From Shannon to Quantum Information Science

1. Ideas and Techniques

Rajiah Simon

Rajiah Simon is a Professor at the Institute of Mathematical Sciences, Chennai. His primary interests are in classical and quantum optics, geometric phases, group theoretical techniques and quantum information science.

Quantum information science is a young and vigorously growing area of research which promises enormous potential and opportunities. This article, which will appear in two parts, gives a quantitative description of some of the ideas and techniques fundamental to this discipline. Considerations involving mixed states are reserved for Part 2.

Introduction

Quantum Information Science (QIS) is an emerging field which has the promise to cause revolutionary advances in fields of science and engineering involving computation, communication, precision measurement, and fundamental quantum theory. Basic to this new science, which started in the eighties and has been experiencing an explosive growth over the past ten years, is an appreciation of the fact that *information is physical*, and so also are all acts of information processing, thereby implying that both are controlled by the fundamental quantum laws of nature. Thus QIS encompasses all information processing situations wherein the underlying quantum nature becomes appreciable either in the way information is encoded in physical signals or in the way the processing units act on the signal. This field cuts across the traditional disciplines of physics, mathematics, computer science, and engineering, and explores how the quantum laws of nature can be harnessed to dramatically improve the acquisition, transmission, and processing of information.

Keywords
Teleportation, entanglement, information theory, quantum channels, no-cloning theorem.

Several considerations have acted as compelling stimuli for the effort in this field. The well-known Moore's

law captures the empirical fact that miniaturization and performance of electronic circuitry have been doubling every two years. Extrapolation will tell us that this shrinkage will soon reach atomic dimensions, implying that this exponential growth we have got used to is bound to break down unless we get prepared to handle this new regime where quantum effects are bound to become dominant. Recent developments leading to sophisticated optical cavities, trapped atomic ions, quantum dots, and nanotechnology have made it possible to contemplate the actual construction of workable quantum logic devices. The need for secret communication whose security will be guaranteed by the laws of nature, rather than by unproven mathematical assumptions, drove the study of quantum communication schemes. Finally, we may mention the realization that quantum effects are no more just a nuisance, but in fact can be exploited to perform important information processing tasks impossible or difficult in the classical scheme of things:

Recent developments leading to sophisticated optical cavities, trapped atomic ions, quantum dots, and nanotechnology have made it possible to contemplate the actual construction of workable quantum logic devices.

- Large integers, which the fastest supercomputers we have today are estimated to take millions of years to factorize, can be factorized by a quantum computer in minutes.

- The time taken by a classical computer to search and locate a particular item from a database of N items is proportional to N, but a quantum computer can do this in time proportional to $\sqrt{N}$.

- It happens often that nontrivial quantum problems cannot be solved analytically, and so one has to resort to simulation on a digital computer. However, such a simulation is *not efficient*, i.e., the resources required for the simulation grows unreasonably fast with the size of the problem. But such problems can be simulated efficiently on a quantum computer.

Peter Shor's demonstration that quantum computers can perform factorization of large integers efficiently has played a major role in shaping the kind of momentum QIS has gathered.

- Quantum effects have already been employed to create unbreakable codes, whose security is assured by the quantum laws of nature.

Peter Shor's demonstration that quantum computers can perform factorization of large integers efficiently has played a major role in shaping the kind of momentum QIS has gathered. It exposed the vulnerability of the popular RSA (Ron Rivest, Adi Shamir and Len Aldeman) cryptosystem which rests on the 'assumption' that factorization is a 'hard' problem. Since RSA is currently vital for secure communication in defence, financial, and other sectors, this demonstration led research funding agencies to take serious note of the emerging field of QIS.

Information coded in quantum systems has properties which appear bizarre and counterintuitive, at least in the first encounter. One such property is *entanglement*, a new kind of correlation having no equivalent in the classical world. When two systems are entangled, the state of the combined system is more pure and less random than the state of either system by itself: the whole is a lot more than simply the aggregate of its parts. A book written in an entangled state will have to be grasped in its entirety as one unit, and not sentence by sentence or paragraph by paragraph. In this respect an entangled state is like a good piece of art! It turns out that entanglement plays a key role in most information processing protocols. Indeed, it is the primary resource and currency in quantum information processing tasks.

In this exposition which is to appear in two parts, I have chosen to give a *quantitative* description of several of the important notions and techniques basic to QIS, rather than giving a verbal account of the achievements in this field. The reason for this choice is my belief that much of this material is accessible to students with an elementary exposure to quantum theory; the only prerequisite

on the reader's part is a willingness to try. My hope is that familiarity with the material presented here will enable and encourage the reader to explore the original publications, and possibly participate, in this extremely vigorous research area. Almost all these publications are available electronically, and free. And this field has the distinction that several of the important original contributions originate from graduate students.

Information theory came into being, in one stroke, through the truly remarkable 1948 work of Shannon.

We will pay particular attention to the fact that certain processes which appear plausible to our classical bent of mind are actually forbidden by the quantum laws of nature. As a more than adequate compensation the same laws permit tasks which would have been impossible in a classical world.

The classical information theory created by Shannon continues to be a guiding principle and goal post for quantum information theory and science. For this reason, I begin by drawing the reader's attention to some of the important features of Shannon's theory.

Shannon Information Theory

Information theory came into being, in one stroke, through the truly remarkable 1948 work of Shannon. As A I Khinchin states in his *Mathematical Foundations of Information Theory* (Dover, 1957): "rarely does it happen in mathematics that a new discipline achieves the character of a mature and developed scientific theory in the first investigation devoted to it. Such in its time was the case with the theory of integral equations, after the fundamental work of Fredholm; so it was with information theory after the work of Shannon". In the past 50 plus years Shannon's information theory has been made more and more precise, has been extended in enormous number of dimensions, and has been applied to several walks of life beyond the various branches of communication science.

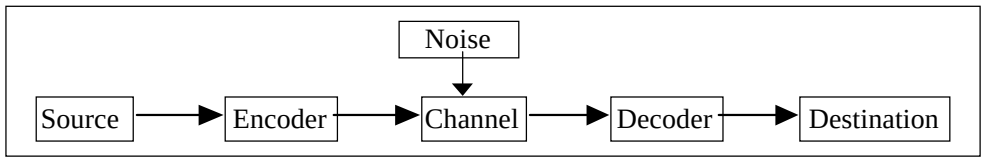

Figure 1. Block diagram of communication system.

The key idea of Shannon is to model communication as a stochastic process, as depicted in the abstract block diagram shown in *Figure* 1. As Shannon puts it "the fundamental problem of communication is of reproducing at one point (destination) either exactly or approximately a message selected at another point (source)". That this is a problem is due to the ever present channel noise which tends to disfigure the message. Earlier attempts concentrated on clever techniques trying to simply 'deconvolve' the original message out of the deformed message at the receiver's end, but Shannon took the innovative route of combating channel noise through preprocessing (encoding) of the message, at the sender's end, prior to transmission through the channel. In crude abstract terms, the idea is to code the message in a direction 'orthogonal' to that of channel noise, thus enabling undoing of the disfiguring. Efficient design of the encoder-decoder blocks thus demands sufficient knowledge of the noise characteristics of the channel.

In Shannon's scheme of things the source, channel, and destination should be viewed as *fixed* parts of a communication system; the coder and decoder are the *variable* parts that should be optimized in an attempt to improve reliability, increase the data rate, or decrease the cost. Indeed, the distinguishing characteristic of Shannon's theory is "a primary concern with the encoder and decoder, both in terms of their functional roles and in terms of the existence (or nonexistence) of encoders and decoders that achieve a given level of performance" (R G Gallager [1]). For this reason, information theory has become almost synonymous with coding theory; books dealing with information theory are often titled *Coding Theory.*

Shannon gave a mathematical definition for *measure of information*, and then formulated two *fundamental coding theorems*. A remarkable feature of Shannon's theorems is a clear separation between source coding and channel coding.

The channel need not, in every case, be an extent of physical space. In an electronic memory system, for instance, the source and the destination correspond to storage and retrieval respectively and, therefore, the channel to simply the passage of time.

Shannon's definition of information clearly discriminates between data and information.

Shannon's Measure of Information

The simplest model source emits every τ_s seconds a symbol from its finite alphabet A containing N symbols: $A = (a_1, a_2, \cdots, a_N)$. Let $p_1 = \Pr(a_1)$, $p_2 = \Pr(a_2)$, $\cdots$, $p_N = \Pr(a_N)$ be the probabilities for emission of these symbols. We assume that this discrete source is memoryless, i.e., these probabilities are the same for every emission, and are independent of earlier (or later) emissions. Such a source is called a *discrete memoryless source*. Both these qualifications on the model source can however be relaxed.

Shannon's definition of information clearly discriminates between data and information. The information content of a book will be measured by how better informed one becomes after reading the book. How uncertain a message was, before it was received, is a measure of the information content of the message. In the presence of 'match-fixing', the result of a match is certainly a datum, but this has zero information content!

Returning to our model source, and concentrating on one emission, suppose the symbol a_k was emitted. We may associate with this event an information content $\log(1/p_k)$. This matches the intuition, motivated in the last paragraph, that occurrence of a less likely event conveys more information than that of a more likely one. It

We say that the source emits bits at regular time intervals.

matches also the expectation that independence, which leads to multiplication of probabilities, should lead to additivity of information. The unit of information is determined by the logarithm base. Two common units are bits (base 2) and nats (base e). Clearly, n bits $= n \log_e 2$ nats. Averaging $\log(1/p_k)$ over the source alphabet A, we obtain the information per emission of our (statistical) source:

$$H(A) = \sum_{k=1}^{N} p_k \log_2 (1/p_k) \text{ bits.}$$

This is Shannon's definition of information. The expression $H(A)$ is often known as the Shannon entropy. It has the same structure as the Boltzmann entropy in statistical thermodynamics.

It is clear that the information has no dependence on the nature of the actual symbols constituting the alphabet: these can be integers, characters of the Tamil language, beasts, or aliens! $\mathcal{H}(A)$ is a function of the probabilities alone. Thus, in the case of a binary source we can take, without loss of generality, the alphabet to be the set $B = \{0, 1\}$, even when it is known that the source actually emits the live and dead states of the proverbial cat. We say that the source emits bits at regular time intervals (here a bit means a thing which can assume one of two possible states, a two-level system). If the binary source emits 0 with probability p (and hence 1 with probability $1 - p$), the Shannon entropy of the source is

$$H(p) = -[\, p \log_2 p + (1-p) \log_2 (1-p) \,] \text{ bits.}$$

Clearly, the maximum possible value of $H(p)$ is one bit, and this value obtains for $p = 0.5$. So a bit of data can carry one bit of information or less. Similarly, a source with $N = 2^x$ symbols in its alphabet can give out no more than x bits of information per emission, the limiting rate being realized when $p_1 = p_2 = \cdots = p_N = 1/N$.

Shannon's Fundamental Theorems

The two fundamental theorems of Shannon deal with redundancy, but they work in opposite directions.

The two fundamental theorems of Shannon deal with redundancy, but they work in opposite directions. The first theorem deals with compression (stripping message/data of its redundancy), and sets a limit to compression. This may be roughly explained, in the specific context of the binary memoryless source, by considering n successive emissions (this corresponds, by virtue of memorylessness, to $nH(p)$ bits of information), where n is a large integer. That is, we have a n bit word or sequence from the set of 2^n possible words. Memorylessness of the source means that the probability of a bit of the word assuming 0 is the same for every bit, as in n Bernoulli trials, and is independent of the values the other bits assume.

Now a *typical sequence* will have approximately np 0's and $n(1-p)$ 1's. The number of such typical strings is of order the binomial coefficient $\binom{n}{np}$. Using Stirling's approximation for $\log n!$ we have

$$\binom{n}{np} \approx 2^{nH(p)}.$$

The law of large numbers tells us that atypical sequences are unlikely. The key idea of Shannon, therefore, is that we need to find codewords only for the typical sequences, order $2^{nH(p)}$ in number, and not for all the 2^n possible sequences. Let m be the smallest integer not less than $nH(p)$. It is clear that the $2^{nH(p)}$ n-bit typical sequences can be coded into m-bit codewords, resulting in data compression whenever $p \neq 0.5$. Thus, $H(p)$ is the asymptotic limit of message compression achievable by these block codes.

In short: a sufficiently long message can be compressed to the level of its bare information content. This crude rendering of Shannon's *source coding theorem* or *noise-*

Shannon's second theorem is called the *channel coding theorem* or *noisy channel theorem*.

less channel theorem can be made more precise by the use of δ's and ϵ's.

We may rephrase the situation as follows. A discrete memoryless source emitting τ_s^{-1} bits of data per second generates $R = \tau_s^{-1} H(p)$ bits of information per second, and Shannon's source coding theorem guarantees that this source can be replaced, by block coding, by a source emitting 0's and 1's at random, with *equal* probability, $\tau_s^{-1}[H(p)+\delta]$ times per second. Here $\delta > 0$ can be made as small as desired by choosing n, the size of the blocks, sufficiently large.

Shannon's second theorem is called the *channel coding theorem* or *noisy channel theorem*. The idea is to introduce deliberate redundancy into the source coded message in order to combat channel noise. The simplest model channel is the *binary symmetric channel*. The effect of noise in this channel is characterized by a small number $\epsilon > 0$: a 0 (or 1) at the input of the channel is mapped into a 1 (or 0) at the output with probability ϵ, and into a 0 (or 1) with probability $(1 - \epsilon)$. A naive channel coding scheme in this case is this: introduce a five-fold redundancy by coding every 0 into 00000 and every 1 into 11111. The corresponding decoding can be *decision by majority voting*, i.e., 11111, 10111, 11001, etc. at the receiver's end should be interpreted as distortions of 11111 representing 1. For $\epsilon = 0.1$, the probability that more than two of the five bits get reversed by this channel is 0.856%, and thus this channel coding scheme gives 99% reliability. Channel codes are often referred to as *error correcting codes*.

With every channel is associated a characteristic quantity called *channel capacity*, and denoted C bits/sec. Let R bits/sec be the rate at which information flows into the input of the channel. Shannon's channel coding theorem asserts that if $R < C$, i.e., if the source information rate is less than the channel capacity, and given a

desired level of reliability $1-\delta$, there exists a channel encoding-decoding system which will achieve this level of reliability, for any $\delta > 0$. This was a truly surprising result, for it disproved the firm belief of those days that the reliability will decrease with R. Shannon's theorem showed that this was not the case, as long as $R < C$.

The channel coding theorem can also be taken as the very definition of channel capacity, a notion that remained undefined until Shannon.

The channel coding theorem can also be taken as the very definition of channel capacity, a notion that remained undefined until Shannon.

The Quantum Measurement Postulate

It happens often with (elementary) courses in quantum theory that the (von Neumann) measurement postulate is listed along with the other basic principles of quantum mechanics, but rarely returned to in the rest of the course. This postulate has three parts:

- Measurement of a dynamical variable x always results in one of the eigenvalues $\{\lambda_k\}$ of the corresponding self-adjoint (hermitian) operator $\hat{x}$.

- With the state $|\psi\rangle$ under consideration written as a linear combination of the *complete* orthonormal set of eigenvectors $\{|\phi_{k,\alpha}\rangle\}$ of $\hat{x}$,

$$|\psi\rangle = \sum_{k,\alpha} c_{k,\alpha}\,|\phi_{k,\alpha}\rangle\,,$$

  the probability that the value λ_k will obtain is given by $\sum_\alpha |\,c_{k,\alpha}\,|^2$. Here the label α runs within *degenerate* sets of eigenvectors of $\hat{x}$.

- In the event of the value λ_k occurring, the system collapses ('jumps' or projects) to the new state $N\sum_\alpha c_{k,\alpha}\,|\phi_{k,\alpha}\rangle$, where $N^{-1} = \sqrt{\sum_\alpha |\,c_{k,\alpha}\,|^2}$.

The measurement postulate concentrates on eigenvalues, but typical courses stress the coarse grained version of expectation values. This is understandable; while

It is only recently that individual microscopic systems have become accessible to experiments.

the postulate applies to individual systems (like individual atoms), experiments were traditionally performed on bulks (ensembles), and so the actual measurement corresponded to the coarse grained version. It is only recently that individual microscopic systems have become accessible to experiments.

Quantum information processing takes place almost exclusively in this domain! Thus, one should not be surprised that consequences of the measurement postulate have proved to be of fundamental importance to quantum information processing. Some of these are immediate.

Some Implications

In the traditional interpretation, a sufficiently large supply of identically prepared systems (ensemble), all in the same state $|\psi\rangle$, is assumed to be available, thereby enabling estimation of the unknown state $|\psi\rangle$ through measurements. Suppose the state $|\psi\rangle$ was prepared by someone who decides to give us only a single copy, possibly because preparation of the state involves a cost. Assume that we face the challenge of determining what this state is. Our immediate temptation could be to make enough additional copies of the state, and then proceed with the estimation; but this is forbidden by the very linearity of quantum mechanics, a fact known as the *no cloning theorem*: an unknown state cannot be duplicated (without the original being lost in the process).

Let another system, similar to the one in state $|\psi\rangle$, be in a reference state $|\text{ref}\rangle$, so that the state of the combined system is $|\psi\rangle \otimes |\text{ref}\rangle$. You can imagine that $|\psi\rangle$ corresponds to a sheet of paper containing an all important message, and $|\text{ref}\rangle$ to a blank sheet of paper. If these two sheets are input into a copying machine, we obtain at the output a perfect copy of the message on the reference sheet in addition to the original message sheet, *whatever the message be*. A corresponding quan-

tum copying machine, if it exists, will be represented by a unitary operator

$$U: \quad U|\psi\rangle \otimes |\text{ref}\rangle = |\psi\rangle \otimes |\psi\rangle .$$

If our quantum machine accomplished a perfect copying job for an arbitrary state, we should have

$$U: \quad U|\phi\rangle \otimes |\text{ref}\rangle = |\phi\rangle \otimes |\phi\rangle ,$$

for every other state $|\phi\rangle$ as well. Taking the inner product of respectively the left and right hand sides of these equations we have $\langle\phi|\psi\rangle = (\langle\phi|\psi\rangle)^2$, showing that our machine cannot act as a perfect copying machine for non- orthogonal states! There exists a cousin of the no-cloning theorem. It is called the quantum no-deleting principle, but I shall not go into its details.

Now consider a challenge which may appear to be much simpler than the original challenge. Alice and Bob have agreed that Alice will send to Bob one of the two states $|\psi\rangle = |1\rangle$, $|\phi\rangle = (|1\rangle + |2\rangle)/\sqrt{2}$; Bob has to determine which of these states he received. We will grant Bob unlimited access to sophisticated (even ideal) experiments. Suppose Bob measures an observable which has as its eigenstates the two orthogonal vectors $|1\rangle$, $|2\rangle$ with eigenvalues λ_1, λ_2 respectively. If the eigenvalue λ_2 clicks, then Bob knows with certainty that he received the state $|\phi\rangle$, since the state $|\psi\rangle$ will always click λ_1. On the other hand if λ_1 clicked (this will happen in 75% of the trials), Bob *will not know* if he received $|\phi\rangle$ or $|\psi\rangle$. Bob can think of cleverer and more involved experiments, but he cannot win this game. *There is no way one can reliably distinguish between nonorthogonal states.* It follows that an unknown state cannot be determined, given just one copy of the state. A quantum state is automatically *copy righted* in the strong sense that complete knowledge of the state is accessible only to the one who prepared it. This conclusion and the no cloning theorem have obvious relevance to cryptography.

A quantum state is automatically *copy righted* in the strong sense that complete knowledge of the state is accessible only to the one who prepared it.

It is useful to compare the state spaces in classical and quantum theories.

A less ambitious and more pragmatic question is this: given m copies of a state from an n dimensional Hilbert space, to what extent can one determine the state? This issue has received the attention of several researchers, and much progress has been achieved. It is transparent that one's estimate of the unknown state will improve with increasing number of copies being made available.

State Space in Classical and Quantum Theory

It is useful to compare the state spaces in classical and quantum theories. In information processing situations we are often interested in systems which, in the classical case, have only a finite number of states. Of particular interest is the (classical) *bit* which has just two states, generically denoted $|0\rangle$ and $|1\rangle$. Given a classical system with n states, the corresponding quantum system has as allowed states not only the classical states but also *all* possible superpositions of these states. Thus the construction of the quantum state space of such a system begins with a Hilbert space $\mathcal{H}$ for which the classical set of states becomes an orthonormal basis. When a classical bit is quantized in this manner we get a *qubit*, which is governed by a two-dimensional Hilbert space. Distincts vectors of $\mathcal{H}$, however, do not correspond to distinct states. Indeed, all nonzero vectors in $\mathcal{H}$ which differ from one another by a multiplicative scalar represent one and the same quantum state. In other words states are represented in the quantum theory by *equivalence classes* of (nonzero) vectors of the Hilbert space. Thus one finds that distinct quantum states of the qubit are described by the (representative) vectors

$$|\theta,\, \phi\rangle = \begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} \cos\frac{\theta}{2} \\ \mathrm{e}^{\mathrm{i}\phi}\sin\frac{\theta}{2} \end{pmatrix}, 0 \leq \theta \leq \pi, 0 \leq \phi < 2\pi.$$

This collection is not a cylinder of unit radius and height π, with ϕ running around the cylinder and θ running along the axis on the cylinder. Since all values of ϕ get identified for $\theta = 0$ and for $\theta = \pi$, the circles at the top

and bottom boundaries have to be collapsed into a point each. Thus, we see that the state space of a qubit is the sphere S^2 for which θ, ϕ are the polar coordinates, pairs of orthogonal states being represented by diametrically opposite points.

Though the state space has a continuum of distinct states, three states are never perfectly distinguishable, and two states are perfectly distinguishable if and only if they correspond to antipodal points! This may be surprising at first sight, but remember that the entire continuum came about from just two states.

According to Feynman, superposition (interference) is the *only* mystery of quantum mechanics.

Composite Systems: Quantum Entanglement

According to Feynman, superposition (interference) is the *only* mystery of quantum mechanics. Entanglement is a striking manifestation of superposition in situations wherein the system of interest consists of two (or more) subsystems. This term, which means intricate and confused involution, is a translation of the German *Verschränktheit*, coined by Schrödinger in 1935. Entanglement is at the heart of the well known EPR (Einstein–Podolsky–Rosen) paradox. In those early days, entanglement was viewed more or less as a bizarre, and even embarrassing, feature the then new quantum theory would have to live with. Today it has become, as we will see, an indispensable resource needed to perform otherwise impossible tasks of information processing and computation.

Let a bipartite system S consist of subsystems A and B. If $\mathcal{H}_A$ and $\mathcal{H}_B$ are the Hilbert spaces of the subsystems, the Hilbert space of S is the tensor product $\mathcal{H}_S = \mathcal{H}_A \otimes \mathcal{H}_B$, which differs from the Cartesian product in that in addition to containing ordered pairs or *product vectors*, denoted $|\psi\rangle_A \otimes |\psi\rangle_B$ with $|\psi\rangle_A \in \mathcal{H}_A$ and $|\psi\rangle_B \in \mathcal{H}_B$, the tensor product space $\mathcal{H}_S$ contains all linear combinations of product vectors. Let $\dim \mathcal{H}_A = m$ and $\dim \mathcal{H}_B = n$.

A state is said to be *entangled* if it cannot be represented by a product vector.

If $\{ |\psi_j\rangle \}$ is an orthonormal basis (O.N.B.) in $\mathcal{H}_A$, and $\{ |\phi_\alpha\rangle \}$ an O.N.B. in $\mathcal{H}_B$, then $\{ |\Psi_{j\alpha}\rangle = |\psi_j\rangle \otimes |\phi_\alpha\rangle \}$ is an O.N.B. in the mn-dimensional space $\mathcal{H}_A \otimes \mathcal{H}_B$. This means any vector $|\Psi\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B$ can be written in the form

$$|\Psi\rangle = \sum_{j,\,\alpha} C_{j\alpha} |\psi_j\rangle \otimes |\phi_\alpha\rangle.$$

It follows that $|\Psi\rangle$ is a product vector if and only if the expansion coefficients $C_{j\alpha}$ have the product form $C_{j\alpha} = x_j\, y_\alpha$, i.e., if and only if the $m \times n$ coefficient matrix C is the outer product of two vectors. States represented by product vectors are called *product states*. A state is said to be *entangled* if it cannot be represented by a product vector.

If L, M are matrices representing linear transformations in $\mathcal{H}_A$ and $\mathcal{H}_B$ respectively, the mn-dimensional tensor product matrix $L \otimes M$ represents a linear transformation in $\mathcal{H}_A \otimes \mathcal{H}_B$. The vectors of the orthonormal product basis $\{ |\Psi_{j\alpha}\rangle \}$ can be conveniently arranged in the sequence $|\Psi_{11}\rangle, \cdots, |\Psi_{1n}\rangle;\ |\Psi_{21}\rangle, \cdots, |\Psi_{2n}\rangle; \cdots \cdots;$ $|\Psi_{m1}\rangle, \cdots, |\Psi_{mn}\rangle$. Then,

$$\Omega \equiv L \otimes M = \begin{pmatrix} \Omega_{11} & \Omega_{12} & \cdots & \Omega_{1m} \\ \Omega_{21} & \Omega_{22} & \cdots & \Omega_{2m} \\ \vdots & \vdots & \cdots & \vdots \\ \Omega_{m1} & \Omega_{m2} & \cdots & \Omega_{mm} \end{pmatrix},$$

where Ω_{ij} has the same size as the matrix M, for each $i,\ j$; indeed Ω_{ij}, the ij-th block of Ω, is simply the matrix M multiplied by the scalar L_{ij}. If both L and M are hermitian (positive semidefinite), then $L \otimes M$ is hermitian (positive semidefinite); the trace (determinant) of $L \otimes M$ is the product of the traces (determinants) of L and M; if $\lambda_j,\ j = 1, 2, \cdots, m$ are the eigenvalues of L and $\mu_k,\ k = 1, 2, \cdots, n$ are those of M, then $\{ \omega_{jk} = \lambda_j \mu_k \}$ are the eigenvalues of Ω; and the eigenvectors of Ω are (tensor) products of the eigenvectors

of L, M. A generic linear operator acting on $\mathcal{H}_A \otimes \mathcal{H}_B$ is not a tensor product, but a superposition of tensor products, in precisely the same way a generic vector in $\mathcal{H}_A \otimes \mathcal{H}_B$ is a superposition of product vectors.

The integer *r* is known as the *Schmidt rank* of |*y*ñ

Now consider independent *local* changes of bases in $\mathcal{H}_A$ and $\mathcal{H}_B$, described respectively by unitary matrices $V \in U(m)$ and $W \in U(n)$:

$$|\psi'_k\rangle = \sum_j V^*_{jk} |\psi_j\rangle, \qquad |\phi'_\beta\rangle = \sum_\alpha W^*_{\alpha\beta} |\psi_\alpha\rangle \, .$$

Under these local changes of bases the coefficient matrix C undergoes the change $C \rightarrow C' = V\, C\, W^{\mathrm{T}}$. It follows that V and W can be chosen so as to render C' 'diagonal' with nonnegative entries along the 'diagonal', i.e., $C'_{j\alpha} = \sqrt{\lambda_j}\, \delta_{j\alpha}$, $\lambda_j \geq 0$. We have thus proved:

Schmidt Decomposition Theorem: Given a vector $|\Psi\rangle$ in the tensor product space $\mathcal{H}_S = \mathcal{H}_A \otimes \mathcal{H}_B$, it can be written in the canonical form

$$|\Psi\rangle = \sum_{j=1}^{r} \sqrt{\lambda_j}\, |\psi'_j\rangle \otimes |\phi'_j\rangle,$$

where $r \leq \min(m, n)$, $\lambda_j > 0$, $\sum_j \lambda_j = 1$, and $\{ |\psi'_j\rangle \}$ and $\{ |\phi'_j\rangle \}$ are vectors from an O.N.B. in $\mathcal{H}_A$ and $\mathcal{H}_B$, respectively.

Later we will make use of this canonical form involving a *single* sum of product vectors. The integer r is known as the *Schmidt rank* of $|\Psi\rangle$. It is clear that r is a *local invariant*, and so are the λ_j's. It is also transparent that $|\Psi\rangle$ is a product (separable) vector if and only if it is of unit Schmidt rank. The Schmidt rank thus acts as an *entanglement witness* for bipartite state vectors.

Quantum Teleportation

Suppose that Alice has to communicate to Bob a state $|\psi\rangle$ prepared by Charlie. With just one copy in her possession there is no way she can determine what state has

The process, which uses quantum *entanglement* as an essential resource, has become one of the basic operations of quantum information theory.

been given to her, and therefore she cannot convert it into a bunch of numbers and then transmit these numbers to Bob over a conventional communication channel; she has to resort to *intact* transport of $|\psi\rangle$ itself. Quantum *teleportation* provides a way of accomplishing this task. This process, which uses quantum *entanglement* as an essential resource, has become one of the basic operations of quantum information theory.

Let us assume that Alice and Bob share a pair of qubits A, B in the (maximally) entangled state $|\Psi\rangle_{AB} = \frac{1}{\sqrt{2}}$ $(\,|0\rangle_A \otimes |0\rangle_B + |1\rangle_A \otimes |1\rangle_B\,)$ (which they prepared when they were together last), and that currently they are in different cities, with the qubit labeled A in Alice's possession and that labeled B in the possession of Bob. Suppose Charlie has given to Alice a qubit C which he prepared in the state $|\psi\rangle_C = u|0\rangle_C + v|1\rangle_C$, with the request that it be communicated to Bob. Alice has no knowledge of the state $|\psi\rangle_C$, i.e., Charlie has not disclosed to her the values of the complex parameters $(\,u,\,v\,)$ which correspond to a point on S^2.

We will now show how Alice can do intact transportation of this state to her willing partner Bob, at the cost of the entanglement she shares with him. Clearly, Alice's mission would get accomplished when Bob's qubit attains the state $|\psi\rangle_B = u|0\rangle_B + v|1\rangle_B$ (and decoupled from A and C), whatever $(u,\,v)$ be.

To begin with, the three qubit system is in the pure state

$$|\Psi\rangle_{ABC} = \frac{1}{\sqrt{2}}(\,|0\rangle_A \otimes |0\rangle_B + |1\rangle_A \otimes |1\rangle_B\,) \otimes |\psi\rangle_C\,.$$

This state may be rewritten in the suggestive form

$$|\Psi\rangle_{ABC} = \tfrac{1}{2}[\,|\Phi_0\rangle_{AC} \otimes (\,u\,|0\rangle_B + v\,|1\rangle_B\,) + |\Phi_1\rangle_{AC} \otimes (\,v\,|0\rangle_B + u\,|1\rangle_B\,) + |\Phi_2\rangle_{AC} \otimes (\,v\,|0\rangle_B - u\,|1\rangle_B\,) + |\Phi_3\rangle_{AC} \otimes (\,u\,|0\rangle_B - v\,|1\rangle_B\,)\,],$$

where

$$
\begin{aligned}
|\Phi_0\rangle_{AC} &= \frac{1}{\sqrt{2}}(\,|0\rangle_A \otimes |0\rangle_C + |1\rangle_A \otimes |1\rangle_C\,),\\
|\Phi_1\rangle_{AC} &= \frac{1}{\sqrt{2}}(\,|0\rangle_A \otimes |1\rangle_C + |1\rangle_A \otimes |0\rangle_C\,),\\
|\Phi_2\rangle_{AC} &= \frac{1}{\sqrt{2}}(\,|0\rangle_A \otimes |1\rangle_C - |1\rangle_A \otimes |0\rangle_C\,),\\
|\Phi_3\rangle_{AC} &= \frac{1}{\sqrt{2}}(\,|0\rangle_A \otimes |0\rangle_C - |1\rangle_A \otimes |1\rangle_C\,).
\end{aligned}
$$

Since both the qubits A and C are in Alice's lab, she can make a joint measurement on them. Further, the four maximally entangled states $|\Phi_0\rangle_{AC}$, $|\Phi_1\rangle_{AC}$, $|\Phi_2\rangle_{AC}$, and $|\Phi_3\rangle_{AC}$ are orthogonal (the sense in which these states are maximally entangled will be clarified later). Therefore, there exists a *local* dynamical variable Ω_{AC} of the AC system for which these four states are the eigenstates with *distinct* eigenvalues λ_0, λ_1, λ_2, λ_3. When Alice measures this variable she will obtain (with equal probability) one of these eigenvalues. Assume that the value λ_0 obtains. By the measurement postulate, the state $|\Psi\rangle_{ABC}$ projects to the state $|\Phi_0\rangle_{AC} \otimes (\,u\,|0\rangle_B + v\,|1\rangle_B\,)$, and Alice (*but not Bob*) knows instantly that Bob's qubit has attained the desired state $u|0\rangle_B + v|1\rangle_B$ (and, of course, decoupled from A and C). On the other hand if the value λ_1 obtains, Alice knows instantly that Bob's qubit has attained the state $v|0\rangle_B + u|1\rangle_B$. But this is not the intended state, and so Alice will have to request Bob in that case to apply the unitary Pauli matrix σ_1 to his qubit, so that it attains the desired state. Alice will similarly request Bob to apply the Pauli matrix σ_2 (or σ_3) if the value λ_2 (or λ_3) obtains.

Define $\sigma_0 = \mathrm{Id}_{2\times 2}$, the two-dimensional unit matrix. Based on the result of her measurement on the AC system Alice will communicate to her partner Bob over a *classical side channel* (telephone, for instance), a classical two-bit number 0, 1, 2, or 3. On receipt of this

A classical bit (two-state system) can carry just one bit of information.

number k, Bob will apply to his qubit the corresponding unitary transformation σ_k, inducing his qubit to the desired state $u|0\rangle_B + v|1\rangle_B$, and thus completing the teleportation protocol.

It is transparent that teleportation is not superluminal: the side channel classical communication acts as the bottle-neck for its speed. It does not contradict the no-cloning theorem either, for Alice has at the end of the protocol no copy of the state supplied by Charlie. The entanglement that existed between Alice and Bob has been consumed in the process. Each one of the maximally entangled states $|\Phi_k\rangle$ of a pair of qubits, listed in the last equation, is said to have one *e-bit* of entanglement. We may thus say that teleportation of a qubit costs one e-bit of entanglement and two bits of classical information.

In some sense the qubit A is a winner all the way in this game. While it lost its entanglement with B, it has got entangled with C, and this entanglement can certainly be used as a resource in future quantum processing applications. In this sense, the above teleportation protocol can be viewed as entanglement assisted transfer of a qubit at the cost of two bits of classical information. To close this discussion we mention that teleportation has been achieved recently in several labs.

Super Dense Coding

A classical bit (two-state system) can carry just one bit of information. While a qubit has a continuum of distinct states, we have seen that no three states are perfectly distinguishable. It is therefore of interest to ask if a qubit carries, in any sense, more than one bit of information. A simple variation of the teleportation protocol describes a situation in which transmission of one qubit amounts to transfer of two bits of information.

Assume again that Alice and Bob share an entangled

pair of qubits in the state $|\Phi_0\rangle_{AB}$. Suppose Alice wishes to send Bob two bits of classical information, i.e., one of the integers 0, 1, 2, or 3. She subjects the qubit at her end to the corresponding unitary evolution σ_k, so that the pair goes to the state $|\Phi_k\rangle_{AB}$. Alice now sends her qubit to Bob. With both qubits in his possession, Bob measures the dynamical variable Ω_{AB} referred to in the teleportation protocol. It is clear that he will obtain the eigenvalue λ_k, *with certainty*, and thereby infer the two-bit number Alice wished to transmit.

Though this process, like most quantum information processing protocols, is assisted by entanglement created in advance, the point to note is that Alice interacted with just one qubit. This entanglement assisted transmission of two bits of classical information through the transfer of just one qubit has come to be known as *super dense coding*. In an obvious sense, it is the reverse of the teleportation process.

Suggested Reading

[1] R G Gallager, *Information Theory and Reliable Communication*, John Wiley, New York, p.1, 1968.

[2] Most of the papers dealing with QIS are to be found with the LANL e-Print Archive (quantum physics), as also with its mirror: http:\\xxx.imsc.ernet.in

[3] Shannon's 1948 work, 'A Mathematical Theory of Communication', is now available electronically: http:\\cm.bell-labs.com/cm/ms/what/shannonday/paper.html

[4] A valuable NSF report on 'Quantum Information Science' can be found at http:\\www.nsf.gov/pubs/2000/nsf00101/nsf00101.htm

[5] Two readable reviews are: C H Bennett and P Shor, 'Quantum Information Theory', and P Shor, 'Quantum Information Theory: Results and open problems', Both can be found at http:\\www.research.att. com/~shor/papers/index.html

[6] John Preskill has a valuable set of extensive lecture notes on Quantum Information and Computation: http:\\www.theory.caltech.edu/people/preskill/ph219

[7] Most aspects of QIS are covered in the carefully written book M A Nielsen and I L Chuang, *Quantum Computation and Quantum Information*, Cambridge Univ. Press, 2000.

Address for Correspondence
Rajiah Simon
The Institute of Mathematical Sciences, Tharamani
Chennai 600 113, India.
Email: simon@imsc.ernet.in