PHYSICAL REVIEW
LETTERS

VOLUME 77 29 JULY 1996 NUMBER 5

Error Correcting Codes in Quantum Theory

A. M. Steane
Clarendon Laboratory, Parks Road, Oxford, OX1 3PU, England

(Received 4 October 1995)

A new type of uncertainty relation is presented, concerning the information-bearing properties of a discrete quantum system. A natural link is then revealed between basic quantum theory and the linear error correcting codes of classical information theory. A subset of the known codes is described, having properties which are important for error correction in quantum communication. It is shown that a pair of states which are, in a certain sense, "macroscopically different," can form a superposition in which the interference phase between the two parts is measurable. This provides a highly stabilized "Schrödinger cat" state. [S0031-9007(96)00779-X]

PACS numbers: 03.65.Bz, 03.75.Dg, 89.70.+c

This Letter discusses fundamental questions concerning quantum interference among many particles in a group. It will be shown that such questions are linked with the properties of the error correcting codes arising in classical information theory [1]. The possibility of error correction in quantum systems has been considered recently because of its importance in the theory of quantum computation [2] and quantum cryptography [3]. The present work provides the answers to fundamental questions in this area. First, a new way of expressing the Heisenberg uncertainty principle is presented. Here it describes a limit on the degree of robustness with which information can be encoded in a quantum state which is to be analyzed in either of two mutually rotated bases. In brief, if multiple error correction is possible in one basis, then it is ruled out in the other. The precise meaning of this sentence will be elucidated below. This gives a simple way of understanding the well-known instability of the phase relationship between quantum states expressing macroscopically different physical situations. Next, the *linear codes* of classical information theory are shown to have a remarkable property (Theorem 3 below) in the quantum mechanical context. This establishes a previously unremarked link between these two mathematical edifices. The new insights gained enable one to construct states which are both macroscopically distinguishable, in a technical sense to be described, and which also can be observed to show stable quantum mechanical interference. This has important implications for the possibility of quantum computation and is a new development in the understanding of the famous "Schrödinger's cat" experiment [4].

Consider a quantum system having a Hilbert space of 2^n dimensions (with positive integer n). For example, this could be a set of n two-state systems, such as n spin one-half particles, or n two-level atoms. Such systems can model the behavior of any other quantum system [5], including macroscopic objects such as measuring devices.

The two orthogonal states of each particle are written $|0\rangle$ and $|1\rangle$, and a product state such as $|0\rangle \otimes |0\rangle \otimes |1\rangle$ is written $|001\rangle$, where it is understood that the first binary digit (0 or 1) refers to the state of the first particle, the second digit the second particle, and so on. A general state of n particles can be written as a sum (entanglement) of product states. The singlet state of two particles, for example, is $(|10\rangle - |01\rangle)/\sqrt{2}$. In what follows, the notation will be simplified by omitting the overall normalization factor in such expressions. This will not affect the argument, and the factor can be reintroduced easily if necessary.

The states $|0\rangle$ and $|1\rangle$ form a basis, hereafter called "basis 1." We will be concerned with the state of the system as expressed using the states of basis 1, and also those of a rotated basis, "basis 2." For example, the two bases could be those corresponding to a vertical or horizontal choice of quantization axis, in the case of the spin state of spin-half particles. The basis states of basis

1 will be written using a plain $|0\rangle$ and $|1\rangle$; those of basis 2 will be written using bold fond $|\mathbf{0}\rangle$ and $|\mathbf{1}\rangle$. Thus ignoring normalization as already remarked, $|0\rangle \equiv |\mathbf{0}\rangle + |\mathbf{1}\rangle$, $|1\rangle \equiv |\mathbf{0}\rangle - |\mathbf{1}\rangle$, $|00\rangle \equiv |\mathbf{00}\rangle + |\mathbf{01}\rangle + |\mathbf{10}\rangle + |\mathbf{11}\rangle$, and so on. It will be convenient to have a shorthand for referring to the individual product states making up a superposition. Since a product state is identified by a unique string of bits, it will be referred to as a *word.*

A state which is equal to a superposition of words in basis 1 is equal to some other superposition in basis 2. Some basic relationships between the two bases will now be stated.

Theorem 1. *The word $|000\cdots0\rangle$ consisting of all zeros in basis 1 is equal to a superposition of all 2^n possible words in basis 2, with equal coefficients.*

Theorem 2. *If the jth bit of each word is complemented* ($0 \leftrightarrow 1$) *in basis 1, then all words in basis 2 in which the jth bit is set (is a* **1***) change sign.* For example,

$$|000\rangle + |111\rangle \equiv |\mathbf{000}\rangle + |\mathbf{011}\rangle + |\mathbf{101}\rangle + |\mathbf{110}\rangle,$$

$$|001\rangle + |110\rangle \equiv |\mathbf{000}\rangle - |\mathbf{011}\rangle - |\mathbf{101}\rangle + |\mathbf{110}\rangle.$$

Corollary. *If all the words are complemented in basis 1, then all words of odd parity change sign in basis 2, and vice versa.* (Odd parity means having an odd number of 1's.)

These theorems are easy to prove by writing each word in basis 1 as a product of bits, converting each bit to the form ($|\mathbf{0}\rangle \pm |\mathbf{1}\rangle$), and multiplying out the products.

Next some of the standard results and notation of coding theory will be described. This is very basic material but is necessary in order to make the argument widely accessible.

In coding theory, information takes the form of a string of bits, or "words." A *code* is a set of words, all of the same *length* (number of bits). Words in the code are *code words.* The Hamming *distance* between two words (of the same length) is the number of places where they differ, i.e., the number of positions where one has a 0 and the other a 1. The *minimum distance* of a code is the smallest Hamming distance between any two code words in the code. A single error is the erroneous complementing of a single bit of a word, for example, when the word is transmitted or stored. *A code of minimum distance d allows $\lfloor (d-1)/2 \rfloor$ errors to be corrected.* This is because if less than $d/2$ errors occur, then the correct original code word, which gave rise to the erroneous received word, can be identified as the only code word at a distance less than $d/2$ from the received word. The price of this error correction is that only code words (i.e., a subset of the 2^n possible n-bit words) may be transmitted.

The fundamental problem of coding theory is to find codes having the maximum number of code words for given length n and minimum distance d. Let $A(n,d)$ be defined as this maximum number of words. The problem is notoriously difficult and has no general solution.

The notation $[n,k,d]$ refers to a set of 2^k code words, each of length n, with minimum distance d, and having the property of being a *linear code.* This means that if the EXCLUSIVE-OR operation is carried out bitwise between any two code words, then the resulting word is also a member of the code. (Not all codes are linear.) If C is a code, then the *dual* code $C^\perp$ is the set of all words u for which $u \cdot v$ has even parity for all $v \in C$, where the dot signifies the bitwise AND operation. The dual of a linear $[n,k,d]$ code is a linear $[n, n-k, d^\perp]$ code. In general, there is no simple precise relationship between d and $d^\perp$, though they are related indirectly through a theorem due to MacWilliams [1]. If d is large, then $d^\perp$ is small.

A linear code C is completely specified by its $n \times (n-k)$ *parity check matrix H,* or equivalently by its $n \times k$ *generator matrix G.* The rows of these matrices are n-bit words. The code C is the set of all words u for which $H_i \cdot u$ has even parity, for all rows H_i of H. Also the code C is the set of all linear combinations (by bitwise EXCLUSIVE-OR) of the rows of G. It can be shown that *the parity check matrix of a code C is the generator matrix of the dual code $C^\perp$.* This property will be used below and ends the present list of standard results.

In the context of the set of n quantum bits (two-state systems), the sets of words which express a given state in bases 1 and 2 are related through the basis rotation operation, which is a Hadamard transform. Just as the properties of the continuous Fourier transform lead to the Heisenberg uncertainty principle $\Delta x \Delta p \geq \hbar/2$ where x and p are conjugate continuous variables, so also for the discrete case the basis rotation operation implies a limit on the way a given state can be expressed in two mutually rotated bases.

Suppose a state can be written as a superposition of m_1 of the product states of basis 1 and as a superposition of m_2 of the product states of basis 2. Then

$$m_1 m_2 \geq 2^n. \tag{1}$$

Proof. Inequality (1) is subsumed by the "entropic uncertainty relation" introduced by Bialynicki-Birula and Mycielski [6] and by Deutsch [7], as improved by Maassen and Uffink [8].

Now suppose we wish to find a state which is expressed in basis 1 by a set of words of minimum Hamming distance d_1, and simultaneously in basis 2 by a set of words of minimum Hamming distance d_2. By definition, $m_1 \leq A(n,d_1)$ and $m_2 \leq A(n,d_2)$; therefore, using inequality (1), we have

$$A(n,d_1)A(n,d_2) \geq 2^n. \tag{2}$$

This "error correction uncertainty relation" places a limit on the highest minimum distance simultaneously achievable in bases 1 and 2. If d_1 is large, then $A(n,d_1)$ is necessarily small, which means, by (2), that $A(n,d_2)$ must be large, which in turn means that d_2 must be small. Thus we have a complementarity between d_1 and d_2. Its implications will be described below.

For odd d, Hamming [9] derived the Hamming or "sphere-packing bound" $A(n,d) \leq 2^n / \sum_{i=0}^{(d-1)/2} \binom{n}{i}$, where $\binom{n}{i}$ is the binomial coefficient $n!/i!(n-i)!$. Substituting in (2), one obtains, for odd d_1 and d_2,

$$\sum_{i=0}^{(d_1-1)/2} \binom{n}{i} \sum_{i=0}^{(d_2-1)/2} \binom{n}{i} \leq 2^n. \tag{3}$$

It is not generally possible to find codes which satisfy the upper limit of the Hamming bound, but it can be shown that for large enough n, codes exist which allow d_2 to exceed any value for any given d_1.

We will now consider the state

$$|\psi\rangle = |000\cdots0\rangle + e^{i\phi}|111\cdots1\rangle, \tag{4}$$

where the two words are those of all zeros or all ones, in basis 1. Such a state can be shown to violate a Bell-type inequality by an amount that grows exponentially with the number of bits n [10]. If n is large, then we have a superposition of two states representing macroscopically different situations (somewhat like a cat alive or dead [11]). However, the presence of both parts of the superposition, rather than simply of one part *or* the other, can be revealed only in experiments whose outcome depends on the value of ϕ. In practice, technological difficulties make ϕ extremely difficult to measure with an experimental uncertainty less than $\pm\pi$. In other words, experimental observation of the quantum interference is prevented by the sensitivity of ϕ to random errors.

A simple way to understand the state $|\psi\rangle$ is revealed by Theorem 1 and the corollary to Theorem 2. When $\phi = 0$, it is easy to see that in basis 2, $|\psi\rangle$ is equal to a superposition of all words having even parity (even number of **1**'s), while if $\phi = \pi$, the state is a superposition of all words having odd parity. Therefore to distinguish the cases $\phi = 0$ and $\phi = \pi$ experimentally one must find out whether the state in basis 2 has even or odd parity. However, a single error (complementing of a bit) in basis 2 is sufficient to destroy this information. If the probability of an error in any one bit is p, then the probability that no errors occur, enabling ϕ to be deduced, is $(1-p)^n$, which falls off exponentially with n [12]. For example, if $n = 1001$, $p = 0.02$, then $(1-p)^n \sim 10^{-9}$.

In the state just discussed, the code obtained in basis 2 (that consisting of all words of even parity) is the dual of the code appearing in basis 1 [Eq. (4)]. This is an example of a more general property which will now be stated.

Theorem 3. *When the quantum state of the system forms a linear code C in basis 1, in a superposition with equal coefficients, then in basis 2 the words appearing in the superposition are those of the dual code $C^\perp$.*

Proof. We will construct a code in basis 1 having generator matrix G and show that in basis 2 the code of which G is the parity check matrix appears.

Consider first the state $|\{0\}\rangle = |000\cdots0\rangle$ consisting of all zeros in basis 1. In basis 2, all 2^n possible words are superposed (Theorem 1), each with positive sign. Let G be a generator matrix, that is, a matrix of k_1 rows, each row being a word n bits long. Take the first row G_1 of G and form the corresponding word $|G_1\rangle$ in basis 1 by starting from $|\{0\}\rangle$ and applying Theorem 2 once for each nonzero bit in G_1. These successive applications of Theorem 2 show that the state $|G_1\rangle$ is one for which all 2^n possible words appear in basis 2, and all those, and only those, words in basis 2 change sign which do not satisfy the parity check G_1.

Now form the state $|\{0\}\rangle + |G_1\rangle$. By the argument just given, when the sum is formed, all words in basis 2 which do not satisfy the parity check G_1 disappear. Therefore at this stage of the argument, G_1 is the (single-row) generator matrix of the code in basis 1 and also the parity check matrix of the code in basis 2.

Now take the next row G_2 of G, and form the pair of words $|G_2\rangle + |G_1 \oplus G_2\rangle$ by applying Theorem 2 the necessary number of times to the state $|\{0\}\rangle + |G_1\rangle$. Here $\oplus$ signifies the bitwise addition modulo 2 (EXCLUSIVE-OR) operation. By Theorem 2 again, all those and only those words in basis 2 change sign which do not satisfy the parity check G_2. Therefore the state $|\{0\}\rangle + |G_1\rangle + |G_2\rangle + |G_1 \oplus G_2\rangle$ has the property that the first two rows of G form the generator matrix of the code in basis 1 and also the parity check matrix of the code in basis 2.

The above process is continued for the rest of the rows of G, and the theorem is proved.

Since the dual of an $[n,k,d]$ code is an $[n, n-k, d^\perp]$ code, Theorem 3 shows that the linear codes satisfy the lower bound of inequality (1). However, it seems unlikely that nonlinear codes should do so; therefore we may conjecture that the linear codes approach the lower bound of the error correction uncertainty relation (2) more closely. In this case the MacWilliams theorem also yields a limit on d_1 and d_2, though often one must use tables of known codes to find the smallest length n which permits the distances d_1 and d_2 to attain given sizes simultaneously [1].

Error correction in quantum computation might be suggested through the use of the simple repetition code. A bit value 0 is represented by $|000\rangle$, and a value 1 by $|111\rangle$, for example. This allows single error correction in basis 1. However, the possibility of superpositions such as $|000\rangle \pm |111\rangle$ is fundamentally important to quantum computation, and as we have just seen, the sign in such superpositions is highly sensitive to errors in basis 2. It will now be shown how to find state $|a\rangle$ and $|b\rangle$ such that error correction is possible in both bases, in the following sense. In basis 1, the Hamming distance between $|a\rangle$ and $|b\rangle$ will be greater than 2, while in basis 2 the Hamming distance between $|c\rangle = |a\rangle + |b\rangle$ and $|d\rangle = |a\rangle - |b\rangle$ will be greater than 2. The Hamming distance between two states, in a given basis, is here defined to be the smallest distance between any word appearing in the first state and any word appearing in the second.

Let $|a\rangle$ be expressed by the [7, 3, 4] simplex code in basis 1:

$$\begin{aligned}|a\rangle = {} & |0000000\rangle + |1010101\rangle + |0110011\rangle \\ & + |1100110\rangle + |0001111\rangle + |1011010\rangle \\ & + |0111100\rangle + |1101001\rangle .\end{aligned}$$

This code has the following properties: It can be augmented to produce a code of minimum distance 3, and its dual code (the [7, 4, 3] Hamming code) has minimum distance 3. The process of augmentation consists of adding to the code the complement of each of its words (equivalent to adding a row of 1's to the generator matrix). Therefore if we let $|b\rangle$ be the complement of $|a\rangle$ in basis 1,

$$\begin{aligned}|b\rangle = {} & |1111111\rangle + |0101010\rangle + |1001100\rangle \\ & + |0011001\rangle + |1110000\rangle + |0100101\rangle \\ & + |1000011\rangle + |0010110\rangle ,\end{aligned}$$

then the desired properties are obtained. For in basis 1, $|a\rangle$ and $|b\rangle$ are nonoverlapping subsets of a distance 3 code, which means the distance between them is at least 3, and in basis 2, $|c\rangle = |a\rangle + |b\rangle$ contains just the even parity words of a [7, 4, 3] code, while $|d\rangle = |a\rangle - |b\rangle$ contains just the odd parity words of the same code. Since these are nonoverlapping subsets of a distance 3 code, the distance in basis 2 between $|c\rangle$ and $|d\rangle$ is at least 3.

Thus a method for error correction of quantum bits has been found, which enables both the bits themselves to be encoded robustly in basis 1 and the values of the signs appearing in superpositions in basis 1 to be encoded robustly.

The above argument can be extended to higher Hamming distances, which leads to the possibility of macroscopic—or at least mesoscopic—superpositions with measurable interference phase. For example, the case was considered of the Schrödinger cat state $|\psi\rangle$ of Eq. (4) involving $n = 1001$ two-state systems. The two parts of the superposition were "macroscopically different" in the sense that any property proportional to the sum of the bits in basis 1 would have a mean value in the state $|000\cdots 0\rangle$ very different from its mean value in the state $|111\cdots 1\rangle$. However, the spirit of Schrödinger's thought experiment can also be retained by arguing that two states are macroscopically different if a macroscopic number of errors would have to occur in order to make it possible to mistake one state for the other. Now suppose we use $n = 5000$ and seek two states $|a\rangle$ and $|b\rangle$ separated by Hamming distance $d_1 = 1001$ in basis 1. The uncertainty relation (3) then implies $d_2 \leq 1213$, and it should be possible to find a dual pair of linear codes of which one is capable of augmentation and $d_1 = 1002$, $d_2 \geq 241$. If so, then subcodes of the augmented code are used to produce $|a\rangle$ and $|b\rangle$ as before, and we consider the superposition $|a\rangle + |b\rangle$. Quantum interference between $|a\rangle$ and $|b\rangle$ can be demonstrated if it can be shown experimentally that the sign in this superposition state is positive and not negative. To do this, measurements are carried out in basis 2. This measurement is the experimental method by which quantum interference between $|a\rangle$ and $|b\rangle$ is observed. Now, by construction, the state $|a\rangle + |b\rangle$ will be mistaken for only $|a\rangle - |b\rangle$ if at least $(d_2 - 1)/2 \geq 120$ errors occur. If these errors are independent, then the probability that the sign is revealed correctly in each experimental run (in which all the bits are measured) is

$$\sum_{i=0}^{120} \binom{n}{i} p^i (1-p)^{n-i} \simeq 0.98 , \qquad (5)$$

where the error per bit $p = 0.02$ as before. This is to be compared with the result of order 10^{-9} obtained for the Schrödinger cat state of the type given in Eq. (4), having the same Hamming distance between its two parts in basis 1. In fact, the error per bit in a real experiment is likely to increase somewhat with n, but as long as $p < 0.055$, then a number n can always be found with makes the interference observable between states separated by a given distance d_1; this is proved in [13,14]. Also it is not always true that errors in different quantum bits are independent. However, situations can be found in which the errors are independent, and in such cases the above argument applies.

In conclusion, a new type of uncertainty relation has been presented in which a discrete quantum system is regarded as an information-bearing entity, with limitations on the degree to which it can store information robustly. The interference phase between two product states separated by a large Hamming distance in one basis is a particularly fragile piece of information because it is expressed by the value of a parity check covering a large number of bits in the rotated basis.

A method has been presented for finding codes which enable error correction in both of two mutually rotated bases. This type of correction does not arise in the classical context, but is important for quantum bits. The argument enables states to be identified in which interferences involving a macroscopic number of particles may be observable. The experimental production of such states is, however, a demanding task which remains to be addressed.

The author is supported by the Royal Society.

Note Added.—During resubmission of this Letter, related work [15] on quantum coding has become known to me. In addition, the coding method introduced in this letter has now been generalized and shown to be fully applicable to quantum communication in that general errors affecting general states of many information qubits can be corrected [13,14].

[1] F.J. MacWilliams and N.J.A. Sloane, *The Theory of Error-Correcting Codes* (North-Holland, Amsterdam, 1977).

[2] For reviews and references, see, e.g., A. Ekert, in *Atomic Physics 14*, edited by D.J. Wineland, C.E. Wieman, and S.J. Smith (AIP Press, New York, 1995); A. Ekert and R. Jozsa (to be published).

[3] For reviews, see, e.g., R.J. Hughes, D.M. Alde, P. Dyer, G.G. Luther, G.L. Morgan, and M. Schauer, Contemp. Phys. **36**, 149 (1995); S.J.D. Phoenix and P.D. Townsend, Contemp. Phys. **36**, 165 (1995).

[4] E. Schrödinger, Naturwissenschaften **23**, 807 (1935); translated in *Quantum Theory and Measurement*, edited by J.A. Wheeler and W.H. Zurek (Princeton University, Princeton, NJ, 1983).

[5] D. Deutsch, Proc. R. Soc. London A **400**, 97 (1985).

[6] I. Bialynicki and J. Mycielski, Commun. Math. Phys. **44**, 129 (1975).

[7] D. Deutsch, Phys. Rev. Lett. **50**, 631 (1983).

[8] H. Maassen and J.B.M. Uffink, Phys. Rev. Lett. **60**, 1103 (1988).

[9] R.W. Hamming, Bell Syst. Tech. J **29**, 147 (1950).

[10] N. David Mermin, Phys. Rev. Lett. **65**, 1838 (1990).

[11] A. Peres, *Quantum Theory: Concepts and Methods* (Kluwer Academic, Dordrecht, 1993).

[12] More precisely, the probability that ϕ is revealed in the correct one of the two ranges $(-\pi/2,\ldots,\pi/2)$, $(\pi/2,\ldots,3\pi/2)$ in a given experimental run is $P = \sum_{i=0}^{\lfloor n/2 \rfloor} \binom{n}{2i} p^{2i}(1-p)^{n-2i}$ (the probability that zero or an even number of errors occur), which approaches $1/2$ very rapidly as n increases. For example, if $p = 0.02$, $n = 1001$ then $|P - 1/2| \sim 10^{-14}$. When $P = 1/2$ the experimental results bear no correlation to ϕ.

[13] A.R. Calderbank and P.W. Shor, Phys. Rev. A (to be published).

[14] A.M. Steane, Proc. R. Soc. London A (to be published).

[15] P.W. Shor, Phys. Rev. A **52**, R2493 (1995).