 \lambda_{m_B} \sim 1.7889$$

$$S\left(\rho^{(/AB)}\right) = -\sum_{m_{AB}} \lambda_{m_{AB}} \log_2 \lambda_{m_{AB}} \sim 1$$

so

$$I(A;B)_{IBM} = 1 + 1.7889 - 1 = 1.7889 \tag{230}$$

If we want to compare to the ideal case to better grasp this result, we could get the mutual information for the ideal value which would have been of a pure maximally entangled GHZ, and we get

$$I(A;B)_{Ideal} = 1 + 1 - 0 = 2 \tag{231}$$

This result is not surprising, since we've already discussed in section (1.2.5), how noise, for example that associated with the non-ideal quantum computer environment, begets decoherences, and in turn loss of entanglement. With our experimental runs we could infer an average of $\sim 10.6\%$ with error correction.

Conclusions and outlook

Throughout our work, we thoroughly went over the mathematical basics behind the techniques involved in extracting entanglement entropy from a given quantum state. The importance of density matrices in painting the picture of quantum states was investigated with several examples of its utility, including its relation to expectation values, purity; mixedness and representing correlations. Afterwards we applied the idea of density matrices to bipartite systems, this was achievable through extending its utility to reduced density matrixes, partial traces and using those properties to get expectation values of operators on bipartite systems, which we extended to multipartite systems with the introduction of combined indices. These basics allowed us to attempt to quantize entanglement in pure bipartite systems, where the von Neumann entropy was the first candidate, however we did need concepts like mutual information to extend this quantization to the mixed state, a tool we needed later on when we dealt with the decoherence in the resulting states from the quantum computer circuit. This was where the informational value of entanglement was discussed and analogy to Shannon entropy and the informational value of entangled qubits was illustrated. To start applying the concepts covered we had a look at tensor networks, as they offered a way to tackle area laws of entanglement entropy, the physical laws forcing nature to lurk in a small corner of Hilbert space. Tapping into that corner Hilbert space was one way of dealing with the exponential growth of the state vector in multi-body quantum systems. With tensor diagrams and tensor networks were able to; explain how tensor networks deal with local gapped Hamiltonians and the relation to area laws of entanglement entropy; Break down wavefunctions into tensor networks; Explore Matrix product states (MPS's) in detail, and going over the incremental SVD algorithm behind right, left and mixed canonical MPS's; give basic examples of product states, a singlet and a GHZ state MPS and build Matrix product operators (MPO's) for the Ising and Heisenberg Hamiltonians; and most importantly explain how tensor networks and specifically MPS's and PEPS compress the cost of running numerical operations on state vectors once they are rewritten as tensor networks. This brought us to the final phase of the study, which was applying the concepts we've learnt along the way to extract entanglement entropies and run practical operations, this allowed us to achieve the following:

Using the ITensor C++ Library

- Build a singlet state and calculating its energy under the Heisenberg Hamiltonian
- Constructing a 20-body MPO for the Heisenberg Hamiltonian and using DMRG and SVD to extract von Neumann entanglement entropy from the output MPS

Using IBM's Q-experience and quantum computers:

- Designed an MPS and MPO inspired circuit to realize a GHZ state where we emphasize the one-to-one mapping of quantum circuits to tensor networks and MPS
- Ran the circuit for a quantum computer generated 4-qubit GHZ state using an IBM 5-qubit quantum computer
- Ran state tomography on the simulated circuit with noise and error correction to derive pre-measurement density matrices for entanglement entropy comparisons.
- Compared the Von Neumann entropy of the Ideal GHZ state vs. the decoherence and mutual Information from noisy quantum circuit generated GHZ states

After laying down such rigorous mathematical basics, we were able to directly compare theory with practical results, and our analysis indeed turned out as expected and might be a very instructive entry point for one who is new to the fields of many-body quantum physics, particularly if they are interested in tensor network theory in many-body physics and how it relates to quantum information and computing.

Appendix 1: Mathematica Codes

Dynamic Bloch Sphere code:

```
Manipulate[
 {Graphics3D[{Arrowheads[0.015], {Opacity[0.17], Sphere[{0, 0, 0}, 1.13]},
    {EdgeForm[{Red}], FaceForm[None], Cylinder[{{0, 0, -.001}, {0, 0, .001}}, 1.13]}, {Line[{{0, 0, 0}, {1, 0, 0}}]},
    {Line[{{0, 0, 0}, {0, 1, 0}}]}, {Line[{{0, 0, 0}, {0, 0, 1}}]},

    {Dashed, Thickness → 0.004, Purple,
     Line[{{1.13 Cos[θ] Sin[φ], 1.13 Sin[θ] Sin[φ], 1.13 Cos[φ]}, {1.13 Cos[θ] Sin[φ], 1.13 Sin[θ] Sin[φ], 0}, {0, 0, 0}}]},



    Arrow[{{1, 0, 0}, {1.3, 0, 0}}], Arrow[{{0, 0, 0}, {-1.3, 0, 0}}],
    Arrow[{{0, 1, 0}, {0, 1.3, 0}}], Arrow[{{0, 0, 0}, {0, -1.3, 0}}],
    Arrow[{{0, 0, 1}, {0, 0, 1.3}}], Arrow[{{0, 0, 1}, {0, 0, -1.3}}],
    Text[Style["|x+⟩", 18], {1.5, 0, 0}], Text[Style["|x-⟩", 18], {-1.5, 0, 0}],
    Text[Style["|y+⟩", 18], {0, 1.5, 0}], Text[Style["|y-⟩", 18], {0, -1.5, 0}],
    Text[Style["|z+⟩", 18], {0, 0, 1.5}], Text[Style["|z-⟩", 18], {0, 0, -1.5}],
    Text[Style["n̂", 18, Red], .5 {Cos[θ] Sin[φ], Sin[θ] Sin[φ], Cos[φ]}],
    {Arrowheads[0.03], Red, Arrow[{{0, 0, 0}, {1.13 Cos[θ] Sin[φ], 1.13 Sin[θ] Sin[φ], 1.13 Cos[φ]}}]}]},

   Boxed → False, ImageSize → 500],
  Text[Style["ψ =  " <> ToString[NumberForm[Cos[φ/2], {4, 4}]] <> " |0⟩ +" , 18]
    Text[Style[NumberForm[Exp[I (θ)] "|1⟩ " Sin[φ/2], {4, 4}] , 18]]]
 },
 {θ, 0, 2 π}, {{φ, .3}, 0, π}

]
```

Appendix 2: ITensor C++ Code using the ITensor Library [14]

Energy of a Singlet acted on by the Heisenberg Hamiltonian

```
#include "itensor/all.h"

using namespace itensor;

////////////////////// ~ PART 1~ ////////////////////////////////////
///////////// The singlet state and the Heisenberg model ///////////////

ITensor
// making the operators for the Hysenberg Model, the S1_z,S2_z, and the S+ and S-
makeSp(Index const& s)
    {
    auto Sp = ITensor(s,prime(s));
    Sp.set(s=2,prime(s)=1, 1);
    return Sp;
    }

ITensor
makeSm(Index const& s)
    {
    auto Sm = ITensor(s,prime(s));
    Sm.set(s=1,prime(s)=2,1);
    return Sm;
    }

ITensor
makeSz(Index const& s)
    {
    auto Sz = ITensor(s,prime(s));
    Sz.set(s=1,prime(s)=1,+0.5);
    Sz.set(s=2,prime(s)=2,-0.5);
    return Sz;
    }


int main()
    {

    // Initial product state of the most general two site wavefunction


    // Building the two physical indices sticking out of the tensor

    //defining the two physical indices individually as s1 and s2

    auto s1 = Index(2,"s1");
    auto s2 = Index(2,"s2");


    //combining the two physical indices into one big tensor
```

```
    auto psi = ITensor(s1,s2);

    //making a singlet state, one of the most common two site states

    psi.set(s1(1),s2(2),+1./sqrt(2));
    psi.set(s1(2),s2(1),-1./sqrt(2));


    PrintData(psi);


    /*defining the Single-site local operators using the operators we ..
    ..made before the main function for the Heisenberg model*/

    auto Sz1 = makeSz(s1);
    auto Sz2 = makeSz(s2);
    auto Sp1 = makeSp(s1);
    auto Sp2 = makeSp(s2);
    auto Sm1 = makeSm(s1);
    auto Sm2 = makeSm(s2);


    /* Using a sum of local spin operators to...
    .. construct a Two-site Heisenberg Hamiltonian */

    auto H = Sz1*Sz2 + 0.5*Sp1*Sm2 + 0.5*Sm1*Sp2;

    // Initial energy expectation value

    auto initEn = elt(dag(prime(psi)) * H * psi);
    printfln("\nInitial energy = %.10f",initEn);


    return 0;
    }
```

DMRG and extracting entanglement entropy of an MPS from a 20-unit spin chain after applying a Heisenberg MPO

```
#include "itensor/all.h"

using namespace itensor;

int main()
    {

    ////////////////////////// ~ PART 2~ ////////////////////////////////////
    ////////////////////// SVD calculation and von Neumann//////////////////////
```

```
// we construct a 20-body spin half chain, so 20=4 and specify that it's spin half
auto N = 20;
auto sites = SpinHalf(N);

//Here we initiate a Matrix Product Operator MPO

auto ampo = AutoMPO(sites);

/* Next we construct a sum of local operators for our local Hamiltoninan...
 ....in this case the Heisenberg Hamiltonian*/

for( auto j : range1(N-1) )
    {
    ampo += 0.5,"S+",j,"S-",j+1;
    ampo += 0.5,"S-",j,"S+",j+1;
    ampo +=     "Sz",j,"Sz",j+1;
    }

// we turn the Heisenberg Hamiltonian into an MPO

auto H = toMPO(ampo);

// we set all the initial states to UP, and the odd sites rewritten to down

auto state = InitState(sites,"Up");
for( auto n : range1(N) ) if( n%2==0 ) state.set(n,"Dn");


//we run DMRG to reach an approximate ground state `psi`
//number of sweeps is 5,the maximum bond dimension of every sweep is also set
//The truncation error cutoff is also set at a modest value here it's near exact

auto sweeps = Sweeps(5);
sweeps.maxdim() = 10,9,8,7,6;
sweeps.cutoff() = 1E-10;

auto [energy,psi] = dmrg(H,randomMPS(state),sweeps,"Silent");

//Given an MPS called "psi" and can identify any of the (N-1) bonds "b"
//The von neumann entropy is measured across this bond b
auto b = 4;

//"Gauge" the MPS to site b, so we select the bond site.

psi.position(b);

//SVD this wavefunction around the bond (b) to get the singular value spectrum

auto l = leftLinkIndex(psi,b);
auto s = siteIndex(psi,b);
auto [U,S,V] = svd(psi(b),{l,s});
auto u = commonIndex(U,S);


//Apply von Neumann formula using the singular values
```

```
Real SvN = 0.;
for(auto n : range1(dim(u)))
    {
    auto Sn = elt(S,n,n);
    auto p = sqr(Sn);
    /* we print the diagonal matrix entries to look at the singular values before we use
them to
    extract the entropy*/
    Print(p);
    if(p > 1E-12) SvN += -p*log(p);
    }

printfln("Across bond b=%d, SvN = %.10f",b,SvN);



    return 0;
    }
```

Jupyter notebook with Qiskit code for deriving the density matrices and reduced density matrices of the experimental 4-Qubit GHZ state

```
from qiskit import *
import numpy as np
from qiskit.quantum_info import DensityMatrix
from qiskit.aqua.algorithms import NumPyEigensolver
from qiskit.quantum_info import Statevector
from sympy import Matrix
from qiskit.aqua.utils import get_subsystem_density_matrix
```

Density matrix of an IBM quantum computer generated GHZ state:

we will use the distribution of states given by running the 4-qubit GHZ circuit to create a statevector which we use to create density matrices and run partial traces to describe subsystem reduced density matrices.

```
from qiskit.ignis.verification.tomography import state_tomography_circuits, StateTomographyFitter
import qiskit.ignis.mitigation.measurement as mc
import qiskit.quantum_info as qi
import time

from qiskit.ignis.verification.tomography import process_tomography_circuits, ProcessTomographyFitter
from qiskit.ignis.verification.tomography import gateset_tomography_circuits, GatesetTomographyFitter
import qiskit.ignis.mitigation.measurement as mc
```

```
qr = QuantumRegister(4)
GHZ_circ = QuantumCircuit(qr)
GHZ_circ.h(qr[0])
GHZ_circ.cx(qr[0], qr[1])
GHZ_circ.cx(qr[1], qr[2])
GHZ_circ.cx(qr[2], qr[3])

print(GHZ_circ)

target_state = qi.Statevector.from_instruction(GHZ_circ)
print(target_state)
```

```
q48_0: ┤ H ├──■─────────────
q48_1: ─────┤ X ├──■────────
q48_2: ──────────┤ X ├──■───
q48_3: ─────────────────┤ X ├
```

```
Statevector([0.70710678+0.j, 0.        +0.j, 0.        +0.j,
             0.        +0.j, 0.        +0.j, 0.        +0.j,
             0.        +0.j, 0.        +0.j, 0.        +0.j,
             0.        +0.j, 0.        +0.j, 0.        +0.j,
             0.        +0.j, 0.        +0.j, 0.        +0.j,
             0.70710678+0.j],
            dims=(2, 2, 2, 2))
```

```
from qiskit import IBMQ
IBMQ.load_account()
provider = IBMQ.get_provider('ibm-q')
```

```
E:\ANACONDA\envs\Qiskit\lib\site-packages\qiskit\providers\ibmq\ibmqfactory.py:192: UserWarning: Timestamps in IBMQ backend properties, jobs, and job results are all now in local time instead of UTC.
  warnings.warn('Timestamps in IBMQ backend properties, jobs, and job results '
ibmqfactory.load_account:WARNING:2020-12-07 04:25:03,464: Credentials are already in use. The existing account in the session will be replaced.
```

```
from qiskit.providers.aer import noise
from qiskit.compiler import assemble

noise_model = noise.NoiseModel()
for qubit in range(4):
    read_err = noise.errors.readout_error.ReadoutError([[0.75, 0.25],[0.3,0.7]])
    noise_model.add_readout_error(read_err,[qubit])
```

```
qcomp = provider.get_backend('ibmq_qasm_simulator')
t = time.time()

meas_calibs, state_labels = mc.complete_meas_cal(qubit_list=[0,1,2,3])
job_cal = qiskit.execute(meas_calibs, backend=qcomp, shots=4000, noise_model=noise_model)

# Generate the state tomography circuits.
qst_GHZ = state_tomography_circuits(GHZ_circ, [qr[0], qr[1],qr[2],qr[3]])

# Execute
job = qiskit.execute(qst_GHZ, backend= qcomp, shots=4000,noise_model=noise_model)

print('Time taken:', time.time() - t)
```

```
Time taken: 4.456462621688843
```

```
from qiskit.tools.monitor import job_monitor
job_monitor(job)
```

```
Job Status: job has successfully run
```

```
# Fit result
meas_fitter = mc.CompleteMeasFitter(job_cal.result(),state_labels)
tomo_fitter_GHZ = StateTomographyFitter(job.result(), qst_GHZ)
```

```
# Perform the tomography fit
# which outputs a density matrix
rho_fit_GHZ = tomo_fitter_GHZ.fit(method='lstsq')
F_GHZ = qi.state_fidelity(rho_fit_GHZ, target_state)
print('State Fidelity: F = {:.5f}'.format(F_GHZ))
```

```
State Fidelity: F = 0.16564
```

```
correct_tomo_results = meas_fitter.filter.apply(job.result(), method='least_squares')
tomo_fitter_GHZ_mit = StateTomographyFitter(correct_tomo_results, qst_GHZ)
rho_fit_GHZ_mit = tomo_fitter_GHZ_mit.fit(method='lstsq')
F_GHZ_mit = qi.state_fidelity(rho_fit_GHZ_mit, target_state)
print('State fidelity (w/ correction): F = {:.5f}'.format(F_GHZ_mit))
```

```
State fidelity (w/ correction): F = 0.72549
```

```
Matrix(rho_fit_GHZ);
```

```
Matrix(rho_fit_GHZ_mit);
```

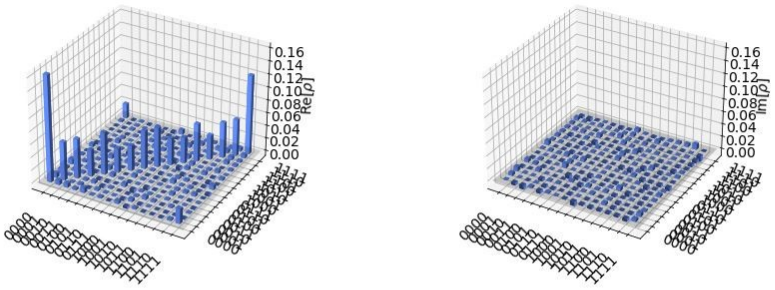
```
Corrected_Tomography_state_2 = (np.round(rho_fit_GHZ_mit, 4))

Matrix(Corrected_Tomography_state_2);
```

```
from qiskit.quantum_info import DensityMatrix
My_corrected_GHZ_density_matrix = DensityMatrix(rho_fit_GHZ_mit)
My_corrected_GHZ_density_matrix.purity()
```

```
(0.6247226250717085+1.1669846660010034e-19j)
```

```
from qiskit.visualization import plot_state_city
plot_state_city(rho_fit_GHZ)
```




```
My_noisy_GHZ_density_matrix = DensityMatrix(rho_fit_GHZ)
My_noisy_GHZ_density_matrix.purity()
```

```
(0.08383273278756478-1.012451200326913e-22j)
```

```
Tomography_state = (np.round(rho_fit_GHZ, 4))

Matrix(Tomography_state);
```

```
plot_state_city(rho_fit_GHZ_mit)
```

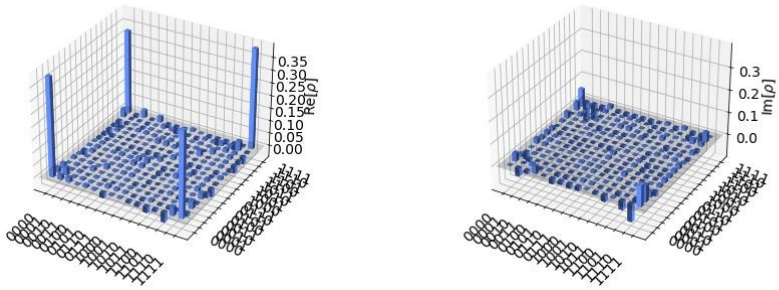
```
My_noisy_GHZ_density_matrix = DensityMatrix(rho_fit_GHZ)
My_noisy_GHZ_density_matrix.purity()
```

```
(0.08383273278756478-1.012451200326913e-22j)
```

```
Tomography_state = (np.round(rho_fit_GHZ, 4))

Matrix(Tomography_state);
```

```
plot_state_city(rho_fit_GHZ_mit)
```



References:

[1] B. Zeng, X. Chen, D.-L. Zhou, and X.-G. Wen, *Quantum Information Meets Quantum Matter -- From Quantum Entanglement to Topological Phase in Many-Body Systems*. Springer Publishing Company, 2018.

[2] J. Preskill, *Preskill Notes - Chapter 10: Quantum Shannon Theory*. Institute for Quantum Information and Matter, Caltech, 2018.

[3] M. D. Barrett *et al.*, "Deterministic quantum teleportation of atomic qubits," *Nature*, vol. 429, no. 6993, pp. 737–739, 2004.

[4] N. Gisin, G. Ribordy, W. Tittel, and H. Zbinden, "Quantum cryptography," *Rev. Mod. Phys.*, vol. 74, no. 1, pp. 145–195, Mar. 2002.

[5] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information*, 10th Anniv. Cambridge University Press, 2011.

[6] J. I. Cirac, "Entanglement in many-body quantum systems," in *Many-Body Physics with Ultracold Gases*, vol. 94, Oxford University Press, 2012, pp. 161–188.

[7] J. Hauschild and F. Pollmann, "Efficient numerical simulations with Tensor Networks: Tensor Network Python (TeNPy)," *SciPost Phys. Lect. Notes*, pp. 1–38, 2018.

[8] P. A. M. Dirac, "Quantum mechanics of many-electron systems," *Proc. R. Soc. London. Ser. A, Contain. Pap. a Math. Phys. Character*, vol. 123, no. 792, pp. 714–733, Apr. 1929.

[9] R. Orús, "A practical introduction to tensor networks: Matrix product states and projected entangled pair states," *Ann. Phys. (N. Y).*, vol. 349, pp. 117–158, 2014.

[10] G. T. Landi, "Quantum Information and Quantum Noise," University of Sao Paulo, 2018.

[11] J. Maziero, "Computing partial traces and reduced density matrices," *Int. J. Mod. Phys. C*, vol. 28, no. 1, pp. 1–12, 2017.

[12] U. Schollwöck, "The density-matrix renormalization group in the age of matrix product states," *Ann. Phys. (N. Y).*, vol. 326, no. 1, pp. 96–192, 2011.

[13] S. R. White, "Density-matrix algorithms for quantum renormalization groups," *Phys. Rev. B*, vol. 48, no. 14, pp. 10345–10356, 1993.

[14] S. R. White, E. M. Stoudenmire, and M. Fishman, "I-Tensor C++ V3 Library." [Online]. Available: https://itensor.org.

[15] glS, "How to draw an interactive, mouse clickable, 3D Bloch sphere?," 2016. [Online]. Available: https://mathematica.stackexchange.com/a/125986/54959.

[16] A. W. Harrow, "Entanglement, density matrices and decoherence- 8.06 Quantum Physics III." Massachusetts Institute of Technology: MIT OpenCourseWare, 2016.

[17] J. C. Bridgeman and C. T. Chubb, "Hand-waving and interpretive dance: An introductory course on tensor networks," *J. Phys. A Math. Theor.*, vol. 50, no. 22, 2017.

[18] J. Eisert, M. Cramer, and M. B. Plenio, “Colloquium: Area laws for the entanglement entropy,” *Rev. Mod. Phys.*, vol. 82, no. 1, pp. 277–306, 2010.

[19] J. Eisert, *Entanglement and tensor network states*. Berlin: Dahlem Center for Complex Quantum Systems Freie Universitat Berlin, 2013.

[20] G. Evenbly, “TensorTrace: an application to contract tensor networks,” *arXiv Prepr. arXiv1911.02558*, Nov. 2019.

[21] R. Gawatz, “Matrix Product State Based Algorithms for Ground States and Dynamics,” University of Copenhagen, 2017.

[22] “The Tensor Network open-source review article,” 2018. [Online]. Available: https://tensornetwork.org.

[23] S. J. Ran, “Encoding of matrix product states into quantum circuits of one- A nd two-qubit gates,” *Phys. Rev. A*, vol. 101, no. 3, pp. 1–7, 2020.

[24] K. Woolfe, “Matrix Product Operator Simulations of Quantum Algorithms,” The University of Melbourne, 2015.

[25] A. McCaskey, E. Dumitrescu, M. Chen, D. Lyakh, and T. Humble, “Validating quantum-classical programming models with tensor network simulations,” *PLoS One*, vol. 13, no. 12, pp. 1–19, 2018.

[26] IBM Q Experience, “A field guide to quantum computing.” [Online]. Available: https://quantum-computing.ibm.com/docs/guide/.

[27] C. Schön, E. Solano, F. Verstraete, J. I. Cirac, and M. M. Wolf, “Sequential generation of entangled multiqubit states,” *Phys. Rev. Lett.*, vol. 95, no. 11, pp. 1–4, 2005.

[28] A. J. *et al.*, “Quantum Algorithm Implementations for Beginners,” no. April, 2018.

[29] W. Huggins, P. Patil, B. Mitchell, K. Birgitta Whaley, and E. Miles Stoudenmire, “Towards quantum machine learning with tensor networks,” *Quantum Sci. Technol.*, vol. 4, no. 2, pp. 1–12, 2019.