

# Chapter 1

## Concentration of Measure And Quantum Entanglement

As the future version of me might forgot everything we have over the summer, as I did for now, I will make a review again from the simple definition to recall the necessary information to tell you why we are here and how we are going to proceed.

First, we will build the mathematical model describing the behavior of quantum system and why they makes sense for physicists and meaningful for general publics.

### 1.1 Motivation

First, we introduce a motivation for introducing non-commutative probability theory to the study of quantum mechanics. This section is mainly based on the book [**kummer1998elements**].

#### 1.1.1 Light polarization and the violation of Bell's inequality

The light which comes through a polarizer is polarized in a certain direction. If we fix the first filter and rotate the second filter, we will observe the intensity of the light will change.

The light intensity decreases with  $\alpha$  (the angle between the two filters). The light should vanish when  $\alpha = \pi/2$ .

However, for a system of 3 polarizing filters  $F_1, F_2, F_3$ , having directions  $\alpha_1, \alpha_2, \alpha_3$ , if we put them on the optical bench in pairs, then we will have three random variables  $P_1, P_2, P_3$ .

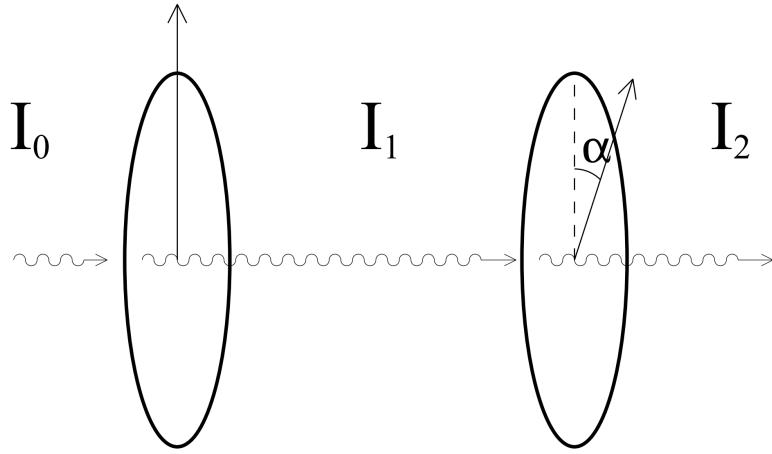


FIG. 1

Figure 1.1: The light polarization experiment, image from [\[kummer1998elements\]](#)

**Theorem 1.** *Bell's 3 variable inequality:*

For any three random variables  $P_1, P_2, P_3$  in a classical probability space, we have

$$\text{Prob}(P_1 = 1, P_3 = 0) \leq \text{Prob}(P_1 = 1, P_2 = 0) + \text{Prob}(P_2 = 1, P_3 = 0)$$

*Proof.* By the law of total probability there are only two possibility if we don't observe any light passing the filter pair  $F_i, F_j$ , it means the photon is either blocked by  $F_i$  or  $F_j$ , it means

$$\begin{aligned} \text{Prob}(P_1 = 1, P_3 = 0) &= \text{Prob}(P_1 = 1, P_2 = 0, P_3 = 0) \\ &\quad + \text{Prob}(P_1 = 1, P_2 = 1, P_3 = 0) \\ &\leq \text{Prob}(P_1 = 1, P_2 = 0) + \text{Prob}(P_2 = 1, P_3 = 0) \end{aligned}$$

□

However, according to our experimental measurement, for any pair of polarizers  $F_i, F_j$ , by the complement rule, we have

$$\begin{aligned} \text{Prob}(P_i = 1, P_j = 0) &= \text{Prob}(P_i = 1) - \text{Prob}(P_i = 1, P_j = 1) \\ &= \frac{1}{2} - \frac{1}{2} \cos^2(\alpha_i - \alpha_j) \\ &= \frac{1}{2} \sin^2(\alpha_i - \alpha_j) \end{aligned}$$

This leads to a contradiction if we apply the inequality to the experimental data.

$$\frac{1}{2} \sin^2(\alpha_1 - \alpha_3) \leq \frac{1}{2} \sin^2(\alpha_1 - \alpha_2) + \frac{1}{2} \sin^2(\alpha_2 - \alpha_3)$$

If  $\alpha_1 = 0, \alpha_2 = \frac{\pi}{6}, \alpha_3 = \frac{\pi}{3}$ , then

$$\begin{aligned} \frac{1}{2} \sin^2(-\frac{\pi}{3}) &\leq \frac{1}{2} \sin^2(-\frac{\pi}{6}) + \frac{1}{2} \sin^2(\frac{\pi}{6} - \frac{\pi}{3}) \\ \frac{3}{8} &\leq \frac{1}{8} + \frac{1}{8} \\ \frac{3}{8} &\leq \frac{1}{4} \end{aligned}$$

Other revised experiments (e.g., Aspect's experiment, calcium entangled photon experiment) are also conducted and the inequality is still violated.

### 1.1.2 The true model of light polarization

The full description of the light polarization is given below:

State of polarization of a photon:  $\psi = \alpha|0\rangle + \beta|1\rangle$ , where  $|0\rangle$  and  $|1\rangle$  are the two orthogonal polarization states in  $\mathbb{C}^2$ .

Polarization filter (generalized 0,1 valued random variable): orthogonal projection  $P_\alpha$  on  $\mathbb{C}^2$  corresponding to the direction  $\alpha$  (operator satisfies  $P_\alpha^* = P_\alpha = P_\alpha^2$ ).

The matrix representation of  $P_\alpha$  is given by

$$P_\alpha = \begin{pmatrix} \cos^2(\alpha) & \cos(\alpha) \sin(\alpha) \\ \cos(\alpha) \sin(\alpha) & \sin^2(\alpha) \end{pmatrix}$$

Probability of a photon passing through the filter  $P_\alpha$  is given by  $\langle P_\alpha \psi, \psi \rangle$ ; this is  $\cos^2(\alpha)$  if we set  $\psi = |0\rangle$ .

Since the probability of a photon passing through the three filters is not commutative, it is impossible to discuss  $\text{Prob}(P_1 = 1, P_3 = 0)$  in the classical setting.

The main vector space we are interested in is  $\mathbb{C}^n$ ; therefore, all the linear operators we defined are from  $\mathbb{C}^n$  to  $\mathbb{C}^n$ .

We denote a vector in vector space as  $|\psi\rangle = (z_1, \dots, z_n)$  (might also be infinite dimensional, and  $z_i \in \mathbb{C}$ ).

A natural inner product space defined on  $\mathbb{C}^n$  is given by the Hermitian inner product:

$$\langle \psi | \varphi \rangle = \sum_{i=1}^n z_i z_i^*$$

This satisfies the following properties:

1.  $\langle \psi | \sum_i \lambda_i | \varphi \rangle = \sum_i \lambda_i \langle \psi | \varphi \rangle$  (linear on the second argument. Note that in physics [Nielsen·Chuang·2010] we use linear on the second argument and conjugate linear on the first argument. But in math, we use linear on the first argument and conjugate linear on the second argument [Axler·2024]. As promised in the beginning, we will use the physics convention in this report.)
2.  $\langle \varphi | \psi \rangle = (\langle \psi | \varphi \rangle)^*$
3.  $\langle \psi | \psi \rangle \geq 0$  with equality if and only if  $|\psi\rangle = 0$

Here  $\psi$  is just a label for the vector, and you don't need to worry about it too much. This is also called the ket, where the counterpart:

- $\langle \psi \rangle$  is called the bra, used to denote the vector dual to  $\psi$ ; such an element is a linear functional if you really want to know what that is.
- $\langle \psi | \varphi \rangle$  is the inner product between two vectors, and  $\langle \psi | A | \varphi \rangle$  is the inner product between  $A | \varphi \rangle$  and  $\langle \psi |$ , or equivalently  $A^\dagger \langle \psi |$  and  $|\varphi\rangle$ .
- Given a complex matrix  $A = \mathbb{C}^{n \times n}$ ,

1.  $A^*$  is the complex conjugate of  $A$ . i.e.,

$$A = \begin{bmatrix} 1+i & 2+i & 3+i \\ 4+i & 5+i & 6+i \\ 7+i & 8+i & 9+i \end{bmatrix}, A^* = \begin{bmatrix} 1-i & 2-i & 3-i \\ 4-i & 5-i & 6-i \\ 7-i & 8-i & 9-i \end{bmatrix}$$

2.  $A^\top$  denotes the transpose of  $A$ . i.e.,

$$A = \begin{bmatrix} 1+i & 2+i & 3+i \\ 4+i & 5+i & 6+i \\ 7+i & 8+i & 9+i \end{bmatrix}, A^\top = \begin{bmatrix} 1+i & 4+i & 7+i \\ 2+i & 5+i & 8+i \\ 3+i & 6+i & 9+i \end{bmatrix}$$

3.  $A^\dagger = (A^*)^\top$  denotes the complex conjugate transpose, referred to as the adjoint, or Hermitian conjugate of  $A$ . i.e.,

$$A = \begin{bmatrix} 1+i & 2+i & 3+i \\ 4+i & 5+i & 6+i \\ 7+i & 8+i & 9+i \end{bmatrix}, A^\dagger = \begin{bmatrix} 1-i & 4-i & 7-i \\ 2-i & 5-i & 8-i \\ 3-i & 6-i & 9-i \end{bmatrix}$$

4.  $A$  is unitary if  $A^\dagger A = A A^\dagger = I$ .

5.  $A$  is hermitian (self-adjoint in mathematics literature) if  $A^\dagger = A$ .

## Motivation of Tensor product

Recall from the traditional notation of product space of two vector spaces  $V$  and  $W$ , that is,  $V \times W$ , is the set of all ordered pairs  $(|v\rangle, |w\rangle)$  where  $|v\rangle \in V$  and  $|w\rangle \in W$ .

The space has dimension  $\dim V + \dim W$ .

We want to define a vector space with the notation of multiplication of two vectors from different vector spaces.

That is

$$(|v_1\rangle + |v_2\rangle) \otimes |w\rangle = (|v_1\rangle \otimes |w\rangle) + (|v_2\rangle \otimes |w\rangle)$$

$$|v\rangle \otimes (|w_1\rangle + |w_2\rangle) = (|v\rangle \otimes |w_1\rangle) + (|v\rangle \otimes |w_2\rangle)$$

and enables scalar multiplication by

$$\lambda(|v\rangle \otimes |w\rangle) = (\lambda|v\rangle) \otimes |w\rangle = |v\rangle \otimes (\lambda|w\rangle)$$

And we wish to build a way to associate the basis of  $V$  and  $W$  with the basis of  $V \otimes W$ . That makes the tensor product a vector space with dimension  $\dim V \times \dim W$ .

**Definition 2.** *Definition of linear functional*

*A linear functional is a linear map from  $V$  to  $\mathbb{F}$ .*

Note the difference between a linear functional and a linear map.

A generalized linear map is a function  $f : V \rightarrow W$  satisfying the condition.

- $f(|u\rangle + |v\rangle) = f(|u\rangle) + f(|v\rangle)$
- $f(\lambda|v\rangle) = \lambda f(|v\rangle)$

**Definition 3.** *A bilinear functional is a bilinear function  $\beta : V \times W \rightarrow \mathbb{F}$  satisfying the condition that  $|v\rangle \rightarrow \beta(|v\rangle, |w\rangle)$  is a linear functional for all  $|w\rangle \in W$  and  $|w\rangle \rightarrow \beta(|v\rangle, |w\rangle)$  is a linear functional for all  $|v\rangle \in V$ .*

The vector space of all bilinear functionals is denoted by  $\mathcal{B}(V, W)$ .

**Definition 4.** *Let  $V, W$  be two vector spaces.*

*Let  $V'$  and  $W'$  be the dual spaces of  $V$  and  $W$ , respectively, that is  $V' = \{\psi : V \rightarrow \mathbb{F}\}$  and  $W' = \{\phi : W \rightarrow \mathbb{F}\}$ ,  $\psi, \phi$  are linear functionals.*

*The tensor product of vectors  $v \in V$  and  $w \in W$  is the bilinear functional defined by  $\forall(\psi, \phi) \in V' \times W'$  given by the notation*

$$(v \otimes w)(\psi, \phi) := \psi(v)\phi(w)$$

*The tensor product of two vector spaces  $V$  and  $W$  is the vector space  $\mathcal{B}(V', W')$*

*Notice that the basis of such vector space is the linear combination of the basis of  $V'$  and  $W'$ , that is, if  $\{e_i\}$  is the basis of  $V'$  and  $\{f_j\}$  is the basis of  $W'$ , then  $\{e_i \otimes f_j\}$  is the basis of  $\mathcal{B}(V', W')$ .*

*That is, every element of  $\mathcal{B}(V', W')$  can be written as a linear combination of the basis.*

*Since  $\{e_i\}$  and  $\{f_j\}$  are bases of  $V'$  and  $W'$ , respectively, then we can always find a set of linear functionals  $\{\phi_i\}$  and  $\{\psi_j\}$  such that  $\phi_i(e_j) = \delta_{ij}$  and  $\psi_j(f_i) = \delta_{ij}$ .*

Here  $\delta_{ij} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{otherwise} \end{cases}$  is the Kronecker delta.

$$V \otimes W = \left\{ \sum_{i=1}^n \sum_{j=1}^m a_{ij} \phi_i(v) \psi_j(w) : \phi_i \in V', \psi_j \in W' \right\}$$

Note that  $\sum_{i=1}^n \sum_{j=1}^m a_{ij} \phi_i(v) \psi_j(w)$  is a bilinear functional that maps  $V' \times W'$  to  $\mathbb{F}$ .

This enables basis-free construction of vector spaces with proper multiplication and scalar multiplication.

This vector space is equipped with the unique inner product  $\langle v \otimes w, u \otimes x \rangle_{V \otimes W}$  defined by

$$\langle v \otimes w, u \otimes x \rangle = \langle v, u \rangle_V \langle w, x \rangle_W$$

In practice, we ignore the subscript of the vector space and just write  $\langle v \otimes w, u \otimes x \rangle = \langle v, u \rangle \langle w, x \rangle$ .

This introduces a new model in mathematics explaining quantum mechanics: the non-commutative probability theory.

## 1.2 Non-commutative probability theory

The non-commutative probability theory is a branch of generalized probability theory that studies the probability of events in non-commutative algebras.

There are several main components of the generalized probability theory; let's see how we can formulate them, comparing with the classical probability theory.

First, we define the Hilbert space in case one did not make the step from the linear algebra courses like me.

**Definition 5.** *Hilbert space:*

*A Hilbert space is a complete inner product space.*

That is, a vector space equipped with an inner product that is complete (every Cauchy sequence converges to a limit).

To introduce an example of Hilbert space we use when studying quantum mechanics, we need to introduce a common inner product used in  $\mathbb{C}^n$ .

**Definition 6.** *Hermitian inner product:*

*On  $\mathbb{C}^n$ , the Hermitian inner product is defined by*

$$\langle u, v \rangle = \sum_{i=1}^n \overline{u_i} v_i$$

**Proposition 7.** *The Hermitian inner product on the complex vector space  $\mathbb{C}^n$  makes it a Hilbert space.*

*Proof.* We first verify that the Hermitian inner product

$$\langle u, v \rangle = \sum_{i=1}^n \overline{u_i} v_i$$

on  $\mathbb{C}^n$  satisfies the axioms of an inner product:

1. **Conjugate symmetry:** For all  $u, v \in \mathbb{C}^n$ ,

$$\langle u, v \rangle = \sum_{i=1}^n \overline{u_i} v_i = \overline{\sum_{i=1}^n \overline{v_i} u_i} = \overline{\langle v, u \rangle}.$$

2. **Linearity:** For any  $u, v, w \in \mathbb{C}^n$  and scalars  $a, b \in \mathbb{C}$ , we have

$$\langle u, av + bw \rangle = \sum_{i=1}^n \overline{u_i} (av_i + bw_i) = a\langle u, v \rangle + b\langle u, w \rangle.$$

3. **Positive definiteness:** For every  $u = (u_1, u_2, \dots, u_n) \in \mathbb{C}^n$ , let  $u_j = a_j + b_j i$ , where  $a_j, b_j \in \mathbb{R}$ .

$$\langle u, u \rangle = \sum_{j=1}^n \overline{u_j} u_j = \sum_{i=1}^n (a_i^2 + b_i^2) \geq 0,$$

with equality if and only if  $u = 0$ .

Therefore, the Hermitian inner product is an inner product.

Next, we show that  $\mathbb{C}^n$  is complete with respect to the norm induced by this inner product:

$$\|u\| = \sqrt{\langle u, u \rangle}.$$

Since  $\mathbb{C}^n$  is finite-dimensional, every Cauchy sequence (with respect to any norm) converges in  $\mathbb{C}^n$ . This is a standard result in finite-dimensional normed spaces, which implies that  $\mathbb{C}^n$  is indeed complete.

Therefore, since the Hermitian inner product fulfills the inner product axioms and  $\mathbb{C}^n$  is complete, the complex vector space  $\mathbb{C}^n$  with the Hermitian inner product is a Hilbert space.  $\square$

Another classical example of Hilbert space is  $L^2(\Omega, \mathcal{F}, P)$ , where  $(\Omega, \mathcal{F}, P)$  is a measure space ( $\Omega$  is a set,  $\mathcal{F}$  is a  $\sigma$ -algebra on  $\Omega$ , and  $P$  is a measure on  $\mathcal{F}$ ). The  $L^2$  space is the space of all functions on  $\Omega$  that is

1. **square integrable:** square integrable functions are the functions  $f : \Omega \rightarrow \mathbb{C}$  such that

$$\int_{\Omega} |f(\omega)|^2 dP(\omega) < \infty$$

with inner product defined by

$$\langle f, g \rangle = \int_{\Omega} \overline{f(\omega)} g(\omega) dP(\omega)$$

2. **complex-valued**: functions are complex-valued measurable.  $f = u + vi$  is complex-valued if  $u$  and  $v$  are real-valued measurable.

**Proposition 8.**  $L^2(\Omega, \mathcal{F}, P)$  is a Hilbert space.

*Proof.* We check the two conditions of the Hilbert space:

- Completeness: Let  $(f_n)$  be a Cauchy sequence in  $L^2(\Omega, \mathcal{F}, P)$ . Then for any  $\epsilon > 0$ , there exists an  $N$  such that for all  $m, n \geq N$ , we have

$$\int_{\Omega} |f_m(\omega) - f_n(\omega)|^2 dP(\omega) < \epsilon^2$$

This means that  $(f_n)$  is a Cauchy sequence in the norm of  $L^2(\Omega, \mathcal{F}, P)$ .

- Inner product: The inner product is defined by

$$\langle f, g \rangle = \int_{\Omega} \overline{f(\omega)} g(\omega) dP(\omega)$$

This is a well-defined inner product on  $L^2(\Omega, \mathcal{F}, P)$ . We can check the properties of the inner product:

- Linearity:

$$\langle af + bg, h \rangle = a\langle f, h \rangle + b\langle g, h \rangle$$

- Conjugate symmetry:

$$\langle f, g \rangle = \overline{\langle g, f \rangle}$$

- Positive definiteness:

$$\langle f, f \rangle \geq 0$$

□

Let  $\mathcal{H}$  be a Hilbert space.  $\mathcal{H}$  consists of complex-valued functions on a finite set  $\Omega = \{1, 2, \dots, n\}$ , and the functions  $(e_1, e_2, \dots, e_n)$  form an orthonormal basis of  $\mathcal{H}$ . (We use Dirac notation  $|k\rangle$  to denote the basis vector  $e_k$  [parthasarathy1992quantum].)

As an analog to the classical probability space  $(\Omega, \mathcal{F}, \mu)$ , which consists of a sample space  $\Omega$  and a probability measure  $\mu$  on the state space  $\mathcal{F}$ , the non-commutative probability space  $(\mathcal{H}, \mathcal{P}, \rho)$  consists of a Hilbert space  $\mathcal{H}$  and a state  $\rho$  on the space of all orthogonal projections  $\mathcal{P}$ .

The detailed definition of the non-commutative probability space is given below:

**Definition 9.** Non-commutative probability space:

A non-commutative probability space is a pair  $(\mathcal{B}(\mathcal{H}), \mathcal{P})$ , where  $\mathcal{B}(\mathcal{H})$  is the set of all **bounded** linear operators on  $\mathcal{H}$ .

A linear operator on  $\mathcal{H}$  is **bounded** if for all  $u$  such that  $\|u\| \leq 1$ , we have  $\|Au\| \leq M$  for some  $M > 0$ .

$\mathcal{P}$  is the set of all orthogonal projections on  $\mathcal{B}(\mathcal{H})$ .

The set  $\mathcal{P} = \{P \in \mathcal{B}(\mathcal{H}) : P^* = P = P^2\}$  is the set of all orthogonal projections on  $\mathcal{B}(\mathcal{H})$ .

Recall from classical probability theory, we call the initial probability distribution for possible outcomes in the classical probability theory as our *state*, similarly, we need to define the *state* in the non-commutative probability theory.

**Definition 10.** Non-commutative probability state:

A state on  $(\mathcal{B}(\mathcal{H}), \mathcal{P})$  is a map  $\rho : \mathcal{P} \rightarrow [0, 1]$ , (commonly named as density operator) such that:

- $\rho(O) = 0$ , where  $O$  is the zero projection, and  $\rho(I) = 1$ , where  $I$  is the identity projection.
- If  $P_1, P_2, \dots, P_n$  are pairwise disjoint orthogonal projections, then  $\rho(P_1 + P_2 + \dots + P_n) = \sum_{i=1}^n \rho(P_i)$ .

An example of a density operator can be given as follows:

If  $(|\psi_1\rangle, |\psi_2\rangle, \dots, |\psi_n\rangle)$  is an orthonormal basis of  $\mathcal{H}$  consisting of eigenvectors of  $\rho$ , for the eigenvalues  $p_1, p_2, \dots, p_n$ , then  $p_j \geq 0$  and  $\sum_{j=1}^n p_j = 1$ .

We can write  $\rho$  as

$$\rho = \sum_{j=1}^n p_j |\psi_j\rangle\langle\psi_j|$$

(Under basis  $|\psi_j\rangle$ , it is a diagonal matrix with  $p_j$  on the diagonal.)

The counterpart of the random variable in the non-commutative probability theory is called an observable, which is a Hermitian operator on  $\mathcal{H}$  (for all  $\psi, \phi$  in the domain of  $A$ , we have  $\langle A\psi, \phi \rangle = \langle \psi, A\phi \rangle$ ). This kind of operator ensures that our outcome interpreted as probability is a real number).

**Definition 11.** Observable:

Let  $\mathcal{B}(\mathbb{R})$  be the set of all Borel sets on  $\mathbb{R}$ .

A random variable on the Hilbert space  $\mathcal{H}$  is a projection-valued map (measure)  $P : \mathcal{B}(\mathbb{R}) \rightarrow \mathcal{P}$ .

With the following properties:

- $P(\emptyset) = O$  (the zero projection)
- $P(\mathbb{R}) = I$  (the identity projection)
- For any sequence  $A_1, A_2, \dots, A_n \in \mathcal{B}(\mathbb{R})$ , the following holds:
  - $P(\bigcup_{i=1}^n A_i) = \bigvee_{i=1}^n P(A_i)$
  - $P(\bigcap_{i=1}^n A_i) = \bigwedge_{i=1}^n P(A_i)$
  - $P(A^c) = I - P(A)$
  - If  $A_j$  are mutually disjoint (that is  $P(A_i)P(A_j) = P(A_j)P(A_i) = O$  for  $i \neq j$ ), then  $P(\bigcup_{j=1}^n A_j) = \sum_{j=1}^n P(A_j)$

**Definition 12.** *Probability of a random variable:*

For a system prepared in state  $\rho$ , the probability that the random variable given by the projection-valued measure  $P$  is in the Borel set  $A$  is  $\text{Tr}(\rho P(A))$ .

When operators commute, we recover classical probability measures.

**Definition 13.** *Definition of measurement:*

A measurement (observation) of a system prepared in a given state produces an outcome  $x$ ,  $x$  is a physical event that is a subset of the set of all possible outcomes. For each  $x$ , we associate a measurement operator  $M_x$  on  $\mathcal{H}$ .

Given the initial state (pure state, unit vector)  $u$ , the probability of measurement outcome  $x$  is given by:

$$p(x) = \|M_x u\|^2$$

Note that to make sense of this definition, the collection of measurement operators  $\{M_x\}$  must satisfy the completeness requirement:

$$1 = \sum_{x \in X} p(x) = \sum_{x \in X} \|M_x u\|^2 = \sum_{x \in X} \langle M_x u, M_x u \rangle = \langle u, (\sum_{x \in X} M_x^* M_x) u \rangle$$

$$\text{So } \sum_{x \in X} M_x^* M_x = I.$$

**Proposition 14.** *Proposition of indistinguishability:*

Suppose that we have two systems  $u_1, u_2 \in \mathcal{H}_1$ , the two states are distinguishable if and only if they are orthogonal.

*Proof.* Ways to distinguish the two states:

1. Set  $X = \{0, 1, 2\}$  and  $M_i = |u_i\rangle\langle u_i|$ ,  $M_0 = I - M_1 - M_2$
2. Then  $\{M_0, M_1, M_2\}$  is a complete collection of measurement operators on  $\mathcal{H}$ .
3. Suppose the prepared state is  $u_1$ , then  $p(1) = \|M_1 u_1\|^2 = \|u_1\|^2 = 1$ ,  $p(2) = \|M_2 u_1\|^2 = 0$ ,  $p(0) = \|M_0 u_1\|^2 = 0$ .

If they are not orthogonal, then there is no choice of measurement operators to perfectly distinguish the two states.

□

*Intuitively, if the two states are not orthogonal, then for any measurement (projection) there exists non-zero probability of getting the same outcome for both states.*

Here is Table 1.1 summarizing the analog of classical probability theory and non-commutative (quantum) probability theory [Feres]:

Table 1.1: Analog of classical probability theory and non-commutative (*quantum*) probability theory

Classical probability	Non-commutative probability
Sample space $\Omega$ , cardinality $ \Omega  = n$ , example: $\Omega = \{0, 1\}$	Complex Hilbert space $\mathcal{H}$ , dimension $\dim \mathcal{H} = n$ , example: $\mathcal{H} = \mathbb{C}^2$
Common algebra of $\mathbb{C}$ valued functions	Algebra of bounded operators $\mathcal{B}(\mathcal{H})$
$f \mapsto \bar{f}$ complex conjugation	$P \mapsto P^*$ adjoint
Events: indicator functions of sets	Projections: space of orthogonal projections $\mathcal{P} \subseteq \mathcal{B}(\mathcal{H})$
functions $f$ such that $f^2 = f = \bar{f}$	orthogonal projections $P$ such that $P^* = P = P^2$
$\mathbb{R}$ -valued functions $f = \bar{f}$	self-adjoint operators $A = A^*$
$\mathbb{I}_{f^{-1}(\{\lambda\})}$ is the indicator function of the set $f^{-1}(\{\lambda\})$	$P(\lambda)$ is the orthogonal projection to eigenspace
$f = \sum_{\lambda \in \text{Range}(f)} \lambda \mathbb{I}_{f^{-1}(\{\lambda\})}$	$A = \sum_{\lambda \in \text{sp}(A)} \lambda P(\lambda)$
Probability measure $\mu$ on $\Omega$	Density operator $\rho$ on $\mathcal{H}$
Delta measure $\delta_\omega$	Pure state $\rho =  \psi\rangle\langle\psi $
$\mu$ is non-negative measure and $\sum_{i=1}^n \mu(\{i\}) = 1$	$\rho$ is positive semi-definite and $\text{Tr}(\rho) = 1$
Expected value of random variable $f$ is $\mathbb{E}_\mu(f) = \sum_{i=1}^n f(i)\mu(\{i\})$	Expected value of operator $A$ is $\mathbb{E}_\rho(A) = \text{Tr}(\rho A)$
Variance of random variable $f$ is $\text{Var}_\mu(f) = \sum_{i=1}^n (f(i) - \mathbb{E}_\mu(f))^2 \mu(\{i\})$	Variance of operator $A$ is $\text{Var}_\rho(A) = \text{Tr}(\rho A^2) - \text{Tr}(\rho A)^2$
Covariance of random variables $f$ and $g$ is $\text{Cov}_\mu(f, g) = \sum_{i=1}^n (f(i) - \mathbb{E}_\mu(f))(g(i) - \mathbb{E}_\mu(g))\mu(\{i\})$	Covariance of operators $A$ and $B$ is $\text{Cov}_\rho(A, B) = \text{Tr}(\rho A \circ B) - \text{Tr}(\rho A) \text{Tr}(\rho B)$
Composite system is given by Cartesian product of the sample spaces $\Omega_1 \times \Omega_2$	Composite system is given by tensor product of the Hilbert spaces $\mathcal{H}_1 \otimes \mathcal{H}_2$
Product measure $\mu_1 \times \mu_2$ on $\Omega_1 \times \Omega_2$	Tensor product of space $\rho_1 \otimes \rho_2$ on $\mathcal{H}_1 \otimes \mathcal{H}_2$
Marginal distribution $\pi_* v$	Partial trace $\text{Tr}_2(\rho)$

### 1.3 Concentration of measure phenomenon

**Definition 15.**  $\eta$ -Lipschitz function

Let  $(X, \text{dist}_X)$  and  $(Y, \text{dist}_Y)$  be two metric spaces. A function  $f : X \rightarrow Y$  is said to be  $\eta$ -Lipschitz if there exists a constant  $L \in \mathbb{R}$  such that

$$\text{dist}_Y(f(x), f(y)) \leq L \text{dist}_X(x, y)$$

for all  $x, y \in X$ . And  $\eta = \|f\|_{\text{Lip}} = \inf_{L \in \mathbb{R}} L$ .

That basically means that the function  $f$  should not change the distance between any two pairs of points in  $X$  by more than a factor of  $L$ .

**Lemma 16.** Isoperimetric inequality on the sphere:

Let  $\sigma_n(A)$  denote the normalized area of  $A$  on the  $n$ -dimensional sphere  $S^n$ . That is,  $\sigma_n(A) := \frac{\text{Area}(A)}{\text{Area}(S^n)}$ .

Let  $\epsilon > 0$ . Then for any subset  $A \subset S^n$ , given the area  $\sigma_n(A)$ , the spherical caps minimize the volume of the  $\epsilon$ -neighborhood of  $A$ .

Suppose  $\sigma^n(\cdot)$  is the normalized volume measure on the sphere  $S^n(1)$ , then for any closed subset  $\Omega \subset S^n(1)$ , we take a metric ball  $B_\Omega$  of  $S^n(1)$  with  $\sigma^n(B_\Omega) = \sigma^n(\Omega)$ . Then we have

$$\sigma^n(U_r(\Omega)) \geq \sigma^n(U_r(B_\Omega))$$

where  $U_r(A) = \{x \in X : d(x, A) < r\}$

Intuitively, the lemma means that the spherical caps are the most efficient way to cover the sphere. Here, the efficiency is measured by the epsilon-neighborhood of the boundary of the spherical cap. To prove the lemma, we need to have a good understanding of the Riemannian geometry of the sphere. For now, let's just take the lemma for granted.

### 1.3.1 Levy's concentration theorem

**Theorem 17.** *Levy's concentration theorem:*

*An arbitrary 1-Lipschitz function  $f : S^n \rightarrow \mathbb{R}$  concentrates near a single value  $a_0 \in \mathbb{R}$  as strongly as the distance function does.*

*That is,*

$$\mu\{x \in S^n : |f(x) - a_0| \geq \epsilon\} < \kappa_n(\epsilon) \leq 2 \exp\left(-\frac{(n-1)\epsilon^2}{2}\right)$$

*where*

$$\kappa_n(\epsilon) = \frac{\int_{-\epsilon}^{\epsilon} \cos^{n-1}(t) dt}{\int_{-\pi/2}^{\pi/2} \cos^{n-1}(t) dt}$$

$a_0$  is the **Levy mean** of function  $f$ , that is, the level set  $f^{-1} : \mathbb{R} \rightarrow S^n$  divides the sphere into equal halves, characterized by the following equality:

$$\mu(f^{-1}(-\infty, a_0]) \geq \frac{1}{2} \text{ and } \mu(f^{-1}[a_0, \infty)) \geq \frac{1}{2}$$

We will prove the theorem via the Maxwell-Boltzmann distribution law. [shioya2014metricmeasuregeometry]

**Definition 18.** *Gaussian measure:*

*We denote the Gaussian measure on  $\mathbb{R}^k$  as  $\gamma^k$ .*

$$d\gamma^k(x) := \frac{1}{\sqrt{2\pi}^k} \exp\left(-\frac{1}{2}\|x\|^2\right) dx$$

$x \in \mathbb{R}^k$ ,  $\|x\|^2 = \sum_{i=1}^k x_i^2$  is the Euclidean norm, and  $dx$  is the Lebesgue measure on  $\mathbb{R}^k$ .

Basically, you can consider the Gaussian measure as the normalized Lebesgue measure on  $\mathbb{R}^k$  with standard deviation 1.

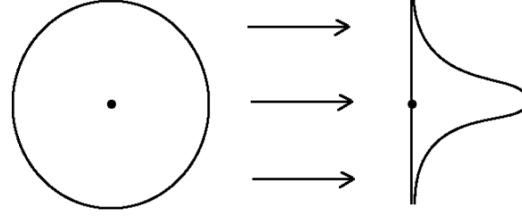
It also has another name, the Projective limit theorem. [romanvershyni]

If  $X \sim \text{Unif}(S^n(\sqrt{n}))$ , then for any fixed unit vector  $x$  we have  $\langle X, x \rangle \rightarrow N(0, 1)$  in distribution as  $n \rightarrow \infty$ .

**Lemma 19.** *Maxwell-Boltzmann distribution law:*

*For any natural number  $k$ ,*

$$\frac{d(\pi_{n,k})_* \sigma^n(x)}{dx} \rightarrow \frac{d\gamma^k(x)}{dx}$$



**Figure 3.9** The projective central limit theorem: the projection of the uniform distribution on the sphere of radius  $\sqrt{n}$  onto a line converges to the normal distribution  $N(0, 1)$  as  $n \rightarrow \infty$ .

Figure 1.2: Maxwell-Boltzmann distribution law, image from [romanvershyni]

where  $(\pi_{n,k})_*\sigma^n$  is the push-forward measure of  $\sigma^n$  by  $\pi_{n,k}$ .

In other words,

$$(\pi_{n,k})_*\sigma^n \rightarrow \gamma^k \text{ weakly as } n \rightarrow \infty$$

*Proof.* We denote the  $n$ -dimensional volume measure on  $\mathbb{R}^k$  as  $\text{vol}_k$ .

Observe that  $\pi_{n,k}^{-1}(x), x \in \mathbb{R}^k$  is isometric to  $S^{n-k}(\sqrt{n - \|x\|^2})$ , that is, for any  $x \in \mathbb{R}^k$ ,  $\pi_{n,k}^{-1}(x)$  is a sphere with radius  $\sqrt{n - \|x\|^2}$  (by the definition of  $\pi_{n,k}$ ).

So,

$$\begin{aligned} \frac{d(\pi_{n,k})_*\sigma^n(x)}{dx} &= \frac{\text{vol}_{n-k}(\pi_{n,k}^{-1}(x))}{\text{vol}_k(S^n(\sqrt{n}))} \\ &= \frac{(n - \|x\|^2)^{\frac{n-k}{2}}}{\int_{\|x\| \leq \sqrt{n}} (n - \|x\|^2)^{\frac{n-k}{2}} dx} \end{aligned}$$

as  $n \rightarrow \infty$ .

Note that  $\lim_{n \rightarrow \infty} (1 - \frac{a}{n})^n = e^{-a}$  for any  $a > 0$ .

$$(n - \|x\|^2)^{\frac{n-k}{2}} = \left(n(1 - \frac{\|x\|^2}{n})\right)^{\frac{n-k}{2}} \rightarrow n^{\frac{n-k}{2}} \exp(-\frac{\|x\|^2}{2})$$

So

$$\begin{aligned} \frac{(n - \|x\|^2)^{\frac{n-k}{2}}}{\int_{\|x\| \leq \sqrt{n}} (n - \|x\|^2)^{\frac{n-k}{2}} dx} &= \frac{e^{-\frac{\|x\|^2}{2}}}{\int_{x \in \mathbb{R}^k} e^{-\frac{\|x\|^2}{2}} dx} \\ &= \frac{1}{(2\pi)^{\frac{k}{2}}} e^{-\frac{\|x\|^2}{2}} \\ &= \frac{d\gamma^k(x)}{dx} \end{aligned}$$

□

Now we can prove Levy's concentration theorem, the proof is from [shioya2014metricmeasuregeometry].

*Proof.* Let  $f_n : S^n(\sqrt{n}) \rightarrow \mathbb{R}$ ,  $n = 1, 2, \dots$ , be 1-Lipschitz functions.

Let  $x$  and  $x'$  be two given real numbers and  $\gamma^1(-\infty, x] = \bar{\sigma}_\infty[-\infty, x']$ , suppose  $\sigma_\infty\{x'\} = 0$ , where  $\{\sigma_i\}$  is a sequence of Borel probability measures on  $\mathbb{R}$ .

We want to show that, for all non-negative real numbers  $\epsilon_1$  and  $\epsilon_2$ .

$$\sigma_\infty[x' - \epsilon_1, x' + \epsilon_2] \geq \gamma^1[x - \epsilon_1, x + \epsilon_2]$$

Consider the two spherical cap  $\Omega_+ := \{f_{n_i} \geq x'\}$  and  $\Omega_- := \{f_{n_i} \leq x\}$ . Note that  $\Omega_+ \cup \Omega_- = S^{n_i}(\sqrt{n_i})$ .

It is sufficient to show that,

$$U_{\epsilon_1}(\Omega_+) \cup U_{\epsilon_2}(\Omega_-) \subset \{x' - \epsilon_1 \leq f_{n_i} \leq x' + \epsilon_2\}$$

By 1-Lipschitz continuity of  $f_{n_i}$ , we have for all  $\zeta \in U_{\epsilon_1}(\Omega_+)$ , there is a point  $\xi \in \Omega_+$  such that  $d(\zeta, \xi) \leq \epsilon_1$ . So  $U_{\epsilon_1}(\Omega_+) \subset \{f_{n_i} \geq x' - \epsilon_1\}$ . With the same argument, we have  $U_{\epsilon_2}(\Omega_-) \subset \{f_{n_i} \leq x + \epsilon_2\}$ .

So the push-forward measure of  $(f_{n_i})_*\sigma^{n_i}$  of  $[x' - \epsilon_1, x' + \epsilon_2]$  is

$$\begin{aligned} (f_{n_i})_*\sigma^{n_i}[x' - \epsilon_1, x' + \epsilon_2] &= \sigma^{n_i}(x' - \epsilon_1 \leq f_{n_i} \leq x' + \epsilon_2) \\ &\geq \sigma^{n_i}(U_{\epsilon_1}(\Omega_+) \cap U_{\epsilon_2}(\Omega_-)) \\ &= \sigma^{n_i}(U_{\epsilon_1}(\Omega_+)) + \sigma^{n_i}(U_{\epsilon_2}(\Omega_-)) - 1 \end{aligned}$$

By the lemma 16, we have

$$\sigma^{n_i}(U_{\epsilon_1}(\Omega_+)) \geq \sigma^{n_i}(U_{\epsilon_1}(B_{\Omega_+})) \quad \text{and} \quad \sigma^{n_i}(U_{\epsilon_2}(\Omega_-)) \geq \sigma^{n_i}(U_{\epsilon_2}(B_{\Omega_-}))$$

By the lemma 19, we have

$$\sigma^{n_i}(U_{\epsilon_1}(\Omega_+)) + \sigma^{n_i}(U_{\epsilon_2}(\Omega_-)) \rightarrow \gamma^1[x' - \epsilon_1, x' + \epsilon_2] + \gamma^1[x - \epsilon_1, x + \epsilon_2]$$

Therefore,

$$\begin{aligned} \sigma_\infty[x' - \epsilon_1, x' + \epsilon_2] &\geq \liminf_{i \rightarrow \infty} (f_{n_i})_*\sigma^{n_i}[x' - \epsilon_1, x' + \epsilon_2] \\ &\geq \gamma^1[x' - \epsilon_1, \infty) \cap \gamma^1(-\infty, x + \epsilon_2] - 1 \\ &= \gamma^1[x - \epsilon_1, x + \epsilon_2] \end{aligned}$$

□

The full proof of Levy's concentration theorem requires more digestion for cases where  $\bar{\sigma}_\infty \neq \delta_{\pm\infty}$  but I don't have enough time to do so. This section may be filled in the next semester.

## 1.4 The application of the concentration of measure phenomenon in non-commutative probability theory

In quantum communication, we can pass classical bits by sending quantum states. However, by the indistinguishability (Proposition 14) of quantum states, we cannot send an infinite number of classical bits over a single qubit. There exists a bound for zero-error classical communication rate over a quantum channel.

**Theorem 20.** *Holevo bound:*

*The maximal amount of classical information that can be transmitted by a quantum system is given by the Holevo bound.  $\log_2(d)$  is the maximum amount of classical information that can be transmitted by a quantum system with  $d$  levels (that is, basically, the number of qubits).*

The proof of the Holevo bound can be found in [Nielsen·Chuang·2010]. In current state of the project, this theorem is not heavily used so we will not make annotated proof here.

### 1.4.1 Quantum communication

To surpass the Holevo bound, we need to use the entanglement of quantum states.

**Definition 21.** *Bell state:*

*The Bell states are the following four states:*

$$\begin{aligned} |\Phi^+\rangle &= \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle), & |\Phi^-\rangle &= \frac{1}{\sqrt{2}}(|00\rangle - |11\rangle) \\ |\Psi^+\rangle &= \frac{1}{\sqrt{2}}(|01\rangle + |10\rangle), & |\Psi^-\rangle &= \frac{1}{\sqrt{2}}(|01\rangle - |10\rangle) \end{aligned}$$

*These are a basis of the 2-qubit Hilbert space.*

### 1.4.2 Superdense coding and entanglement

The description of the superdense coding can be found in [gupta2015functionalanalysisquantuminformation] and [Hayden].

Suppose  $A$  and  $B$  share a Bell state (or other maximally entangled state)  $|\Phi^+\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$ , where  $A$  holds the first part and  $B$  holds the second part.

$A$  wishes to send 2 **classical bits** to  $B$ .

$A$  performs one of four Pauli unitaries (some fancy quantum gates named X, Y, Z, I) on the combined state of entangled qubits  $\otimes$  one qubit. Then  $A$  sends the resulting one qubit to  $B$ .

This operation extends the initial one entangled qubit to a system of one of four orthogonal Bell states.

$B$  performs a measurement on the combined state of the one qubit and the entangled qubits he holds.

$B$  decodes the result and obtains the 2 classical bits sent by  $A$ .

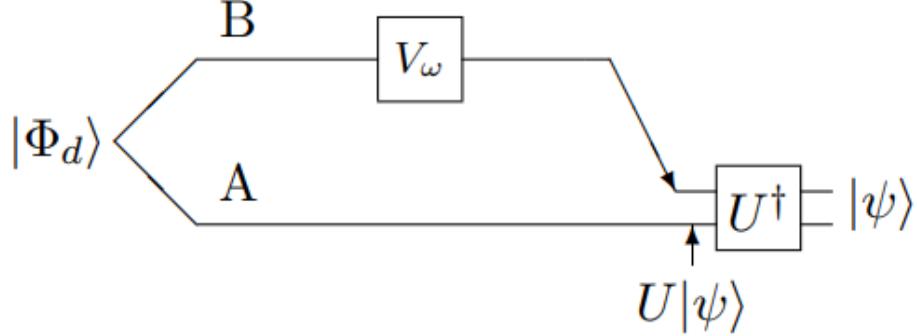


Figure 1.3: Superdense coding, image from [Hayden]

Note that superdense coding is a way to send 2 classical bits of information by sending 1 qubit with 1 entangled qubit. **The role of the entangled qubit** is to help them to distinguish the 4 possible states of the total 3 qubits system where 2 of them (the pair of entangled qubits) are mathematically the same.

Additionally, no information can be gained by measuring a pair of entangled qubits. To send information from  $A$  to  $B$ , we need to physically send the qubits from  $A$  to  $B$ . That means, we cannot send information faster than the speed of light.

#### 1.4.3 Hayden's concentration of measure phenomenon

The application of the concentration of measure phenomenon in the superdense coding can be realized in random sampling the entangled qubits [Hayden]:

It is a theorem connecting the following mathematical structure:

$$\begin{array}{ccc}
 \mathcal{P}(A \otimes B) & \longleftrightarrow & \mathbb{C}P^{d_A d_B - 1} \\
 \downarrow \text{Tr}_B & \searrow f & \\
 S_A & \xrightarrow{H(\psi_A)} & [0, \infty) \subset \mathbb{R}
 \end{array}$$

Figure 1.4: Mathematical structure for Hayden's concentration of measure phenomenon

- The red arrow is the concentration of measure effect.  $f = H(\text{Tr}_B(\psi))$ .

- $S_A$  denotes the mixed states on  $A$ .

To prove the concentration of measure phenomenon, we need to analyze the following elements involved in figure 1.4:

First, we need to define what is a random state in a bipartite system. In fact, for pure states, there is a unique uniform distribution under Haar measure that is unitarily invariant.

$U(n)$  is the group of all  $n \times n$  **unitary matrices** over  $\mathbb{C}$ ,

$$U(n) = \{A \in \mathbb{C}^{n \times n} : A^*A = AA^* = I_n\}$$

The uniqueness of such measurement came from the lemma below [Elizabeth book]

**Lemma 22.** *Let  $(U(n), \|\cdot\|, \mu)$  be a metric measure space where  $\|\cdot\|$  is the Hilbert-Schmidt norm and  $\mu$  is the measure function.*

*The Haar measure on  $U(n)$  is the unique probability measure that is invariant under the action of  $U(n)$  on itself.*

*That is, fixing  $B \in U(n)$ ,  $\forall A \in U(n)$ ,  $\mu(A \cdot B) = \mu(B \cdot A) = \mu(B)$ .*

*The Haar measure is the unique probability measure that is invariant under the action of  $U(n)$  on itself.*

The existence and uniqueness of the Haar measure is a theorem in compact lie group theory. For this research topic, we will not prove it.

A random pure state  $\varphi$  is any random variable distributed according to the unitarily invariant probability measure on the pure states  $\mathcal{P}(A)$  of the system  $A$ , denoted by  $\varphi \in_R \mathcal{P}(A)$ .

It is trivial that for the space of pure state, we can easily apply the Haar measure as the unitarily invariant probability measure since the space of pure state is  $S^n$  for some  $n$ . However, for the case of mixed states, that is a bit complicated and we need to use partial tracing to defined the rank- $s$  random states.

**Definition 23.** *Rank- $s$  random state.*

*For a system  $A$  and an integer  $s \geq 1$ , consider the distribution on the mixed states  $\mathcal{S}(A)$  of  $A$  induced by the partial trace over the second factor from the uniform distribution on pure states of  $A \otimes \mathbb{C}^s$ . Any random variable  $\rho$  distributed as such will be called a rank- $s$  random states; denoted as  $\rho \in_R \mathcal{S}_s(A)$ . And  $\mathcal{P}(A) = \mathcal{S}_1(A)$ .*

Due to time constrains of the projects, the following lemma is demonstrated but not investigated thoroughly through the research:

**Lemma 24.** *Page's lemma for expected entropy of mixed states*

*Choose a random pure state  $\sigma = |\psi\rangle\langle\psi|$  from  $A' \otimes A$ .*

*The expected value of the entropy of entanglement is known and satisfies a concentration inequality known as Page's formula [Pages conjecture; Pages conjecture's simple proof; Bengtsson'yczkowski'2017]. The detailed proof is not fully explored in this project and is intended to be done in the next semester.*

$$\mathbb{E}[H(\psi_A)] \geq \log_2(d_A) - \frac{1}{2 \ln(2)} \frac{d_A}{d_B}$$

It basically provides a lower bound for the expected entropy of entanglement. Experimentally, we can have the following result (see Figure 1.5):

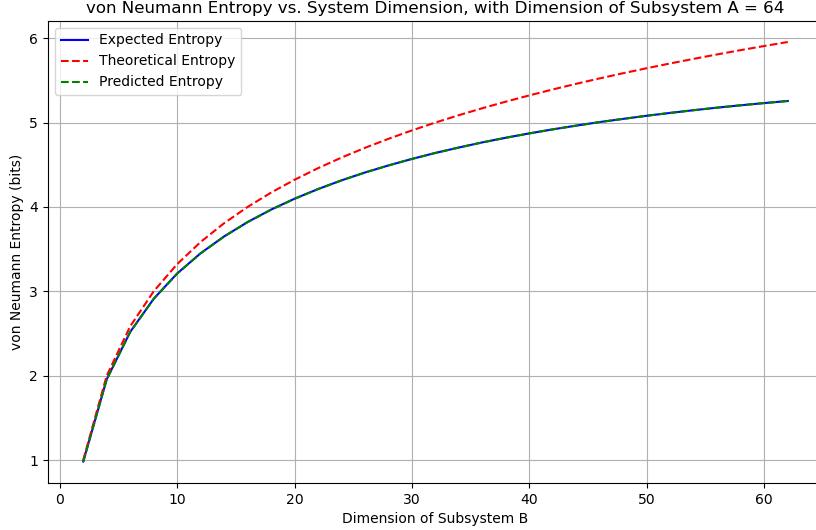


Figure 1.5: Entropy vs dimension

Then we have bound for Lipschitz constant  $\eta$  of the map  $H(\varphi_A)$

**Lemma 25.** *The Lipschitz constant  $\eta$  of  $S(\varphi_A)$  is upper bounded by  $\sqrt{8} \log_2(d_A)$  for  $d_A \geq 3$ .*

From Levy's lemma, we have

If we define  $\beta = \frac{1}{\ln(2)} \frac{d_A}{d_B}$ , then we have

$$\Pr[H(\psi_A) < \log_2(d_A) - \alpha - \beta] \leq \exp\left(-\frac{1}{8\pi^2 \ln(2)} \frac{(d_A d_B - 1)\alpha^2}{(\log_2(d_A))^2}\right)$$

where  $d_B \geq d_A \geq 3$  [Hayden 2006].

Experimentally, we can have the following result:

As the dimension of the Hilbert space increases, the chance of getting an almost maximally entangled state increases (see Figure 1.6).

In Hayden's work, the result is also extended to the multiparty case [Hayden], and the result is still under research and I will show the result in the final report if I have enough time.

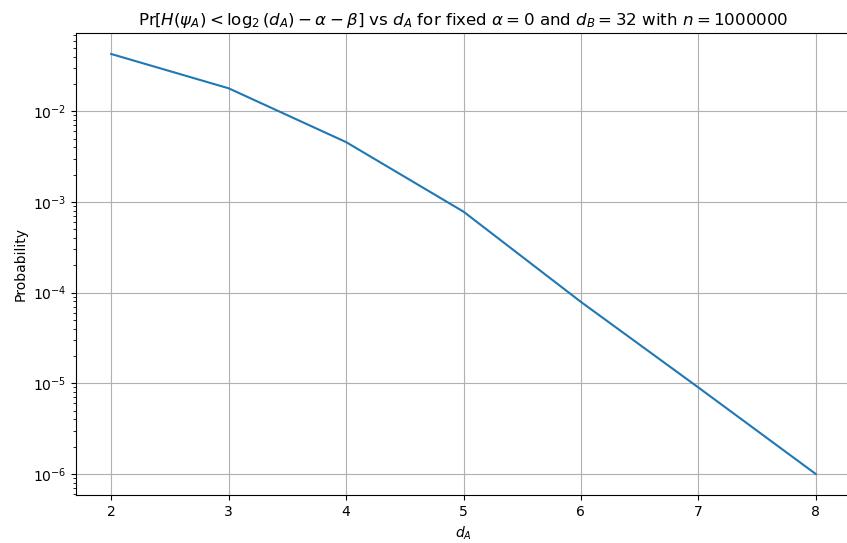


Figure 1.6: Entropy vs  $d_A$