9 Quantum Complexity Theory I

Just as the theory of computability had its foundations in the Church-Turing thesis, computational complexity theory rests upon a modern strengthening of this thesis, which asserts that any "reasonable" model of computation can be efficiently simulated on a probabilistic Turing machine (by "efficient" we mean here a runtime that is bounded by a polynomial in the runtime of the simulated machine). For example, computers that can operate on arbitrary length words in unit time, or that can exactly compute real numbers with infinite precision are unreasonable models, since it seems clear that they cannot be pysically implemented. It had been argued that the Turing machine model is the inevitable choice once we assume that we can implement only finite precision computational primitives. Given the widespread belief that NP $\neq$ BPP, this would seem to put a wide range of important computational problems (the NP-hard problems) well beyond the capability of computers.

However, the Turing machine is an inadequate model for all physically realizable computing devices for a fundamental reason: the Turing machine is based on a classical physics model of the universe, whereas current physical theory asserts that the universe is quantum physical. Can we get inherently new kinds of (discrete) computing devices based on quantum physics? The first indication that such a device might potentially be more powerful than a probabilistic Turing machine appeared in a paper by Feynman about two decades ago. In that paper, Feynman pointed out a very curious problem: it appears to be impossible to simulate a general quantum physical system on a probabilistic Turing machine without an exponential slowdown. The difficulty with the simulation has nothing to do with the problem of simulating a continuous system with a discrete one – we may assume that the quantum physical system to be simulated is discrete, some kind of quantum cellular automaton. It has rather to do with a phenomenon in quantum physics that allows different outcomes of a quantum physical effect to interfere with each other in a strange way. In view of Feynman's observation, we must re-examine the foundations of computational complexity theory and the complexity-theoretic form of the Church-Turing thesis, and study the computational power of computing devices based on quantum physics.

A first precise model of a quantum physical computer, called the quantum Turing machine, was formulated by Deutsch. This model may be thought of as a quantum physical analogue of a probabilistic Turing machine: it has an infinite tape and a finite state control, and the actions of the machine are local and completely specified by this finite state control. Furthermore, on a given input a quantum Turing machine produces a random outcome according to a probability distribution. However, instead of a probabilistic Turing machine, the quantum Turing machine can generate probability distributions in which the amplitudes of certain configurations can have complex instead of just positive real numbers.

9.1 Introduction to quantum physics

Before defining a quantum computing model, some basic notions of quantum mechanics must be introduced. As done by Berthiaume, we will introduce the basic rules of quantum mechanics through a series of principles. Young's celebrated two-slit experiment will serve as a background to illustrate these principles.

In Young's experiment (Figure 1), light coming out of a hole in the left wall must go through two small holes in the center wall. A detector on the right wall measures the light intensity at

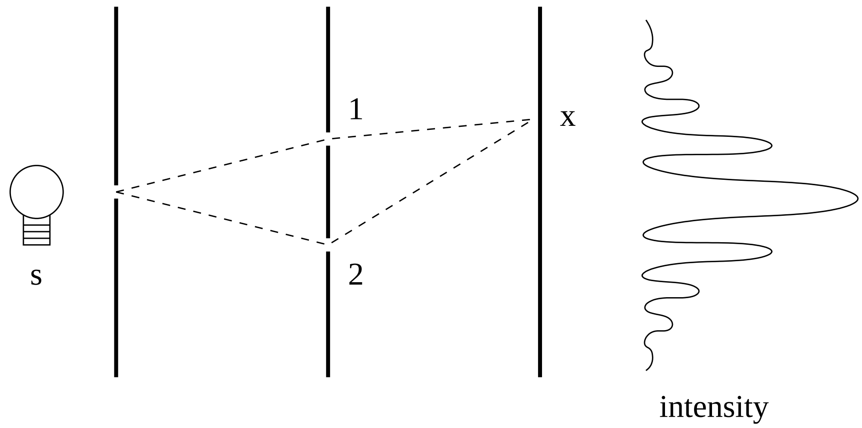

Figure 1: Young's two-slit experiment.

different positions along the length of the wall. If only one hole is open, the intensity reaches its maximum at a position directly in line with that hole and the source s. As the detector moves away from that position, the intensity slowly fades and eventually vanishes. When *both* holes are open, the intensity pattern is *not* the sum of the two one-hole intensities, as one would expect, but an alternation of bright and dark fringes. This effect is caused by the *interference* of the light coming out from both holes. Surprisingly, the interference persists even when the source s is dim enough to send only one photon at a time: if many runs are made and a photon count is kept for various positions, the same pattern of bright and dark fringes appears. Each photon seems to interfere with itself.

The self-interference appearing in Young's experiment is just one example illustrating that classical intuition cannot be applied to quantum systems. The purpose of this subsection is not to explain *why* such strange behavior appears at the quantum level but merely to state the rules for these behaviors. Following Feynman's example, these rules are presented as *principles* of quantum mechanics.

For a given experiment, an *event* is a set of initial and final conditions. For example, an event in Young's experiment is "a photon leaves the source s and arrives in the detector at position x". The goal of quantum mechanics is to predict whether an event can happen or not. The first principle of quantum mechanics defines the probability of an event actually happening.

First Principle: The probability p of an event is given by the square norm of a complex number α (called a *probability amplitude* or simply *amplitude*).

$$p = ||\alpha||^2 = \alpha_r^2 + \alpha_i^2$$

The probability amplitude of an event will be noted as follows:

$$\langle \text{final condition} \mid \text{initial condition} \rangle$$

For example, the above event can be written as

$$\langle \text{a photon is detected at position } x \mid \text{a photon leaves } s\rangle$$

or more succinctly $\langle x|s\rangle$. The bracket notation, due to Dirac, is reminiscent of conditional probabilities and can be read as: $\langle x|s\rangle$ is the amplitude of detecting a photon at position x given that a photon left the source s. In fact, amplitudes can be treated in the same way as probabilities (with the difference that the sum of their square norms has to be equal to 1). Consider again Young's experiment. If a photon leaves the source and arrives at the detector, it must do so by going through the holes in the wall. That is to say, the event $\langle x|s\rangle$ can be broken down in two sequential events: the photon first leaves the source s and arrives at the middle wall, then the photon comes out of this middle wall and arrives at the detector. But to go through the middle wall, the photon has two options: either through hole 1 or hole 2. The following two principles indicate that the laws for addition and multiplication of probabilities also apply to amplitudes.

Second Principle: If an event can be divided into two sequential subevents, the amplitude of the event is the product of the amplitudes for each of the subevents.

Third Principle: If an event can occur in several alternative ways, then the amplitude of the event is the sum of the amplitudes for each alternative taken separately.

From these two principles,

$$\begin{aligned} \langle x|s\rangle &= \langle x|\text{wall}\rangle\langle \text{wall}|s\rangle && (1) \\ &= \langle x|1\rangle\langle 1|s\rangle + \langle x|2\rangle\langle 2|s\rangle && (2) \end{aligned}$$

where $\langle x|i\rangle$ is the amplitude of a photon arriving at x given it came out of hole i and $\langle i|s\rangle$ is the amplitude of a photon entering hole i given that it left the source s. Equation 2 implicitly considers it impossible to enter hole 1 and leave hole 2 ($\langle 2|1\rangle$) or vice versa ($\langle 1|2\rangle$). That is, we assumed that $\langle x|1\rangle\langle 1|2\rangle\langle 2|s\rangle$ and $\langle x|2\rangle\langle 2|1\rangle\langle 1|s\rangle$ are equal to 0. One can verify experimentally that whenever a photon is detected entering one hole, it is never detected coming out of the opposite hole. Equations 3 and 4 express this situation in terms of amplitudes of events.

$$\langle 1|1\rangle = \langle 2|2\rangle = 1 \quad (3)$$

$$\langle 1|2\rangle = \langle 2|1\rangle = 0 \quad (4)$$

No matter where the detector is positioned, the amplitude $\langle x|s\rangle$ is completely determined by the amplitudes to transit to and from the two holes. For this reason, the holes are natural elements for expressing events. This leads to the following definition.

Definition 9.1 *The set $B = \{i \mid i \text{ is the label of some condition}\}$ is a set of* basis states *if for all $i, j \in B$*

$$\langle i|j\rangle = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{otherwise} \end{cases}$$

and for any initial condition Y and final condition X, we have

$$\langle X|Y\rangle = \sum_{i\in B}\langle X|i\rangle\langle i|Y\rangle \ ,$$

that is, the set B is complete.

Fourth Principle: Any event can be described in terms of a set of basis states by giving the transition amplitudes to and from those basis states.

Young's experiment seems to have only one possible set of basis states (namely $B = \{1, 2\}$) but actually, for any experiment, there is an infinite number of sets of basis states. Some appear naturally, such as the two holes in this experiment, and others less so. The notion of basis states can be better understood through an analogy with the three dimensional real vector space V and the dot product in that space. Let the three vectors

$$\vec{e}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \ \vec{e}_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \text{ and } \vec{e}_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$$

be the canonical basis of V and let

$$\vec{A} = \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} \text{ and } \vec{B} = \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix}$$

be two vectors in V. The following is a non-standard yet valid expression for the dot product of $\vec{A}$ and $\vec{B}$:

$$\begin{aligned} \vec{A}\cdot\vec{B} &= a_1b_1 + a_2b_2 + a_3b_3 \\ &= (\vec{A}\cdot\vec{e}_1)(\vec{e}_1\cdot\vec{B}) + (\vec{A}\cdot\vec{e}_2)(\vec{e}_2\cdot\vec{B}) + (\vec{A}\cdot\vec{e}_3)(\vec{e}_3\cdot\vec{B}) \\ &= \sum_{i=1}^{3}(\vec{A}\cdot\vec{e}_i)(\vec{e}_i\cdot\vec{B}) \ . \end{aligned}$$

This last line is similar to the equation for $\langle X|Y\rangle$ in Definition 9.1. The vectors $\vec{A}$ and $\vec{B}$ correspond to the two conditions X and Y, and the set of canonical basis vectors corresponds to the set of basis states. In this sense, $\langle X|Y\rangle \equiv \langle X|\cdot|Y\rangle$. The analogy can be pushed further still. Consider the following:

$$\begin{aligned} \vec{B} &= b_1\vec{e}_1 + b_2\vec{e}_2 + b_3\vec{e}_3 \\ &= \sum_{i=1}^{3}\vec{e}_i b_i \\ &= \sum_{i=1}^{3}\vec{e}_i(\vec{e}_i\cdot\vec{B}) \ . \end{aligned}$$

This simply is the previous equation for $\vec{A} \cdot \vec{B}$ where we abstracted $\vec{A}$. In other words, $\vec{B}$ is the vector sum of its components along each of the basis vectors $\vec{e}_1$, $\vec{e}_2$, and $\vec{e}_3$. Similarly, the initial states Y of $\langle X|Y\rangle$ can be expressed in terms of the set of basis states by abstracting X.

$$|Y\rangle = \sum_{i \in B} |i\rangle\langle i|Y\rangle \ . \tag{5}$$

(Note that $\langle\cdot|\cdot\rangle$ is just a scalar number, so $|i\rangle\langle i|Y\rangle = \langle i|Y\rangle|i\rangle$, but the second form is preferred.) The left and right halves of $\langle\cdot|\cdot\rangle$ are named *bra* and *ket* respectively. Equation 5 defines the state of an initial condition Y as a function of B, the set of basis states. Similarly,

$$\langle X| = \sum_{j \in B} \langle X|j\rangle\langle j|$$

defines the state of a final condition X as function of B. The $\langle X|$ and $|Y\rangle$ could also be written using another set of basis states just as B can be written in a basis other than the canonical one. For a basis set B, the state $|Y\rangle$ is said to be in *quantum superposition* of the basis states in B if more than one $\langle i|Y\rangle$ is non-zero.

By the first principle, the square norm of an amplitude gives a probability of the corresponding event. By the fourth principle, any event can be expressed as a function of the amplitude corresponding to each basis state. Since the probabilities must add up to 1, there should be some constraint on the amplitudes:

Fifth Principle: For any set of basis states B and for any initial condition Y:

$$\sum_{i \in B} ||\langle i|Y\rangle||^2 = 1 \ .$$

This follows from the definition of a set of basis states which has to be complete (all possibilities are accounted for) and orthogonal ($\langle i|j\rangle = 0$ if $i \neq j$, and 1 otherwise).

One can also show the following theorem (recall that α^* is the complex conjugate of α).

Theorem 9.2 *For any condition A and B, $\langle B|A\rangle = \langle A|B\rangle^*$.*

From the theorem we derive that if $|\varphi\rangle = \sum_{i \in B} \alpha_i |i\rangle$ then

$$\langle\varphi| = \sum_{i \in B} \alpha_i^* \langle i| \ .$$

All these principles can be put together in the following situation: consider a system whose initial condition is expressed by the state vector

$$|Y\rangle = \sum_{i \in B} \beta_i |i\rangle$$

where $\beta_i = \langle i|Y\rangle$, the amplitude of the system under consideration in each of the basis states i. Now consider a general final condition

$$\langle X| = \sum_{j \in B} \alpha_j \langle j|$$

where $\alpha_j = \langle X|j\rangle$. What is the amplitude of finding the system in condition X given that it was initially prepared in Y?

$$\langle X|Y\rangle = \left(\sum_{j\in B} \alpha_j \langle j|\right) \left(\sum_{i\in B} \beta_i |i\rangle\right) = \sum_{i\in B} \alpha_i \beta_i$$

by Definition 9.1.

Combining the properties of basis states we specified so far, we can view the ket $|\varphi\rangle = \sum_{i\in B} \alpha_i |i\rangle$ as a state vector in a complex vector space (the so-called *Hilbert space*) spanned by the basis vectors associated with the basis states in B. The Dirac operator $\langle X|Y\rangle$ represents in this vector space an inner product $H \times H \rightarrow \mathbb{C}$ with the properties that

- $\langle v|w\rangle = \langle w|v\rangle^*$,
- $\langle v|\alpha w\rangle = \alpha\langle v|w\rangle$ for all $\alpha \in \mathbb{C}$, and
- $\langle v + w|z\rangle = \langle v|z\rangle + \langle w|z\rangle$.

For a given vector $|v\rangle$ in H, $\langle v|$ represents its covector, which is the conjugate transpose of v. Using the inner product, the square norm of a vector v can be defined as $||v||^2 = \langle v|v\rangle$. From this definition and Principle 5 it follows that every initial or final condition must be a unit-length vector in H.

Quantum mechanics imposes very strict rules as to how to extract information out of a quantum state. This is done through a mathematical construct called an *observable*. Let $|\varphi\rangle$ be the state of a quantum system. We need to define the effect of observing a certain property of φ.

Definition 9.3 *Let H be the Hilbert space used to represent the state vectors of a quantum system. An* observable $\mathcal{O}$ *is a set of subspaces $E_1, E_2, \ldots, E_k \subseteq H$ such that these subspaces completely partition H, i.e.*

$$E_1 \times E_2 \times \ldots \times E_k = H$$

and

$$\forall i, j \in \{1, \ldots, k\},\ i \neq j :\ E_i \perp E_j\ .$$

An observable is the mathematical representation of observing a quantum state.

Sixth Principle: Let $|\varphi\rangle$ be the state vector in a state space H and let $\mathcal{O} = \{E_1, E_2, \ldots, E_k\}$ be an observable. Since $\mathcal{O}$ partitions H, $|\varphi\rangle$ can be expressed as a linear superposition of its components along each of the E_i's:

$$|\varphi\rangle = \sum_{i=1}^{k} \alpha_i |\varphi_{E_i}\rangle \ .$$

where $|\varphi_{E_i}\rangle$ is the part of $|\varphi\rangle$ in E_i. Observing the state $|\varphi\rangle$ will cause the following:

1. One of the E_i will be selected with probability $||\alpha_i||^2$.
2. The state $|\varphi\rangle$ will collapse to $|\varphi_{E_i}\rangle$ (with renormalized amplitudes).
3. The only information given by $\mathcal{O}$ is which subspace i was selected. All information not in $|\varphi_{E_i}\rangle$ is lost.

For example, if the set of basis states represents the set of all possible observable events, then observing $|\varphi\rangle = \sum_{i\in B} \alpha_i |i\rangle$ will cause it collapse to some $i \in B$. Since

$$\langle i|\varphi\rangle = \langle i| \sum_{j\in B} \alpha_j |j\rangle = \alpha_i$$

the probability for this will be equal to $||\alpha_i||^2$. Any observable, such as direction, position, spin, etc. is in principle allowed for observing a quantum state. Whether the physical apparatus that corresponds to that specific observable is easy to build is a different matter entirely.

So far, we have only dealt with elementary computations in quantum mechanics and the effect of observing a quantum physical system. Next we will deal with the question of how a quantum system can evolve. That is, we are interested in the question "what is the amplitude for the final condition X given that the initial condition Y went through apparatus A?". In the bracket notation, this can be written as

$$\langle X| \cdot A \cdot |Y\rangle \quad \text{or simply} \quad \langle X|A|Y\rangle \ .$$

In order to specify what A can do, we introduce the notion of unitary operators. A unitary operator $U : H \to H$ is a linear mapping whose conjugate transpose is its inverse: $U^\dagger = U^{-1}$. Instead of a mapping, we can also view U as a matrix with the property that $UU^\dagger = U^\dagger U = I$ (where I represents the identity matrix). The important property of unitary operators is that they are norm preserving, that is,

$$||Uv||^2 = \langle v|U^\dagger U|v\rangle = \langle v|v\rangle = ||v||^2 \ .$$

Hence, they preserve the condition that state vectors are of unit length (and therefore a valid representation of a quantum mechanical system). As a consequence, we obtain the following principle.

Seventh Principle: State vectors are transformed by unitary operators.

In principle, any unitary transformation on a quantum state is allowed, but constructing a physical device corresponding to any given matrix U might pose some technical problems.

9.2 The quantum Turing machine

We are now able to define a quantum Turing machine.

Definition 9.4 *A* quantum Turing machine *(QTM) is denoted by*

$$M = (Q, \Sigma, \Gamma, \delta, q_0, B, F)$$

where $Q, \Sigma, \Gamma, q_0, B, F$ are defined as for a probabilistic Turing machine. However, the transition function δ is now a mapping of the form

$$\delta : Q \times \Gamma \times Q \times \Gamma \times \{L, N, R\} \to \mathbb{C}$$

where $\delta(p, x, q, y, d)$ gives the amplitude with which the machine in state p reading a will write b, enter state q, and move in direction d. Furthermore, δ has to be well-formed, *that is, it always preserves the square norm.*

If we choose the set B of basis states to consist of all possible configurations of M, then δ specifies a linear mapping in the infinite dimensional Hilbert space of superpositions of configurations. This linear mapping can be represented as a *transition matrix* $A_\delta = (\alpha_{c_1,c_2})_{c_1,c_2 \in B}$ in which $\alpha_{c_1,c_2} \in \mathbb{C}$ specifies the amplitude of going from configuration c_1 to c_2. Recall that a matrix M is unitary if $M^\dagger M = MM^\dagger = I$. It is not hard to see that a matrix M preserves the square norm if and only if $M^\dagger M = I$, and that M is finite-dimensional this is in turn equivalent to M being unitary. However, for infinite dimensional matrices, $M^\dagger M = I$ is not equivalent to $MM^\dagger = I$. Nevertheless, the transition functions of quantum Turing machines have a special structure, and in fact these two conditions are equivalent for A_δ.

Theorem 9.5 *A quantum Turing machine is well-formed if and only if its transition matrix is unitary.*

Using this theorem, one can obtain a local characterization of the well-formedness constraint.

Theorem 9.6 *If a transition function δ is well-formed, then it satisfies the following:*

- *The amplitudes leaving any state-symbol pair have unit total squared magnitude:*

  $$\forall p \in Q \; \forall x \in \Gamma : \quad \sum_{q,y,d} ||\delta(p, x, q, y, d)||^2 = 1 \; .$$

  (This ensures the transitions for every (p, x) represent a probability distribution.)

- *The superpositions of written character, new state, and direction leaving any two different state-symbol pairs are orthogonal:*

  $$\forall (p, x) \neq (p', x') : \quad \sum_{q,y,d} \delta(p, x, q, y, d) \cdot \delta^*(p', x', q, y, d) = 0 \; .$$

  (This ensures that the transitions for different pairs (p, x) do not interfere with each other.)

The proof of the theorem is left as an exercise. Another way of viewing the two conditions above is that the first ensures that energy is preserved and the second ensures that energy is not moved from one state-symbol pair to the successors of another state-symbol pair.

Since δ must be well-defined, every transition of a quantum Turing machine is reversible (simply use $A_\delta^\dagger$ after using A_δ, and we would be back at the same state before A_δ). Note that classical computations are usually not reversible. For instance, the simple command of setting a variable back to 0 may have the effect that several configurations fall together, and therefore it is not possible to say afterwards from which configuration the Turing machine originally came. We formally define reversibility for the deterministic Turing machine as follows.

Definition 9.7 *A Turing machine M is* reversible *if reversing its arrows gives a deterministic mapping of configurations T that undoes the computation of M: for any pair of configurations c_1 and c_2, $c_1 \stackrel{M}{\to} c_2$ if and only if $c_2 \stackrel{T}{\to} c_1$.*

Note that the arrow reversal of a reversible Turing machine M is not necessarily a Turing machine, since the reverse of a transition that both writes and moves must write in a non-local cell. Hence, we only spoke about a mapping T instead of a Turing machine T in the definition.

Now, suppose we have a Turing machine M which computes a function f, i.e. it outputs $f(x)$ on input x. If f is not injective, then there can clearly be no reversible Turing machine that computes f (i.e. has $f(x)$ on its output tape and all other work tapes erased). However, Lecerf and independently Bennett showed that for any deterministic Turing machine M computing a function f there is a reversible Turing machine M' such that M' computes $x\,f(x)$ on input x. The runtime of M' is within a constant factor of the runtime of M.

A reversibility condition can be also formulated for probabilistic and quantum Turing machines: the transition function δ' resulting from δ by reversing the arrows and conjugating the amplitudes of δ undoes the computation of δ. That is, for any superpositions v_1 and v_2, $v_1 \stackrel{\delta}{\to} v_2$ if and only if $v_2 \stackrel{\delta'}{\to} v_1$. Since this is a generalization of reversibility for deterministic Turing machines and quantum Turing machines fulfill this reversibility condition (unless observations are made), it is not surprising that Benioff and Deutsch were able to show that quantum Turing machines can efficiently simulate classical reversible Turing machines.

Theorem 9.8 *Any function f that can be computed efficiently by a classical (deterministic or probabilistic) Turing machine can also be computed efficiently by a quantum Turing machine.*

Theorem 9.8 implies that quantum computing is at least as powerful as classical computing. But does there exist a quantum computer that can execute arbtrary quantum programs in an efficient way? Or in other words: is there a universal quantum Turing machine that can simulate any specific quantum Turing machine in an efficient way? The answer to this was given by Bernstein and Vazirani. They showed the following theorem.

Theorem 9.9 *There exists a universal quantum Turing machine U which when given a description of any quantum Turing machine M and any $t > 0$, simulates M for t steps with accuracy ϵ and with slowdown polynomial in t and $1/\epsilon$.*

Hence, in principle it is possible to construct a quantum computer that can execute arbitrary quantum programs. However, how susceptible would such a machine be to errors in the computation? Bernstein and Vazirani showed in their proof of Theorem 9.9 that $O(t)$ bits of accuracy are sufficient to simulate t steps of computation. However, Lipton pointed out that in order to claim that a model of computation is truly discrete, it is necessary that any constants in the specification should only be required to be accurate up to $O(\log t)$ bits of accuracy. This is because a constant that is accurate up to k bits can lie in one of 2^k distinct ranges, and the task of verifying that the machine meets the specifications seems to require effort that is proportional to 2^k. The following theorem shows that we need only $O(\log t)$ bits of precision in the specification of each of the complex numbers defining a single-step transition function of a quantum Turing machine. The theorem is due to Bennett, Bernstein, Brassard, and Vazirani. It uses the notation that the superposition of a quantum Turing machine M at time t is denoted by $|\phi_t^M\rangle$.

Theorem 9.10 *For any two quantum Turing machines M and $\hat{M}$ with the same state set and tape alphabet and the property that for every transition amplitude λ of M, the corresponding transition amplitude $\hat{\lambda}$ of $\hat{M}$ satisfies $||\hat{\lambda}-\lambda||^2 \leq \epsilon$, it holds that for all inputs, $|||\phi_t^M\rangle - |\phi_t^{\hat{M}}\rangle||^2 = O(\epsilon \cdot t)$ for any time step t.*

Proof. Let such a quantum Turing machine $\hat{M}$ be given for some constant c. Let U be the unitary transformation specified by M, and $\hat{U}$ be the unitary transformation specified by $\hat{M}$. We will first prove that for any superposition $|\phi\rangle$, $U|\phi\rangle$ is close to $\hat{U}|\phi\rangle$. Let

$$|\phi\rangle = \sum_i \alpha_i |c_i\rangle$$

be some superposition over configurations c_i. Then,

$$U|\phi\rangle - \hat{U}|\phi\rangle = \sum_i \left(\sum_{j \in P(i)} (\lambda_{j,i} - \hat{\lambda}_{j,i})\alpha_j \right) |c_i\rangle$$

where $P(i)$ is the set of j such that configuration $|c_j\rangle$ can lead to $|c_i\rangle$ in a single step of M or $\hat{M}$. Note that every j can occur in at most c different sets $P(i)$, where c is some constant depending only on the number of states and tape symbols used in the definition of M. So we have,

$$\begin{aligned}
||U|\phi\rangle - \hat{U}|\phi\rangle||^2 &= \sum_i || \sum_{j \in P(i)} (\lambda_{j,i} - \hat{\lambda}_{j,i})\alpha_j ||^2 \;\leq\; \sum_i \sum_{j \in P(i)} ||(\lambda_{j,i} - \hat{\lambda}_{j,i})\alpha_j||^2 \\
&\leq \sum_i \sum_{j \in P(i)} ||\lambda_{j,i} - \hat{\lambda}_{j,i}||^2 ||\alpha_j||^2 \;\leq\; \sum_i \sum_{j \in P(i)} \epsilon ||\alpha_j||^2 \\
&= \epsilon \sum_i \sum_{j \in S(i)} ||\alpha_j||^2 \;\leq\; \epsilon \cdot c \sum_j ||\alpha_j||^2 \\
&\leq \epsilon \cdot c\, .
\end{aligned}$$

We will use this to prove that M and $\hat{M}$ accept the same n-bit strings.

Let x be an arbitrary n-bit string. Then we will show that the superpositions of M and $\hat{M}$ at time t, $|\phi_t^M\rangle$ and $|\phi_t^{\hat{M}}\rangle$, are close together.

$$\begin{aligned}
\hat{U}^t|x\rangle &= \hat{U}^{t-1}(U|x\rangle + |\Delta_1\rangle) \\
&= \hat{U}^{t-1}(U|x\rangle) + \hat{U}^{t-1}|\Delta_1\rangle \\
&= \hat{U}^{t-2}(U^2|x\rangle) + \hat{U}^{t-2}|\Delta_2\rangle + \hat{U}^{t-1}|\Delta_1\rangle \\
&= U^t|x\rangle + \sum_{i=1}^{t} \hat{U}^{t-i}|\Delta_i\rangle
\end{aligned}$$

where each $|\Delta_i\rangle$ has the property that $||\Delta_i||^2 \leq \epsilon \cdot c$. Therefore,

$$\begin{aligned}
|||\phi_t^M\rangle - |\phi_t^{\hat{M}}\rangle||^2 &= ||U^t|x\rangle - \hat{U}^t|x\rangle||^2 \;\leq\; ||\sum_{i=1}^{t} \hat{U}^{t-i}|\Delta_i\rangle||^2 \\
&\leq \sum_{i=1}^{t} ||\hat{U}^{t-i}|\Delta_i\rangle||^2 \;=\; \sum_{i=1}^{t} ||\Delta_i||^2 \\
&\leq t \cdot \epsilon \cdot c\,.
\end{aligned}$$

□

The theorem shows that in order to obtain a sufficiently good error bound after t steps, the error in each step has to be of order $O(1/t)$. Using error correcting codes, Shor showed that it is even possible to tolerate an error of $O(1/\log^c t)$ for some constant c in each step to obtain a sufficiently small overall error after t steps.

9.3 Quantum complexity classes

Similar to P we define the time complexity class QP to be the set of all languages L for which there exists a quantum Turing machine M with the property that

- for all $x \in L$, Pr[M accepts x] $= 1$ and
- for all $x \notin L$, Pr[M accepts x] $= 0$.

Furthermore, the class BQP is defined as the set of all languages L for which there exists a quantum Turing machine M with the property that

- for all $x \in L$, Pr[M accepts x] $\geq 2/3$ and
- for all $x \notin L$, Pr[M accepts x] $\leq 1/3$.

From the previous subsection we know that

$$\mathrm{P} \subseteq \mathrm{QP} \quad \text{and} \quad \mathrm{BPP} \subseteq \mathrm{BQP}\,.$$

Whether or not P = QP or BPP = BQP is still wide open. However, there is some evidence that the quantum complexity classes are more powerful than the classical counterparts. We will investigate this in more detail in the next section. An upper bound for the complexity of QP and BQP is provided by the following theorem. Its proof is left as an exercise.

Theorem 9.11 BQP $\subseteq$ PSPACE.

Only slighty better upper bounds are known for the complexity of BQP to date.

9.4 References

- C.H. Bennett. Local reversibility of computations. *IBM Journal of Research and Development* 17:525–532, 1973.

- C.H. Bennett, E. Bernstein, G. Brassard, and U. Vazirani. Strengths and weaknesses of quantum computing. *SIAM Journal on Computing* 26(5):1510–1523, 1997.

- E. Bernstein and U. Vazirani. Quantum complexity theory. In *Proc. of the 25th ACM Symp. on Theory of Computing*, pp. 11–20, 1993.

- A. Berthiaume. Quantum Computation. In *Complexity Theory Retrospective II*, Springer Verlag, 1996

- D. Deutsch. Quantum theory, the Church-Turing principle and the universal quantum computer. *Proc. of the Royal Society*, London, A400:97–117, 1985.

- R.P. Feynman. Simulating physics with computers. *International Journal of Theoretical Physics* 21(6/7):467–488, 1982.

- Y. Lecerf. Machines de Turing réversibles. Récursive insolubilité en $n \in \mathbb{N}$ de l'équation $u = \Theta^n$ où Θ est un isomorphisme de codes. In *Comptes rendus de l'Académie française ds sciences* 257:2597–2600, 1963.

- P. Shor. Fault-tolerant quantum computation. In *Proc. of the 37th IEEE Symp. on Foundations of Computer Science*, pp. 56–65, 1996.