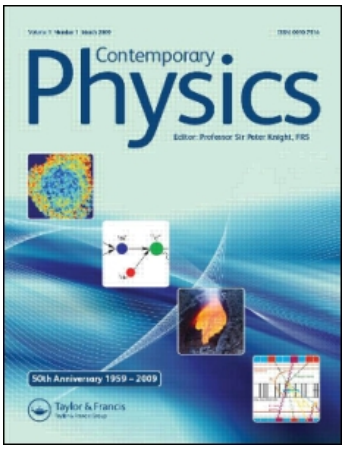

Contemporary Physics
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713394025

Quantum machines

Mark Hillery [a]; Vladimír Bužek [bc]
[a] Department of Physics, Hunter College of CUNY, New York, NY, USA [b] Research Center for Quantum Information, Institute of Physics, Slovak Academy of Sciences, Bratislava, Slovakia [c] Quniverse, Bratislava, Slovakia

Online Publication Date: 01 September 2009

To cite this Article Hillery, Mark and Bužek, Vladimír(2009)'Quantum machines',Contemporary Physics,50:5,575 — 586
To link to this Article: DOI: 10.1080/00107510902924786
URL: http://dx.doi.org/10.1080/00107510902924786

Contemporary Physics
Vol. 50, No. 5, September–October 2009, 575–586

Taylor & Francis
Taylor & Francis Group

Quantum machines

Mark Hillery[a]* and Vladimír Bužek[b,c†]

[a]*Department of Physics, Hunter College of CUNY, 695 Park Avenue, New York, NY 10061, USA;* [b]*Research Center for Quantum Information, Institute of Physics, Slovak Academy of Sciences, Dúbravská cesta 9, 845 11, Bratislava, Slovakia;* [c]*Quniverse, Líščie údolie 116, 841 04, Bratislava, Slovakia*

(*Received 19 December 2008; final version received 26 March 2009*)

We discuss quantum information processing machines. We start with single purpose machines that either redistribute quantum information or identify quantum states. We then move on to machines that can perform a number of functions, with the function they perform being determined by a program, which is itself a quantum state. Examples of both deterministic and probabilistic programmable machines are given, and we conclude with a discussion of the utility of quantum programs.

Keywords: quantum information theory; quantum control; programmable quantum processors; quantum cloners

1. Introduction

Quantum information is information stored in a quantum mechanical system [1]. The systems themselves are either collections of qubits, two-dimensional systems, or qudits, D-dimensional systems. For example, a qubit exists in a two-dimensional space spanned by two orthonormal states, which are denoted by $|0\rangle$ and $|1\rangle$. While a classical bit can be either 0 or 1, a qubit can be in any superposition of $|0\rangle$ and $|1\rangle$, i.e. $|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$, where α and β are complex numbers satisfying $|\alpha|^2 + |\beta|^2 = 1$. This fact leads to large differences in the properties of classical and quantum information. Once one has information in the form of qubits of qudits, one would like to do something with it. Ultimately, in order to extract the information stored in the system, it will be necessary to measure it, but before doing so, it is usually useful to perform some operations on the system. For example, one might want to (approximately) copy it, compare it with known quantum states or with information stored in other quantum systems, or perform an operation on it. The devices that accomplish these tasks are quantum machines. They can either be single-purpose or programmable and able to perform many tasks.

In the case of programmable systems, the program can be either quantum or classical. Examples of classical programs are a sequence of laser pulses to put a molecule into a particular quantum state or a sequence of radio frequency pulses to control the dynamics of a system of nuclear spins [2]. Here, we want to consider programs that are themselves states of quantum systems, i.e. we want to perform quantum programming. Quantum programs have some properties that classical ones do not. For example, they can exist in superposition states, which means that a quantum processor can perform several programs simultaneously. In addition, quantum programs are necessary when the information on which the program is based is, in fact, quantum information. We shall provide examples of each of these situations.

As we shall see, in a number of cases, it is not possible to perform the desired task perfectly. For example, the no-cloning theorem prohibits the perfect copying of quantum information [3]. We are then faced with deciding how to perform the task as best as we can. This usually amounts to adopting one of two possible strategies. The first is to produce an output state that is as close as possible to the ideal output, that is we approximately perform the desired task. The second is to perform the task with some probability in such a way that we know whether the proper task has been performed or not. In that case, the task has been performed probabilistically.

We shall begin by discussing single-purpose machines, in particular cloners and machines that perform the discrimination of quantum states. Next, we shall move on to a general discussion of programmable machines and outline the deterministic

*Corresponding author. Email: mhillery@hunter.cuny.edu
†Email: buzek@savba.sk

ISSN 0010-7514 print/ISSN 1366-5812 online

DOI: 10.1080/00107510902924786
http://www.informaworld.com

and probabilistic approaches to them. We will present a no-go theorem that shows that a deterministic programmable machine cannot be universal. We then consider the problem of implementing a one-parameter unitary group on a programmable machine and discuss the approximate and probabilistic strategies. We also show how, in the probabilistic case, to increase the probability of success by increasing the size of the program space. Finally, we make a case for why quantum programs are useful.

2. Singe-purpose machines

2.1. Cloners

As mentioned in the Introduction, the no-cloning theorem prohibits the perfect copying of quantum information. The theorem states that if we have a quantum system in the state $|\psi\rangle$, we cannot build a machine whose output will be $|\psi\rangle|\psi\rangle$, for all $|\psi\rangle$. The proof relies simply on the linearity of quantum mechanics [3]. If the machine clones each vector of a basis for the input space, then its action is completely determined, and it is incompatible with the transformation $|\psi\rangle \rightarrow |\psi\rangle|\psi\rangle$.

An approximate cloner for qubits can be constructed from four Controlled-NOT gates and three qubits (see Figure 1) [4–6]. A Controlled-NOT gate is a two-qubit gate. The first qubit is the control qubit, and the second is the target qubit. If the state of the control qubit is $|0\rangle$, then nothing happens to the states of either the control or the target qubits. However, if the state of the control qubit is $|1\rangle$, then the state of the control qubit is again unchanged, but the state of the target qubit is flipped, i.e. if it was $|0\rangle$ it becomes $|1\rangle$, and vice versa. Returning now to the cloner, we have that the first qubit, in the state $|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$, where $\{|0\rangle, |1\rangle\}$ is the orthonormal qubit basis, is the one to be copied, and the second qubit will become the copy.

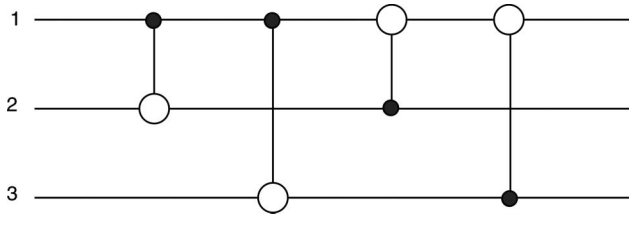

Figure 1. A circuit for an approximate quantum cloner with three qubits and four Controlled-NOT gates. The qubit to be copied goes into input 1, and the copies come out in outputs 1 and 2. The anti-clone comes out of output 3. In the symbol for the Controlled-NOT gate, the filled circles indicate the control qubit, and the open circles indicate the target qubit.

In order to see how this works, define the two two-qubit states

$$|\Xi_{00}\rangle = \frac{1}{2^{1/2}}(|0\rangle|0\rangle + |1\rangle|1\rangle);$$
$$|\Xi_{0x}\rangle = \frac{1}{2^{1/2}}|0\rangle(|0\rangle + |1\rangle). \tag{1}$$

We now note that if qubit 1 is in the state $|\psi\rangle_1$ and qubits 2 and 3 are in one of the two states above, then the cloning circuit will implement the following transformations

$$|\psi\rangle_1|\Xi_{00}\rangle_{23} \rightarrow |\psi\rangle_1|\Xi_{00}\rangle_{23};$$
$$|\psi\rangle_1|\Xi_{0x}\rangle_{23} \rightarrow |\psi\rangle_2|\Xi_{00}\rangle_{13}. \tag{2}$$

Examining these equations, we see that in the first the quantum information from the first qubit appears in output 1, and in the second it appears in output 2. This suggests that if instead of sending either $|\Xi_{00}\rangle$ or $|\Xi_{0x}\rangle$ into inputs 2 and 3, we send in a linear combination of them, some of the quantum information from qubit 1 will appear in output 1 and some of it will appear in output 2, thereby cloning the state. This is, in fact, exactly what happens. If we choose

$$|\Psi\rangle_{23} = c_0|\Xi_{00}\rangle_{23} + c_1|\Xi_{0x}\rangle_{23}, \tag{3}$$

as the input state for qubits 2 and 3, with c_0 and c_1 real for simplicity, then the reduced density matrices for outputs 1 and 2 are

$$\rho_1^{(\mathrm{out})} = (c_1^2 + c_0c_1)|\psi\rangle\langle\psi| + \frac{c_1^2}{2}\mathbb{1}$$
$$\rho_2^{(\mathrm{out})} = (c_0^2 + c_0c_1)|\psi\rangle\langle\psi| + \frac{c_0^2}{2}\mathbb{1}. \tag{4}$$

Note that by choosing c_0 and c_1 we can control how much information about $|\psi\rangle$ goes to which output. In particular, if we choose $c_0 = c_1 = 1/3^{1/2}$, then the information is divided equally, and we find that

$$\rho_1^{(\mathrm{out})} = \rho_2^{(\mathrm{out})} = \frac{5}{6}|\psi\rangle\langle\psi| + \frac{1}{6}|\psi^\perp\rangle\langle\psi^\perp|, \tag{5}$$

where $|\psi^\perp\rangle$ is the qubit state orthogonal to $|\psi\rangle$. Therefore, the fidelity of the cloner output $\rho_1^{(\mathrm{out})}$ (or $\rho_2^{(\mathrm{out})}$, since they are the same in this case) to the ideal output, $|\psi\rangle$, which is given by $\langle\psi|\rho_1^{(\mathrm{out})}|\psi\rangle$, is 5/6. A fidelity of one would imply perfect cloning, so what we have here is a device that produces two copies of the input qubit that are pretty good approximations to it. Note that the fidelity does not depend on the input state, that is all states are cloned equally well. This feature of this cloning machine is known as

universality. Note that the cloner employs three qubits, and we have only discussed the final state of two of them. One might wonder if the output state of the third qubit is of interest. The answer is 'yes'. Its state is the best approximation to the state orthogonal to that of the input qubit that can be realised. A machine that sends a qubit in an arbitrary input state $|\psi\rangle$ into the state orthogonal to it, $|\psi^{\perp}\rangle$, is known as a universal-NOT, or UNOT, gate, and this transformation is also impossible to perform exactly [7,8]. It can, however, be performed approximately, and the best fidelity that can be obtained is 2/3. This can be achieved by measuring the original qubit along an arbitrary axis and then producing an output qubit whose state is orthogonal to the state obtained as the result of the measurement. For example, if one measured along the z axis and found the result $+z$, one would create a qubit in the $-z$ direction (for a discussion of the optimal estimation of the state of a single qubit see Appendix 1). The same result can be achieved by taking the third qubit, the one which is not a clone, from the output of a quantum cloning machine. This output qubit is sometimes referred to as an anticlone.

There have been a number of realisations of a quantum cloning machine, most based on a device known as an optical parametric amplifier [9–11]. This device takes one photon at frequency 2ω and converts it into two photons at frequency ω. A strong beam at 2ω will amplify a weaker beam at frequency ω via stimulated emission. When this device is used as a cloner, the qubits are the polarisation states of the photons. A photon in an arbitrary polarisation state at frequency ω will produce three photons, all at frequency ω, at the output. Two of them will approximate clones, and the third will be an approximate anticlone.

Let us see how this works in more detail [12]. If the pump beam at 2ω is described classically, the Hamiltonian for the two modes at frequency ω is given by

$$H = g(a_v^{\dagger} b_h^{\dagger} - a_h^{\dagger} b_v^{\dagger}) + h.c. \tag{6}$$

Here, the creation operators create photons in what are called the signal ($a_v^{\dagger}$ and $a_h^{\dagger}$) and idler ($b_v^{\dagger}$ and $b_h^{\dagger}$) modes, where the indices v and h refer to vertical and horizontal polarisation states, respectively, and g is a constant that describes the strength of the interaction (it is proportional to the pump field). This Hamiltonian is invariant under simultaneous rotations of the polarisation of the signal and idler modes, which makes cloning by this Hamiltonian universal. Because all input states are copied equally well, we can see how this cloner works, by choosing a specific one. Suppose we wish to clone a signal photon in the vertical polarisation state. To first order in g we find (after dropping the zeroth order term and normalising the state)

$$a_v^{\dagger}|0\rangle \rightarrow \left(\frac{2}{3}\right)^{1/2} \frac{(a_v^{\dagger})^2}{2^{1/2}} b_h^{\dagger}|0\rangle - \left(\frac{1}{3}\right)^{1/2} a_v^{\dagger} a_h^{\dagger} b_v^{\dagger}|0\rangle\,, \tag{7}$$

where $|0\rangle$ is the vacuum state. We see that with a probability of 2/3 we obtain two photons in the input state, and with a probability of 1/3 we obtain one, for an overall fidelity of 5/6. Both of the experiments cited in the previous paragraph obtained fidelities of 0.81, which is very close to the theoretical limit.

A second cloning strategy is a probabilistic one [13]. In this case, one wants to clone a quantum state that is selected from a known, finite set of states. For simplicity, let us assume that this set contains two elements, $|\psi_1\rangle$ and $|\psi_2\rangle$. Our machine is then to do the following. Given an input qubit that is in either the state $|\psi_1\rangle$ or $|\psi_2\rangle$, we don't know which, it is to produce two copies of the input state. If the two input states are not orthogonal, this cannot be done perfectly. It can, however, be done probabilistically. The machine either produces two perfect copies of the input, or it fails, and it tells us which of these two possibilities has occurred. The probability of successfully cloning the input is given by

$$p_{\mathrm{succ}} = \frac{1}{1 + |\langle\psi_1|\psi_2\rangle|}. \tag{8}$$

Note that this is one if the states are orthogonal, and decreases as their overlap increases.

2.2. *State discriminators*

The problem in quantum state discrimination is, given a particle in an unknown state selected from a known set of states, to determine the quantum state of the particle [14,15]. If the set of possible states contains states that are not orthogonal to each other, then this cannot be done perfectly. Again, for simplicity, let us assume that our set of possible states contains only two states, $|\psi_1\rangle$ and $|\psi_2\rangle$.

We shall again explore two strategies for accomplishing this task. The first is the minimum-error strategy, which is a strategy that approximately discriminates the two states [16]. It can make mistakes, but the probability of making a mistake is minimised. A machine that implements this strategy is given an input, which is equally likely to be $|\psi_1\rangle$ or $|\psi_2\rangle$, and it then tells us which of the two states it was given. The probability of the output being incorrect is

$$p_{\mathrm{err}} = \frac{1}{2}\left(1 - (1 - |\langle\psi_1|\psi_2\rangle|^2)^{1/2}\right). \tag{9}$$

Note that when the states are orthogonal, this is zero, but that it increases as the overlap between the states increases. The second strategy is known as unambiguous state discrimination [17–19]. In this case our machine has three outputs, one corresponding to state 1, one corresponding to state 2, and a third corresponding to failure. This machine will never incorrectly identify a state, but it may fail. For example, if the input is in $|\psi_1\rangle$, the machine will either tell us the input was in state $|\psi_1\rangle$, or fail, but it will never tell us the input was in state $|\psi_2\rangle$. Assuming that each input state is equally likely, the probability of successfully identifying the state is

$$p_{\text{succ}} = 1 - |\langle\psi_1|\psi_2\rangle|\,. \tag{10}$$

As has been the case before, the probability of successfully identifying the input state is one if the states are orthogonal, and decreases as the overlap of the states increases.

3. Programmable machines

We now want to consider programmable quantum machines, which we shall often refer to as quantum processors [20]. These have two inputs, one for the data, which is to be acted upon, and one for the program, which will specify the operation to be performed on the data. Both the data and the program are quantum states. In particular, the processor is a unitary operator acting on the Hilbert space $\mathcal{H}_{\text{d}} \otimes \mathcal{H}_{\text{p}}$, where $\mathcal{H}_{\text{d}}$ is the data Hilbert space and $\mathcal{H}_{\text{p}}$ is the program Hilbert space. The machine can act in either a deterministic or probabilistic fashion. In the case of a deterministic machine (see Figure 2), we always accept the output, and the action of the machine on the data state is described by a trace-preserving completely positive map, which is a result of tracing out the program state output. In the case of a probabilistic machine (see Figure 3), we measure the program state output, and only accept the data state output if a particular result is obtained. We shall examine both scenarios. It is, perhaps, best to begin with an example [21]. Let us go back and consider the three-qubit circuit for the approximate cloner. Qubit 1 will now be our data state, and qubits 2 and 3 will be our program. We will denote the data state by $|\psi\rangle_1$ and the program state by $|\Xi\rangle_{23}$. Define the two-qubit Bell states to be

$$|\Psi_\pm\rangle = \frac{1}{2^{1/2}}(|00\rangle \pm |11\rangle);$$

$$|\Phi_\pm\rangle = \frac{1}{2^{1/2}}(|01\rangle \pm |10\rangle). \tag{11}$$

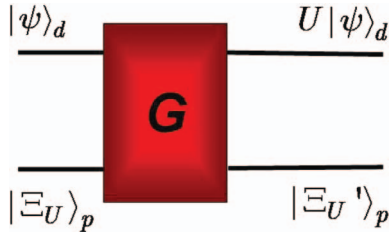

Figure 2. A model of a deterministic quantum processor G that implements a unitary operation U on the data register.

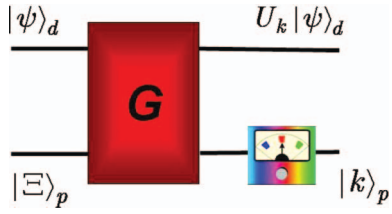

Figure 3. A model of a probabilistic general quantum processor. On the output of the program register a measurement is performed.

If these states are used as programs in our processor, we find that

$$\begin{aligned}
&|\psi\rangle_1|\Psi_+\rangle_{23} \to |\psi\rangle_1|\Psi_+\rangle_{23}\,;\\
&|\psi\rangle_1|\Psi_-\rangle_{23} \to \sigma_z|\psi\rangle_1|\Psi_-\rangle_{23}\,;\\
&|\psi\rangle_1|\Phi_+\rangle_{23} \to \sigma_x|\psi\rangle_1|\Phi_+\rangle_{23}\,;\\
&|\psi\rangle_1|\Psi_-\rangle_{23} \to (-\mathrm{i}\sigma_y)|\psi\rangle_1|\Psi_-\rangle_{23},
\end{aligned} \tag{12}$$

where σ_x, σ_y, and σ_z are the Pauli matrices. If we choose the program

$$|\Xi\rangle_{23} = c_0|\Psi_+\rangle_{23} + c_1|\Phi_+\rangle_{23} + c_2|\Phi_-\rangle_{23} + c_3|\Psi_-\rangle_{23}, \tag{13}$$

then operating our machine in the deterministic mode, by tracing out the program state output, we obtain for the data state output

$$\begin{aligned}
\rho_1^{(\text{out})} = {} & |c_0|^2\rho_1^{(\text{in})} + |c_1|^2\sigma_x\rho_1^{(\text{in})}\sigma_x\\
& + |c_2|^2\sigma_y\rho_1^{(\text{in})}\sigma_y + |c_3|^2\sigma_z\rho_1^{(\text{in})}\sigma_z.
\end{aligned} \tag{14}$$

In the above equation, we have set $\rho_1{}^{(\text{in})} = |\psi\rangle_1\langle\psi|$. Examining the output state, we see that this circuit can implement a number of quantum channels: the bit-flip channel ($c_2 = c_3 = 0$), which flips a bit, with a certain probability, the phase-flip channel ($c_1 = c_2 = 0$), which sends $|0\rangle \to |0\rangle$ and $|1\rangle \to -|1\rangle$, with a certain probability, and the depolarising channel ($c_1 = c_2 = c_3$), in which the input state is replaced by the completely

mixed state, with a certain probability [1]. The same processor can be used in the probabilistic mode. Suppose we want to implement the operator $A = I - 2|\phi\rangle\langle\phi|$ on the data state, where $|\phi\rangle$ is a specified one-qubit state. The operator A is similar to σ_z, but instead of flipping the phase of the state $|1\rangle$, it flips the phase of the state $|\phi\rangle$. Defining the two-qubit operator, U_0,

$$U_0|00\rangle = -|10\rangle\, U_0|10\rangle = -|11\rangle;$$
$$U_0|01\rangle = |00\rangle\, U_0|11\rangle = |01\rangle, \qquad (15)$$

we choose for our program state

$$|\Xi\rangle_{23} = \frac{1}{2^{1/2}}\, U_0(|\phi\rangle_2|\phi^{\perp}\rangle_3 + |\phi^{\perp}\rangle_2|\phi\rangle_3), \qquad (16)$$

where $|\phi^{\perp}\rangle$ is the qubit state orthogonal to $|\phi\rangle$. At the program state output, we project onto the state $(|\Phi_+\rangle_{23} + |\Phi_-\rangle_{23} + |\Psi_-\rangle_{23})/3^{1/2}$, and if we get one, we keep the data state output. This will happen with a probability of 1/3, independent of the state $|\phi\rangle$. If we do get one, then the data state output will be in the state $A|\psi\rangle$.

Now, let us return to deterministic processors and examine the resources that are necessary in order to implement a given set of operations on the data. Suppose that our data state is a qubit, and we want to implement a one-parameter unitary group $U(\alpha) = \exp(i\alpha\sigma_z)$, where $0 \leq \alpha < 2\pi$, on it. We want to encode the angle α in the program state. It turns out that this cannot be done with a finite-dimensional program space, due to a no-go theorem due to Nielsen and Chuang [22]. It states that if the program $|\Xi_1\rangle$ implements the unitary operator U_1 on the data state, and $|\Xi_2\rangle$ implements the unitary operator U_2, then $\langle\Xi_1|\Xi_2\rangle = 0$. This implies that for every unitary operator that the processor can implement on the data state, we need an extra dimension in the program space. The proof of this theorem is given in Appendix 2. In the case of our one-paramenter group, there are an infinite number of operators, so it clearly cannot be implemented on a processor with a finite-dimensional program space.

Given this result, we can adopt the same strategies we did in the case of single-purpose machines that were prohibited by a no-go theorem. We have already seen in our example, that a probabilistic machine has no problem implementing an infinite number of operations. The only remaining issue in that case is figuring out how to make the success probability as large as possible. This we shall address shortly. The other strategy is to construct a machine that carries out a set of operations approximately. It is this type of machine we shall discuss now [23].

We have seen that deterministic processors implement trace-preserving, completely positive maps. Therefore, when considering an approximate deterministic processor, we need to have some kind of a measure of how close two such maps are. We shall use the process fidelity [24], which has a number of useful properties [25]. Let T_1 and T_2 be two trace-preserving, completely positive maps, mapping the space, $\mathcal{B}(\mathcal{H})$, of linear operators on a D-dimensional Hilbert space, $\mathcal{H}$, into itself. The Jamiolkowski isomorphism associates a density matrix on $\mathcal{H}\otimes\mathcal{H}$ to each trace-preserving, completely positive map on $\mathcal{B}(\mathcal{H})$. Letting $\{|j\rangle | j = 1, 2, \ldots, D\}$ be an orthonormal basis for $\mathcal{H}$, define the maximally entangled state, $|\Phi\rangle$ in $\mathcal{H}\otimes\mathcal{H}$

$$|\Phi\rangle = \frac{1}{D^{1/2}} \sum_{j=1}^{D} |j\rangle|j\rangle. \qquad (17)$$

The density matrix associated with the trace-preserving, completely positive map, T, is

$$\rho = (\mathcal{I}\otimes T)(|\Phi\rangle\langle\Phi|)\,, \qquad (18)$$

where $\mathcal{I}$ is the identity map. If ρ_1 is the density matrix associated with T_1 and ρ_2 is the density matrix associated with T_2, the process fidelity between T_1 and T_2 is

$$F_{\mathrm{proc}}(T_1, T_2) = \left[\mathrm{Tr}\left(\rho_1^{1/2}\rho_2\rho_1^{1/2}\right)^{1/2}\right]^2, \qquad (19)$$

that is, just the standard fidelity between the density matrices ρ_1 and ρ_2. The process fidelity is related to the average fidelity of the outputs of the two maps for identical inputs [25,26]. If we define

$$F_{\mathrm{ave}}(T_1, T_2) = \int d\psi\, F(T_1(|\psi\rangle\langle\psi|), T_2(|\psi\rangle\langle\psi|)), \qquad (20)$$

where $d\psi$ denotes the Haar measure on a d-dimensional state space, and F is the standard fidelity for two density matrices, then

$$F_{\mathrm{ave}}(T_1, T_2) = \frac{1}{d+1}\left[F_{\mathrm{proc}}(T_1\,, T_2)d + 1\right]. \qquad (21)$$

We will not discuss the case of a general approximate processor, but will look at a specific type. Suppose we have a processor that is a controlled-U gate. That means that if our program space, $\mathcal{H}_{\mathrm{p}}$, has dimension N, there is an orthonormal basis of $\mathcal{H}_{\mathrm{p}}$, $\{|j\rangle | j = 1, 2, \ldots, N\}$ such that the processor acts as follows

$$|\psi\rangle|j\rangle \rightarrow U_j|\psi\rangle|j\rangle\,, \qquad (22)$$

where U_j is a unitary operator on the data space. Therefore, this processor implements the set of unitary

operators $S_u = \{U_j | j = 1,2, \ldots, N\}$ perfectly. Now suppose we want to use this processor to approximate another unitary operator U, which is not in S_u. We want to choose a program state that maximises the process fidelity between the map it generates and U. What one finds is that the best program is one of the basis states $|j\rangle$, and it is the one for which $|Tr(U^\dagger U_j)|$ is a maximum. In this case, one simply chooses the unitary operator in S_u that is closest to U and implements that operator. Using a program state that is a superposition of different basis vectors does not help.

Let us look at an example of this situation. We wish to implement the operator, for $0 \leq \theta < 2\pi$,

$$U(\theta) = \exp\left[\frac{i\pi}{2}(\exp(-i\theta)\sigma^+ + \exp(i\theta)\sigma^-)\right], \quad (23)$$

on our data state, which is a qubit. Here $0 \leq \theta < 2\pi$, and $\sigma^\pm = (\sigma_x \pm i\sigma_y)/2$. Our program state has dimension N, and an orthonormal basis $\{|j\rangle | j = 1, 2, \ldots, N\}$. Define the operators $E_\pm$ on the program space by $E_\pm |j\rangle = |j \pm 1\rangle$, where the addition and subtraction are modulo N. Now let the overall processor unitary operator, which acts on the tensor product of the data and program spaces, be

$$G = \exp\left[\frac{i\pi}{2}(\sigma^+ E_- + \sigma^- E_+)\right], \quad (24)$$

and consider the program states

$$|\theta\rangle = \frac{1}{N^{1/2}} \sum_{j=0}^{N-1} \exp(-ij\theta)|j\rangle . \quad (25)$$

When $\theta = \theta_m = 2\pi m/N$, for m an integer between 0 and $N-1$, we find that

$$G(|\psi\rangle|\theta_m\rangle) = U(\theta_m)|\psi\rangle|\theta_m\rangle , \quad (26)$$

where $|\psi\rangle$ is a general qubit state. Therefore, this processor implements the operations $U(\theta_m)$ perfectly. Now suppose we want to implement $U(\theta)$ for a value of θ that is not one of the θ_m. The optimal strategy is to find the θ_m closest to θ and to send in the program state $|\theta_m\rangle$ corresponding to that value. If we do so we find that

$$F_{proc}(U(\theta), U(\theta_m)) > \cos^2\left(\frac{\pi}{N}\right) \sim 1 - \left(\frac{\pi}{N}\right)^2 . \quad (27)$$

Rather than determining which θ_m is the best one to use, a simpler procedure is just to use the program state $|\theta\rangle$. There should be some cost to doing this, and, indeed, we find that the process fidelity in this case is approximately $1-(2/N)$. The optimal program has an error that goes like $1/N^2$ while the simpler procedure gives an error of $1/N$. Determining whether the extra accuracy is worth the extra work in determining the best program would depend on the application.

Now let us return to probabilistic programmable devices. Suppose our data system is a qubit, and we want to implement the one parameter group we mentioned earlier, $U(\alpha) = \exp(i\alpha\sigma_z)$, where $0 \leq \alpha < 2\pi$. This can be accomplished with a success probability of $1/2$ by using a qubit program and a controlled-NOT gate. As was noted before, the controlled-NOT gate has two inputs, a control input and a target input. The state of the control qubit is not changed, and if the state of the control qubit is $|0\rangle$, neither is the state of the target qubit. However, if the control qubit is in the state $|1\rangle$, then the operator σ_x is applied to the target qubit. It is, in fact, a controlled-U gate with the two unitary operators being the identity and σ_x. In our case, the target qubit is the program and the control qubit is the data. The program states are

$$|\Xi(\alpha)\rangle = \frac{1}{2^{1/2}}[\exp(i\alpha)|0\rangle + \exp(-i\alpha)|1\rangle] . \quad (28)$$

If the data state input is $|\psi\rangle$, the output of this processor is then

$$|\Psi_{out}\rangle = \frac{1}{2^{1/2}}(U(\alpha)|\psi\rangle|0\rangle + U^{-1}(\alpha)|\psi\rangle|1\rangle) . \quad (29)$$

By measuring the program state output in the basis $\{|0\rangle, |1\rangle\}$, and keeping the result only if we get $|0\rangle$, which happens with a probability of $1/2$, we obtain the data state output $U(\alpha)|\psi\rangle$, which is the desired result.

A closely related programmable device has been recently realised experimentally [27]. It carries out the transformation

$$\alpha|0\rangle + \beta|1\rangle \rightarrow \alpha|0\rangle + \exp(i\phi)\beta|1\rangle , \quad (30)$$

where the angle ϕ is encoded in a second qubit. The qubits are polarisation states of photons, with $|H\rangle$ representing a horizontally polarised photon and $|V\rangle$ representing a vertically polarised one. A polarising beam splitter, which transmits horizontally polarised photons and reflects vertically polarised ones is the main component of the device. The beam splitter has two input modes, which we shall label 1 and 2, and two output modes, which we shall also denote as 1 and 2. For a photon incident in input mode 1 we would have $|H\rangle_1 \rightarrow |H\rangle_1$ and $|V\rangle_1 \rightarrow |V\rangle_2$. Input mode 2 behaves similarly. If two photons, one in the state $\alpha|H\rangle_1 + \beta|V\rangle_1$ (data) and the other in the state $(1/2^{1/2})$

$(|H\rangle_2 + \exp(i\phi)|V\rangle_2)$ (program) are incident on the polarising beam splitter, then in the cases in which a single photon emerges from each output, which happens with a probability of 1/2, the conditional output state is

$$|\psi_{out}\rangle = \frac{1}{2^{1/2}}(\alpha|H\rangle_1|H\rangle_2 + \beta\exp(i\phi)|V\rangle_1|V\rangle_2). \quad (31)$$

If we measure the second photon in the $|\pm\rangle = (1/2^{1/2})(|H\rangle \pm |V\rangle)$ basis, then the remaining photon is in either the state $(1/2^{1/2})(|H\rangle + \exp(i\phi)|V\rangle)$, if our measurement result was $|+\rangle$, and $(1/2^{1/2})(|H\rangle - \exp(i\phi)|V\rangle)$, if our measurement result was $|-\rangle$. If we obtain the result $|-\rangle$ we can apply a correcting operation on the remaining qubit that sends $|H\rangle \rightarrow |H\rangle$ and $|V\rangle \rightarrow -|V\rangle$. The final result is that this device implements the transformation

$$\alpha|H\rangle + \beta|V\rangle \rightarrow \alpha|H\rangle + \exp(i\phi)\beta|V\rangle, \quad (32)$$

with a probability of 1/2. It should be noted, however, that in the experiment the gate is implemented in such a way that the photons cannot be used once they have gone through it. The conditional measurement necessary to make the gate work is performed when both of the photons are detected at the output of the device, thereby destroying them.

Suppose that we want to increase the probability of a successful outcome. One possibility is to try again if we get the wrong result for our measurement on the program state [28,29]. If we obtained the result $|1\rangle$ from our measurement, then the data qubit is in the state $U^{-1}(\alpha)|\psi\rangle$. We can take this qubit and run it through the processor again, but this time use the program $|\Xi(2\alpha)\rangle$. If we do so, the output state is

$$|\Psi'_{out}\rangle = \frac{1}{2^{1/2}}(U(\alpha)|\psi\rangle|0\rangle + U^{-1}(3\alpha)|\psi\rangle|1\rangle). \quad (33)$$

We again measure the program state and keep the result if we get $|0\rangle$. This again happens with a probability of 1/2. Adding this second step has increased our overall success probability to 3/4, and the procedure can be repeated to bring the success probability as close to one as we wish. What we need to do this, however, is a collection of qubits in the proper program states, that is, besides a qubit in the state $|\Xi(\alpha)\rangle$, we need an additional one in the state $|\Xi(2\alpha)\rangle$.

We can also accomplish the same thing by enlarging our program space [29]. Our data space still consists of one qubit, but the program space now contains two qubits. Let us label the three inputs, input 1 being the data input, input 2 the first program input and input 3 the second program input. The processor now consists of two gates. The first is a controlled-NOT gate whose control qubit is qubit 1 and whose target qubit is qubit 2. The second gate is a Toffoli gate. This gate has two control qubits and one target qubit. The states of the control qubits are not changed, and if they are in the states $|0\rangle|0\rangle$, $|0\rangle|1\rangle$, or $|1\rangle|0\rangle$, neither is the state of the target qubit. However, if they are in the state $|1\rangle|1\rangle$, then σ_x is applied to the target qubit. In our processor, qubits 1 and 2 are the control qubits and qubit 3 is the target qubit. The input state is $|\psi\rangle_1|\Xi(\alpha)\rangle_2|\Xi(2\alpha)\rangle_3$, and the output state is

$$|\Psi''_{out}\rangle = \frac{1}{2}[U(\alpha)|\psi\rangle_1(|0\rangle_2|0\rangle_3 + |0\rangle_2|1\rangle_3 + |1\rangle_2|0\rangle_3) + U^{-1}(3\alpha)|\psi\rangle|1\rangle_2|1\rangle_3] \quad (34)$$

At the output we measure the program qubits in the computational basis and keep the data state output if we get $|0\rangle|0\rangle$, $|0\rangle|1\rangle$, or $|1\rangle|0\rangle$. If we do, the data output is in the state $U(\alpha)|\psi\rangle$, and we have achieved our goal. This happens with a probability of 3/4. By increasing the dimension of the program space further, we can increase our probability of success. We have, therefore, two strategies for increasing the success probability for a probabilistic processor (for details see [30]).

4. Why quantum programs?

The programs in the quantum processors we have been discussing have been quantum states. One might wonder whether this is necessary and whether classical programs would suffice. That is, one could have gates that can perform a number of operations, but the selection of which operation they do perform is governed by a classical input. Do quantum programs provide an advantage? There are several scenarios that suggest themselves for which quantum programs would be useful. One is that the information on which the program is based is intrinsically quantum. We shall explore an example of this situation when we discuss programmable state discriminators. This could also occur if the program is the result of an earlier quantum computation. A second situation is one in which we would like to apply quantum information processing techniques, such as a Grover search, to programs. In that case, the programs must be quantum.

Let us first consider programmable state discriminators. The first such device was proposed by Bužek and Dušek [31]. Here we will discuss a different version, which is a type of universal state discriminator [32]. So far, when discussing state discriminators, we have assumed we knew the set of states we

were trying to discriminate among. This knowledge was built into the discriminator. The resulting discriminator is useful for discriminating states from that particular set, but it is not useful for discriminating among members of other sets of states. Suppose, however, that we would like a discriminator that would work for any set of states, i.e. a universal discriminator. In that case, we have to provide information about the set of possible states as well as the quantum system whose state we want to determine with the machine. The information about the set of possible states will be the program.

Let us consider the simplest version of such a device. It will unambiguously discriminate between two different qubit states. The program consists of two states, one in each of the states we want to discriminate between, which we shall call $|\psi_1\rangle$ and $|\psi_2\rangle$. The data qubit is in either $|\psi_1\rangle$ or $|\psi_2\rangle$, and we would like to know which. What the machine does is implement a POVM, which takes advantage of the symmetry of the three-qubit input state. Let us call the program inputs a and b, and the data input c. Our task is to discriminate between the states

$$|\Psi_1\rangle = |\psi_1\rangle_a|\psi_2\rangle_b|\psi_1\rangle_c;$$
$$|\Psi_2\rangle = |\psi_1\rangle_a|\psi_2\rangle_b|\psi_2\rangle_c. \tag{35}$$

Note that in $|\Psi_1\rangle$ the first and third qubits are in the same state, while in $|\Psi_2\rangle$ the second and third qubits are in the same state. Therefore, if we project the three-qubit input state onto the antisymmetric subspace of qubits a and c, and we get a non-zero result, then we know that qubit c was in the state $|\psi_2\rangle$. Similarly, if we project qubits b and c onto the antisymmetric subspace of two qubits, and we get a non-zero result, then we know that qubit c was in the state $|\psi_1\rangle$. There will also be a 'don't know' result in which the measurement fails, and we want to minimise the probability of obtaining this result. If the two states are equally likely, and averaging over $|\psi_1\rangle$ and $|\psi_2\rangle$, since we do not know what they are, we find that the optimal probability of identifying the input data state is 1/6. Note that in this case, the information contained in the program was quantum information, in particular, it consisted of examples of quantum states, and this necessitated the program itself being quantum.

Now let us look at an example in which it is useful to apply quantum information processing techniques to quantum programs. In order to do so, we first need to explain the quantum search algorithm due to Grover [33]. We have a black box that evaluates a Boolean function. A Boolean function is one whose value is either zero or one. We send in an input, which is an n-digit binary number, x, and the output of the box is $f(x)$. This particular function is zero on all inputs except one, which we shall call x_0, and $f(x_0) = 1$. Our object is to find x_0 with a minimum number of uses of the black box.

Classically, we simply send in different inputs to the black box until we find one that gives one as an output. On average we will have to make $2^{(n-1)}$ tries. The Grover algorithm works in a completely different way, and its result is a considerable improvement over the classical one. It starts with an input state that is an equal superposition of all possible input values. By successively applying the black box followed by an operator Grover called 'inversion about the mean' approximately $2^{n/2}$ times, the initial state is rotated into the state $|x_0\rangle$, and then one simply measures this state in the computational basis to find out what x_0 is. Note that the black box was only used $2^{n/2}$ times in this case, which means that the number of evaluations in the Grover algorithm is approximately the square root of the number of evaluations that are necessary in the classical case.

Now consider the following problem [34]. We have a set of M permutations on N objects. In particular let $X = \{k|k = 0,1, \ldots, N-1\}$ be the set of objects being permuted, and let $S = \{\sigma_j \,|j = 1,2, \ldots, M\}$ be the set of permutations. For some specified k_0, $k_1 \in X$, we are promised that there is one $\sigma_j \in S$ such that $\sigma(k_0) = k_1$, and we want to find which permutation satisfies this property. A variant of this problem, determining whether there is a $\sigma_j \in S$ such that $\sigma(k_0) = k_1$, can be used to attack the conjugacy problem in group theory. If G is a group, and g_1, $g_2 \in G$, we would like to know whether g_1 and g_2 are conjugate to each other, that is, whether there is an $h \in G$, such that $g_2 = hg_1h^{-1}$. The connection between this problem and the one involving the permutations is provided by realising that the automorphism $\alpha_h : G \to G$ given by $\alpha_h(g) = hgh^{-1}$ is just a permutation on G. Thus, the conjugacy problem is reduced to determining whether there is an α_h such that $\alpha_h(g_1) = g_2$.

We suppose we have a quantum processor, which acts on the Hilbert space $\mathcal{H}_X \otimes \mathcal{H}_S$, where $\mathcal{H}_X$ is spanned by the orthonormal basis $\{|k\rangle_X \,|k = 0,1, \ldots N-1\}$ and $\mathcal{H}_S$ is spanned by the orthornormal basis $\{|j\rangle_S \,|j = 1,2, \ldots M\}$. We regard $\mathcal{H}_S$ as the program space, and $\mathcal{H}_X$ as the data space. The processor acts as follows

$$|j\rangle_S|k\rangle_X \to |j\rangle_S U_j|k\rangle_X, \tag{36}$$

where $U_j|k\rangle_X = |\sigma_j(k)\rangle_X$. Once we have this processor, we can do a Grover search on the programs in order to find the permutation that satisfies $\sigma(k_0) = k_1$. This will

require approximately $M^{1/2}$ uses of the processor, whereas classically M uses would be required. It is the fact that the programs are quantum states that allows us to search among them by using a quantum search procedure.

5. Conclusion

As we have seen, quantum machines have been developed for a number of information processing tasks. Cloners move quantum information around and discriminators allow one to distinguish among nonorthogonal quantum states. Discriminators can be generalised to distinguish between nonorthogonal subspaces as well [35]. In addition, we have seen that it is possible to construct programmable quantum machines, which are capable of performing a number of different tasks.

The capabilities of programmable machines are still not well understood. We concentrated mainly on processors that implement unitary operators, but, as we saw processors can also implement more general maps. Some families of maps, for example, those in Equation (14), can be implemented with a finite-dimensional program space, while others, such as a one-parameter unitary group, cannot. What determines whether a set of maps can be programmed with a finite-dimensional program? Another issue is the equivalence of programmable processors. Suppose we have two processors, both of which can perform the same set of operations but they do so with different programs. This could happen, for example, if the one of the processors differed from the other simply by having a fixed unitary gate at the input to its program register. Given two processors, is there a simple way of telling whether or not the set of operations they can implement is the same? These are only two questions about the properties of quantum processors, and we suspect there are many more.

Acknowledgements

This work was supported by the European Union projects HIP and QAP, by Slovak grant agencies APVV and VEGA via projects RPEU-0014-06 and 2/0092/09, respectively.

Notes on contributors

Mark Hillery is a Professor of Physics at Hunter College and the Graduate Center of the City University of New York.

Vladimír Bužek is the head of the Research Center for Quantum Information of the Physics Institute of the Slovak Academy of Sciences. Both originally worked in quantum optics but became interested in quantum information in the mid-1990s. Their first foray into the field was a paper on quantum cloners, and this led, ultimately, to an interest in quantum information processing machines in general.

References

[1] M.A. Nielsen and I.L. Chuang, *Quantum Computation and Quantum Information*, Cambridge University Press, Cambridge, 2000.
[2] D. D'Alessandro, *Introduction to Quantum Control and Dynamics*, Chapman & Hall/CRC, Boca Raton, 2008.
[3] W.K. Wooters and W.H. Zurek, *A single quantum cannot be cloned*, Nature 299 (1982), p. 802.
[4] V. Bužek and M. Hillery, *Quantum copying: beyond the nonclonning theorem*, Phys. Rev. A 54 (1996), pp. 1844–1852.
[5] S.L. Braunstein, V. Bužek, and M. Hillery, *Quantum information distributors: quantum network for symmetric and asymmetric cloning in arbitrary dimension and continuous limit*, Phys. Rev. A 63 (2001), 052313.
[6] For reviews of quantum cloning see, V. Scarani, S. Iblisdir, N. Gisin, and A. Acín, *Quantum cloning*, Rev. Mod. Phys. 77 (2005), pp. 1225–1256; N.J. Cerf and J. Fiurášek, *Optical quantum cloning – a review*, Progr. Opt. 49 (2006), p. 455.
[7] V. Bužek, M. Hillery, and R.F. Werner, *Optimal manipulations with qubits: universal NOT gate*, Phys. Rev. A 60 (1999), R2626.
[8] V. Bužek, M. Hillery, and R.F. Werner, *Universal-NOT gate*, J. Mod. Opt. 47 (2000), pp. 211–232.
[9] F. de Martini, V. Mussi, and F. Bovino, *Schrödinger cat states and optimum universal quantum cloning by entangled parametric amplification*, Opt. Commun. 179 (2000), pp. 581–589.
[10] F. De Martini, V. Bužek, F. Sciarrino, and C. Sias, *Experimental realization of the quantum universal NOT gate*, Nature 419 (2002), pp. 815–818.
[11] A. Lamas-Linares, C. Simon, J.C. Howell, and D. Bouwmeester, *Experimental quantum cloning of single photons*, Science 296 (2002), pp. 712–714.
[12] C. Simon, G. Weihs, and A. Zeilinger, *Optimal quantum cloning and universal NOT without quantum gates*, J. Mod. Opt. 47 (2000), pp. 233–246.
[13] L.-M. Duan and G.-C. Guo, *Probabilistic cloning and identification of linearly independent quantum states*, Phys. Rev. Lett. 80 (1998), pp. 4999–5002.
[14] For a review of the theory of state discrimination see, J.A. Bergou, U. Herzog, and M. Hillery, Discrimination of quantum states, in *Quantum State Estimation*, M.G.A. Paris, and J. Řehaček, eds., Springer Verlag, Berlin, 2004, p. 417; A. Chefles, *Quantum state discrimination*, Contemp. Phys. 41 (2000), pp. 401–424.
[15] For a review of experimental state discrimination see A. Chefles, Quantum states: discrimination and classical information transmission. A review of experimental progress, in *Quantum State Estimation*, M.G.A. Paris, and J. Řehaček, eds., 2004. Springer-Verlag, Berlin, 2004, p. 467.
[16] C.W. Helstrom, *Quantum Detection and Estimation Theory*, Academic Press, New York, 1976.
[17] I.D. Ivanovic, *How to differentiate between non-orthogonal states*, Phys. Lett. A 123 (1987), pp. 257–259.
[18] D. Dieks, *Overlap and distinguishability of quantum states*, Phys. Lett. A 126 (1988), pp. 303–306.
[19] A. Peres, *How to differentiate between non-orthogonal states*, Phys. Lett. A 128 (1988), p. 19.
[20] For a detailed review of programmable quantum processors see, V. Bužek, M. Hillery, M. Ziman, and M. Roško, *Programmable quantum processors*, Quantum Informat. Process. 5 (2006), pp. 313–420.

[21] M. Hillery, V. Bužek, and M. Ziman, *Probabilistic implementation of universal quantum processors*, Phys. Rev. A 65 (2002), 022301.
[22] M.A. Nielsen and I.L. Chuang, *Programmable quantum gate arrays*, Phys. Rev. Lett. 79 (1997), pp. 321–324.
[23] M. Hillery, M. Ziman, and V. Bužek, *Approximate programmable quantum processors*, Phys. Rev. A 73 (2006), 022345.
[24] M. Raginsky, *A fidelity measure for quantum channels*, Phys. Lett. A 290 (2001), pp. 11–18.
[25] A. Gilchrist, N.K. Langford, and M.A. Nielsen, *Distance measures to compare real and ideal quantum processes*, Phys. Rev. A 71 (2005), 062310.
[26] M. Horodecki, P. Horodecki, and R. Horodecki, *General teleportation channel, singlet fraction, and quasidistillation*, Phys. Rev. A 60 (1999), pp. 1888–1898.
[27] M. Micuda, M. Ježek, M. Dušek, and J. Fiurášek, *Experimental realization of a programmable quantum gate*, Phys. Rev. A 78 (2008), 062311.
[28] J. Preskill, *Reliable quantum computers*, Proc. Roy. Soc. Lond. A 454 (1998), pp. 385–410.
[29] G. Vidal, L. Masanes, and J.I. Cirac, *Storing quantum dynamics in quantum states: a stochastic programmable gate*, Phys. Rev. Lett. 88 (2002), 047905.
[30] A. Brazier, V. Bužek, and P.L. Knight, *Probabilistic programmable quantum processors with multiple copies of program states*, Phys. Rev. A 71 (2005), 032306.
[31] V. Bužek and M. Dušek, *Quantum multimeters: a programmable state discriminator*, Phys. Rev. A 66 (2002), 0022112.
[32] J. Bergou and M. Hillery, *Universal programmable quantum state discriminator that is optimal for unambiguously distinguishing between unknown states*, Phys. Rev. Lett. 94 (2005), 160501.
[33] L.K. Grover, *Quantum mechanics helps in searching for a needle in a haystack*, Phys. Rev. Lett. 79 (1997), pp. 325–328.
[34] M. Bonanome, M. Hillery, and V. Bužek, *Application of quantum algorithms to the study of permutations and group automorphisms*, Phys. Rev. A 76 (2007), 012324.
[35] J. Bergou, E. Feldman, and M. Hillery, *Optimal unambiguous discrimination of two subspaces as a case in mixed-state discrimination*, Phys. Rev. A 73 (2006), 032107.
[36] A. Holevo, *Probabilistic and Statistical Aspects of Quantum Theory*, North Holland, Amsterdam, 1982.
[37] S. Massar and S. Popescu, *Optimal extraction of information from finite ensembles*, Phys. Rev. Lett. 74 (1995), pp. 1259–1263.
[38] R. Derka, V. Bužek, and A. Ekert, *Universal algorithm for optimal state estimation from finite ensembles*, Phys. Rev. Lett. 80 (1998), pp. 1571–1575.
[39] S. Massar and S. Popescu, *Amount of information obtained by a quantum measurement*, Phys. Rev. A 62 (2000), 062303.
[40] S. Massar and R.D. Gill, *State estimation for large ensembles*, Phys. Rev. A 61 (2000), 042312.
[41] V. Bužek and R. Derka, *Quantum observations*, in *Coherence and Statistics of Photons and Atoms*, J. Peřina, ed., Wiley, New York, 2001, pp. 198–261.
[42] D. Bruß, A. Ekert, and C. Machiavello, *Optimal universal quantum cloning and state estimation*, Phys. Rev. Lett. 81 (1998), pp. 2598–2601.
[43] N. Gisin and S. Massar, *Optimal quantum cloning machines*, Phys. Rev. Lett. 79 (1997), pp. 2153–2156.
[44] R.F. Werner, *Optimal cloning of pure states*, Phys. Rev. A 58 (1998), pp. 1827–1832.

Appendix 1. Basics of Bayesian inference

To understand the bounds on the optimal manipulation of quantum information let us turn our attention to the problem of the optimal estimation of states of quantum systems. Let us consider a finite ensemble of N qubits all prepared in the same pure state $|\psi\rangle$. If the state is totally unknown, i.e. we have no *a priori* information about its preparation, then we have to assume that all pure states are equally probable. This corresponds to a uniform probability distribution on the state space of a given system, i.e. in the case of qubits – the Bloch sphere (see Figure 4). It is well known [16,36–40] (for a review see [41]) that there exists an optimal measurement of a finite set of N qubits by means of which the *best* possible estimation of the state $|\psi\rangle$ can be performed. Holevo [36] has shown that it is possible to realise the best estimation via a so-called *covariant* measurement, which is a continuous POVM measurement performed on the whole finite ensemble. Obviously, in this case the problem is, that physically it is difficult to perform experimentally a measurement with a continuous number of observables. Later it has been shown by Massar and Popescu [37] and Derka et al. [38], that the optimal measurement on a finite ensemble of qubits can be realised via a finite-dimensional POVM. Such POVM can be realised when we imagine projective measurements performed on the whole set of N qubits (that is the qubits are not measured sequentially, but simultaneously, in one 'shot'). Once this optimal measurement is performed then the best possible estimate

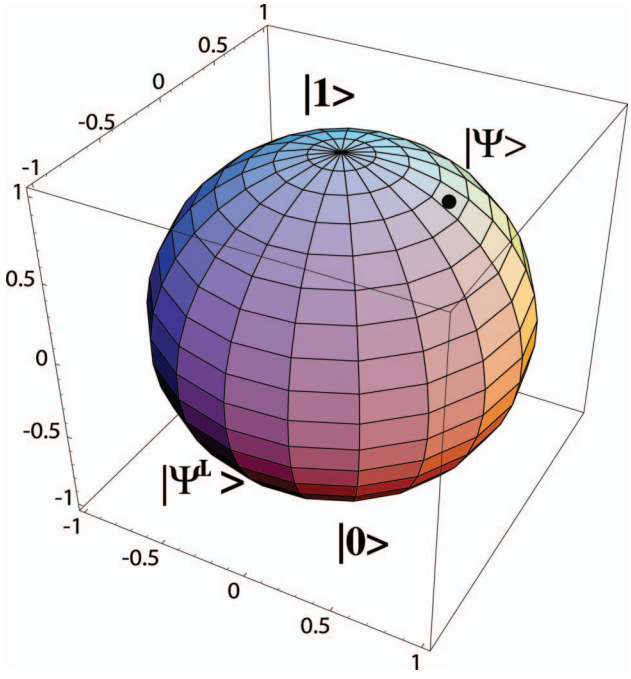

Figure 4. The state space of a qubit is a Bloch sphere. Pure states $|\psi\rangle$ are represented by points on the sphere, while statistical mixtures are points inside the sphere. The state $|\psi^{\perp}\rangle$ that is orthogonal to $\psi\rangle$, i.e. $\langle\psi|\psi^{\perp}\rangle = 0$, is its antipode.

of the measured state can be expressed in the form of the density operator

$$\rho^{(\mathrm{est})} = s_N \rho + \frac{1 - s_N}{2} \mathbb{1}\,, \tag{37}$$

where the 'scaling' factor s_N is given by the expression

$$s_N = \frac{N}{N+2}\,, \tag{38}$$

and is directly related to the mean fidelity

$$\overline{\mathcal{F}} = \int \mathrm{d}\Omega_\rho \langle\psi|\rho^{(\mathrm{est})}|\psi\rangle\,, \tag{39}$$

where the integration is performed over all input states ρ and $\mathrm{d}\,\Omega_\rho = \sin\vartheta\ \mathrm{d}\vartheta\ \mathrm{d}\varphi/4\pi$ is the integration measure associated with the state space, i.e. the Bloch sphere. When we insert $\rho^{(\mathrm{est})}$ given by Equation (37) into Equation (39) we find

$$\overline{\mathcal{F}} = s_N + \frac{1 - s_N}{2} = \frac{N+1}{N+2}\,. \tag{40}$$

There does not exist a measurement that would give us more information than is given by this POVM (for more details see the review article [41]).

A1.1. Single qubit case

In the case of a single qubit, a simple projective measurement optimal. Specifically, the *optimal* way to estimate the state, is to measure it along a *randomly* (we have no prior knowledge about the state) chosen direction in the two-dimensional Hilbert space [36–38]. So let us choose a random vector $|\eta\rangle$, where

$$|\eta\rangle = \cos(\vartheta'/2)|0\rangle + \exp(\mathrm{i}\varphi')\sin(\vartheta'/2)|1\rangle\,, \tag{41}$$

and measure $|\psi\rangle$ along it. If the result is positive, then the output is taken to be $|\eta\rangle$, and if negative, the output is $|\eta^\perp\rangle$. This gives us the best estimate of the input state given the result of the measurement.

To evaluate the fidelity of the estimation we present a statistical picture of the measurement. Firstly, let us average over all possible orientations of the measurement apparatus. In order to do so let us write down a single-qubit density operator

$$\rho^{(\mathrm{meas})}(\eta) = |\langle\psi|\eta\rangle|^2\,|\eta\rangle\langle\eta| + |\langle\psi|\eta^\perp\rangle|^2\,|\eta^\perp\rangle\langle\eta^\perp|\,. \tag{42}$$

which describes statistics of the measurements for a given orientation of the measurement apparatus. To get the final output density matrix we average (42) over all possible choices of the measurement (i.e. over all vectors $|\eta\rangle$)

$$\rho^{(\mathrm{est})} = \int \mathrm{d}\Omega_\eta \rho^{(\mathrm{meas})}(\eta)\,, \tag{43}$$

where $\mathrm{d}\Omega_\eta = (1/4\pi)\ \sin\vartheta'\ \mathrm{d}\vartheta'\ \mathrm{d}\varphi'$ is the integration measure on the state space of the 'measurement' apparatus. After the integration is performed we find

$$\rho^{(\mathrm{est})} = s\rho + \frac{1 - s}{2}\mathbb{1}\,, \tag{44}$$

where for a single input qubit we have $s = 1/3$ and $\rho = |\psi\rangle\langle\psi|$.

In order to find the mean fidelity of the estimation itself we have to average the fidelity, i.e. $\langle\psi|\rho^{(\mathrm{est})}|\psi\rangle$ over all possible preparations, i.e.

$$\overline{\mathcal{F}} = \int \mathrm{d}\Omega_\rho \langle\psi|\rho^{(\mathrm{out})}|\psi\rangle = \frac{2}{3}\,. \tag{45}$$

Obviously, instead of projective measurements one can consider some other optimal generalised measurement to be performed on the input qubit. We can even consider a continuous POVM. Nevertheless, since in the given case the projective measurement, described above, is the optimal one, no other measurement can give us more information about the input state $|\psi\rangle$.

Now it is clear that the quantum cloning can be represented as a specific generalised POVM measurement. It is a particular physical realisation of the Naimark theorem [36] – the information contained in the original qubit (i.e. the state $|\psi\rangle$) is spread among many clones. But when the optimal measurement on these clones is performed [42] the mean fidelity of the estimation is again equal to 2/3. In other words we cannot generate information via cloning. The argument can be generalised when the optimal $N \to N + M$ cloning is considered [43,44]. Information about the input qubit(s) cannot be 'generated'. It only can be redistributed [5].

Appendix 2. Proof of the Nielsen–Chuang no-go theorem

We want to prove the no-go theorem for deterministic programmable quantum processors [22]. We assume the processor is represented by a unitary operator, G, acting $\mathcal{H}_\mathrm{d} \otimes \mathcal{H}_\mathrm{p}$, where $\mathcal{H}_\mathrm{d}$ is the data space and $\mathcal{H}_\mathrm{p}$ is the program space. We suppose that we have a program $|\Xi_1\rangle_\mathrm{p} \in \mathcal{H}_\mathrm{p}$ that implements the unitary operator U_1 on $\mathcal{H}_\mathrm{d}$, in particular

$$G(|\psi\rangle_\mathrm{d} \otimes |\Xi_1\rangle_\mathrm{p}) = U_1|\psi\rangle_\mathrm{d} \otimes |\Xi_1'\rangle_\mathrm{p}. \tag{46}$$

Now it could be the case that the output in the program space depends on the state $|\psi\rangle_d$ that is sent into the data input. In order to show that this is not the case, assume that

$$\begin{aligned} G(|\psi_1\rangle_\mathrm{d} \otimes |\Xi_1\rangle_\mathrm{p}) &= U_1|\psi\rangle_\mathrm{d} \otimes |\Xi_1'\rangle_\mathrm{p};\\ G(|\psi_2\rangle_\mathrm{d} \otimes |\Xi_1\rangle_\mathrm{p}) &= U_1|\psi\rangle_\mathrm{d} \otimes |\Xi_1''\rangle_\mathrm{p}. \end{aligned} \tag{47}$$

Taking the inner products of the left-hand sides of the above equations and equating that to the inner product of the right-hand sides, and assuming that $\langle\psi_1|\psi_2\rangle \neq 0$, gives us $\langle\Xi_1'|\,\Xi_1''\rangle = 1$, thereby implying that the program state outputs are identical.

Now suppose that the program state $|\Xi_1\rangle$ implements the operator U_1 and the program state $|\Xi_2\rangle$ implements U_2. We then have that

$$\begin{aligned} G(|\psi\rangle_\mathrm{d} \otimes |\Xi_1\rangle_\mathrm{p}) &= U_1|\psi\rangle_\mathrm{d} \otimes |\Xi_1'\rangle_\mathrm{p};\\ G(|\psi\rangle_\mathrm{d} \otimes |\Xi_2\rangle_\mathrm{p}) &= U_2|\psi\rangle_\mathrm{d} \otimes |\Xi_2'\rangle_\mathrm{p}. \end{aligned} \tag{48}$$

Taking inner products we find

$$\langle \Xi_2|\Xi_1\rangle = \langle\psi|U_2^{-1}U_1|\psi\rangle\langle\Xi_2'|\Xi_1'\rangle. \tag{49}$$

We will examine both the case $\langle\Xi_2'|\Xi_1'\rangle \neq 0$ and the case $\langle\Xi_2'|\Xi_1'\rangle = 0$. If $\langle\Xi_2'|\Xi_1'\rangle \neq 0$, we have

$$\frac{\langle\Xi_2|\Xi_1\rangle}{\langle\Xi_2'|\Xi_1'\rangle} = \langle\psi|U_2^{-1}U_1|\psi\rangle, \tag{50}$$

and we note that the left-hand side does not depend on $|\psi\rangle$, so the right-hand side cannot either. That implies that $U_2^{-1}U_1$ is a multiple of the identity, and since both of the operators are unitary, we must have $U_2 = \exp(i\theta)U_1$ for some θ between 0 and 2π. Now if, on the other hand, $\langle\Xi_2'|\Xi_1'\rangle = 0$, then we see that we must also have that $\langle\Xi_2|\Xi_1\rangle = 0$. Summarising, what we have found is that if U_1 and U_2 are different, that is, they are not multiples of each other, then they must correspond to orthogonal program states. Therefore, the dimension of the program space must be at least as great as the number of unitary operators that the processor can perform.

Similar reasoning can be employed to show that a deterministic scheme employing measurement is also impossible. We can call this a measure-and-correct scheme. Suppose that we send a program and data into our processor, and at the output measure the program state in a fixed basis. Each measurement outcome corresponds to a different unitary operator being applied to the data state, but for each program state the resulting operators are related to each other in the same way. That means that for any program state, if we do not obtain the desired measurement result, we can correct the resulting output state by applying an operator that does not depend on the program state.

Let us look at a simple example. Suppose that both the data and program spaces are two-dimensional, and that our processor acts as follows:

$$G(|\psi\rangle_d \otimes |\Xi_1\rangle_p) = \frac{1}{2^{1/2}}(U_1|\psi\rangle_d \otimes |0\rangle_p + VU_1|\psi\rangle_d \otimes |1\rangle_p);$$
$$G(|\psi\rangle_d \otimes |\Xi_2\rangle_p) = \frac{1}{2^{1/2}}(U_2|\psi\rangle_d \otimes |0\rangle_p + VU_2|\psi\rangle_d \otimes |1\rangle_p). \tag{51}$$

Here, V is a fixed unitary operator. Such a processor is capable of deterministically applying four different unitary operators to the data state, U_1, VU_1, U_2 and VU_2. For example, suppose we want to apply U_1. We use the program $|\Xi_1\rangle$, and then measure the program state in the basis $\{|0\rangle, |1\rangle\}$. If we obtain $|0\rangle$ we are done, and if we obtain $|1\rangle$, then we can apply V^{-1} to the data state. In either case, we obtain the output state $U_1|\psi\rangle_d$. We will also be able to deterministically obtain the superpositions $c_1U_1 + c_2U_2$ and $c_1VU_1 + c_2VU_2$, where c_1 and c_2 are complex numbers. It appears that we have beaten the no-go theorem, because we are able to deterministically realise four unitary operators with a two-dimensional program space. Unfortunately, it will not work. If we take the inner products of the two equations above, we find that

$$\langle\Xi_1|\Xi_2\rangle = \langle\psi|U^{-1}U_2|\psi\rangle. \tag{52}$$

The left-hand side does not depend on $|\psi\rangle$, which, as before, implies that U_1 and U_2 are related by a phase factor, and that the program states are multiples of each other. Therefore, we can only realise two operators in this way, U_1 and VU_1, and we have not gained anything.