Three Approaches to the Quantitative Definition of Information in an Individual Pure Quantum State

Paul Vitanyi*
CWI and University of Amsterdam

Abstract

In analogy of classical Kolmogorov complexity we develop a theory of the algorithmic information in bits contained in any one of continuously many pure quantum states: quantum Kolmogorov complexity. Classical Kolmogorov complexity coincides with the new quantum Kolmogorov complexity restricted to the classical domain. Quantum Kolmogorov complexity is upper bounded and can be effectively approximated from above. With high probability a quantum object is incompressible. There are two alternative approaches possible: to define the complexity as the length of the shortest qubit program that effectively describes the object, and to use classical descriptions with computable real parameters.

1 Introduction

While Kolmogorov complexity is the accepted absolute measure of information content in a *classical* individual finite object, a similar absolute notion is needed for the information content of a pure quantum state. [1] Quantum theory assumes that every complex vector, except the null vector, represents a realizable pure quantum state.[2] This leaves open the question of how to design the equipment that prepares such a pure state. While there are continuously many pure states in a finite-dimensional complex vector space—corresponding to all vectors of unit length—we can finitely describe only a countable subset. Imposing effectiveness on such descriptions leads to constructive procedures. The most general such procedures satisfying universally agreed-upon logical principles of effectiveness are quantum Turing machines, [2]. To define quantum Kolmogorov complexity by way of quantum Turing machines leaves essentially two options:

1. We want to describe every quantum superposition exactly; or
2. we want to take into account the number of bits/qubits in the specification as well the accuracy of the quantum state produced.

We have to deal with three problems:

- There are continuously many quantum Turing machines;
- There are continously many pure quantum states;
- There are continuously many qubit descriptions.

There are uncountably many quantum Turing machines only if we allow arbitrary real rotations in the definition of machines. Then, a quantum Turing machine can only be universal in the sense that it can approximate the computation of an arbitrary machine, [2]. In descriptions using universal quantum Turing machines we would have to account for the closeness of approximation, the number of steps required to get this precision, and the like. In contrast, if we fix the rotation of all contemplated machines to a single primitive rotation θ with $\cos\theta = 3/5$ and $\sin\theta = 4/5$ then there are only countably many Turing machines and the universal machine simulates the others exactly [1]. Every quantum Turing machine computation using arbitrary real rotations can be approximated to any precision by machines with fixed rotation θ but in general cannot be simulated exactly—just like in the case of the simulation of arbitrary quantum Turing machines by a universal quantum Turing machine. Since exact simulation is impossible by a fixed universal quantum Turing machine anyhow, but arbitrarily close approximations are possible by Turing machines using a fixed rotation like θ, we are motivated to fix $Q_1, Q_2, \ldots$ as a standard enumeration of quantum Turing machines using only rotation θ.

Our next question is whether we want programs (descriptions) to be in classical bits or in qubits? The intuitive

*Partially supported by the EU fifth framework project QAIP, IST–1999–11234, the NoE QUIPROCONE IST–1999–29064, the ESF QiT Programmme, and ESPRIT BRA IV NeuroCOLT II Working Group EP 27150. Part of this work was done during the author's 1998 stay at Tokyo Institute of Technology, Tokyo, Japan, as Gaikoku-Jin Kenkyuin at INCOCSAT. A preliminary version was archived as quant-ph/9907035.
Address: CWI, Kruislaan 413, 1098 SJ Amsterdam, The Netherlands. Email: paulv@cwi.nl

[1] For definitions and theory of Kolmogorov complexity consult [4], and for quantum theory consult [5].

[2] That is, every complex vector that can be normalized to unit length.

0-7695-0674-7/00 $10.00 © 2000 IEEE

notion of computability requires the programs to be classical. Namely, to prepare a quantum state requires a physical apparatus that "computes" this quantum state from classical specifications. Since such specifications have effective descriptions, every quantum state that can be prepared can be described effectively in descriptions consisting of classical bits. Descriptions consisting of arbitrary pure quantum states allows noncomputable (or hard to compute) information to be hidden in the bits of the amplitudes. In Definition 2 we call a pure quantum state *directly computable* if there is a (classical) program such that the universal quantum Turing machine computes that state from the program and then halts in an appropriate fashion. In a computational setting we naturally require that directly computable pure quantum states can be prepared. By repeating the preparation we can obtain arbitrarily many copies of the pure quantum state. [3] Restricting ourselves to an effective enumeration of quantum Turing machines and classical descriptions to describe by approximation continuously many pure quantum states is reminiscent of the construction of continuously many real numbers from Cauchy sequences of rational numbers, the rationals being effectively enumerable.

The second approach considers the shortest effective qubit description of a pure quantum state. This can also be properly formulated in terms of the conditional version of the first approach. An advantage of this version is that the upper bound on the complexity of a pure quantum state is immediately given by the number of qubits involved in the literal description of that pure quantum state. The status of incompressibility and degree of uncomputability is as yet unknown and potentially a source of problems with this approach.

The third approach is to give programs for the 2^{n+1} real numbers involved in the precise description of the n-qubit state. Then the question reduces to the problem of describing lists of real numbers.

In the classical situation there are also several variants of Kolmogorov complexity that are very meaningful in their respective settings: plain Kolmogorov complexity, prefix complexity, monotone complexity, uniform complexity, negative logarithm of universal measure, and so on [4]. It is therefore not surprising that in the more complicated situation of quantum information several different choices of complexity can be meaningful and unavoidable in different settings.

[3] See the discussion in [5], pp. 49–51. If descriptions are not effective then we are not going to use them in our algorithms except possibly on inputs from an "unprepared" origin. Every quantum state used in a quantum computation arises from some classically preparation or is possibly captured from some unknown origin. If the latter, then we can consume it as conditional side-information or an oracle.

2 Classical Descriptions

The complex quantity $\langle x|z\rangle$ is the inner product of vectors $|x\rangle$ and $|z\rangle$. Since pure quantum states $|x\rangle, |z\rangle$ have unit length, $|\langle x|z\rangle| = |\cos\theta|$ where θ is the angle between vectors $|x\rangle$ and $|z\rangle$ and $|\langle x|z\rangle|^2$ is the probability of outcome $|x\rangle$ being measured from state $|z\rangle$, [5]. The idea is as follows. A *von Neumann measurement* is a decomposition of the Hilbert space into subspaces that are mutually orthogonal, for example an orthonormal basis is an observable. Physicists like to specify observables as Hermitian matrices, where the understanding is that the eigenspaces of the matrices (which will always be orthogonal) are the actual subspaces. When a measurement is performed, the state is projected into one of the subspaces (with probability equal to the square of the projection). So the subspaces correspond to the possible *outcomes* of a measurement. In the above case we project $|z\rangle$ on outcome $|x\rangle$ using projection $|x\rangle\langle x|$ resulting in $\langle x|z\rangle|x\rangle$.

Our model of computation is a quantum Turing machine with classical binary program p on the input tape and a quantum auxiliary input on a special conditional input facility. We think of this auxiliary input as being given as a pure quantum state $|y\rangle$ (in which case it can be used only once), as a mixture density matrix ρ, or (perhaps partially) as a classical program from which it can be computed. In the last case, the classical program can of course be used indefinitely often.[4] It is therefore not only important *what* information is given conditionally, but also *how* it is described—like this is the sometimes the case in the classical version of Kolmogorov complexity for other reasons that would additionally hold in the quantum case. We impose the condition that the set of *halting programs* $\mathcal{P}_y = \{p : T(p|y) < \infty\}$ is *prefix-free*: no program in $\mathcal{P}_y$ is a proper prefix of another program in $\mathcal{P}_y$. Put differently, the Turing machine scans all of a halting program p but never scans the bit following the last bit of p: it is *self-delimiting*. [5] [6]

[4] We can even allow that the conditional information y is infinite or noncomputable, or an oracle. But we will not need this in the present paper.

[5] One can also use a model were the input p is delimited by distinguished markers. Then the Turing machine always knows where the input ends. In the self-delimiting case the endmarker must be implicit in the halting program p itself. This encoding of the endmarker carries an inherent penalty in the form of increased length: typically a prefix code of an n-length binary string has length about $n + \log n + 2\log\log n$ bits, [4].

[6] There are two possible interpretations for the computation relation $Q(p, y) = |x\rangle$. In the narrow interpretation we require that Q with p on the input tape and y on the conditional tape halts with $|x\rangle$ on the output tape. In the wide interpretation we can define pure quantum states by requiring that for every precision $\delta > 0$ the computation of Q with p on the input tape and y on the conditional tape and δ on a tape where the precision is to be supplied halts with $|x'\rangle$ on the output tape and $|\langle x|x'\rangle|^2 \geq 1 - \delta$. Such a notion of "computable" or "recursive" pure quantum states is similar to Turing's notion of "computable numbers." In

0-7695-0674-7/00 $10.00 © 2000 IEEE

DEFINITION 1 The *(self-delimiting) complexity* of $|x\rangle$ with respect to quantum Turing machine Q with y as conditional input given for free is

$$K_Q(|x\rangle|y) := \min_p\{l(p)+\lceil -\log(|\langle z|x\rangle|^2)\rceil : Q(p,y) = |z\rangle\}$$

where $l(p)$ is the number of bits in the specification p, y is an input quantum state and $|z\rangle$ is the quantum state produced by the computation $Q(p,y)$, and $|x\rangle$ is the target state that one is trying to describe.

THEOREM 1 *There is a universal machine* [7] *U such that for all machines Q there is a constant c_Q (the length of the description of the index of Q in the enumeration) such that for all quantum states $|x\rangle$ we have $K_U(|x\rangle|y) \leq K_Q(|x\rangle|y) + c_Q$.*

PROOF. There is a universal quantum Turing machine U in the standard enumeration $Q_1, Q_2, \ldots$ such that for every quantum Turing machine Q in the enumeration there is a self-delimiting program i_Q (the index of Q) and $U(i_Q p, y) = Q(p,y)$ for all p, y. Setting $c_Q = l(i_Q)$ proves the theorem. □

We fix once and for all a reference universal quantum Turing machine U and define the quantum Kolmogorov complexity as

$$K(|x\rangle|y) := K_U(|x\rangle|y),$$
$$K(|x\rangle) := K_U(|x\rangle|\epsilon),$$

where ϵ denotes the absence of any conditional information. The definition is continuous: If two quantum states are very close then their quantum Kolmogorov complexities are very close. Furthermore, since we can approximate every (pure quantum) state $|x\rangle$ to arbitrary closeness, [2], in particular, for every constant $\epsilon > 0$ we can compute a (pure quantum) state $|z\rangle$ such that $|\langle z|x\rangle|^2 > 1 - \epsilon$. [8] For this definition to be useful it should satisfy:

- The complexity of a pure state that can be directly computed should be the length of the shortest program that computes that state. (If the complexity is less then this may lead to discontinuities when we restrict quantum Kolmogorov complexity to the domain of classical objects.)
- The quantum Kolmogorov complexity of a classical object should equal the classical Kolmogorov complexity of that object (up to a constant additive term).
- The quantum Kolmogorov complexity of a quantum object should have an upper bound. (This is necessary for the complexity to be approximable from above, even if the quantum object is available in as many copies as we require.)
- Most objects should be "incompressible" in terms of quantum Kolmogorov complexity.
- In a probabilistic ensemble the expected quantum Kolmogorov complexity should be about equal (or have another meaningful relation) to the von Neumann entropy. [9]

For a quantum system $|z\rangle$ the quantity $P(x) := |\langle z|x\rangle|^2$ is the probability that the system passes a test for $|x\rangle$, and vice versa. The term $\lceil -\log(|\langle z|x\rangle|^2)\rceil$ can be viewed as the code word length to redescribe $|x\rangle$ given $|z\rangle$ and an orthonormal basis with $|x\rangle$ as one of the basis vectors using the well-known Shannon-Fano prefix code. This works as follows: For every state $|z\rangle$ in $N := 2^n$-dimensional Hilbert space with basis vectors $\mathcal{B} = \{|e_0\rangle, \ldots, |e_{N-1}\rangle\}$ we have $\sum_{i=0}^{N-1} |\langle e_i|z\rangle|^2 = 1$. If the basis has $|x\rangle$ as one of the basis vectors, then we can consider $|z\rangle$ as a random variable that assumes value $|x\rangle$ with probability $|\langle x|z\rangle|^2$. The Shannon-Fano code word for $|x\rangle$ in the probabilistic ensemble $\mathcal{B}, (|\langle e_i|z\rangle|^2)_i$ is based on the probability $|\langle x|z\rangle|^2$ of $|x\rangle$ given $|z\rangle$ and has length $\lceil -\log(|\langle x|z\rangle|^2)\rceil$. Considering a canonical method of constructing an orthonormal basis $\mathcal{B} = |e_0\rangle, \ldots, |e_{N-1}\rangle$ from a given basis vector, we can choose $\mathcal{B}$ such that $K(\mathcal{B}) = \min_i\{K(|e_i\rangle)\} + O(1)$. The Shannon-Fano code is appropriate for our purpose since it is optimal in that it achieves the least expected code word length—the expectation taken over the probability of the source words—up to 1 bit by Shannon's Noiseless Coding Theorem.

2.1 Consistency with Classical Complexity

Our proposal would not be useful if it were the case that for a directly computable object the complexity is less than the shortest program to compute that object. This would imply that the code corresponding to the probabilistic component in the description is possibly shorter than the difference in program lengths for programs for an approximation of the object and the object itself. This would penalize definite description compared to probabilistic description and in case of classical objects would make quantum Kolmogorov complexity less than classical Kolmogorov complexity.

the remainder of this section we use the narrow interpretation.

[7] We use "U" to denote a universal (quantum) Turing machine rather than a unitary matrix.

[8] We can view this as the probability of the possibly noncomputable outcome $|x\rangle$ when executing projection $|x\rangle\langle x|$ on $|z\rangle$ and measuring outcome $|x\rangle$.

[9] In the classical case the average self-delimiting Kolmogorov complexity equals the Shannon entropy up to an additive constant depending on the complexity of the distribution concerned.

THEOREM 2 *Let U be the reference universal quantum Turing machine and let $|x\rangle$ be a basis vector in a directly computable orthonormal basis $\mathcal{B}$ given y: there is a program p such that $U(p,y) = |x\rangle$. Then $K(|x\rangle|y) = \min_p\{l(p) : U(p,y) = |x\rangle\}$ up to $K(\mathcal{B}|y) + O(1)$.*

PROOF. Let $|z\rangle$ be such that

$$K(|x\rangle|y) = \min_q\{l(q) + \lceil -\log(|\langle z|x\rangle|^2)\rceil : U(q,y) = |z\rangle\}.$$

Denote the program q that minimizes the righthand side by $q_{\min}$ and the program p that minimizes the expression in the statement of the theorem by $p_{\min}$.

By running U on all binary strings (candidate programs) simultaneously dovetailed-fashion [10] one can enumerate all objects that are directly computable given y in order of their halting programs. Assume that U is also given a $K(\mathcal{B}|y)$ length program b to compute $\mathcal{B}$—that is, enumerate the basis vectors in $\mathcal{B}$. This way $q_{\min}$ computes $|z\rangle$, the program b computes $\mathcal{B}$. Now since the vectors of $\mathcal{B}$ are mutually orthogonal

$$\sum_{|e\rangle\in\mathcal{B}} |\langle z|e\rangle|^2 = 1.$$

Since $|x\rangle$ is one of the basis vectors we have $-\log|\langle z|x\rangle|^2$ is the length of a prefix code (the Shannon-Fano code) to compute $|x\rangle$ from $|z\rangle$ and $\mathcal{B}$. Denoting this code by r we have that the concatenation $q_{\min}br$ is a program to compute $|x\rangle$: parse it into $q_{\min}, b$, and r using the self-delimiting property of $q_{\min}$ and b. Use $q_{\min}$ to compute $|z\rangle$ and use b to compute $\mathcal{B}$, determine the probabilities $|\langle z|e\rangle|^2$ for all basis vectors $|e\rangle$ in $\mathcal{B}$. Determine the Shannon-Fano code words for all the basis vectors from these probabilities. Since r is the code word for $|x\rangle$ we can now decode $|x\rangle$. Therefore,

$$l(q_{\min}) + \lceil -\log(|\langle z|x\rangle|^2)\rceil \geq l(p_{\min}) - K(\mathcal{B}|y) - O(1)$$

which was what we had to prove. □

COROLLARY 1 On classical objects (that is, the natural numbers or finite binary strings that are all directly computable) the quantum Kolmogorov complexity coincides up to a fixed additional constant with the self-delimiting Kolmogorov complexity since $K(\mathcal{B}|n) = O(1)$ for the standard classical basis $\mathcal{B} = \{0,1\}^n$. [11] (We assume that the information about the dimensionality of the Hilbert space is given conditionally.)

REMARK 1 Fixed additional constants are no problem since the complexity also varies by fixed additional constants due to the choice of reference universal Turing machine. ◇

2.2 Upper Bound on Complexity

A priori, in the worst case $K(|x\rangle|n)$ is possibly ∞. We show that the worst-case has a $2n$ upper bound.

LEMMA 1 *For all n-qubit quantum states $|x\rangle$ we have $K(|x\rangle|n) \leq 2n + O(1)$.*

PROOF. For every state $|x\rangle$ in $N := 2^n$-dimensional Hilbert space with basis vectors $|e_0\rangle, \ldots, |e_{N-1}\rangle$ we have $\sum_{i=0}^{N-1} |\langle e_i|x\rangle|^2 = 1$. Hence there is an i such that $|\langle e_i|x\rangle|^2 \geq 1/N$. Let p be a $K(i|n) + O(1)$-bit program to construct a basis state $|e_i\rangle$ given n. Then $l(p) \leq n + O(1)$. Then $K(|x\rangle|n) \leq l(p) - \log(1/N) \leq 2n + O(1)$. □

2.3 Computability

In the classical case Kolmogorov complexity is not computable but can be approximated from above by a computable process. The non-cloning property prevents us from copying an unknown pure quantum state given to us. Therefore, an approximation from above that requires checking every output state against the target state destroys the latter. To overcome the fragility of the pure quantum target state one has to postulate that it is available as an outcome in a measurement.

THEOREM 3 *Let $|x\rangle$ be the pure quantum state we want to describe.*

(i) *The quantum Kolmogorov complexity $K(|x\rangle)$ is not computable.*

(ii) *If we can repeatedly execute the projection $|x\rangle\langle x|$ and perform a measurement with outcome $|x\rangle$, then the quantum Kolmogorov complexity $K(|x\rangle)$ can be approximated from above by a computable process with arbitrarily small probability of error α of giving a too small value.*

PROOF. The uncomputability follows a fortiori from the classical case. The semicomputability follows because we have established an upper bound on the quantum Kolmogorov complexity, and we can simply enumerate all halting classical programs up to that length by running their computations dovetailed fashion. The idea is

[10] A *dovetailed* computation is a method related to Cantor's diagonalization to run all programs alternatingly in such a way that every program eventually makes progress. On an list of programs $p_1, p_2, \ldots$ one divides the overall computation into stages $k := 1, 2, \ldots$. In stage k of the overall computation one executes the ith computation step of every program p_{k-i+1} for $i := 1, \ldots, k$.

[11] This proof does not show that it coincide up to an additive constant term with the original plain complexity defined by Kolmogorov, [4], based on Turing machines where the input is delited by distinguished markers. The same proof for the plain Kolmogorov complexity shows that it coincides up to a logarithmic additive term.

as follows: Let the target state be $|x\rangle$ of n qubits. Then, $K(|x\rangle|n) \leq 2n + O(1)$. (The unconditional case $K(|x\rangle)$ is similar with $2n$ replaced by $2(n + \log n)$.) We want to identify a program x^* such that $p := x^*$ minimizes $l(p) - \log|\langle x|U(p,n)\rangle|^2$ among all candidate programs. To identify it in the limit, for some fixed k satisfying (2) below for given n, α, ϵ, repeat the computation of every halting program p with $l(p) \leq 2n + O(1)$ at least k times and perform the assumed projection and measurement. For every halting program p in the dovetailing process we estimate the probability $q := |\langle x|U(p,n)\rangle|^2$ from the fraction m/k: the fraction of m positive outcomes out of k measurements. The probability that the estimate m/k is off from the real value q by more than an ϵq is given by Chernoff's bound: for $0 \leq \epsilon \leq 1$,

$$P(|m - qk| > \epsilon qk) \leq 2e^{-\epsilon^2 qk/3}. \quad (1)$$

This means that the probability that the deviation $|m/k - q|$ exceeds ϵq vanishes exponentially with growing k. Every candidate program p satisfies (1) with its own q or $1 - q$. There are $O(2^{2n})$ candidate programs p and hence also $O(2^{2n})$ outcomes $U(p,n)$ with halting computations. We use this estimate to upper bound the probability of error α. For given k, the probability that *some* halting candidate program p satisfies $|m - qk| > \epsilon qk$ is at most α with

$$\alpha \leq \sum_{U(p,n)<\infty} 2e^{-\epsilon^2 qk/3}.$$

The probability that *no* halting program does so is at least $1 - \alpha$. That is, with probability at least $1 - \alpha$ we have

$$(1 - \epsilon)q \leq \frac{m}{k} \leq (1 + \epsilon)q$$

for every halting program p. It is convenient to restrict attention to the case that all q's are large. Without loss of generality, if $q < \frac{1}{2}$ then consider $1 - q$ instead of q. Then,

$$\log \alpha \leq 2n - (\epsilon^2 k \log e)/6 + O(1). \quad (2)$$

The approximation algorithm is as follows:

Step 0: Set the required degree of approximation $\epsilon < 1/2$ and the number of trials k to achieve the required probability of error α.

Step 1: Dovetail the running of all candidate programs until the next halting program is enumerated. Repeat the computation of the new halting program k times

Step 2: If there is more than one program p that achieves the current minimum then choose the program with the smaller length (and hence least number of successfull observations). If p is the selected program with m successes out of k trials then set the current approximation of $K(|x\rangle)$ to

$$l(p) - \log \frac{m}{(1 + \epsilon)k}.$$

This exceeds the proper value of the approximation based on the real q instead of m/k by at most 1 bit for all $\epsilon < 1$.

Step 3: Goto **Step 1**. □

2.4 Incompressibility

DEFINITION 2 A pure quantum state $|x\rangle$ is *computable* if $K(|x\rangle) < \infty$. Hence all finite-dimensional pure quantum states are computable. We call a pure quantum state *directly computable* if there is a program p such that $U(p) = |x\rangle$.

The standard orthonormal basis—consisting of all n-bit strings—of the 2^n-dimensional Hilbert space $\mathcal{H}_N$ has at least $2^n(1-2^{-c})$ basis vectors $|e_i\rangle$ that satisfy $K(|e_i\rangle|n) \geq n - c$. This is the standard counting argument in [4]. But what about nonclassical orthonormal bases?

LEMMA 2 *There is a (possibly nonclassical) orthonormal basis of the 2^n-dimensional Hilbert space $\mathcal{H}_N$ such that at least $2^n(1 - 2^{-c})$ basis vectors $|e_i\rangle$ satisfy $K(|e_i\rangle|n) \geq n - c$.*

PROOF. Every orthonormal basis of $\mathcal{H}_N$ has 2^n basis vectors and there are at most $m \leq \sum_{i=0}^{n-c-1} 2^i = 2^{n-c} - 1$ programs of length less than $n - c$. Hence there are at most m programs available to approximate the basis vectors. We construct an orthonormal basis satisfying the lemma: The set of directly computed pure quantum states $|x_0\rangle, \ldots, |x_{m-1}\rangle$ span an m'-dimensional subspace $\mathcal{A}$ with $m' \leq m$ in the 2^n-dimensional Hilbert space $\mathcal{H}_N$ such that $\mathcal{H}_N = \mathcal{A} \oplus \mathcal{A}^\perp$. Here $\mathcal{A}^\perp$ is a $(2^n - m')$-dimensional subspace of $\mathcal{H}_N$ such that every vector in it is perpendicular to every vector in $\mathcal{A}$. We can write every element $|x\rangle \in \mathcal{H}_N$ as

$$\sum_{i=0}^{m'-1} \alpha_i |a_i\rangle + \sum_{i=0}^{2^n - m' - 1} \beta_i |b_i\rangle$$

where the $|a_i\rangle$'s form an orthonormal basis of $\mathcal{A}$ and the $|b_i\rangle$'s form an orthonormal basis of $\mathcal{A}^\perp$ so that the $|a_i\rangle$'s and $|b_i\rangle$'s form an orthonormal basis K for $\mathcal{H}_N$. For every directly computable state $|x_j\rangle \in \mathcal{A}$ and basis vector $|b_i\rangle \in A^\perp$ we have $|\langle x_j|b_i\rangle|^2 = 0$ implying that $K(|x_j\rangle|n) - \log|\langle x_j|b_i\rangle|^2 = \infty$ and therefore $K(|b_i\rangle|n) > n - c$ ($0 \leq j < m, 0 \leq i < 2^n - m'$). This proves the lemma. □

We generalize this lemma to arbitrary bases:

THEOREM 4 *Every orthonormal basis $|e_0\rangle, \ldots, |e_{2^n-1}\rangle$ of the 2^n-dimensional Hilbert space $\mathcal{H}_N$ has at least $2^n(1 - 2^{-c})$ basis vectors $|e_i\rangle$ that satisfy $K(|e_i\rangle|n) \geq n - c$.*

PROOF. Use the notation of the proof of Lemma 2. Assume to the contrary that there are $> 2^{n-c}$ basis vectors

$|e_i\rangle$ with $K(|e_i\rangle|n) < n - c$. Then at least two of them, say $|e_0\rangle$ and $|e_1\rangle$ and some pure quantum state $|x\rangle$ directly computed from a $< (n-c)$-length program satisfy

$$K(|e_i\rangle|n) = K(|x\rangle|n) + \lceil -\log|\langle e_i|x\rangle|^2 \rceil. \quad (3)$$

$(i = 0, 1)$. This means that $K(|x\rangle|n) < n - c - 1$ since not both $|e_0\rangle$ and $|e_1\rangle$ can be equal to $|x\rangle$. Hence for every directly computed pure quantum state of complexity $n-c-1$ there is at most one basis state of the same complexity (in fact only if that basis state is identical with the directly computed state.) Now eliminate all directly computed pure quantum states $|x\rangle$ of complexity $n - c - 1$ together with the basis states $|e\rangle$ that stand in relation Equation 3. We are now left with $> 2^{n-c-1}$ basis states that stand in relation of Equation 3 with the remaining at most $2^{n-c-1} - 1$ remaining directly computable pure quantum states of complexity $\leq n - c - 2$. Repeating the same argument we end up with > 1 basis vector that stand in relation of Equation 3 with 0 directly computable pure quantum states of complexity ≤ 0 which is impossible. □

COROLLARY 2 *The uniform probability* $\Pr\{|x\rangle : K(|x\rangle|n) \geq n - c\} \geq 1 - 1/2^c$.

EXAMPLE 1 We elucidate the role of the $-\log|\langle x|z\rangle|^2$ term. Let x be a random classical string with $K(x) \geq l(x)$ and let y be a string obtained from x by complementing one bit. It is known (Exercise 2.2.8 in [4]) that for every such x of length n there is such a y with complexity $K(y|n) = n - \log n + O(1)$. Now let $|z\rangle$ be a pure quantum state which has classical bits except the difference bit between x and y that has equal probabilities of being observed as "1" and as "0." We can prepare $|z\rangle$ by giving y and the position of the difference bit (in $\log n$ bits) and therefore $K(|z\rangle|n) \leq n + O(1)$. Since from $|z\rangle$ we have probability $\frac{1}{2}$ of obtaining x by observing the particular bit in superposition and $K(x|n) \geq n$ it follows $K(|z\rangle|n) \geq n + O(1)$ and therefore $K(|z\rangle|n) = n + O(1)$. From $|z\rangle$ we have probability $\frac{1}{2}$ of obtaining y by observing the particular bit in superposition which (correctly) yields that $K(y|n) \leq n + O(1)$. ◇

2.5 Conditional Complexity

We have used the conditional complexity $K(|x\rangle|y)$ to mean the minimum sum of the length of a classical program to compute $|z\rangle$ plus the negative logarithm of the probability of outcome $|x\rangle$ when executing projection $|x\rangle\langle x|$ on $|z\rangle$ and measuring, given the pure quantum state y as input on a separate input tape. In the quantum situation the notion of inputs consisting of pure quantum states is subject to very special rules.

Firstly, if we are given an unknown pure quantum state $|y\rangle$ as input it can be used only once, that is, it is irrevocably consumed and lost in the computation. It cannot be copied or cloned without destroying the original [5]. This phenomenon is subject to the so-called *no-cloning theorem* and means that there is a profound difference between giving a directly computable pure quantum state as a classical program or giving it literally. Given as a classical program we can prepare and use arbitrarily many copies of it. Given as an (unknown) pure quantum state in superposition it can be used as start of a computation only once—unless of course we deal with an identity computation in which the input state is simply transported to the output state. This latter computation nonetheless destroys the input state.

If an unknown state $|y\rangle$ is given as input (in the conditional for example) then the no-cloning theorem of quantum computing says it can be used only *once*. Thus, for a non-classical pure quantum state $|x\rangle$ we have

$$K(|x\rangle, |x\rangle||x\rangle) \leq K(|x\rangle) + O(1)$$

rather than $K(x, x|x) = O(1)$ as in the case for classical objects x. This holds even if $|x\rangle$ is directly computable but is given in the conditional in the form of an unknown pure quantum state. However, if $|x\rangle$ is directly computable and the conditional is a classical program to compute this directly computable state, then that program can be used over and over again.

In the previous example, if the conditional $|x\rangle$ is directly computable, for example by a classical program p, then we have both $K(|x\rangle|p) = O(1)$ and $K(|x\rangle, |x\rangle|p) = O(1)$. In particular, for a classical program p that computes a directly computable state $|x\rangle$ we have

$$K(|x\rangle, |x\rangle|p) = O(1).$$

It is important here to notice that a classical program for computing a directly computable quantum state carries *more information* than the directly computable quantum state itself—much like a shortest program for a classical object carries more information than the object itself. In the latter case it consists in partial information about the halting problem. In the quantum case of a directly computable pure state we have the additional information that the state is directly computable *and* in case of a shortest classical program additional information about the halting problem.

2.6 Sub-Additivity

Quantum Kolmogorov complexity of directly computable pure quantum states in simple orthonormal bases is *sub-additive*:

LEMMA 3 *For directly computable $|x\rangle, |y\rangle$ both of which belong to (possibly different) orthonormal bases of Kolmogorov complexity $O(1)$ we have*

$$K(|x\rangle, |y\rangle) \leq K(|x\rangle||y\rangle) + K(|y\rangle)$$

up to an additive constant term.

PROOF. By Theorem 2 we there is a program p_y to compute $|y\rangle$ with $l(p) = K(|y\rangle)$ and a program $p_{y\to x}$ to compute $|x\rangle$ from $|y\rangle$ with $l(p_{y\to x}) = K(|x\rangle||y\rangle)$ up to additional constants. Use p_y to construct two copies of $|y\rangle$ and $p_{y\to x}$ to construct $|x\rangle$ from one of the copies of $|y\rangle$. The separation between these concatenated binary programs is taken care of by the self-delimiting property of the subprograms. The additional constant term takes care of the couple of $O(1)$-bit programs that are required. □

REMARK 2 In the classical case we have equality in the theorem (up to an additive logarithmic term). The proof of the remaining inequality, as given in the classical case, doesn't hold directly for the quantum case. It would require a decision procedure that establishes equality between two pure quantum states without error. While the sub-additivity property holds in case of directly computable states, is easy to see that for the general case of pure states the subadditivity property fails due to the "non-cloning" property. For example for pure states $|x\rangle$ that are not "clonable" we have:

$$K(|x\rangle, |x\rangle) > K(|x\rangle||x\rangle) + K(|x\rangle) = K(|x\rangle) + O(1).$$

◇

We additionally note:

LEMMA 4 *For all directly computable pure states $|x\rangle$ and $|y\rangle$ we have $K(|x\rangle, |y\rangle) \leq K(|y\rangle) - \log |\langle x|y\rangle|^2$ up to an additive logarithmic term.*

PROOF. $K(|x\rangle||y\rangle) \leq -\log |\langle x|y\rangle|^2$ by the proof of Theorem 2. Then, the lemma follows by Lemma 3. □

3 Qubit Descriptions

One way to avoid two-part descriptions as we used above is to allow qubit programs as input. This leads to the following definitions, results, and problems.

DEFINITION 3 The *qubit complexity* of $|x\rangle$ with respect to quantum Turing machine Q with y as conditional input given for free is

$$KQ_Q(|x\rangle|y) := \min_p \{l(|p\rangle) : Q(|p\rangle, y) = |x\rangle\}$$

where $l(|p\rangle)$ is the number of qubits in the qubit specification $|p\rangle$, $|p\rangle$ is an input quantum state, y is given conditionally, and $|x\rangle$ is the quantum state produced by the computation $Q(|p\rangle, y)$: the target state that one describes.

Note that here too there are two possible interpretations for the computation relation $Q(|p\rangle, y) = |x\rangle$. In the narrow interpretation we require that Q with $|p\rangle$ on the input tape and y on the conditional tape halts with $|x\rangle$ on the output tape. In the wide interpretation we require that for every precision $\delta > 0$ the computation of Q with $|p\rangle$ on the input tape and y on the conditional tape and δ on a tape where the precision is to be supplied halts with $|x'\rangle$ on the output tape and $|\langle x|x'\rangle|^2 \geq 1 - \delta$. Additionally one can require that the approximation finishes in a certain time, say, polynomial in $l(|x\rangle)$ and $1/\delta$. In the remainder of this section we can allow either interpretation (note that the "narrow" complexity will always be at least as large as the "wide" complexity). Fix an enumeration of quantum Turing machines like in Theorem 1, this time with Turing machines that use qubit programs. Just like before it is now straightforward to derive an Invariance Theorem:

THEOREM 5 *There is a universal machine U such that for all machines Q there is a constant c (the length of a self-delimiting encoding of the index of Q in the enumeration) such that for all quantum states $|x\rangle$ we have $KQ_U(|x\rangle|y) \leq KQ_Q(|x\rangle|y) + c$.*

We fix once and for all a reference universal quantum Turing machine U and express the *qubit quantum Kolmogorov complexity* as

$$KQ(|x\rangle|y) := KQ_U(|x\rangle|y),$$
$$KQ(|x\rangle) := KQ_U(|x\rangle|\epsilon),$$

where ϵ indicates the absence of conditional information (the conditional tape contains the "quantum state" with 0 qubits). We now have immediately:

LEMMA 5 $KQ(|x\rangle) \leq l(|x\rangle) + O(1)$.

PROOF. Give the reference universal machine $|1^n 0\rangle \otimes |x\rangle$ as input where n is the index of the identity quantum Turing machine that transports the attached pure quantum state $|x\rangle$ to the output. □

It is possible to define unconditional KQ-complexity in terms of conditional K-complexity as follows: Even for pure quantum states that are not directly computable from effective descriptions we have $K(|x\rangle||x\rangle) = O(1)$. This naturaly gives:

LEMMA 6 *The qubit quantum Kolmogorov complexity of $|x\rangle$ satisfies*

$$KQ(|x\rangle) = \min_p \{l(|p\rangle) : K(|x\rangle||p\rangle)\} + O(1),$$

where $l(|p\rangle)$ denotes the number of qubits in $|p\rangle$.

PROOF. Transfer the conditional $|p\rangle$ to the input using an $O(1)$-bit program. □

We can generalize this definition to obtain conditional KQ-complexity.

3.1 Potential Problems of Qubit Complexity

While it is clear that (just as with the previous aproach) the qubit complexity is not computable, it is unknown to the author whether one can approximate the qubit complexity from above by a computable process in any meaningful sense. In particular, the dovetailing approach we used in the first approach now doesn't seem applicable due to the non-countability of the potentential qubit program candidates. While it is clear that the qubit complexity of a pure quantum state is at least 1, why would it need to be more than one qubit since the probability amplitude can be any complex number? In case the target pure quantum state is a classical binary string, as observed by Harry Buhrman, Holevo's theorem [5] tells us that on average one cannot transmit more than n bits of classical information by n-qubit messages (without using entangled qubits on the side). This suggests that for every n there exist classical binary strings of length n that have qubit complexity at least n. This of course leaves open the case of the non-classical pure quantum states—a set of measure one—and of how to prove incompressibility of the overwhelming majority of states. These matters have since been investigated by A. Berthiaume, S. Laplante, and W. van Dam (paper in preparation).

4 Real Descriptions

A final version of quantum Kolmogorov complexity uses computable real parameters to describe the pure quantum state with complex probability amplitudes. This requires two reals per complex probability amplitude, that is, for n qubits one requires 2^{n+1} real numbers in the worst case. Since every computable real number may require a separate program, a computable n qubit state may require 2^{n+1} finite programs. While this approach does not allow the development of a clean theory in the sense of the previous approaches, it can be directly developed in terms of algorithmic thermodynamics—an extension of Kolmogorov complexity to randomness of infinite sequences (such as binary expansions of real numbers) in terms of coarse-graining and sequential Martin-Löff tests, completely analogous to Peter Gács theory [3, 4].

Acknowledgement

The ideas presented in this paper were developed from 1995 through early 1998. Other interests prevented me from earlier publication. I thank Harry Buhrman, Richard Cleve, Wim van Dam, Barbara Terhal, John Tromp, and Ronald de Wolf for discussions and comments on QKC.

References

[1] L.M. Adleman, J. Demarrais, M.-D. A. Huang, Quantum computability, *SIAM J. Comput.*, 26:5(1997), 1524–1540.

[2] E. Bernstein and U. Vazirani, Quantum complexity theory, *SIAM J. Comput.*, 26:5(1997), 1411–1473.

[3] P. Gács, The Boltzmann entropy and randomness tests, *Proc. IEEE Physics and Computation Conf.*, IEEE Comp. Soc. Press, 1994, 209–216.

[4] M. Li and P. Vitanyi, *An Introduction to Kolmogorov Complexity and Its Applications*, Springer, New York, 1997 (2nd Edition).

[5] A. Peres, *Quantum Theory: Concepts and Methods*, Kluwer, 1995.