Chapter 14

Entropy and information

In classical information theory the Shannon entropy plays quite an important role. In quantum information theory, this role is played by the **von-Neumann entropy**

$$S(\hat{\varrho}) \equiv -\mathrm{tr}\{\hat{\varrho}\ln\hat{\varrho}\}\,. \tag{14.1}$$

The von-Neumann entropy is bounded

$$0 \leq S(\hat{\varrho}) \leq \ln \dim\mathcal{H}\,, \tag{14.2}$$

with

$$S(\hat{\varrho}) = 0 \Leftrightarrow \mathrm{tr}\{\hat{\varrho}^2\} = 1 \Leftrightarrow \hat{\varrho} \text{ a pure state}\,, \tag{14.3}$$

and

$$S(\hat{\varrho}) = \ln \dim\mathcal{H} \Leftrightarrow \hat{\varrho} \text{ a complete mixture}\,. \tag{14.4}$$

Expressed in terms of the eigenvalues of $\hat{\varrho}$, see Eq. (13.10), the von-Neumann entropy is given by,

$$S(\hat{\varrho}) = -\sum_{\nu} \varrho_\nu \ln \varrho_\nu\,. \tag{14.5}$$

Note that the von-Neumann entropy may in general *not* be identified with the the Shannon entropy.

Example Consider a memoryless source which emits qubits either in state $|\uparrow_a\rangle$ or in state $|\uparrow_b\rangle$ with equal probability $p_a = p_b = 0.5$. The directions $\vec{a}$, $\vec{b}$ are distinct; the states $|\uparrow_a\rangle$, $|\uparrow_b\rangle$ are not necessarily orthogonal. As $\vec{a}$, $\vec{b}$ occur as true alternatives, the Shannon entropy of the production is

$$H = -p_a \mathrm{ld} p_a - p_b \mathrm{ld} p_b = 1\mathrm{bit}\,. \tag{14.6}$$

The state operator

$$\hat{\varrho} = p_a |\uparrow_a\rangle\langle\uparrow_a| + p_b |\uparrow_b\rangle\langle\uparrow_b| \tag{14.7}$$

admits the spectral representation

$$\hat{\varrho} = \varrho_+ |\uparrow_c\rangle\langle\uparrow_c| + \varrho_- |\downarrow_c\rangle\langle\downarrow_c| \tag{14.8}$$

with eigenvalues (recall $p_a = p_b = 1/2$)

$$\varrho_\pm = \tfrac{1}{2}\left(1 \pm |\langle\uparrow_a|\uparrow_b\rangle|\right)\,. \tag{14.9}$$

The eigenvectors are spin states with respect to a quantization axis $\vec{c} = (\vec{a} + \vec{b})/|\vec{a} + \vec{b}|$.

The von-Neumann entropy

$$S(\hat{\varrho}) = -\varrho_+ \ln \varrho_+ - \varrho_- \ln \varrho_-\,, \tag{14.10}$$

in generally differs from the Shannon entropy, $S(\hat{\varrho}) \leq 1\mathrm{bit}$.

The von-Neumann entropy of the predicted post-measurement state $\hat{\varrho}' = \Phi(\hat{\varrho})$ is given by $S' = S(\Phi(\hat{\varrho}))$. Its value depends on both, the in-state $\hat{\varrho}$, and the type of measurement – the orientation of the Stern-Gerlach magnet, say. For a SGM with orientation $\vec{a}$,

$$S' = -q_+ \ln q_+ - q_- \ln q_- \,, \tag{14.11}$$

where $q_+ = \langle \uparrow_a |\hat{\varrho}| \uparrow_a \rangle$, $q_- = \langle \downarrow_a |\hat{\varrho}| \downarrow_a \rangle$. The *entropy-of-measurement* S' obeys the inequality

$$0 \leq S \leq S' \leq 1 \,. \tag{14.12}$$

with equality $S = S'$ iff the orientation $\vec{a}$ concides with the polarization of the qubit $\vec{s}$.

Unless one choses $\vec{a} \propto \vec{s}$, the act of measurement (i) increases the impurity of a qubit's state, and (ii) the uncertainty about its state. This sounds quite amazing, as one usually associates the act of measurement with a removal of uncertainty. The paradox resolves, if we consider the SGM as part of a communication channel. Misalignment adds channel noise, which in turn increases the effective entropy $S \to S'$ at the end of the communication line. Note that this channel noise is not due to some technological imperfection of the channel, but rather results from the basic laws of quantum mechanics.

14.1 State estimation

Given a *single* qubit, the state of which is unknown. How to ascertain its state?

If we had many qubits, all in the same unknown state, the answer would be simple: just estimate $\vec{s}$ in Eq. () by measuring the three expectation values $\langle \hat{\sigma}_i \rangle$, $i = x, y, z$. But this route is closed, since we have only one quibt, i.e. the measurement consists of one click only.

The question is in fact quite tricky. Its answer heavily relies on the actual meaning of "unknown", i.e. whether "unknown" means "incomplete knowledge" or whether it means "complete ignorance".

Complete ignorance is equivalent to not knowing anything about the recipe of its preparation. The a priori state of the qubit is therefore a complete mixture, $\hat{\varrho} = \frac{1}{2}\hat{1}$. As there are infinitely many distinct recipes (choice of qubit polarization $\vec{s}$), all of which are macroscopically different, our initial level of ignorance is infinte. In any measurement, which is specified by an orientation $\vec{b}$, say, the particle will leave the device either through the upper, or the lower channel. Whatever it does, the only thing we can say about the particle is that it was with certainty not prepared in the pure state which labels the other channel. In state space, that state is a point of measure zero, and thus knowing that it was not prepared in that state does not affect our level of ignorance.

Incomplete knowledge is equivalent to knowing something but not enough. For example, we could know that Alice used a SGM with orientation along the z-axis for preparation, but we do not know whether she send the particle in state $|\uparrow_z\rangle$ or in state $|\downarrow_z\rangle$. Again the apriori state of the particle is a complete mixture,

$$\hat{\varrho} = \frac{1}{2}|\uparrow_z\rangle\langle\uparrow_z| + \frac{1}{2}|\downarrow_z\rangle\langle\downarrow_z| = \frac{1}{2}\hat{1}\,, \tag{14.13}$$

but equipped with our partial knowledge, we can easily determine Alice's actual preparation without making any error: just send the particle through a SGM with orientation in the z-direction. If it leaves the device through the upper channel we know with certainty that the particle was in fact prepared in state $|\uparrow_z\rangle$, otherwise we know with certainty that it was prepared in $|\downarrow_z\rangle$.

The juxtaposition of complete ignorance and incomplete knowledge clearly demonstrates that information can not be assigned to a state, but is rather associated with a state in a specific context: in the above examples, the state is the same – a

complete mixture – but the context is different.

Also the above examples indicate, that the von-Neumann entropy is *not* a suitable measure for information gain. For both examples $S = S' = 1\text{bit}$, but the information gain is clearly different.

14.2 Information gain

Alice sends Bob a qubit. Bob knows that in preparing the qubit, Alice used one out of N different procedures, each of which prepares the qubit in a certain state $\hat{\varrho}_i$, $i = 1, \ldots, N$. Bob knows the context, that is he knows the preparation procedures (he knows the set of possible states), and he knows the probabilities of their occurences p_i, $i = 1, \ldots, N$. His task is to give an estimate as reliable as possible of the actual state prepared by Alice.

The a priori state of the source (Alice) is given by the convex sum

$$\hat{\varrho} = \sum_{i=1}^{N} p_i \hat{\varrho}_i \,. \tag{14.14}$$

Let μ label the event "click in channel m ($m = \pm 1$) for a fixed orientation $\vec{b}$". Since Bob knows the individual ϱ_i, he can compute the conditional probability that μ occurs, given ϱ_i was actually prepared,

$$P_{\mu|i} = \langle \mu | \hat{\varrho}_i | \mu \rangle \tag{14.15}$$

and – since he also knows the apriori probabilities p_i – he can compute the apriori probability that μ occurs,

$$q_\mu = \sum_{i=1}^{N} P_{\mu|i} p_i \,. \tag{14.16}$$

From these two quantities Bob can compute

$$Q_{i|\mu} \equiv P_{\mu|i} p_i / q_\mu \,, \tag{14.17}$$

which according to Baye denotes the likelihood (or "a-posteriori probability") that $\hat{\varrho}_i$ was in fact prepared given μ occured.

Before the measurement, Bob's level of ignorance is given by the Shannon entropy of the source,

$$H_{|\text{initial}} = -\sum_{i=1}^{N} p_i \text{ld} p_i \,. \tag{14.18}$$

After the occurence of μ, his level of ignorance is given by

$$H_\mu = -\sum_{i=1}^{N} Q_{i|\mu} \text{ld} Q_{i|\mu} \,. \tag{14.19}$$

But since μ is not known in advance, his ignorance after the measurement is in the average

$$H_{\text{final}} = \sum_\mu p_\mu H_\mu \tag{14.20}$$

where the sum-over-μ is a summation over the channels $m = \pm 1$ for fixed orientation $\vec{b}$. Bob's average information gain is given by

$$\begin{aligned} I_{\text{av}} &\equiv H_{\text{initial}} - H_{\text{final}} && (14.21) \\ &= -\sum_i p_i \text{ld} p_i - \sum_\mu q_\mu H_\mu \,. && (14.22) \end{aligned}$$

For given context $\{\hat{\varrho}_i, p_i | i = 1, \ldots, N\}$, the average information gain I_{av} strongly depends on the measurement strategy (choice of $\vec{b}$), $I_{\text{av}} = I_{\text{av}}(\vec{b})$.

Example Alice prepares a qubit either in state $\hat{\varrho}_1 = |\uparrow_a\rangle\langle\uparrow_a|$, or in state $\hat{\varrho}_2 = |\downarrow_a\rangle\langle\downarrow_a|$. The probabilities that she does the one or the other are given by $p_1 = p_2 = 0.5$. Thus the apriori state of the qubit is a complete mixture, $\hat{\varrho} = \frac{1}{2}\hat{1}$. Bob's initial level of ignorance (the Shannon entropy of the source) is one bit.

Although Bob knows Alice's choice $\vec{a}$, he measures with orientation $\vec{b}$ (just to see how much he could learn); the two outcomes are labeled + and −.

$$P_{\pm|1} = \frac{1}{2}\left(1 \pm \vec{b}\cdot\vec{a}\right), \tag{14.23}$$

$$P_{\pm 1|2} = \frac{1}{2}\left(1 \mp \vec{b}\cdot\vec{a}\right), \tag{14.24}$$

$$q_\pm = 0.5\,, \tag{14.25}$$

$$Q_{i|\pm} = P_{\pm|i}\,, \quad i = 1, 2\,. \tag{14.26}$$

$$H_\pm = -c^2 \mathrm{ld} c^2 - (1-c^2)\mathrm{ld}(1-c^2)\,, \quad c^2 = \frac{1}{2}\left(1 + \vec{b}\cdot\vec{a}\right) \tag{14.27}$$

$$H_{\text{final}} = H_+ = H_-\,, \tag{14.28}$$

$$I_{\text{av}}(\vec{b}\cdot\vec{a}) = 1 - \left(-c^2\mathrm{ld}c^2 - (1-c^2)\mathrm{ld}(1-c^2)\right)\,. \tag{14.29}$$

Information gain is maximal (1 bit) for $\vec{b} = \pm\vec{a}$: As Bob already knows that Alice is using $\vec{a}$, he better sticks with it. Information gain is minimal (0 bits gained) if Bob choses orientation in the plane orthogonal to $\vec{a}$: nothing can be learned about the spin z-component in a measurement of an orthogonal component.

This example proves that in an ideal world, where there is no disturbing influence from the enviroment, 100-percent reliable communication is possible

using qubits. The only point sender and receiver must be aware of is to chose equal alignment of their Stern-Gerlach devices. The example also indicates that in this particular context, the maximal amount of information which a single qubit can carry is one bit. In the next chapter we shall see that in a different context, which involves entanglement, one can cram two bits in a single qubit.

Example Alice prepares a qubit in the up-state $|\uparrow_i\rangle$ with respect to one out of three possible quantization axis $\vec{a}_i$, $i = 1, 2, 3$, where the $\vec{a}_i$ form a "Mercedes-Stern",

$$\sum_{i=1}^{3} \vec{a}_i = 0\,. \tag{14.30}$$

The a priori state $\hat{\varrho} = \sum_{i=1}^{3} \frac{1}{3}|\uparrow_i\rangle\langle\uparrow_i|$ is a complete mixture,

$$\hat{\varrho} = \frac{1}{2}\hat{1}\,. \tag{14.31}$$

Bob knows the possible directions $\vec{a}_i$, but he does not know which particular direction Alice has chosen. His initial level of ignorance is given by

$$H_{\text{initial}} = \text{ld}3 \tag{14.32}$$

How much could he expect to learn about Alice's choice, and what is his optimal strategy?

The a priori state being a complete mixture, one would naively expect that any bet on either 1, 2, or 3 has a 1/3 chance of winning. But this turns out to be wrong – one can do much better than that!

Within our general framework

$$P_{\pm|i} = \frac{1}{2}\left(1 \pm \vec{b}\cdot\vec{a}_i\right), \qquad i = 1,2,3\,; \tag{14.33}$$

$$q_{\pm} = \frac{1}{2}\,, \tag{14.34}$$

$$Q_{i|\pm} = \frac{2}{3}P_{\pm|i}\,, \tag{14.35}$$

$$H_{\pm} = -\sum_{i=1}^{3}\tfrac{1}{3}\left(1 \pm \vec{b}\cdot\vec{a}_i\right)\mathrm{ld}\left[\tfrac{1}{3}\left(1 \pm \vec{b}\cdot\vec{a}_i\right)\right], \tag{14.36}$$

$$H_{\rm final} = -\sum_{m=\pm}\tfrac{1}{2}\sum_{i=1}^{3}\tfrac{1}{3}\left(1 + m\vec{b}\cdot\vec{a}_i\right)\mathrm{ld}\left[\tfrac{1}{3}\left(1 + m\vec{b}\cdot\vec{a}_i\right)\right]. \tag{14.37}$$

The average information gain, $I_{\rm av} = H_{\rm initial} - H_{\rm final}$, as function of $\vec{b}$ is plotted in Fig. **??**. Information gain is maximal, $I_{\rm av}^{\rm max} = \frac{1}{2}\mathrm{ld}3 - \frac{1}{3}\mathrm{ld}2 = 0.459(1)bits$ at angles $0, 60, 120, 180, 240, 300$.

Chose $\vec{b} = \vec{a}_1$. In case the particle leaves your device through the lower channel, $m = -1$ you know for sure that $|\uparrow_1\rangle$ was not prepared, leaving you with ignorance ld2, and a successrate of $1/2$ in betting on either 2 or 3. However, click $-$ only occurs with a 50-percent probability. In the remaning 50-percent of the cases, the click will be $+$ in which case states $|\uparrow_2\rangle$ and $|\uparrow_3\rangle$ have been prepared with probability $Q_{2|+} = Q_{3|+} = 1/6$, while $|\uparrow_1\rangle$ was prepared with probability $Q_{1|+} = 2/3$. In this case, you should bet on 1 which will be successfull with probability $2/3$. Thus, your overall successrate will be $1/4 + 1/3 = 7/12$ which is much larger than $1/3$, and even larger than $1/2$!

As an excercise you will demonstrate that, using positive operator valued measures, an even higher successrate can be achieved.

14.3 States of composite systems

Definition A state $\hat{\varrho}$ of a bi-partite composite system is called *uncorrelated*, if it can be written in the form

$$\hat{\varrho} = \hat{\varrho}_A \otimes \hat{\varrho}_B \tag{14.38}$$

Uncorrelated mixed states are the most natural generalization of a pure product state.

Definition A state $\hat{\varrho}$ of a bi-partite composite system is called **seperable** if it can be written in the form of a convex sum of uncorrelated states,

$$\hat{\varrho} = \sum_{i=1}^{N} \lambda_i \hat{\varrho}_A^i \otimes \hat{\varrho}_B^i \,. \tag{14.39}$$

A state is called **n**on-seperable if it can *not* be written in that form.

Note, that both seperable and non-seperable states may give rise to correlations between Alice and Bob, but in the case of seperable states these correlations are purely classical. Non-classical correlations are only contained in non-seperable states, which are the natural generalization of entangled states.

Note that the definition of seperability is a trap-door definition: given the decomposition on the right hand side it is easy to compute $\hat{\varrho}$, but given only $\hat{\varrho}$ it is extremely difficult to decide whether any representation in form of a convex sum exists. In fact, only for the case of two qubits and the for case of one qubit and a qtrit (a spin-1 system) is a criterion, which is both necessary and sufficient, explicitely known.

14.4 Entropy and measure of entanglement

Given a pure state $\hat{\varrho} = |\psi\rangle\langle\psi|$ of a bi-partite system. In tracing over Bob's (Alice's) degrees of freedom we obtain a reduced state in Alice's (Bob's) Hilbert space,

$$\hat{\varrho}_A \equiv \mathrm{tr}_B\{\hat{\varrho}\}\,, \qquad \hat{\varrho}_B \equiv \mathrm{tr}_A\{\hat{\varrho}\}\,. \tag{14.40}$$

By virtue of the Schmidt-decomposition (without loss of generality $M = \dim\mathcal{H}_A \geq N = \dim\mathcal{H}_B$)

$$\hat{\varrho}_A = \sum_{\nu=1}^{N} p_\nu |e_\nu\rangle\langle e_\nu|\,, \qquad \hat{\varrho}_B = \sum_{\nu=1}^{N} p_\nu |f_\nu\rangle\langle f_\nu|\,. \tag{14.41}$$

The reduced states are in general mixed states with equal von-Neumann entropies $S(\hat{\varrho}_A) = S(\hat{\varrho}_B) \equiv S_{\mathrm{red}}$,

$$S_{\mathrm{red}} = -\sum_{\nu=1}^{N} p_\nu \ln p_\nu\,. \tag{14.42}$$

A necessary and sufficient criterion for a pure state to be entangled is now given by the following

Theorem

$$|\psi\rangle \qquad \text{is a product state} \tag{14.43}$$
$$\Leftrightarrow \quad \text{it's Schmidt number } r = 1 \tag{14.44}$$
$$\Leftrightarrow \quad \text{the reduced states } \hat{\varrho}_A,\ \hat{\varrho}_B \text{ are pure states} \tag{14.45}$$

A "natural" measure of entanglement, which is measured in *ebit*, is then provided by the reduced von-Neumann entropy

$$E(|\psi\rangle) = S_{\mathrm{red}} \tag{14.46}$$

The Bell state $|\psi^-\rangle$, for example, carries one ebit of entanglement.

Note that the above measure of entanglement is only defined for pure states. For mixed states the situation is much more complicated.

14.5 Non-Separability

A bipartite state, mixed or not, is non-separable if it can not be written as a convex combination (14.39). That is good a definition, but pretty bad a criterion. For a given bipartite state – what would be a simple criterion to decide?

Asher Peres, in a famous PRL [Asher Peres, "Separability Criterion for Density Matrices", Phys. Rev. Lett. 77, 14131415 (1996)], shows that if a state is separable, its partial transpose is necessarily positive. In short $\mathsf{SEP} \to \mathsf{PPT}$. In the case 2×2 and 2×3 this criterion is also sufficient, i.e. $\mathsf{SEP} \leftrightarrow \mathsf{PPT}$, as has been shown by some members of the Horodecki family [Horodecki, Michał; Horodecki, Paweł; Horodecki, Ryszard: "Separability of mixed states: necessary and sufficient condition", Physics Letters **A223**, 18.]

In general, transposition of an operator $\hat{A} = \sum_{ij} A_{ij}|i\rangle\langle j|$ is a map

$$\begin{array}{rcll} \theta : \mathcal{B}(\mathcal{H}) & \to & \mathcal{B}(\mathcal{H}) & \\ \hat{A} & \mapsto & \hat{A}^T & = \sum_{ij} A_{ij} \left(|i\rangle\langle j|\right)^T \\ & & & = \sum_{ij} A_{ij}|j\rangle\langle i| \\ & & & = \sum_{ij} A_{ji}|i\rangle\langle j| \end{array} \tag{14.47}$$

in matrix-notation "mirror at the diagonal". This map preserves the spectrum of $\hat{A}$, i.e. it is a positive map on state space. But it is not completely positive, i.e. for bi-partite system, the map $I \otimes \theta$, where only one of the components is transposed (here: Bob), is not positiv semi-definit.

For bi-partite state $\rho^{A\&B} = \sum_{ij,kl} p_{ij,kl}|ij\rangle\langle kl|$, the "partial transpose" is a map, where only one of the participating parties is transposed, Bob,say

$$\begin{array}{rcccl} \theta^B : \mathcal{B}(\mathcal{H}^{A\&B}) & \rightarrow & \mathcal{B}(\mathcal{H}^{A\&B}) & & \\ \hat{\rho}^{A\&B} & \mapsto & \hat{\rho}^{A\&TB} & = & (I \otimes \theta)(\hat{\rho}^{A\&B}) \\ & & & = & \sum_{ij,kl} p_{il,kj}|ij\rangle\langle kl| \end{array} \tag{14.48}$$

In matrix notation the bi-partite state could be written

$$\rho^{A\&B} = \begin{pmatrix} A_{11} & A_{12} & \cdots & A_{1n} \\ A_{21} & A_{22} & & \\ \vdots & & \ddots & \\ A_{n1} & & & A_{nn} \end{pmatrix} \tag{14.49}$$

Where $n = \dim\mathcal{H}^A$ and each block is a square matrix of dimension $m = \dim\mathcal{H}^B$. Then the partial transpose is

$$\rho^{A\&TB} = \begin{pmatrix} A_{11}^T & A_{12}^T & \cdots & A_{1n}^T \\ A_{21}^T & A_{22}^T & & \\ \vdots & & \ddots & \\ A_{n1}^T & & & A_{nn}^T \end{pmatrix} \tag{14.50}$$

The Peres criterion states, that if $\rho^{A\&B}$ is separable, then $\rho^{A\&TB}$ is a state, i.e. selfadjoint, normalized and positive-semidefinit.

So lets assume, $\rho^{A\&B}$ is seperable, i.e. has the form (14.39). Now, since transposition is positive (maps states to states), the partial transposition of a separable state yields a separable state, i.e. an operator selfadjoint, normalized and positive-semidefinit.

In the reverse, if the partial transpose does not yield a possible state, the state is not separable. It then contains "non-classical" correlation - but to what extend this

can be exploited for fancy purposes depends on the systems under considerations. For a qubit coupled to a qubit, and a qubit coupled to a qutrit (3-dim Hilbertspace), the Horodocki family has proven to be the Peres necessary and sufficient.

Example: Given the 2-qubit family of Werner States

$$\hat{\rho} = p|\Psi^-\rangle\langle\Psi^-| + (1-p)\frac{\mathrm{id}}{4} \tag{14.51}$$

which is nothing but the convex combination of a spin-singlet and the complete mixture. In matrix notation

$$\hat{\rho}^{A\&B} = \frac{1}{4}\begin{pmatrix} 1-p & 0 & 0 & 0 \\ 0 & p+1 & -2p & 0 \\ 0 & -2p & p+1 & 0 \\ 0 & 0 & 0 & 1-p \end{pmatrix} \tag{14.52}$$

The partial transopose

$$\hat{\rho}^{A\&TB} = \frac{1}{4}\begin{pmatrix} 1-p & 0 & 0 & -2p \\ 0 & p+1 & 0 & 0 \\ 0 & 0 & p+1 & 0 \\ -2p & 0 & 0 & 1-p \end{pmatrix} \tag{14.53}$$

has eigenvalues $\lambda_1 = \frac{1}{4}(1-3p)$, $\lambda_{2,3,4} = \frac{1}{4}(1+p)$, thus the state is non-separable for $\frac{1}{3} < p \leq 1$.

[To be completed]

