Theory of Molecular Machines.
I. Channel Capacity of Molecular Machines
running title: Channel Capacity of Molecular Machines

Thomas D. Schneider *

version = 5.76 of ccmm.tex 2004 Feb 3
Version 5.67 was submitted 1990 December 5
Schneider, T. D. (1991). Theory of molecular machines. I. Channel capacity of molecular machines. *J. Theor. Biol.* **148**, 83-123.
http://www.lecb.ncifcrf.gov/~toms/paper/ccmm

Like macroscopic machines, molecular-sized machines are limited by their material components, their design, and their use of power. One of these limits is the maximum number of states that a machine can choose from. The logarithm to the base 2 of the number of states is defined to be the number of bits of information that the machine could "gain" during its operation. The maximum possible information gain is a function of the energy that a molecular machine dissipates into the surrounding medium (P_y), the thermal noise energy which disturbs the machine (N_y) and the number of independently moving parts involved in the operation (d_{space}): $C_y = d_{space} \log_2(\frac{P_y+N_y}{N_y})$ bits per operation. This "machine capacity" is closely related to Shannon's channel capacity for communications systems.

An important theorem that Shannon proved for communication channels also applies to molecular machines. With regard to molecular machines, the theorem states that if the amount of information which a machine gains is less than or equal to C_y,

*National Cancer Institute, Frederick Cancer Research and Development Center, Laboratory of Mathematical Biology, P. O. Box B, Frederick, MD 21702. Internet address: toms@ncifcrf.gov.

then the error rate (frequency of failure) can be made arbitrarily small by using a sufficiently complex coding of the molecular machine's operation. Thus, the capacity of a molecular machine is sharply limited by the dissipation and the thermal noise, but the machine failure rate can be reduced to whatever low level may be required for the organism to survive.

> *If you want to understand life, don't think about vibrant, throbbing gels and oozes, think about information technology.*
> — Richard Dawkins [1]

1 Introduction and Overview

The most important theorem in Shannon's communication theory guarantees that one can transmit information with very low error rates [2, 3, 4, 5] (Appendix 20). The goal of this paper is to show how Shannon's theorem can be applied in molecular biology. With this theorem in hand we can begin to understand why, under optimal conditions, the restriction enzyme *Eco*RI cuts only at the DNA sequence $5'$ GAATTC $3'$ even though there are 4096 alternative sequences of the same length in random DNA [6, 7]. A general explanation of this and many other feats of precision has eluded molecular biologists [8].

Unfortunately it is not a simple matter to translate Shannon's communications model into molecular biology. For example, his concepts of transmitter, channel, and signal do not obviously correspond to anything that *Eco*RI does or has. Yet, a correspondence exists between a receiver and this molecule since both choose particular states from among several possible alternatives, both dissipate energy to ensure that the correct choice is taken, both must undertake their task in the presence of thermal noise [9], and therefore both fail at a finite rate (Appendix 21). By picking out a specific DNA sequence pattern, *Eco*RI acts like a tiny "molecular machine" capable of making decisions. Once the "molecular machine" concept has been defined, as best as is possible at present, we will begin to construct a general theory of how *Eco*RI and other molecular machines perform their precise actions. In doing this, we will derive a formula for the channel capacity of a molecular machine (or, more correctly, the machine capacity, equation (38)). The derivation has several distinct steps which parallel Shannon's logic [4]. These steps are outlined below.

The lock-and-key analogy in biology draws a correspondence between the fitting of a key in a lock and the stereospecific fit between bio-molecules [10, 11]. It accounts for many specific interactions. We will extend this analogy to include the moving "pins" in a lock, and then focus on each "pin" as if it were an independent particle undergoing Brownian motion.

To understand these motions, we consider simple harmonic motion of a particle, first in a vacuum and then in a thermal bath. The motion of many such particles serves as a model of how the important parts of a molecular machine ("pins") move.

Just as any two numbers define a point on a plane and any three numbers define a single point in three-dimensional space, the set of numbers used to describe the configuration of the machine define a point in a high dimensional "velocity configuration space".

We then show that the set of all possible velocity configurations forms a sphere whose radius equals the square root of the thermal noise energy. Similar spheres appear in statistical mechanics as the Maxwell speed distribution of particles in a gas [12, 13, 14].

When a molecular machine is primed, it gains energy and the sphere expands. When the molecular machine performs its specific action, it dissipates energy and the sphere shrinks while the sphere center moves to a new location. Because the location of the sphere describes the state of the molecular machine, the number of distinct actions that the machine could do depends on how many of the smaller spheres could fit into the bigger sphere without overlapping (Fig. 1). The logarithm of this number is the machine's capacity. Because the ⇐Fig 1
geometrical approach we take is the same as Shannon's approach [4], his theorem about precision also applies to molecular machines. Hence, although molecular machines are tiny and immersed in a thermal maelstrom, they are capable of taking precise actions.

The particular way that a molecular machine has evolved to pack the smaller spheres together corresponds to the way code words are arranged relative to one another in communications systems [15, 16]. This suggests that we should be able to gain insight into how molecular machines work and how to design them by studying information and coding theory.

2 Examples of Molecular Machines

In Jacob's hierarchy of physical, chemical, biological and social objects [17], molecular machines lie just inside the domain of biology, because they perform specific functions for living systems. Molecular biologists continuously unveil lovely examples of molecular machines [18, 19, 20, 21, 22] and many people have pointed out the technological advantages of building these devices ourselves [23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33]. If we were to consider only one kind of molecular machine at a time, we would miss the general features common to all molecular machines. Therefore, throughout this paper we will refer to the following four molecular machines.

1. The genetic material deoxyribonucleic acid (DNA) can act like a simple molecular machine. If DNA is sheared into a heterogeneous population of 400 base-pair long

fragments and then heated (or denatured by other means), the double stranded structure is "melted" into separate single strands. When the solution is slowly cooled, many of the single strands bind to a complementary strand and reform the double helix (Fig. 2a) [34]. ⇐Fig 2

Two characteristics make this reaction machine-like. First, a priming step (denaturation) brings the molecules into a high energy state. Second, the molecules dissipate the energy and anneal to one another in a reasonably precise way by using the complementarity between bases [35]. This "hybridization" reaction can be made so specific that it is widely used as a technique in molecular biology [34, 36, 37]. Base complementarity is also essential to all living things because it is the basis of nucleic-acid replication. For this reason, the degree of base-pairing precision is important in evolution.

2. The restriction enzyme *Eco*RI is a protein which cuts duplex DNA between G and A in the sequence $5'$ GAATTC $3'$ [38, 20, 39]. A single molecule of *Eco*RI performs three machine-like operations [8]. First, it can bind non-specifically to a DNA double helix. Second, after sliding along the DNA until it reaches GAATTC, it will bind specifically to that pattern. Third, it cuts the DNA. In the absence of magnesium, binding is still specific but cutting does not occur, so binding can be distinguished from cutting experimentally. We will focus on the binding operation (Fig. 2b). As with DNA, two characteristics make this reaction machine-like. First, there is a priming operation in which the non-specific binding to DNA places *Eco*RI into a "high" energy state relative to its energy when it is bound specifically. Second, the transition from non-specific to specific binding dissipates this energy so that *Eco*RI is located precisely on a GAATTC sequence. Without a dissipation associated with the specific binding, *Eco*RI would quickly move away from its binding site. After this *local* dissipation, the molecule is obliged to remain in place until it has cut the DNA, or a sufficiently large thermal fluctuation kicks it off again.

   In vivo cellular DNA is protected from *Eco*RI by the actions of another enzyme called the modification methylase. This enzyme attaches a methyl group to the second A in the sequence GAATTC, so that *Eco*RI can no longer cut the sequence. In contrast, invading foreign DNAs are liable to be destroyed because they are unmethylated. The methylase is precise, attaching the methyl only to GAATTC and not to any of the sequences, such as CAATTC, that differ by only one base from GAATTC [40]. So *in vivo Eco*RI is exposed to many hexamer sequences that are *almost* an *Eco*RI site, yet *under optimal conditions* [6, 7, 41] it only cuts at GAATTC. How a single molecule of *Eco*RI can achieve this extraordinary precision has not been understood [8, 42, 43].

3. The retina contains a protein called rhodopsin which detects single photons of light

[44, 45]. Upon capturing a photon, rhodopsin becomes excited and then dissipates the energy. Most of the time this converts rhodopsin into bathorhodopsin. A chemical cascade then amplifies the bathorhodopsin "signal" 400,000 times, leading to a nerve impulse. Because of this enhancement we can see single photons of light.

Why doesn't rhodopsin merely "use the energy" to convert directly into bathorhodopsin? This transformation is not as easy as it first appears, since the high energy state is a chemical transition state from which it is possible to go backwards to rhodopsin, rather than forwards to bathorhodopsin. Rhodopsin must make a "decision" about what to do.

4. Little is known about the exact molecular mechanism of muscles [46, 24, 47, 48]. However, we know that the interaction of the proteins myosin and actin consumes the energy molecule adenosine triphosphate (ATP). We may therefore imagine that the hydrolysis of an ATP molecule primes the actomyosin complex into a high energy state so that as the energy is dissipated a force is generated. As with rhodopsin, the activated actomyosin complex must "choose" whether to go forwards or backwards.

3 Definition of Molecular Machines

In each example given in the previous section, a specific macromolecule is primed from a low energy level—or ground state—into a high energy state. This is followed by a specific action that dissipates the energy and performs a function that is evolutionarily advantageous to the organism that synthesized the macromolecule. There are many other examples of molecular machines that follow this pattern [18, 21]. In general we will not be interested in the priming step, but rather with a precise measure of the specific action taken in exchange for the lost energy. The measure we will use is the number of distinct states which the machine can choose between. If the machine can select from two states, we say that it gains 1 bit of information per operation. Likewise, the selection of one state from amongst 8 corresponds to $\log_2 8 = 3$ bits per operation [5].

1. *A molecular machine is a single macromolecule or macromolecular complex.* In this paper we discuss the microscopic nature of individual molecules, not the macroscopic effects of large numbers of molecules. A molecular machine is not a macroscopic chemical reaction [24]. This does not deny that we can model a solution containing many molecules of *Eco*RI and DNA (without magnesium) by stating that the ratio of specifically bound to non-specifically bound molecules is constant once the reaction has reached equilibrium. This binding constant reflects the energetics of the individual

reactions (ΔG°), but it does not reveal the binding mechanism because that is independent of concentration. A single *Eco*RI molecule will cut a single DNA molecule irrespective of the number of other DNA and *Eco*RI molecules in the solution.

Suppose, for example, that we allow a macroscopic solution of DNA and *Eco*RI (without magnesium) to come to equilibrium at 37°C. Since individual molecules continue to bind and disassociate under these conditions, machine operations take place even after macroscopic equilibrium has been reached [28]. Thus, the operation of a single molecular machine cannot be treated as a macroscopic chemical reaction since that "stops" when equilibrium is reached. For this reason, the molecular machine model does not (and should not) refer to concentrations.

As McClare [24] pointed out, *each molecular machine acts locally as an individual*. Likewise Arrhenius *et al.* [31] distinguish functions at the molecular level from bulk material effects.

It is also worth noting that *Eco*RI alone is not a molecular machine. Only the combination of *Eco*RI *and* DNA is a molecular machine. Likewise, only the combination of a car *and* a road (or other suitable surface) can do useful work.

2. *A molecular machine performs a specific function for a living system.* That is, if the machine did not exist, the organism would be at a competitive disadvantage relative to an organism that had the machine. Thus, a molecular machine must be important for the evolutionary survival of an organism or it will be lost by atrophy. Shannon pointed out that information theory is unable to deal with the meaning or value of a communication [2, 3]. In biology, however, we work with the closely related concepts of function and usefulness, factors which are ultimately defined by natural selection. This part of the definition is important for accounting for the precision of molecular machines. Without a requirement for function, precision—or any other non-deleterious property—does not matter, just as nobody cares whether or not a car on a junk heap works. With a requirement for function, the very survival of the organism is it stake. In practical terms, the requirement for precise function dictates that the states of the molecular machine should be distinct and hence that the spheres represented by gumballs in Fig. 1 should avoid overlap.

   This definition encompasses machines that operate outside cells, such as digestion enzymes, and machines created entirely by humans [25, 30].

   (Even a Rube Goldberg[1] molecular machine's function would be to amuse, to educate, or to attempt to evade this definition.) Unlike simple chemicals like water, molecular

[1]The English equivalent is Heath Robinson.

machines are usually encoded by a genetic material and have the potential to evolve by natural selection.

3. *A molecular machine is usually primed by an energy source.* These include not only photons and ATP, but also thermal motions—as in the case of *Eco*RI separating from a binding site. (DNA heat-denaturation is an artificial method that only appears in the laboratory. Natural priming mechanisms usually do not use this macroscopic heating, although they frequently use the "microscopic heating" provided by thermal fluctuations.) Priming places the machine in an activated *before* state where it is ready to do work. The *before* state corresponds to the large sphere that encases the gumballs in Fig. 1.

   The act of priming is usually, but not always, required for a molecular machine to operate. For example, just after a new molecule of *Eco*RI has been synthesized, it is ready to operate even though it never was in a low energy state before.

4. *A molecular machine dissipates energy as it does something specific.* This phase of the machine's cycle is called its *operation*. Once the operation is completed, the machine is in an *after* state, which is represented by a single gumball in Fig. 1. Since the machine is always subject to thermal noise, an *after* state consists of the set of all possible motions that a single molecular machine could have at low energy. We will call this set an *ensemble*. Likewise the *before* state consists of the set of all possible motions that a single molecular machine could have at high energy, and this also forms an ensemble.

5. *A molecular machine "gains" information by selecting between two or more* after *states.* For example, *Eco*RI chooses one pattern out of $4^6 = 4096$ possible hexanucleotides, so it gains $\log_2 4096 = 12$ bits of information during its operation. Measurements of the amounts of information gained by genetic recognizers have been described in previous papers [49, 50, 51].

6. *Molecular machines are isothermal engines,* not heat engines [52]. They are obliged to operate at a single temperature because they do not have any way to insulate themselves from the huge heat bath that they are embedded in. However, they can use a priming energy to change their conformation to a more flexible one. This is essentially a controlled form of denaturation. After priming, any excess energy is quickly dissipated, leaving the molecule trapped in a flexible *before* state at the ambient temperature. In this state the machine is like a "frustrated" physical system [53] randomly searching through various conformations to find the correct one for the operation.

When this is found, the formerly inaccessible (i.e. potential) energy is quickly dissipated leaving the molecule once again at ambient temperature. This model allows for the evolution of a molecular machine from primitive beginnings because the energy is captured by a denaturation, which is simple and easy to achieve. The model does not require any form of molecular insulation or special vibrational modes which would be difficult if not impossible to evolve.

This paper shows that the number of parts of a machine, the energy dissipated per operation and the thermal energy in the machine determine the largest amount of information a molecular machine can gain (equation (38)). This "channel capacity of a molecular machine" (or, more accurately, "machine capacity") is measured in bits per operation, where one bit is the amount of information necessary to choose cleanly between two distinct machine states. This paper demonstrates that although the machine capacity is sharply limited by the amounts of dissipation and the thermal noise, the accuracy of the machine is not.

4 Lock-and-Key Model of a Molecular Machine

The state of a molecule is defined by the positions and motions of its atoms. To determine the locations of the n atoms in a molecular machine, we first define a coordinate system. Three spatial coordinates are needed to locate each atom, so we need $3n$ numbers. In many cases we won't care if the molecule is tumbling or moving through space, so we can affix the coordinate system to the molecule's center of mass and ignore the six numbers that describe the coordinate system's orientation and position in space. So for the positions we need no more than:

$$d_{space} = 3n - 6 \tag{1}$$

coordinate numbers (**Assumption 1**).[2] These coordinates are called "degrees of freedom". We also need d_{space} numbers to describe the velocities.

A molecular machine can only use a few of these degrees of freedom because many of the atoms are required as structural components. In this context it is useful to extend the lock-and-key analogy of biological interactions [10, 11]. A key opens a pin-tumbler lock by moving a set of two-part pins to positions which allow the two parts to separate when the key is turned [54, 55]. The wrong key will leave one or more pins in a position that blocks the turning, and this will prevent the bolt from being released. **Assumption 1** is that we only need to account for the motions of clusters of atoms—the molecular machine's "pins"—in

[2]The assumptions are listed in section 17 after equation (38). Only after the capacity formula has been constructed can we determine the consequences of relaxing each assumption. In most cases equation (38) remains the upper bound on the machine capacity.

order to describe its operation. Likewise, it is not necessary to keep track of the individual atoms in a lock in order to understand how it works.

A second, closely related assumption is that the parts of a molecular machine move independently (**Assumption 2**). Likewise *the pins in a lock move independently.* Yet because of the design of a lock, the bolt can only move if the pins are all aligned correctly by the key. Thus, although the individual pins are independent, they must "cooperate" for the lock to open. If two pins were not independent, then it would be easier to pick the lock, and it would not carry as much "protective" information because one pin could be set and the position of the other would be determined. For example, two pins fused together would act as one pin. Thus, in this analogy, d_{space} *refers to the number of "pins" used by the molecular machine,* which is quite likely to be much smaller than the degrees of freedom:

$$d_{space} \ll 3n - 6. \tag{2}$$

That is, the important degrees of freedom are not all of the degrees of freedom of the molecule, but only those directly involved in the machine operation. We only need to account for these to describe the machine's operation. Estimates of n and d_{space} for rhodopsin will be discussed later.

5 A Simple Harmonic Oscillator in a Vacuum

To demonstrate the method used in this paper, we first investigate the energetics of an oscillator which executes simple harmonic motion around its mean position without external interferences:

$$h(t) = a\cos(\omega t + \phi), \tag{3}$$

where $h(t)$ is the position of the oscillator as a function of time t, a is the amplitude of oscillation, ω is the frequency of vibration, and ϕ is the phase. This models the motion of a single molecular machine "pin". If we choose $r = -a\omega$, then the velocity is simply:

$$v(t) = \frac{dh(t)}{dt} = r\sin(\omega t + \phi). \tag{4}$$

The velocity has two independent Fourier components [56] with amplitudes x and y:

$$v(t) = x\sin(\omega t) + y\cos(\omega t). \tag{5}$$

From the trigonometric identity $\sin(A+B) = \sin A\cos B + \cos A\sin B$ and equations (4) and (5) we immediately find that $x = r\,\cos\phi$ and $y = r\,\sin\phi$. Fig. 3 represents these quantities ⇐Fig 3 graphically. On this graph, the point (x, y) completely defines the state of the oscillator at any time t. *It is important to keep in mind that x and y have units of velocity.*

In this paper we use the Fourier components (x, y) rather than polar coordinates (r, ϕ) because the Fourier description is symmetrical (x and y have the same units of velocity) whereas polar coordinates are not (they have units of velocity and angle).

The energy of the oscillator can be found from the maximum velocity and the mass:

$$E = \tfrac{1}{2} m v_{max}^2. \tag{6}$$

[57]. The total energy is also the sum of the energies of the two independent sinusoidal components in equation (5) [58], and since according to equation (4) $v_{max} = r$,

$$E = \tfrac{1}{2} m r^2 = \tfrac{1}{2} m x^2 + \tfrac{1}{2} m y^2, \tag{7}$$

so

$$r^2 = x^2 + y^2. \tag{8}$$

This equation shows that in a vacuum, where the total energy E is constant, the radius r is constant and the locus of the point (x, y) is a circle. That is, *at a given energy the set of all possible phase angles ϕ describes a circle of radius $r = \sqrt{\frac{2E}{m}}$ in a two dimensional velocity space whose axes are the amplitudes of the two independent Fourier components of the oscillator.*

6 A Simple Harmonic Oscillator in a Thermal Bath

If a simple harmonic oscillator is immersed in a thermal bath, then impacts with neighboring atoms change the phase and energy in an irregular way. Equipartition of energy between the oscillator and the bath implies that each independent Fourier component of the velocity in (5) has a Boltzmann distribution [14]:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-E_x/2\sigma^2} \tag{9}$$

and

$$f(y) = \frac{1}{\sigma\sqrt{2\pi}} e^{-E_y/2\sigma^2}, \tag{10}$$

where

$$E_x = \tfrac{1}{2} m x^2 \quad \text{and} \quad E_y = \tfrac{1}{2} m y^2. \tag{11}$$

The meaning of σ will be discussed below. We use the Boltzmann distribution to introduce thermal noise into our Newtonian description of an oscillator. Substituting from (11) into (9) and (10) gives:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-mx^2/4\sigma^2} \tag{12}$$

and

$$f(y) = \frac{1}{\sigma\sqrt{2\pi}} e^{-my^2/4\sigma^2}, \tag{13}$$

so the velocities x and y have a normal or Gaussian distribution with a standard deviation proportional to σ. Since the oscillator is surrounded by a huge thermal bath and impacts from the bath are not predictable, the changes in motion of the oscillator are probabilistic. Maxwell's classical model for the velocity distribution of molecules in an ideal gas also uses a Gaussian velocity distribution [12, 13, 14]. The normal distribution is graphed as the $D = 1$ curve in Fig. 4. ⇐Fig 4

What is the probability $f(x,y)$ that the oscillator will have the velocity components x and y? Since x and y are independent, we may write the probability density as

$$f(x,y) = f(x)f(y) = \frac{1}{\sigma^2 2\pi} e^{-m(x^2+y^2)/4\sigma^2} = \frac{1}{\sigma^2 2\pi} e^{-mr^2/4\sigma^2}, \tag{14}$$

where $r = \sqrt{x^2+y^2}$ is the distance in velocity space from the origin to the point (x,y), as in Fig. 3. The probability of finding that the oscillator has velocities in a small region $dxdy$ is $f(x,y)dxdy$. Since $dxdy = rdrd\phi$ [59] we can convert to polar coordinates:

$$f(x,y)dxdy = \frac{1}{\sigma^2 2\pi} re^{-mr^2/4\sigma^2} drd\phi. \tag{15}$$

The total density at the radius r in an interval dr is therefore

$$f_2(r)dr = \int_0^{2\pi} \frac{1}{\sigma^2 2\pi} re^{-mr^2/4\sigma^2} drd\phi = \frac{1}{\sigma^2} re^{-mr^2/4\sigma^2} dr. \tag{16}$$

The subscript "2" in "$f_2(r)$" indicates that two Gaussian distributions were used to obtain the density distribution. This "Rayleigh" distribution is graphed as the $D = 2$ curve in Fig. 4 and shown as a smooth grey scale in Fig. 5. Notice that the distribution is radially symmetric ⇐Fig 5 and that the density in a thin ring around the origin approaches zero at the origin since $r = 0$ there.

We found in the previous section that when an oscillator is in a vacuum the total energy is constant so that the radius r is constant and the set of all possible states with energy r^2 is represented by a circle. In a heat bath the oscillator can exchange energy with the surrounding medium and the distribution is more spread out, according to the Rayleigh distribution. This "open" description of a simple harmonic oscillator allows for energy and phase changes.

7 A Simple Molecular Machine in a Thermal Bath

We will assume that the energies in the independent "pins" of a molecular machine form a Boltzmann distribution (**Assumption 3**) so each "pin" acts like a simple harmonic oscillator in a thermal bath and

$$D = 2d_{space} \tag{17}$$

numbers are required to describe the machine velocities because each "pin" has a phase and an amplitude (*i.e.* two Fourier components x and y). At a given moment the energy of the j^{th} such "pin" component is determined by its velocity and the "pin's" mass:

$$E_j = \tfrac{1}{2} m_j v_j^2. \tag{18}$$

So that we will be able to easily compare "pins" with different masses we combine the velocity with the square root of the mass to define a new variable:

$$y_j = \sqrt{E_j}. \tag{19}$$

The assumption that energies of the "pins" have a Boltzmann distribution implies that

$$f(y_j) = \exp(-\beta E_j)/z, \tag{20}$$

where z is the "partition function", $z = \int_{-\infty}^{\infty} \exp(-\beta E_j) dy_j$ [14]. Dividing by z assures us that the probabilities $f(y_j)$ sum to 1. $\beta = (k_{\rm B} T)^{-1}$ where $k_{\rm B}$ is Boltzmann's constant and T is the absolute temperature. Comparing (20) to (9), we find $\beta = \frac{1}{2\sigma^2}$ so $\sigma^2 = \frac{1}{2} k_{\rm B} T$.

Substituting for the energy by using (19) we find that

$$f(y_j) = \exp(-\beta {y_j}^2)/z. \tag{21}$$

The form of this equation shows that the set of y_j normalized velocity components have a Gaussian distribution.

8 *Y* Space: A High Dimensional Model of Molecular Machines

By placing the magnitudes of the independent y_j numbers at right angles to one another, we form the coordinates of a single point in a space of D orthogonal dimensions. To paraphrase

Shannon: "Essentially we have replaced a complex entity (the velocity configuration of a macromolecule) in a simple environment (three dimensional space) by a simple entity (a point) in a complex environment (D dimensional space)" [4, 60]. The space defined by all possible values of y_j is called Y space.

The reader may feel that such a high dimensional space is difficult to think about. Fortunately, it is always possible to visualize the two or three dimensional cases. We have already done this for the 2D oscillator. It is also worth keeping in mind that a point in a D-dimensional space is defined by nothing more than a list of D numbers [61]. For example, a lock with 10 pins is a 20 dimensional machine because 20 numbers are needed to define the positions and velocities of the pins. Using a high dimensional space enormously simplifies the problem of understanding molecular machines because in such a space both the *before* and *after* states of the machine are represented by hollow spheres.

9 The Energetics and Distribution of Molecular Machines in Y Space

Our next task is to determine the distribution of all possible machine configurations at a given ambient temperature. From here on *we will no longer be discussing just one molecular velocity configuration*, but rather the entire set of configurations that satisfy the distribution given by (20). This collection is called a molecular machine "ensemble" or a "state" of the machine. The probability density in Y space is the product ($\prod$) of the individual independent probabilities:

$$f(y_1, \ldots, y_j, \ldots, y_D) = \prod_{j=1}^{D} f(y_j). \tag{22}$$

Using equation (20) this becomes

$$f(y_1, \ldots, y_j, \ldots, y_D) = \exp(-\beta \sum_{j=1}^{D} E_j)/z^D = \exp(-\beta N_y)/z^D, \tag{23}$$

where N_y is the total thermal noise energy in the "pins" of the molecular machine. If instead of using (20), we combine (21) with (22) we obtain

$$f(y_1, \ldots, y_j, \ldots, y_D) = \exp(-\beta \sum_{j=1}^{D} {y_j}^2)/z^D = \exp(-\beta {r_y}^2)/z^D, \tag{24}$$

where we have used the Pythagorean theorem to collapse the D orthogonal y_j^2 values into a single variable, r_y, which is the radial distance from the origin to one of the possible points describing the motions of the machine. Comparing equation (23) to (24) shows that:

$$r_y = \sqrt{N_y}. \tag{25}$$

The Boltzmann distribution we have assumed for the "pins" implies that the parts of the machine are at equilibrium with each other. At equilibrium the machine is not dissipating energy and the thermal noise N_y is roughly the same for every possible configuration. So r_y is also roughly constant (Appendix 22). Since a constant distance does not imply any particular direction in space, *the set of possible motions of the machine form a sphere in Y space.*

Shannon called such spheres "sharply defined billiard balls" [4], but perhaps the ping-pong ball is a more apt analogy because at high dimensions most of the density of a sphere is close to the surface. This is demonstrated in Fig. 4, where one can see that at higher dimensions the sphere density becomes tightly focused. The derivation of the distributions for the higher dimensions is given in Appendix 22.

Brillouin [62, 60] gave the following simple proof of this curious property. Just as the area of a circle is proportional to the radius squared, and the volume of a sphere is proportional to the radius cubed, the volume of a D-dimensional sphere of radius r is proportional to the radius raised to the dimension D:

$$V = \frac{\pi^{\frac{D}{2}}}{\Gamma\left(\frac{D}{2}+1\right)} r^D, \tag{26}$$

where Γ is the gamma function [63, 4, 64]. Taking the derivative gives us

$$dV = \frac{\pi^{\frac{D}{2}}}{\Gamma\left(\frac{D}{2}+1\right)} D r^{D-1} dr \tag{27}$$

and dividing (27) by (26) gives

$$\frac{dV}{V} = D\frac{dr}{r}. \tag{28}$$

This equation means that a fractional change in the radius (dr/r) is magnified by the dimension (D) to get the fractional change in the volume (dV/V).

Even for a small molecule, D can be enormous. For example, Warshel [65] modeled the light-activated "switch" in rhodopsin, 11-*cis* retinal, with 200 vibrational modes. To emphasize the potential for high dimensions, we will find a minimum for the number of dimensions needed to describe the motion of this vitamin A derivative. A minimum number of atoms to model would be the 20-carbon backbone of retinal, so $n = 20$, $d_{space} = 54$ and $D = 108$.

Is this enough to create a sharply defined sphere? Suppose the radius of the sphere describing retinal is 11 in Y space units. (It doesn't matter what these units are, since they cancel in Equation (28).) Then as the radius increases from 10 to 11 units, the volume increases by $\frac{dV}{V} = 108 \cdot \frac{1}{10} \approx 11$ fold. More than $\frac{dV}{V+dV} = 90\%$ of the volume is already concentrated in the outer 10% of the sphere.

Our estimate for n is conservative because we did not include the 28 hydrogen atoms in retinal nor any part of the 39,048 Dalton opsin protein to which retinal is attached [66, 67], nor the surrounding water and membrane (which undoubtedly have important effects on molecular motions [68, 69]). There are 5511 atoms in rhodopsin, so for rhodopsin alone D could be as high as 33,054. Not all of the atoms can be directly involved in the mechanism (**Assumption 1**), but it is clearly possible for the number of dimensions used by the machine to be large.

Exact calculation of the sphere density as a function of radius (Appendix 22) shows that the sphere surface becomes sharply defined at higher dimensions (Fig. 4), so the entire set of possible motions of even a small molecular machine are well depicted by a ping-pong ball.

10 Simulation of High Dimensional Spheres

We can generate the spheres both by numerical simulation and analytically. The two methods are illustrated together in Fig. 6. For a numerical simulation, the motions of a molecular machine would be determined by the technique of molecular dynamics [70]. If the "pins" have been identified—which probably requires understanding how the machine works—then we could obtain a set of y_j. When the machine parts are at equilibrium, these should have independent Gaussian distributions (**Assumption 2**, **Assumption 3**). So instead of doing molecular dynamics, we can use *any* set of real numbers having a Gaussian distribution. These are easily created by adding together many pseudo-random numbers that have a flat distribution [71]. The central limit theorem assures us that the resulting sum approaches a Gaussian distribution [72]. A set of D such numbers forms the point in Y space. ⇐Fig 6

To see what the distribution of these points looks like, we can map the sphere onto a plane. The method is equivalent to moving cities from their particular latitudes and longitudes on a globe to their longitudes at the equator. In Fig. 6 we have mapped 4-dimensional Y space points onto the page by this method. From equation (24) the radial distance from the origin to a point in Y space is given by:

$$r_y = \sqrt{\sum_{j=1}^{D} {y_j}^2}. \tag{29}$$

When $\sigma = 1$, the distribution has a maximum at $r_{max} = \sqrt{D-1}$ (Appendix 22), so the radius was normalized by dividing by $\sqrt{D-1}$. The direction (angle) of each point was arbitrarily taken from two of the y_j coordinates.

To graph the corresponding smooth analytic function, we chose points on the page and determined their distance from the origin to obtain r_y. We then find the probability density directly from the $f_D(r)$ function given by equation (48) of Appendix 22.

Fig. 7 shows the correspondence between the simulated points and the smooth analytic function. As the dimensionality increases, the spheres become sharper. The concentration of points at a particular radius is a consequence of the enormously increasing volume as the radius increases in higher dimensions. So, although the density is highest in the center according to equation (24), the majority of points are found far from the center. ⇐Fig 7

11 Thermal Noise in Y space

An oscillator in equilibrium with a thermal bath has an average energy of $k_{\mathrm{B}}T$ joules [14]. Since a molecular machine has d_{space} "pins", each of which is assumed to be equivalent to an oscillator, the total energy is

$$N_y = k_{\mathrm{B}}T \times d_{space} \qquad \text{(joules).} \tag{30}$$

This expression also gives the average energy of a molecule with d_{space} degrees of freedom [13]. From equations (30) and (17) we see that

$$N_y = \tfrac{1}{2}k_{\mathrm{B}}T \times D \qquad \text{(joules)} \tag{31}$$

so each of the y_j variables has an average energy of $\frac{1}{2}k_{\mathrm{B}}T$. Combining equation (30) with equation (25), we find

$$r_y = \sqrt{k_{\mathrm{B}}T d_{space}}. \tag{32}$$

Thus, thermal noise displaces the configuration of the machine away from the sphere center by an amount related to the absolute temperature. For this reason we may regard the gumballs of Fig. 1 as representing "thermal noise spheres".

12 Location of Spheres in Y Space

The square of a distance in Y space is equal to the energy required to traverse that distance. Suppose that there are two *after* states of the machine. (That is, two gumball spheres.) If the distance from the first state to the second state is big enough, then the velocity configuration of the machine—which is represented by a point—would almost never be able to jump from one sphere to the other and the two spheres would be separated by an energy "barrier". On the other hand, if the distance between the sphere centers were small, the two spheres would be connected and the machine would often have enough thermal energy to make the transition to the other state.

An analogy is useful to see how the states could become connected. Suppose that we have a coin lying in a tub. There are two states, heads up and tails up. A certain fixed minimum amount of energy is required to lift the edge of the coin in order to flip it over. If we start to vibrate and shake the tub, then the probability that the coin will switch to the other side increases. If we successively replace the coin with each of the 5 regular Platonic solids—tetrahedron (4 sides), cube (6 sides), octahedron (8 sides), dodecahedron (12 sides), and icosahedron (20 sides)—while keeping the mass the same, then switching between sides (states) becomes increasingly easy. With more intense shaking, the states also become less and less distinct.

The tub vibrations correspond to the temperature, which determines the radius of the molecular machine's spheres according to equation (32). Thus, at higher temperatures the sharply defined spheres overlap and the states are no longer distinct. A molecular example is the heat denaturation of double stranded DNA.

Specifying the location of the center of a sphere in Y space specifies the average configuration of the molecule relative to other possible configurations. To be able to discuss several spheres at once, we can represent the shape of the Y space ensemble with a vector notation:

$$\vec{y} = \vec{s} + \vec{N_y}. \tag{33}$$

The center of the sphere is defined by a vector, $\vec{s} = (s_1, \ldots, s_j, \ldots, s_D)$, while the instantaneous radius of the current point on the sphere is defined by the vector $\vec{N_y} = (y_1, \ldots, y_j, \ldots, y_D)$. The magnitude of $\vec{N_y}$ is given by any of the relations (25), (29), (32) or

$$|\vec{N_y}| = \sqrt{N_y}. \tag{34}$$

The s_j and y_j variables play important roles in this paper, since they correspond to the signal samples Shannon used in his theory. The set of variables that define the center of each sphere, s_j, plays the part of DC voltages, while the y_j correspond to AC voltages due to thermal noise (Appendix 21).

13 Molecular Machine Operations

So far we have modeled a molecular machine jiggling at equilibrium, and we found that it can be represented by a sphere in Y space. Now let's investigate what happens when the machine operates. Recall our four example machines. For DNA an operation means base-pairing or hybridization, for a genetic recognizer like *Eco*RI it means locating a binding site, for rhodopsin it means switching to bathorhodopsin, and for muscle it means contracting. We say that *"information is gained"* when a machine changes from an indeterminate state to a more determined state. Decreases in thermal motion corresponding to machine operations have been observed in many molecular machines [18, 73]. In each case, a corresponding energy decrease allows a specific action to be taken. Thus, rhodopsin dissipates the energy of a photon to change states [65, 74, 75] and actomyosin dissipates the energy of a hydrolyzed ATP molecule to generate motion [24, 47]. When DNA becomes double-stranded [34], or when genetic recognizers stick to their binding sites [76, 77, 8], their range of motion becomes restricted by a lower potential energy.

We only need to consider two energetic states of the machine [8]. Before an operation, a machine has some specific amount of energy, while afterwards it has a smaller amount. How the machine attains the activated *before* state (i.e., "priming") is outside the scope of our considerations, though we may note that a photon does this for rhodopsin [74], and ATP hydrolysis does it for actomyosin [47]. Even large (but rare) thermal fluctuations can cause this priming, since they can free repressors and other proteins like *Eco*RI from their binding sites [76]. Likewise, DNA strands may be separated artificially by heat and chemical denaturants, or naturally by helicases using ATP, while bases incorporated into growing nucleic-acid chains are already separate.

14 The *before* and *after* states in Y Space

Let us now consider the energetics of the two states of *Eco*RI. When the molecule is bound to its sites in the *after* state, its "pins" have an energy determined by the thermal noise. Each possible configuration of the machine is represented by a point in Y space, and the set of all such points forms a sphere of radius

$$r_{after} = \sqrt{N_y} \tag{35}$$

according to equation (25).

In the *before* state, *Eco*RI must have an internal energy higher than it does in the *after* state, or it could not stick to the binding site in the *after* state. We will call the extra energy P_y, so that the total energy *before* is $P_y + N_y$. P_y is the energy difference between the states. We will assume that the energy $P_y + N_y$ is equally partitioned between all the degrees of freedom open to the molecule (**Assumption 4**). This is reasonable for *Eco*RI since in the *before* state *Eco*RI wanders by Brownian motion along the DNA. Only when *Eco*RI encounters the sequence GAATTC can the energy P_y be dissipated (**Assumption 5**).

In Y space, the configuration of the machine is represented as a noisy vector displacement from the sphere center (equation (33)). If we add an energy P_y to the machine, the effect in Y space is to add a vector $\vec{P_y}$ of magnitude $\sqrt{P_y}$ to the noise vector $\vec{N_y}$. But in the high dimensional Y space, most of this additional noisy energy will be directed at 90° to the original noise energy because there are so many possible directions in the space. For example, if one were in the center of a three dimensional globe looking north, $2/3$ of the noise would be at 90° to the direction of sight. Likewise, if $D = 100$, then 99% of the noise would be at right angles to any given direction.

Therefore, as shown in Fig. 8, the two vectors $\vec{P_y}$ and $\vec{N_y}$ form a right triangle, whose hypotenuse is $\sqrt{P_y + N_y}$ according to the Pythagorean theorem. Since both $\vec{P_y}$ and $\vec{N_y}$ may point in any direction, *the before state is represented by a sphere* of radius ⇐Fig 8

$$r_{before} = \sqrt{P_y + N_y} \tag{36}$$

with an energy $r^2_{before} = P_y + N_y$, which is the total energy that we defined initially. It is difficult to see this geometry in three dimensions.

15 Machine Operations in Y Space

Once the machine dissipates energy, the vector $\vec{P_y}$ becomes a specific direction relating the *before* and *after* states. Referring to Fig. 8, we see that the *before* sphere has its center at point O, while the *after* sphere has its center at point B. In this two-dimensional diagram, the *after* sphere is represented by the line segment that extends from C to A. (The *after* state is still spherical, but the two-dimensional diagram cannot show it. In three dimensions, the *after* sphere would be represented by a circle at a particular latitude on a globe.) Because of the high dimensionality most of the *after* sphere is "flattened" at 90° *with respect to the specific direction of* $\vec{P_y}$, which is shown as $\vec{OB}$ in the figure.

Therefore, the machine operation corresponds to the motion of the sphere center from O to B with a concomitant collapse of the radius, and loss of energy P_y to the surrounding medium.

Since $\vec{s}$ represents the average configuration of the machine, *a change in the sphere center,* $\Delta\vec{s} = \vec{P_y}$, *corresponds to a change in the average physical configuration of the molecular machine,* and different directions and magnitudes of $\vec{P_y}$ in Y space correspond to different state changes. Furthermore, the location of a small *after* noise sphere within the larger *before* sphere represents only one of several possible states of the machine since there can be several non-intersecting *after* spheres [78, 79, 18]. Placement of the spheres according to equation (33) is called the molecular machine's *coding scheme* because the packing of spheres in space corresponds to the arrangement of code words in a communications system [15]. The total dimensionality, D, determines how sharp-edged the spheres are, and so this controls the intensity of the threshold effects if two spheres overlap [4]. Thus, the precision of a molecular machine depends on its size. If the machine is big enough ($n \gg 1$), then *the noise is predictable* because D may become so large that spheres are sharp-edged. By evolving to be big, even single molecules can have macroscopic stability. If the machine contains enough independent components, then the spheres may also be placed accurately in the space of D dimensions so that they just barely miss contacting each other. Thus, the machine can have distinct *after* states. However, since the spheres are defined by a smooth analytic function ($f_D(r)$, equation (48)), they always overlap and there is always a small probability that a machine in one *after* state can jump into another *after* state. The rate of such transitions (or incorrect transitions from *before* to *after*) is the error rate.

Of course, simply increasing the number of atoms in order to raise the dimensionality does not guarantee accurate placement of the spheres. However, the number of "pins" can be increased during evolution of the machine, so the placement could be refined. This suggests, for example, that many of the amino acids in a large protein could have subtle effects on the sphere placement and coding [42]. These effects could be missed by conventional genetic approaches that are based on the premise of finding "the" major recognition factor. For example, recent X-ray crystal structure determination of a tRNA synthetase bound to its cognate tRNA [80, 81, 82] suggests that the complete set of tRNA recognition factors is spread over a large surface of both molecules [83] (as one would expect from this theory) rather than concentrated in the anti-codon or other small regions.

We should emphasize that the configurations (points in Y space) that we have been considering are in either the *before* or the *after* states. We have not looked at configurations during the operation. Since the energy changes during an operation, a set of such configurations must connect the *before* to the *after* spheres. As we will see in the next section, it is to our advantage to focus only on the simple spherical *before* and *after* states, for together these characterize what the machine is able to do.

16 Derivation of the Machine Capacity

How many distinct *after* states can there be? Certainly the largest number of distinct states that the machine could have after dissipation of its energy cannot be bigger than the maximum number, M_y, of small *after* spheres that can be packed into the volume of the large *before* sphere, as suggested by Fig. 1 (**Assumption 6**, see also the second part of the definition of molecular machines) [15, 61]. We obtain this by dividing the volume of the larger sphere by the volume of a smaller sphere [4]:

$$M_y \leq \frac{V_{before}}{V_{after}} = \left(\sqrt{\frac{P_y + N_y}{N_y}}\right)^{2d_{space}} \tag{37}$$

using equations (26), (36), (35), (17), and the fact that a sphere volume is proportional to the radius raised to the dimension (D) that the sphere is embedded in. The "machine capacity" is the maximum information, $\log_2 M_y$, that could be gained during the operation [2, 3, 5, 4]:

$$C_y = \log_2 M_y \leq d_{space} \log_2 \frac{P_y + N_y}{N_y} \quad \text{(bits per operation).} \tag{38}$$

Aside from a constant due to the nature of the different situations, this equation is identical in form to Shannon's famous channel capacity formula (equation (45) in Appendix 20). In Appendix 21 we discuss how Shannon's precision theorem applies to the case of molecular machines and in Appendix 23 we discuss a more general derivation.

17 Assumptions

Assumption 1 *Only some of the atoms in a molecular machine are involved in an operation.* For example, if the flip of a tyrosine ring in bovine pancreatic trypsin inhibitor has no function [70] or effect on sphere sharpness or placement then d_{space} is effectively less than $3n-6$. In this paper d_{space} is taken to refer only to the number of spatial degrees of freedom involved in the operation. Even with the restriction of equation (2), d_{space} can still be large in a typical macromolecule, so (38) still applies.

Most protein dynamics are well modeled with just the locations of the nuclei, and quantum corrections are small at 300K [84, 85]. If quantum effects were used in a machine operation, d_{space} would be given by the number of independent parameters that are required to describe the system.

Two independent "pins" need not have the same importance to the organism. If we use information content as a measure of "importance", we can see that the "importance" of various bases in a binding site is strongly dependent on the position in the site [49, 50, 86]. Likewise, one pin in a lock could have more "importance" than other pins if it used more distinct levels than the others.

Assumption 2 *The important parts of the molecular machine move independently.* In the lock-and-key analogy, this assumption is that the pins of the lock move independently of one another. However, it is possible for one part of a molecular machine to affect the motions of its neighbors. In communications there are similar phenomena [4]. Regions of a television picture are correlated to one another, and each frame is often similar to the next. Shannon pointed out that this simply reduces the number of independent parameters. So correlations between parts of the machine effectively reduce the dimensionality by confining the machine to surfaces in Y space. If the dimensionality is reduced, then C_y remains the upper bound, as can be seen from equations (37) and (38).

This assumption has a biological rational. It asserts that the components of a molecular machine can become independent through natural selection. For example, where it is important that two successive amino acids in the chain of a protein move independently to satisfy the protein's function, mutational insertions in the gene for the protein will confer a selective advantage. Eventually a flexible segment may evolve that allows the amino acids to move nearly independently.

The linear structure of binding sites on nucleic acids suggests that parts of the binding site recognizers could operate independently in the same sense that lock pins are independent. Three lines of evidence support this idea. First, it is possible to train a linear perceptron to identify ribosome binding sites and splice junctions [87, 88, 89]. Second, it is possible to predict the amount of translational initiation using a linear model of the 12 bases preceding and including the first base of the initiation codon of ribosome binding sites [90, 91, 92]. There are similar data for the Cro, λ and *lac* repressor binding sites [93, 94]. Third, the contribution of individual amino acids to the total association free energy between proteins has been found to be additive in a number of cases [95, 96]. The success of these approaches suggests that at least some parts of molecular machines exhibit independence and that further experimental work may allow us to map the locations of the "pins".

It is possible that a transformation of the descriptive variables is required to reveal independence. For example, if the transformation involved in harmonic analysis provides a good model for a particular molecular machine [57, 97, 98, 99, 100, 101, 85, 102] then the modes are guaranteed to be independent, and the equipartition theorem [14] guarantees that the energy is evenly distributed over all $3n-6$ modes [103]. A molecular machine need not use all of these modes.

The independence assumption has a curious consequence. Since its components are independent, the machine is modeled as an *ideal gas* in Y space and a machine operation is represented by the collapse of this gas. The entropy decrease is simply the log of the ratio of the volumes (equation (37)), as in classical thermodynamics [13]. The decrease in entropy of the molecular machine is proportional to the information it gains.

Assumption 3 *The energetics of molecular machine components ("pins") are described by a Boltzmann distribution* [14, 97, 62]. This is equivalent to assuming that each component is affected by band-limited white Gaussian noise [104, 105, 106, 4, 58, 73] or Brownian motion [107] in which the velocity of a particle is the sum of many small impacts. Atomic fluctuations in proteins are well characterized by Gaussian distributions [85].

Shannon considered the case of the channel capacity with an arbitrary type of noise [4]. He pointed out that white Gaussian noise is the worst possible noise, and that other kinds of noise exist. As Shannon noted for communications systems, the ensemble states of molecular machines are not spherical when the noise is not white Gaussian. This is equivalent to changing the energy function of the "pins". For example, suppose that the energies were related to the maximum velocities x and y by

$$E_x \propto |x|^m \quad \text{and} \quad E_y \propto |y|^m \tag{39}$$

instead of the form $E \propto x^2$, as in equation (6). Then the total energy would be proportional to

$$|x|^m + |y|^m = |r|^m. \tag{40}$$

This may or may not be physically realizable, but we can use it to illustrate the possible properties of non-Gaussian noise. The case of $m = 2$ produces a circle, as in Fig. 3. This represents Gaussian noise. If $m = 1$ then the formula reduces to a line segment in the positive quadrant. This is reflected around the origin by the absolute value functions, to produce a "diamond" shape, as shown in Fig. 9. The figure also shows that there are a set of curves that ⇐Fig 9
lie between $m = 1$ and $m = 2$.

If $m > 2$ then the curve bulges outward and the limit as $m \to \infty$ is a square! These shapes exceed the area of a circle with the same total energy. Now consider how these objects could be packed together. Circles could be packed into a hexagonal array. In contrast, the same hexagonal packing of the rounded squares would cause them to overlap, so circles produce a higher channel capacity. Since a molecular machine could obtain circles by evolving springs that move by simple harmonic motion, the $m > 2$ case could be avoided. This is why white Gaussian noise, where $m = 2$, is the worst possible noise. When $m < 2$ the area is less than that of a circle. At $m = 1$, the shape becomes a diamond and below this the shape is concave and has cusps. Since these spiky shapes can be packed more closely than circles, the capacity can be reduced in the absence of Gaussian noise. Similar effects occur in higher dimensions and with other force functions.

In general, if the effective "entropy power" of a noise N_1 is less than the white Gaussian noise N_y ($N_1 \leq N_y$) then

$$\frac{P_y + N_y}{N_y} \leq \frac{P_y + N_y}{N_1} \tag{41}$$

so the machine capacity is bounded by

$$C_y \leq d_{space} \log_2 \left(\frac{P_y + N_y}{N_1} \right) \tag{42}$$

and the upper bound exceeds the bound given by equation (38) [4]. We can see this geometrically from the example given above. If the shape of the *before* state is spherical (i.e. the radius is $\sqrt{P_y + N_y}$), and the shape of the *after* state is spiky (i.e. the radius is effectively $\sqrt{N_1}$), then we obtain the upper bound of (42).

In this paper we have defined a classical physics benchmark against which we may examine real systems to see how well they do. Can a biological system use quantum effects to circumvent white Gaussian noise? By experimentally investigating the capacity of real molecular machines, it may be possible to answer this question.

Assumption 4 *The* before *state is in equilibrium.* The shape of the machine ensemble is spherical in the *after* state because the machine has reached equilibrium with its surroundings and the "pins" have a Boltzmann distribution (**Assumption 3**). In some cases the *before* state is also in equilibrium because the machine is a "frustrated" physical system [53]. For example, on a time scale far shorter than it takes to find a binding site, a molecule of *Eco*RI should come to equilibrium with the surrounding solution. In contrast, if rhodopsin does not have a "frustrated" state, then one vibrational mode of rhodopsin might absorb more energy from a photon than the other modes, so that the ensemble would become an ellipsoid in Y space. However, of all possible ellipsoids, a sphere contains the largest possible volume given the constraint that the energy is constant. (For an ellipse, $\left(\frac{x}{a}\right)^2 + \left(\frac{y}{b}\right)^2 = r^2$, the area, πab, is maximized when $a = b$.) So if the energies are unequally distributed in the *before* state, the volume will be smaller than that given by equations (36) and (26), M_y will be decreased (equation (37)), and hence the information gain, R, will be below C_y (equation (38)). Thus C_y remains the upper bound. We call this argument "The Ellipsoidal Defense".

It is advantageous for a molecular machine, such as rhodopsin or actomyosin, to operate as close to its capacity as possible, because then it would gain as much information as possible for a given energy dissipation. To operate near capacity, the machine must have, or equilibrate to, a spherical *before* state. In other words, the entropy of the *before* state will tend to be maximized by evolution, and the Ellipsoidal Defense is an argument that it is advantageous to the organism to allow the entropy of the *before* state to be maximized [49, 50]. Indeed, there is evidence for "complete thermal relaxation" in the *before* state of rhodopsin [108, 74]. Complete thermal relaxation could easily be obtained by rhodopsin if it enters a "frustrated" state when excited by a photon. It is possible that this relaxation improves rhodopsin's capacity to detect light.

Assumption 5 *None of the power is wasted.* If only part of P_y is used by the machine to make selections, while the rest is dissipated directly, then the rate that the machine gains information, R (bits per operation), would be lower than the right hand side of equation (38), and C_y would remain the upper bound.

Assumption 6 *The* after *spheres are perfectly packed and do not overlap.* The *after* spheres could overlap. This effectively reduces the number of distinct *after* states M_y, and lowers the capacity according to equation (38). Thus C_y remains the upper bound.

Sphere overlaps represent transitions or isomerizations between semi-distinct states of the machine [109, 85, 110]. To see this, consider two *after* spheres that are so close together that they overlap. A point which is in the overlap region between the spheres could be considered to be part of either sphere. Now recall that each point in Y space represents a velocity configuration of the machine. A moment later the machine has moved, and this corresponds to a point somewhere else on the sphere. If the machine starts out on one of the spheres, and is in the overlap region next, it could easily end up on the other sphere. Since the other sphere represents a different *after* state of the machine, the machine would have two states but they would not be distinct because the machine would keep switching between them. The rate that the machine switches states depends on the volume of the overlap region relative to the size of the spheres.

These conformational substates may exist in either the *before* or the *after* machine states. If the *before* state is broken into several connected conformational substates, one can find a machine with a higher capacity by *joining* the substates, since this increases the volume of the *before* state. In contrast, if an *after* state is broken into several conformational substates, a better machine can be found by *separating* the substates, since this would increase the number of distinct *after* states and so increase the capacity of the machine.

As an example, suppose that an RNA polymerase inserts the four bases at $W_s = 200$ operations per second [111]. Since it performs $R = \log_2 4 = 2$ bits per operation, it operates at $W_s R = 400$ bits per second, which we will take to be its capacity. (See Appendix 23 for a discussion of various forms of the capacity.) Now suppose that the temperature is raised, increasing the thermal noise and swelling the *after* spheres so that they overlap (equation (32)). Suppose that A and G become indistinguishable, that C becomes indistinguishable from U, but that the operating rate is not increased significantly by the temperature increase. Then the machine performs only 1 bit per operation at a rate of 200 bits per second. This shows how blurring the distinction between *after* states decreases the machine's rate of operation below the machine capacity.

18 Toward a Coding Theory for Molecular Machines

If two *after* spheres are placed too close together, then they overlap. Since this leads to semi-distinct substates that decrease the capacity (**Assumption 6**), it is advantageous for a biological system to have a good way to pack the spheres together. With a good packing, less energy needs to be dissipated per operation because the enclosing *before* sphere can be smaller. The placement of the *after* spheres is the *coding scheme* of the molecular machine, because finding a good sphere packing method is the same problem as finding an optimal communications code [15, 61, 112, 113, 114, 16]. Since every molecular machine has its own code, there are many codes in molecular biology besides the genetic code.

In Shannon's communication model (Appendix 20), long delay periods are required to encode and decode the signal. The delay increases the dimensionality of the space (because more numbers are used to describe the signal), so that the spheres become more sharply defined. Sharp-edged high dimensional spheres overlap less than fuzzy low dimensional ones, and this reduces the error rate. Therefore, a long coding period can be used to protect against noise. Surprisingly, this allows a communication system to operate *at* the channel capacity and yet have arbitrarily few errors [4]. A well known example of this kind of coding is the parity check [114, 115].

A simple molecular machine can reach high dimensionality only by using spatial mechanisms since it is not possible for them to remember more than one item at a time (Appendix 23). In a *time-encoding*, the parts of a communications signal that are spread out in time are combined to form a code to protect against errors [15, 61, 112]. In a *space-encoding*, the information from a set of parallel channels is combined to form the code. The simplest molecular machines are obliged to use space-encoding, so their parts must interact during the operation. Indeed, cooperative interactions within a single molecule were recently proposed to explain the high accuracy of the restriction enzyme *Eco*RI [20, 39, 8, 116, 42] and the specific binding of sugars by cell surface receptors [22], while the cooperative nature of DNA and RNA hybridization [34] and oxygen binding by hemoglobin [78] are well known. The frequent appearance of lock-and-key [10, 11] and allosteric mechanisms [117] in molecular biology suggests that space-encoding is used by most if not all molecular machines. Instead of paying for accuracy by using long time periods, molecular machines use large numbers of interacting atoms.

Shannon's channel capacity theorem states that as long as the rate of communication is less than the channel capacity, the error rate may be made arbitrarily small. This theorem also applies to molecular machines because the proof is based only on the geometry of the spheres, and this is the same for both models (see Appendix 21). In terms of molecular machines, the theorem says that:

> *By increasing the number of independently moving parts that can interact cooperatively to make decisions, a molecular machine can reduce the error frequency (rate of incorrect choices) to whatever arbitrarily low level is required for survival of the organism, even when the machine operates near its capacity and dissipates small amounts of power.*

The degree to which this happens during evolution depends, of course, on the requirements for function, the current design, and the evolutionary paths available to the machine. If a good code is found (*i.e.* if there is a good way to have the molecular machine's motions in one state be distinct from the motions when it is in another state), then the molecular machine can operate close to its machine capacity. In other words, the enormous complexity of molecular machines allows them to be accurate, and coding theory should help us to understand the mechanisms, accuracy, and evolution of molecular machines.

19 Summary

In this paper I have defined molecular machines and constructed a mathematical model for them that fits many examples in modern molecular biology. The mathematical description of molecular machine operations uses the methods of information theory, for which the hallmark and yardstick is the bit. According to this theory if a molecular machine is exposed to white Gaussian noise, then it should not be possible for it to gain more information than that given by the capacity formula, equation (38), although it may be able to approach this limit.

A theorem, originally proven by Shannon, shows that molecular machines can act precisely despite the ubiquitous presence of thermal noise. This is not a quantum nor chemical-bonding effect, but rather it arises from the degree of complexity that a molecular machine can attain by evolving a molecular coding scheme. The channel capacity should be a useful criterion for understanding and designing molecular machines [36, 118, 119, 120, 10, 121, 122].

I thank Herb Schneider, John Spouge, Randy Smith, and Peter Basser for many fun and useful discussions; Doris Schneider, Sislin Schneider, Sue Aldor and Maureen Manns for their support; and Stephen Altschul, Steve Garavelli, Rob Harrison, Jim Hofrichter, Andrzej Konopka, Peter Lemkin, Sarah Lesher, David Lipman, Joe Mack, Jake Maizel, Hugo Martinez, Ranjan Muttiah, Howard Nash, Peter Rogan, Denise Rubens, Morton Schultz, Bruce Shapiro, and R. Michael Stephens for critically reading the manuscript. I also thank Larry Gold for supporting the preliminary stages of this project under NIH grant GM28755, Ron Fox for pointing out the form of $f_D(r)$, Richard Pastor for pointing out the need to account for conformational substates, and Gray Abbott for useful discussions on Fourier analysis. The gumball machine was manufactured by Superior Toy & Mfg. Co. Inc., Chicago, IL. I thank Charles Rockwell for the photography of Fig. 1.

20 Appendix 1: Introduction to Information Theory

Forty years ago Shannon published several famous papers that rigorously defined the concept of information so that it could be used in designing communications systems [2, 3, 5]. To allow information from several independent sources to be additive, he and earlier workers chose a logarithmic measure. One "bit" is the amount of information required to distinguish between two equally likely symbols, two bits are required to distinguish one symbol out of 4, and 3 bits are required to distinguish one symbol out of 8. In general, if there are M equally likely symbols to be distinguished, then one needs $\log_2 M$ bits to pick out one of them.

Communication requires at least three components. A *transmitter* sends a signal over a communications *channel* to a *receiver* that collects the signal for further use. The signal consists of a series of symbols, which convey some average amount of information, R, measured in bits per symbol [2]. We follow Shannon and other early workers [2, 126, 127, 128] and take this to be the uncertainty of the receiver before receiving symbols minus the uncertainty after reception:

$$R = H_{before} - H_{after} \quad \text{(bits per symbol)} \tag{43}$$

where an uncertainty H is

$$H = -\sum_{i=1}^{M} p_i \log_2 p_i \quad \text{(bits per symbol)} \tag{44}$$

and p_i is the probability of each symbol i. When the symbols are equally likely, $p_i = 1/M$ and equation (44) simplifies to the form $H_{equal} = \log_2 M$. Likewise, when one symbol is certain, $H = 0$.

To find the maximum information from equation (43), the symbols appearing at the receiver must be equally likely, (so that $H_{before} = \log_2 M$) and every symbol must be exactly identified (no uncertainty left after reception, $H_{after} = 0$). Under these circumstances the information is $R_{maximum} = \log_2 M$. If there is any noise ($H_{after} \neq 0$), or the symbols are not equally likely ($H_{before} < H_{equal}$) then this simple formula must not be used and $R < R_{maximum}$.

If the symbols are sent at a rate of W_s symbols per second, then the channel carries $W_s R$ bits per second.

Shannon defined the "channel capacity" of a communications system and showed that it is:

$$C = W \log_2 \left(\frac{P}{N} + 1 \right) \quad \text{(bits per second)} \tag{45}$$

where the bandwidth W is the range of frequencies used in the communication (in cycles per second or Hertz), and P/N is the "signal-to-noise ratio" [4]. At the receiver a certain amount of signal power P (in joules per second) is required to distinguish the signals from each other in the presence of thermal noise N (also in joules per second).

Shannon proved a remarkable theorem about the channel capacity. One part of the theorem says that we cannot send information at a rate faster than the channel capacity. If we try to do this (*i.e.*, $W_s R > C$), a quantity of noise will be received that limits the rate to C. The other part of the theorem is surprising: *if we transmit at any rate less than or equal to the channel capacity ($W_s R \leq C$), then the transmission is possible with as low an error rate as we may desire*.

There is a price to be paid to get a low error rate: we must carefully encode the signal before transmission and then carefully decode it afterward. Although both steps require a delay, the overall transmission rate can approach C. Unfortunately the derivation of (45) and the proof of the theorem do not tell us how to make codes which allow transmission at rates close to C. Nevertheless, the formula is useful for understanding and designing communication systems, and methods have been found for creating "good" codes [112, 114, 113].

21 Appendix 2: Correspondences between Molecular Machines and Communication Channels

Although the molecular machine and communication channel models are not identical, we may draw several analogies as discussed in the introduction. A molecular machine corresponds to the receiver in Shannon's theory [4] since both gain energy and dissipate it to settle into a specific substate. For most molecular machines, such as the restriction enzymes on DNA, there is no transmitter, nor is there a communication channel. Rather, the formation of correct matches between molecular surfaces usually serves the function of directing the molecular machine to one or another substate. The phrase "signal-to-noise ratio" is not meaningful in the context of simple molecular machines.

The reader may have noticed that for channel capacity, bits were defined as a selection amongst possible symbols, whereas for machine capacity they were defined for selection among states. von Neumann [125] pointed out that so long as we can correlate events (or symbols) with states, these definitions are functionally identical.

Shannon's theory took advantage of the fact that the square of the voltage across a resistor is proportional to the power through the resistor. Likewise, the square of each y_j is the energy in the sine or cosine component of a "pin". Rather than using these mechanical Fourier "potentials", Shannon used voltage potentials in his theory. The mathematical equivalences between mechanical and electrical models are well known [57].

The proof of the channel capacity theorem depends entirely on geometry and not on the system being modeled. However, it is important to show that the geometry applies to molecular machines. In Shannon's Figure 5 [4], which is reproduced here as Fig. 10, the outer circle, with radius $\sqrt{P_y + N_y}$, corresponds to the molecular machine's spherical *before* state. The "received signal" (point A) represents only one of the possible *before* configurations. A spherical noise cloud around the "transmitted signal" (point B) of radius $\sqrt{N_y}$ corresponds to an *after* state. Having received signal point A, the receiver must select the transmitted point B. This corresponds to a machine operation in which the sphere center moves from point O to B, as the radius collapses. Since most of the dimensions are orthogonal to OB, very little noise power extends in the direction OB, and the *after* sphere essentially remains inside the *before* sphere. The shaded region L in Shannon's figure contains centers of small spheres that have the same *after* configuration at A. Shannon's theorem 2 shows that the probability of having a second *after* sphere centered in L—so that two *after* spheres overlap at point A—can be driven as low as desired even if the locations of the *after* spheres are chosen randomly. Thus the molecular machine can choose an *after* state with little probability of error as long as the machine capacity is not exceeded. That this result is obtained from most *random* choices of the coding suggests that the evolution of good codes may be easy. ⇐Fig 10

In Shannon's theory the capacity limit is approached by increasing t, while for the simple molecular machines described in this theory, d_{space} must increase. Molecular receivers, discussed in Appendix 23, could increase either t or d_{space}.

22 Appendix 3: Derivation of the Sphere Density Function

In this appendix we determine the probability density distribution of a set of D independent normally distributed random variables as a function of radial distance in the space defined by those variables. By definition, the probability density along the j^{th} axis in the space is:

$$f(y_j) = \frac{1}{\sigma\sqrt{2\pi}} e^{-y_j^2/2\sigma^2}. \tag{46}$$

To determine the overall probability density function in the space, we integrate over spherical shells. The probability of the machine being in a small shell of volume dV at radius r is

$$\begin{aligned} p(r) &= f(y_1, \ldots, y_j, \ldots, y_D)dV \\ &= \Pi_{j=1}^{D} f(y_j)dV \\ &= \frac{1}{\sigma^D \sqrt{2\pi}^D} e^{-r^2/2\sigma^2} dV. \end{aligned} \tag{47}$$

If $f_D(r)$ is the probability density of the sphere as a function of radius, then the probability of the machine being in a small interval of radius dr is also $p(r) = f_D(r)dr$. Combining the two equations for $p(r)$ with equation (27) we obtain

$$f_D(r) = \frac{r^{D-1} e^{-r^2/2\sigma^2}}{\Gamma\left(\frac{D}{2}\right) \sigma^D 2^{\frac{D}{2}-1}}, \tag{48}$$

which has a maximum at $r_{max} = \sigma\sqrt{D-1}$. If D is sufficiently high, then the $f_D(r)$ function can be approximated by a Gaussian distribution. Since any Gaussian with mean μ and standard deviation σ' has the property that $f(\mu+\sigma')/f(\mu) = e^{-1/2}$, we may estimate the fuzziness or thickness of the shell from the two intercepts with $e^{-1/2}$ in Fig. 4: σ'^- and σ'^+. This can also be calculated by noting that $r^{D-1} = e^{(D-1)\ln r}$, expanding the log by $\ln(x+1) \approx x - x^2/2$ [59] and setting $r_{max} = 1$. This leads to $\sigma' \approx \frac{1}{\sqrt{2(D-1)}}$ when $r_{max} = 1$. For the curves in Fig. 4, $\sigma'^- < \sigma' < \sigma'^+$.

The $f_D(r)$ function is the probability density function of a χ^2 distribution for the variable $x = r^2/\sigma^2$ and D degrees of freedom [129, 123]. Fig. 7 is essentially a series of χ^2 tests. The curves for the lower dimensions are named after well known physicists: $D = 1$ is a Gaussian distribution [12, 14]; $D = 2$ is a Rayleigh distribution [123]; and $D = 3$ is a Maxwellian or Maxwell-Boltzmann speed distribution [12, 13, 14].

23 Appendix 4: General Theory of Molecular Machines

A receiver is a device whose state is determined by an external signal. In contrast, a simple molecular machine such as *Eco*RI is not directed to its *after* state (binding sites) by an external command. Encoding or decoding a communications signal also requires a memory to record the signal as it is being processed. Simple molecular machines don't have the necessary memory. For example, DNA in the groove of *Eco*RI acts like a key in a lock, with the recognition process taking place in parallel over a surface of contact between *Eco*RI and DNA [10, 11, 20]. Since *Eco*RI has no record of its previous bound and unbound states it has no record of its history and cannot handle a time varying communications signal.

However, a time-encoded message could be received, remembered and processed by a combination of simple molecular machines. Such a "molecular receiver" could decode a message of the kind that Shannon's theory is designed to handle. Since they could be made insensitive to thermal noise by appropriate coding, molecular receivers are likely to play an important role as the interface between humans and artificial molecular machines and molecular computers. It is not known if any living organisms contain such devices, although the processes of translation, cell movement, mitosis, embryonic development and circadian rhythms are candidates.

According to Fourier analysis, a time varying signal may be recorded as a series of discrete samples. If t is the period of the recording and W is the highest frequency in the signal's spectrum, then the original signal may be reproduced exactly if at least

$$d_{time} = 2tW \tag{49}$$

samples are recorded [4, 61, 56]. This powerful result is the basis of digital-sound recording methods such as the compact disk [56].

If distinct states of a molecular receiver are determined by an external communications signal, then a high dimensional space consisting of

$$D = d_{space} d_{time} \tag{50}$$

dimensions can be used to describe the coding space of the machine. The machine could take advantage of both the spatial and the time dimensions and would operate in a "space-time" we will call Z space.

As in equation (31), we find that the average total energy for the entire molecular receiver in Z space is

$$\begin{aligned} \langle E_z \rangle &= (\tfrac{1}{2} k_{\mathrm{B}} T) \times D \\ &= t d_{space}(W k_{\mathrm{B}} T) \quad \text{(joules).} \end{aligned} \tag{51}$$

using equations (49) and (50). Dividing both sides of (51) by t gives the total thermal noise for the molecular receiver:

$$\begin{aligned} N_z &\equiv \frac{\langle E_z \rangle}{t} \\ &= d_{space}(W k_{\mathrm{B}} T) \quad \text{(joules per second).} \end{aligned} \tag{52}$$

The probability density is still given by equation (48). The sphere volume, which gives the capacity, depends on the radius raised to the dimension that the sphere is embedded in, so the maximum number of states is:

$$M_z \leq \frac{V_{before}}{V_{after}} = \left(\sqrt{\frac{P_z + N_z}{N_z}} \right)^{d_{space} 2tW} . \tag{53}$$

The definition of the molecular receiver capacity follows Shannon's definition exactly [4]:

$$C_z = \frac{\log_2(M_z)}{t} = d_{space} W \log_2 \left(\frac{P_z + N_z}{N_z} \right) \quad \text{(bits per sec).} \tag{54}$$

The relationship of this general equation to the capacity equations in the other two theories is straightforward. If we set $d_{space} = 1$ to indicate that there is only one spatial degree of freedom, we obtain Shannon's formula (equation (45)), and equation (52) becomes Nyquist's formula for thermal noise in a single wire [104, 124, 5]. If instead we set $tW = 1$ (to indicate a complete lack of long-term memory) and use the time independent capacity definition $C_z = \log_2(M_z)$, we obtain the formula for a simple molecular machine, equation (38), and the thermal energy formula (30) is obtained from (51).

The three theories are summarized in Table 1. ⇐Table 1

The capacity of a molecular receiver is most easily understood as the capacity of d_{space} parallel communications channels (compare (45) to (54)). The method of encoding in space would then correspond to spreading the coding bits across the parallel channels rather than spreading them out over time. From this it is clear that for a given error rate one can reduce the required encoding and decoding time by increasing the parallelism.

References

[1] Dawkins, R. (1986). *The Blind Watchmaker*. W. W. Norton & Co., New York.

[2] Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Tech. J.* **27**, 379–423, 623–656. http://cm.bell-labs.com/cm/ms/what/shannonday/paper.html.

[3] Shannon, C. E. & Weaver, W. (1949). *The Mathematical Theory of Communication*. University of Illinois Press, Urbana.

[4] Shannon, C. E. (1949). Communication in the presence of noise. *Proc. IRE,* **37**, 10–21.

[5] Pierce, J. R. (1980). *An Introduction to Information Theory: Symbols, Signals and Noise*. second edition, Dover Publications, Inc., New York.

[6] Polisky, B., Greene, P., Garfin, D. E., McCarthy, B. J., Goodman, H. M. & Boyer, H. W. (1975). Specificity of substrate recognition by the *Eco*RI restriction endonuclease. *Proc. Natl. Acad. Sci. USA,* **72**, 3310–3314.

[7] Woodhead, J. L., Bhave, N. & Malcolm, A. D. B. (1981). Cation dependence of restriction endonuclease *Eco*RI activity. *Eur. J. Biochem,* **115**, 293–296.

[8] Rosenberg, J. M., McClarin, J. A., Frederick, C. A., Wang, B. C., Boyer, H. W., Grable, J. & Greene, P. (1987). The structure and function of *Eco*RI endonuclease. In *Biological Organization: Macromolecular Interactions at High Resolution*, (Burnett, R. M. & Vogel, H. J., eds), pp. 11–43, Academic Press, Inc., Orlando.

[9] Johnson, H. A. (1987). Thermal noise and biological information. *Quart. Rev. Biol.* **62**, 141–152.

[10] Rastetter, W. H. (1983). Enzyme engineering. *Appl. Biochem. and Biotech.* **8**, 423–436.

[11] Gilbert, S. F. & Greenberg, J. P. (1984). Intellectual traditions in the life sciences. II. Stereocomplementarity. *Perspect. Biol. Med.* **28**, 18–34.

[12] Wannier, G. H. (1966). *Statistical physics*. John Wiley & Sons, Inc., New York.

[13] Castellan, G. W. (1971). *Physical Chemistry*. second edition, Addison-Wesley Publishing Company, Reading, Mass.

[14] Waldram, J. R. (1985). *The Theory of Thermodynamics.* Cambridge University Press, Cambridge.

[15] Sloane, N. J. A. (1984). The packing of spheres. *Sci. Am.* **250** (1), 116–125.

[16] Cipra, B. (1990). Packing your *n*-dimensional marbles. *Science,* **247**, 1035.

[17] Jacob, F. (1977). Evolution and tinkering. *Science,* **196**, 1161–1166.

[18] Porter, R., O'Connor, M. & Whelan, J., eds (1983). *Mobility and function in proteins and nucleic acids, Ciba Foundation Symposium 93*. Pitman Books Ltd, London.

[19] Alberts, B. M. (1984). The DNA enzymology of protein machines. *Cold Spring Harb. Symp. Quant. Biol.* **XLIV**, 1–12.

[20] McClarin, J. A., Frederick, C. A., Wang, B., Greene, P., Boyer, H. W., Grable, J. & Rosenberg, J. M. (1986). Structure of the DNA-*Eco*RI endonuclease recognition complex at 3Å resolution. *Science,* **234**, 1526–1541.

[21] Watson, J. D., Hopkins, N. H., Roberts, J. W., Steitz, J. A. & Weiner, A. M. (1987). *Molecular Biology of the Gene*. fourth edition, The Benjamin/Cummings Publishing Co., Inc., Menlo Park, California.

[22] Vyas, N. K., Vyas, M. N. & Quiocho, F. A. (1988). Sugar and signal-transducer binding sites of the *Escherichia coli* galactose chemoreceptor protein. *Science,* **242**, 1290–1295.

[23] Feynman, R. P. (1961). There's plenty of room at the bottom. In *Miniaturization*, (Gilbert, H. D., ed.), pp. 282–296, Reinhold Publishing Corporation, New York.

[24] McClare, C. W. F. (1971). Chemical machines, Maxwell's demon and living organisms. *J. Theor. Biol.* **30**, 1–34.

[25] Drexler, K. E. (1981). Molecular engineering: an approach to the development of general capabilities for molecular manipulation. *Proc. Natl. Acad. Sci. USA,* **78**, 5275–5278.

[26] Carter, F. L. (1984). The molecular device computer: point of departure for large scale cellular automata. *Physica D,* **10**, 175–194.

[27] Haddon, R. C. & Lamola, A. A. (1985). The molecular electronic device and the biochip computer: present status. *Proc. Natl. Acad. Sci. USA,* **82**, 1874–1878.

[28] Conrad, M. (1985). On design principles for a molecular computer. *Comm. ACM,* **28** (5), 464–480.

[29] Conrad, M. (1986). The lure of molecular computing. *IEEE Spectrum,* **23** (10), 55–60.

[30] Drexler, K. E. (1986). *Engines of Creation*. Anchor Press, Garden City, New York.

[31] Arrhenius, T. S., Blanchard-Desce, M., Dvolaitzky, M., Lehn, J. & Malthete, J. (1986). Molecular devices: caroviologens as an approach to molecular wires—synthesis and incorporation into vesicle membranes. *Proc. Natl. Acad. Sci. USA,* **83**, 5355–5359.

[32] Hong, F. T. (1986). The bacteriorhodopsin model membrane system as a prototype molecular computing element. *BioSystems,* **19**, 223–236.

[33] Hopfield, J. J., Onuchic, J. N. & Beratan, D. N. (1988). A molecular shift register based on electron transfer. *Science,* **241**, 817–820.

[34] Britten, R. J. & Kohne, D. E. (1968). Repeated sequences in DNA. *Science,* **161**, 529–540.

[35] Watson, J. D. & Crick, F. H. C. (1953). Molecular structure of nucleic acids: a structure for deoxyribose nucleic acid. *Nature,* **171**, 737–738.

[36] Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982). *Molecular Cloning, A Laboratory Manual*. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.

[37] Gibbs, R. A., Nguyen, P. & Caskey, C. T. (1989). Detection of single DNA base differences by competitive oligonucleotide priming. *Nucleic Acids Res.* **17**, 2437–2448.

[38] Smith, H. O. (1979). Nucleotide sequence specificity of restriction endonucleases. *Science,* **205**, 455–462.

[39] Rosenberg, J. M., McClarin, J. A., Frederick, C. A., Wang, B. C., Grable, J., Boyer, H. W. & Greene, P. (1987). Structure and recognition mechanism of *Eco*RI endonuclease. *Trend. Bioch. Sci.* **12**, 395–398.

[40] Dugaiczyk, A., Hedgpeth, J., Boyer, H. W. & Goodman, H. M. (1974). Physical identity of the SV40 deoxyribonucleic acid sequence recognized by the *Eco*RI restriction endonuclease and modification methylase. *Biochem.* **13**, 503–511.

[41] Pingoud, A. (1985). Spermidine increases the accuracy of type II restriction endonucleases. Suppression of cleavage at degenerate, non-symmetrical sites. *Eur. J. Biochem,* **147**, 105–109.

[42] Needels, M. C., Fried, S. R., Love, R., Rosenberg, J. M., Boyer, H. W. & Greene, P. J. (1989). Determinants of *Eco*RI endonuclease sequence discrimination. *Proc. Natl. Acad. Sci. USA,* **86**, 3579–3583.

[43] Thielking, V., Alves, J., Fliess, A., Maass, G. & Pingoud, A. (1990). Accuracy of the *Eco*RI restriction endonuclease: binding and cleavage studies with oligodeoxynucleotide substrates containing degenerate recognition sequences. *Biochem.* **29**, 4682–4691.

[44] Lewis, A. & Priore, L. V. D. (1988). The biophysics of visual photoreception. *Physics Today,* **41**, 38–46.

[45] Wessling-Resnick, M., Kelleher, D. J., Weiss, E. R. & Johnson, G. L. (1987). Enzymatic model for receptor activation of GTP-binding regulatory proteins. *Trend. Bioch. Sci.* **12**, 473–477.

[46] Huxley, H. E. (1969). The mechanism of muscular contraction. *Science,* **164**, 1356–1366.

[47] Highsmith, S. & Jardetzky, O. (1983). Actin-induced changes in the dynamics of myosin subfragment-1 detected by nuclear magnetic resonance. In *Mobility and function in proteins and nucleic acids, Ciba Foundation Symposium 93*, (Porter, R., O'Connor, M. & Whelan, J., eds), pp. 156–168, Pitman Books Ltd, London.

[48] Trayer, H. R. & Trayer, I. P. (1988). Fluorescence resonance energy transfer within the complex formed by actin and myosin subfragment 1. Comparison between weakly and strongly attached states. *Biochem.* **27**, 5718–5727.

[49] Schneider, T. D., Stormo, G. D., Gold, L. & Ehrenfeucht, A. (1986). Information content of binding sites on nucleotide sequences. *J. Mol. Biol.* **188**, 415–431. http://www.lecb.ncifcrf.gov/~toms/paper/schneider1986/.

[50] Schneider, T. D. (1988). Information and entropy of patterns in genetic switches. In *Maximum-Entropy and Bayesian Methods in Science and Engineering*, (Erickson, G. J. & Smith, C. R., eds), vol. 2, pp. 147–154, Kluwer Academic Publishers, Dordrecht, The Netherlands.

[51] Schneider, T. D. & Stormo, G. D. (1989). Excess information at bacteriophage T7 genomic promoters detected by a random cloning technique. *Nucleic Acids Res.* **17**, 659–674.

[52] Jaynes, E. T. (1988). The evolution of Carnot's principle. In *Maximum-Entropy and Bayesian Methods in Science and Engineering*, (Erickson, G. J. & Smith, C. R., eds), vol. 1, pp. 267–281, Kluwer Academic Publishers, Dordrecht, The Netherlands. http://bayes.wustl.edu/etj/articles/ccarnot.ps.gz http://bayes.wustl.edu/etj/articles/ccarnot.pdf.

[53] Shakhnovich, E. I. & Gutin, A. M. (1989). Frozen states of a disordered globular heteropolymer. *J. Phys. A: Math. Gen.* **22**, 1647–1659.

[54] Roper, C. A. (1976). *The Complete Handbook of Locks and Locksmithing*. TAB Books, Blue Ridge Summit, Pa. 17214.

[55] Macaulay, D. (1988). *The Way Things Work*. Houghton Mifflin Company, Boston.

[56] Walker, J. S. (1988). *Fourier Analysis*. Oxford University Press, New York.

[57] Feynman, R. P., Leighton, R. B. & Sands, M. (1963). *The Feynman Lectures on Physics*. Addison-Wesley Publishing Company, Reading, Mass.

[58] Stremler, F. G. (1982). *Introduction to Communication Systems*. second edition, Addison-Wesley Publishing Company, Reading, Massachusetts.

[59] Thomas, G. B. (1968). *Calculus and Analytic Geometry*. fourth edition, Addison-Wesley, Reading, Mass.

[60] Callen, H. B. (1985). *Thermodynamics and an Introduction to Thermostatistics*. second edition, John Wiley & Sons, Ltd., N. Y.

[61] Conway, J. H. & Sloane, N. J. A. (1988). *Sphere Packings, Lattices and Groups*. Springer-Verlag, New York.

[62] Brillouin, L. (1962). *Science and Information Theory*. second edition, Academic Press, Inc., New York.

[63] Sommerville, D. M. Y. (1929). *An Introduction to the Geometry of N Dimensions*. E. P. Dutton, NY., NY.

[64] Kendall, M. G. (1961). *A Course in the Geometry of* n *Dimensions*. Hafner Publishing Company, New York.

[65] Warshel, A. (1976). Bicycle-pedal model for the first step in the vision process. *Nature,* **260**, 679–683.

[66] Ovchinnikov, Y. A. (1982). Rhodopsin and bacteriorhodopsin: structure-function relationships. *FEBS Letters,* **148**, 179–191.

[67] Nathans, J. & Hogness, D. S. (1983). Isolation, sequence analysis, and intron-exon arrangement of the gene encoding bovine rhodopsin. *Cell,* **34**, 807–814.

[68] Levitt, M. & Sharon, R. (1988). Accurate simulation of protein dynamics in solution. *Proc. Natl. Acad. Sci. USA,* **85**, 7557–7561.

[69] Brooks III, C. L. & Karplus, M. (1989). Solvent effects on protein motion and protein effects on solvent motion. Dynamics of the active site region of lysozyme. *J. Mol. Biol.* **208**, 159–181.

[70] Karplus, M. & McCammon, J. A. (1986). The dynamics of proteins. *Sci. Am.* **254**, 42–51.

[71] Miller, A. R. (1981). *Pascal Programs for Scientists and Engineers*. Sybex, Inc., Berkeley.

[72] Breiman, L. (1969). *Probability and Stochastic Processes: With a View Toward Applications*. Houghton Mifflin Company, Boston.

[73] Petsko, G. A. & Ringe, D. (1984). Fluctuations in protein structure from X-ray diffraction. *Annu. Rev. Biophys. Biophys. Bioeng.* **13**, 331–371.

[74] Birge, R. R. & Hubbard, L. M. (1980). Molecular dynamics of cis-trans isomerization in rhodopsin. *J. Am. Chem. Soc.* **102**, 2195–2205.

[75] Huber, R. & Bennett, Jr., W. S. (1983). Functional significance of flexibility in proteins. *Biopolymers,* **22**, 261–279.

[76] Weber, I. T. & Steitz, T. A. (1984). A model for the non-specific binding of catabolite gene activator protein to DNA. *Nucleic Acids Res.* **12**, 8475–8487.

[77] Weber, I. T. & Steitz, T. A. (1984). Model of specific complex between catabolite gene activator protein and B-DNA suggested by electrostatic complementarity. *Proc. Natl. Acad. Sci. USA,* **81**, 3973–3977.

[78] Perutz, M. F. (1970). Stereochemistry of cooperative effects in haemoglobin. *Nature,* **228**, 726–734.

[79] Chothia, C. & Lesk, A. M. (1985). Helix movements in proteins. *Trend. Bioch. Sci.* **10**, 116–118.

[80] Rould, M. A., Perona, J. J., Söll, D. & Steitz, T. A. (1989). Structure of *E. coli* Glutaminyl-tRNA synthetase complexed with $tRNA^{Gln}$ and ATP at 2.8Å resolution. *Science,* **246**, 1135–1142.

[81] Rould, M., Perona, J. J. & Steitz, T. A. (1989). High resolution crystal structure of *E. coli* glutaminyl-tRNA synthetase complexed with its cognate tRNA. *Biophys J.* **55**, 49a.

[82] Perona, J. J., Swanson, R., Steitz, T. A. & Söll, D. (1988). Overproduction and purification of *Escherichia coli* $tRNA_2^{Gln}$ and its use in crystallization of the glutaminyl-tRNA synthetase-$tRNA^{Gln}$ complex. *J. Mol. Biol.* **202**, 121–126.

[83] Hou, Y., Francklyn, C. & Schimmel, P. (1989). Molecular dissection of a transfer RNA and the basis for its identity. *Trends. Biochem. Sci.* **14**, 233–237.

[84] Levy, R. M., Perahia, D. & Karplus, M. (1982). Molecular dynamics of an α-helical polypeptide: temperature dependence and deviation from harmonic behavior. *Proc. Natl. Acad. Sci. USA,* **79**, 1346–1350.

[85] Ichiye, T. & Karplus, M. (1987). Anisotropy and anharmonicity of atomic fluctuations in proteins: analysis of a molecular dynamics simulation. *PROTEINS: Structure, Function and Genetics,* **2**, 236–259.

[86] Schneider, T. D. & Stephens, R. M. (1990). Sequence logos: a new way to display consensus sequences. *Nucleic Acids Res.* **18**, 6097–6100. http://www.lecb.ncifcrf.gov/˜toms/paper/logopaper/.

[87] Stormo, G. D., Schneider, T. D., Gold, L. & Ehrenfeucht, A. (1982). Use of the 'Perceptron' algorithm to distinguish translational initiation sites in *E. coli*. *Nucleic Acids Res.* **10**, 2997–3011.

[88] Nakata, K., Kanehisa, M. & DeLisi, C. (1985). Prediction of splice junctions in mRNA sequences. *Nucleic Acids Res.* **13**, 5327–5340.

[89] Brunak, S., Engelbrecht, J. & Knudsen, S. (1990). Cleaning up gene databases. *Nature,* **343**, 123.

[90] Childs, J., Villanueba, K., Barrick, D., Schneider, T. D., Stormo, G. D., Gold, L., Leitner, M. & Caruthers, M. (1985). Ribosome binding site sequences and function. In *Sequence Specificity in Transcription and Translation, UCLA Symposia on Molecular and Cellular Biology,* Vol. 30, (Calendar, R. & Gold, L., eds), pp. 341–350, Alan R. Liss, Inc, New York.

[91] Stormo, G. D., Schneider, T. D. & Gold, L. (1986). Quantitative analysis of the relationship between nucleotide sequence and functional activity. *Nucleic Acids Res.* **14**, 6661–6679.

[92] Barrick, D., Villanueba, K., Childs, J., Kalil, R., Schneider, T. D., Lawrence, C. E., Gold, L. & Stormo, G. D. (1994). Quantitative analysis of ribosome binding sites in *E. coli. Nucleic Acids Res.* **22**, 1287–1295.

[93] Takeda, Y., Sarai, A. & Rivera, V. M. (1989). Analysis of the sequence-specific interactions between Cro repressor and operator DNA by systematic base substitution experiments. *Proc. Natl. Acad. Sci. USA,* **86**, 439–443.

[94] Lehming, N., Sartorius, J., Kisters-Woike, B., von Wilken-Bergmann, B. & Müller-Hill, B. (1990). Mutant *lac* repressors with new specificities hint at rules for protein-DNA recognition. *EMBO J.* **9**, 615–621.

[95] Horovitz, A. (1987). Non-additivity in protein-protein interactions. *J. Mol. Biol.* **196**, 733–735.

[96] Horovitz, A. & Rigbi, M. (1985). Protein-protein interactions: additivity of the free energies of association of amino acid residues. *J. Theor. Biol.* **116**, 149–159.

[97] Colthup, N. B., Daly, L. H. & Wiberley, S. E. (1975). *Introduction to Infrared and Raman Spectroscopy*. second edition, Academic Press, Inc., New York.

[98] Crawford, Jr., B. & Swanson, D. (1976). An introduction to molecular vibrations. In *Infrared and Raman Spectroscopy, Part A*, (Brame, .Jr., E. G. & Grasselli, J. G., eds), pp. 1–69, Marcel Dekker, Inc., New York.

[99] McCammon, J. A., Gelin, B. R. & Karplus, M. (1977). Dynamics of folded proteins. *Nature,* **267**, 585–590.

[100] Karplus, M. & McCammon, J. A. (1979). Protein structural fluctuations during a period of 100 ps. *Nature,* **277**, 578.

[101] Karplus, M. (1987). Molecular dynamics simulations of proteins. *Physics Today,* **40**, 68–72.

[102] Nishikawa, T. & Gō, N. (1987). Normal modes of vibration in bovine pancreatic trypsin inhibitor and its mechanical property. *Proteins: Structure, Function and Genetics,* **2**, 308–329.

[103] Tidor, B., Irikura, K. K., Brooks, B. R. & Karplus, M. (1983). Dynamics of DNA oligomers. *J. Biomol. Struct. Dyn.* **1**, 231–252.

[104] Nyquist, H. (1928). Thermal agitation of electric charge in conductors. *Physical Review,* **32**, 110–113.

[105] Rice, S. O. (1944). Mathematical analysis of random noise. *Bell System Tech. J.* **23**, 282–332.

[106] Rice, S. O. (1945). Mathematical analysis of random noise. *Bell System Tech. J.* **24**, 46–156.

[107] MacDonald, D. K. C. (1962). *Noise and fluctuations: an introduction*. John Wiley & Sons, Inc., New York.

[108] Hurley, J. B., Ebrey, T. G., Honig, B. & Ottolenghi, M. (1977). Temperature and wavelength effects on the photochemistry of rhodopsin, isorhodopsin, bacteriorhodopsin and their photoproducts. *Nature,* **270**, 540–542.

[109] Campbell, I. D., Dobson, C. M. & Williams, R. J. P. (1985). The study of conformational states of proteins by nuclear magnetic resonance. *Biochem. J.* **231**, 1–10.

[110] Frauenfelder, H., Parak, F. & Young, R. D. (1988). Conformational substates in proteins. *Annu. Rev. Biophys. Chem.* **17**, 451–479.

[111] Golomb, M. & Chamberlin, M. (1974). Characterization of T7- specific Ribonucleic Acid Polymerase. IV. Resolution of the major *in vitro* transcripts by gel electrophoresis. *J. Biol. Chem.* **249**, 2858–2863.

[112] Gilbert, E. N. (1966). Information theory after 18 years. *Science,* **152**, 320–326.

[113] Sourlas, N. (1989). Spin-glass models as error-correcting codes. *Nature,* **339**, 693–695.

[114] Hamming, R. W. (1986). *Coding and Information Theory*. Prentice-Hall, Inc., Englewood Cliffs NJ.

[115] Gersting, J. L. (1986). *Mathematical structures for computer science*. second edition, W. H. Freeman and Co., New York.

[116] Rosenberg, J. M., Wang, B. C., Frederick, C. A., Reich, N., Greene, P., Grable, J. & McClarin, J. (1987). Development of a protein design strategy for *Eco*RI endonuclease. In *Protein Engineering*, (Oxender, D. L. & Fox, C. F., eds), pp. 237–250, Alan R. Liss, Inc, New York.

[117] Monod, J., Wyman, J. & Changeux, J. (1965). On the nature of allosteric transitions: a plausible model. *J. Mol. Biol.* **12**, 88–118.

[118] Beaucage, S. L. & Caruthers, M. H. (1981). Deoxynucleoside phosphoramidites - a new class of key intermediates for deoxypolynucleotide synthesis. *Tetrahedron Letters,* **22**, 1859–1862.

[119] Ulmer, K. M. (1983). Protein engineering. *Science,* **219**, 666–671.

[120] Pabo, C. (1983). Molecular technology: designing proteins and peptides. *Nature,* **301**, 200.

[121] Wetzel, R. (1986). What is protein engineering? *Protein Engineering,* **1**, 3–5.

[122] Lesk, A. M. (1988). Introduction: protein engineering. *BioEssays,* **8**, 51–52.

[123] International Telephone and Telegraph Corporation (1956). *Reference data for radio engineers*. fourth edition, International Telephone and Telegraph Corporation, New York.

[124] Johnson, J. B. (1928). Thermal agitation of electricity in conductors. *Physical Review,* **32**, 97–109.

[125] von Neumann, J. (1963). Probabilistic logics and the synthesis of reliable organisms from unreliable components. In *Collected works*, (Taub, A. H., ed.), vol. 5, pp. 341–342, The MacMillan Company, New York.

[126] Brillouin, L. (1951). Physical entropy and information. II. *J. of Applied Physics,* **22**, 338–343.

[127] Tribus, M. & McIrvine, E. C. (1971). Energy and information. *Sci. Am.* **225** (3), 179–188. (Note: the table of contents in this volume incorrectly lists this as volume **224**).

[128] Rothstein, J. (1951). Information, measurement, and quantum mechanics. *Science,* **114**, 171–175.

[129] Weast, R. C., Astle, M. J. & Beyer, W. H. (1988). *CRC Handbook of Chemistry and Physics*. CRC Press, Inc., Boca Raton, Florida.

Figure 1: A gumball machine demonstrates sphere packing.
The enclosing large sphere represents a molecular machine having high energy, while each small sphere (gumball) represents the machine having low energy. There are many possible low energy conformations. The machine or channel capacity is the logarithm of the number of small spheres that can fit into the large sphere.

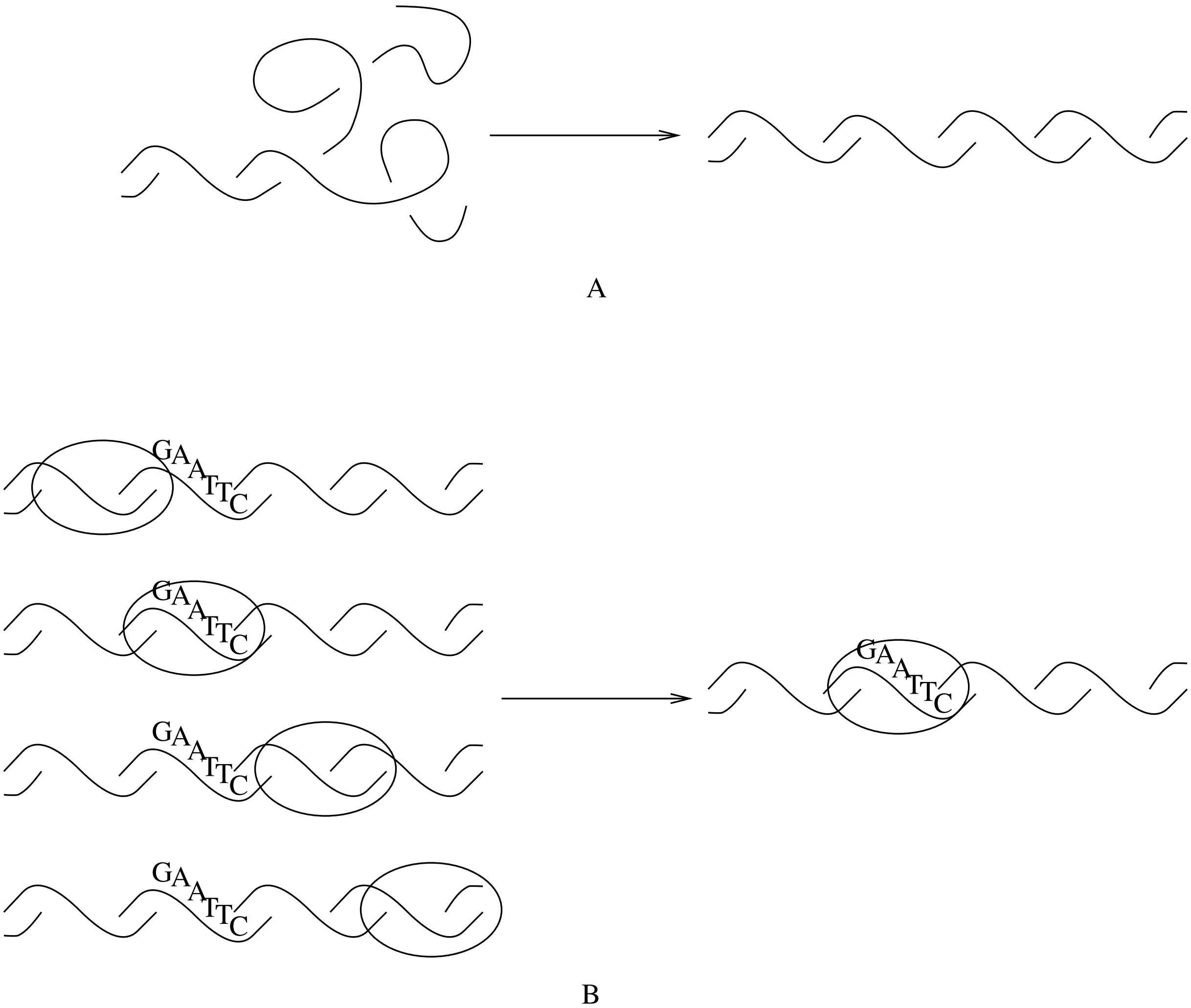

Figure 2: Operations of Two Molecular Machines.
A. Single-stranded DNA will hybridize to become a double-stranded helix.
B. *Eco*RI will scan along a DNA molecule and then bind specifically to the sequence 5′ GAATTC 3′.

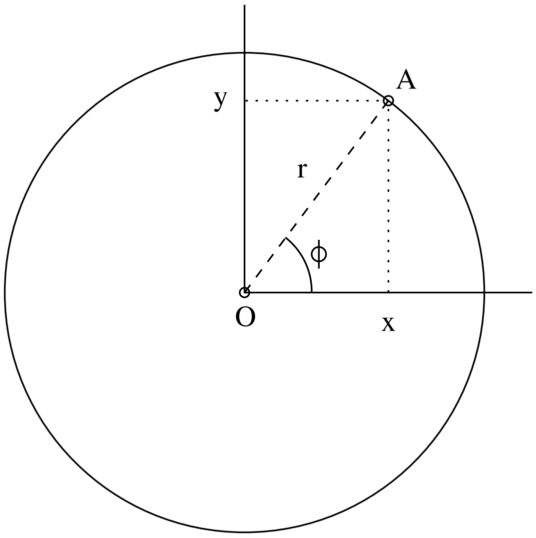

Figure 3: Geometry for a simple harmonic oscillator.
A possible state of a harmonic oscillator is represented by point A. Its maximum velocity is r and its phase is ϕ. This state may also be represented by the coordinate (x, y). Distances in this figure have units of velocity.

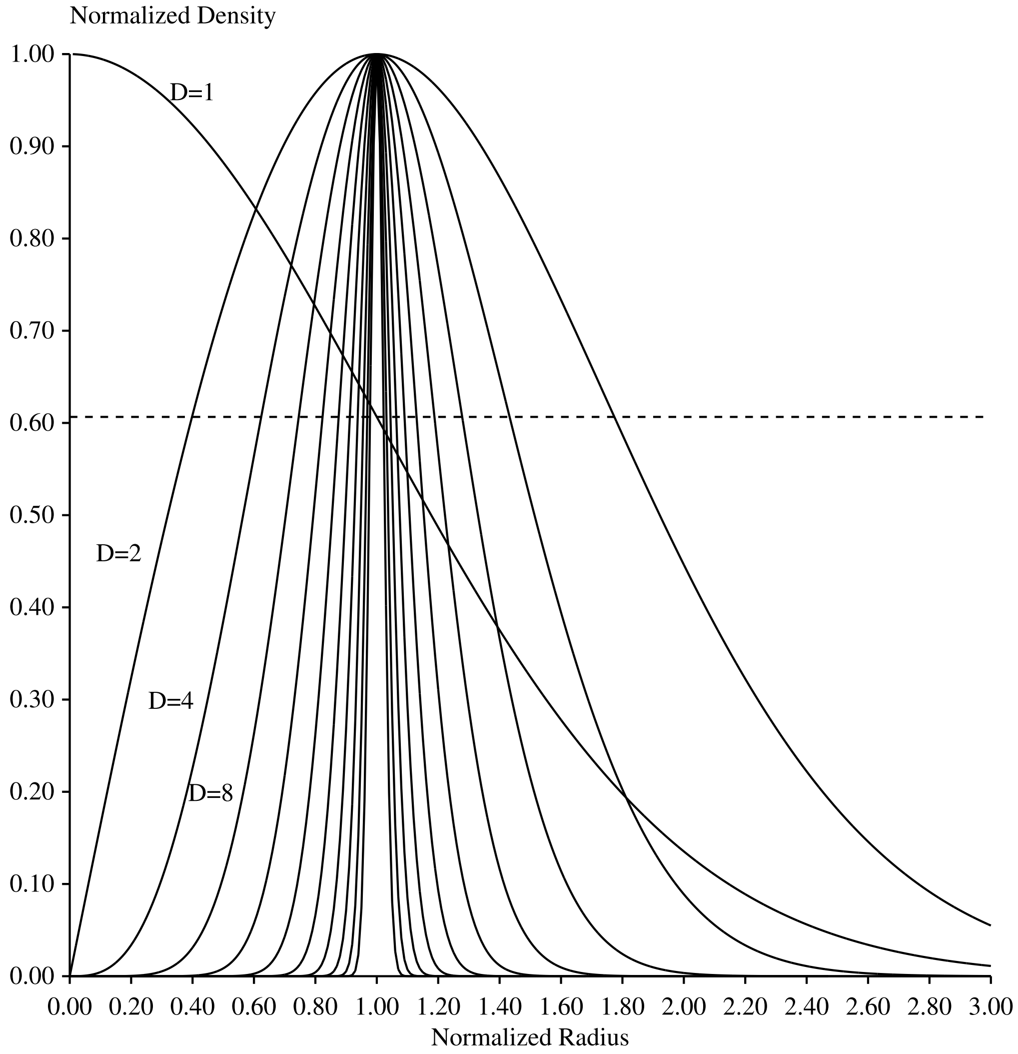

Figure 4: High Dimensional Sphere Density.
The sphere probability density as a function of radius, $f_D(r)$, is drawn for $D = 1, 2, 4, 8, \ldots, 1024$ dimensions (see Appendix 22). Except for the Gaussian curve ($D = 1$), which passes through the point (0,1), the curves are "normalized" so that their peaks pass through (1,1). At higher dimensions the curves approach the Gaussian distribution again and peak sharply. The dashed line is at $e^{-1/2}$, which intercepts "normalized" Gaussian distributions at one standard deviation from the mean.

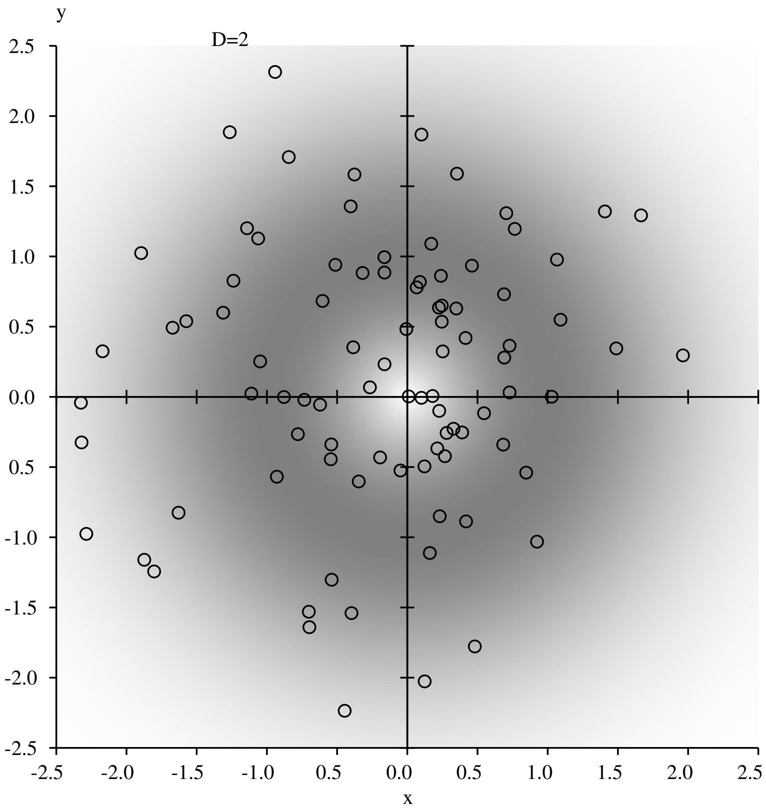

Figure 5: The Rayleigh Distribution

The continuous grey-tone distribution represents the analytic probability density, $f_2(r)$. Each small open circle ($\bigcirc$) represents the coordinates of two normally distributed values with mean 0 and standard deviation 1. Each normally distributed value was the sum of 100 pseudo-random numbers with a flat distribution.

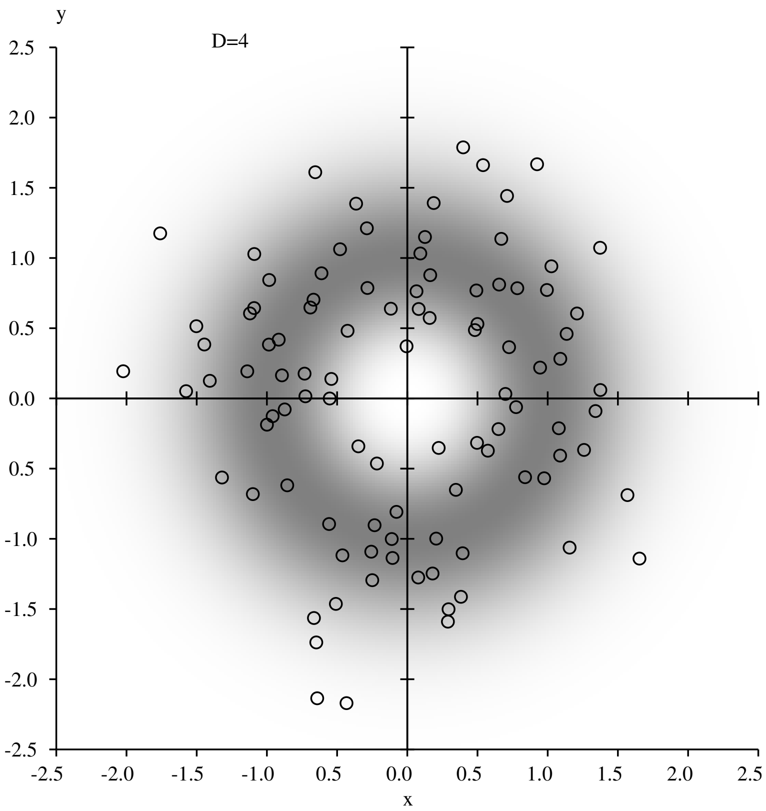

Figure 6: Simulation of a fourth-dimensional sphere.

A four-dimensional Y space was projected onto the two-dimensional space represented by the page. This is equivalent to a plane cross section through the space [4]. The continuous grey-tone distribution represents the analytic probability density, $f_D(r)$ (equation (48) in Appendix 22 and Fig. 4) for $D = 4$ and $\sigma = 1$. Each small open circle (◯) represents a numerical simulation produced from four normally distributed values with mean 0 and standard deviation 1 according to equation (29). Each normally distributed value was the sum of 100 pseudo-random numbers with a flat distribution.

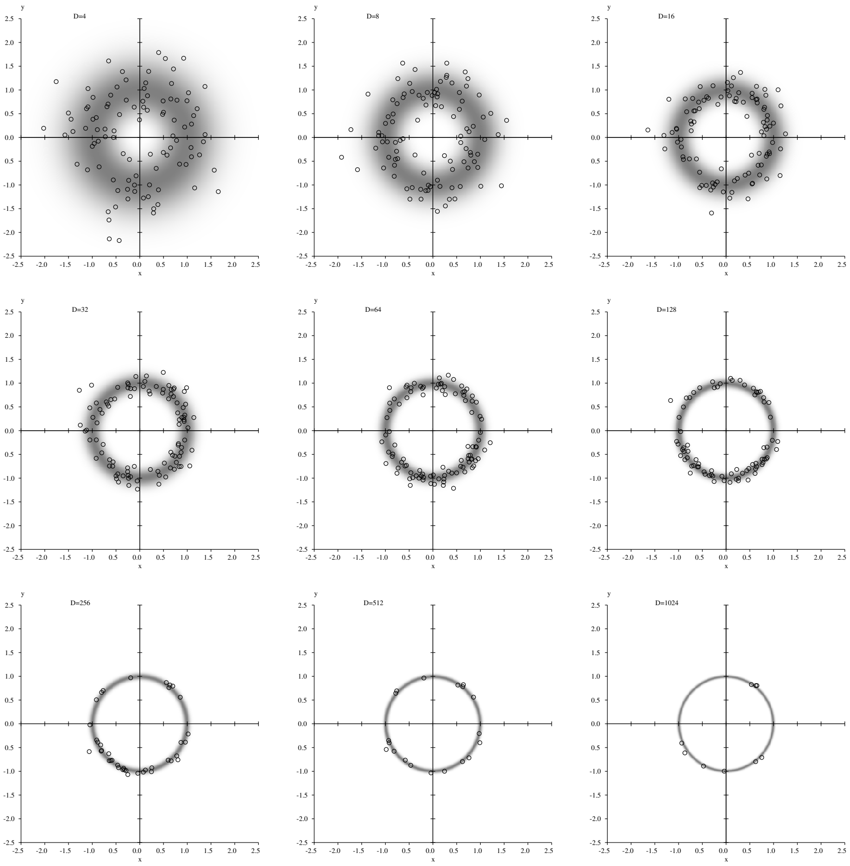

Figure 7: Increase of Sphere Sharpness with Increasing Dimensionality.
A series of sphere projections are shown for $D = 4, 8, \ldots, 1024$ dimensions. The first one is the same as Fig. 6, but reduced in size. Only 10,000 Gaussian values were precalculated, so the number of simulated points that could be calculated decreased as the dimensionality increased.

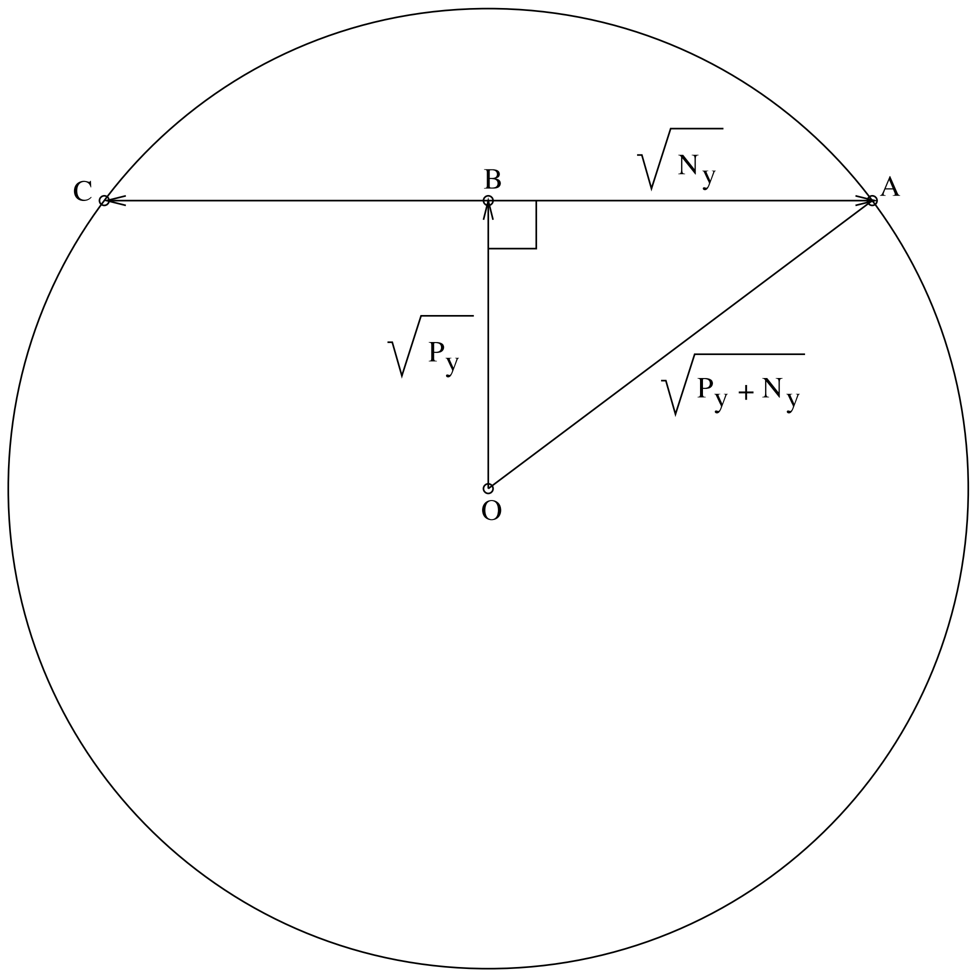

Figure 8: Geometry of Thermal Noise Spheres in High Dimensional Space.
The *before* sphere is represented by the outer circle, while the *after* sphere is represented by the line segment $\overline{CA}$. $\vec{N}_y$ is $\vec{BA}$ or $\vec{BC}$, with $|\vec{N}_y| = \sqrt{N_y}$. $\vec{P}_y$ is $\vec{OB}$ with $|\vec{P}_y| = \sqrt{P_y}$. See the main text for further description. The figure was derived from Shannon [4].

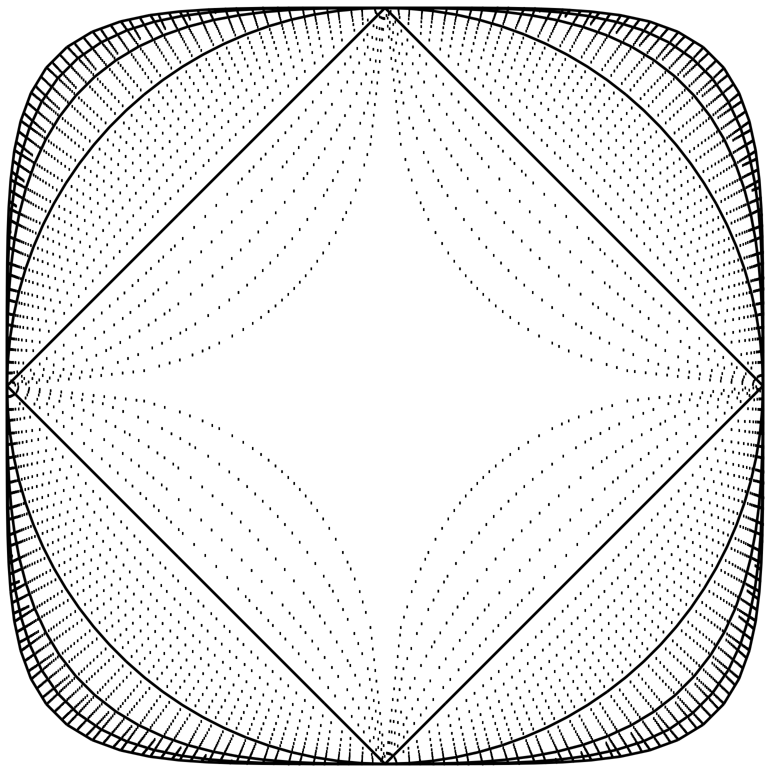

Figure 9: $|x|^m + |y|^m = |r|^m$.

The equation is plotted for $m = 0.5$ to $m = 5$ by increments of 0.1. Integer values of m are indicated by solid curves and other values by dotted curves.

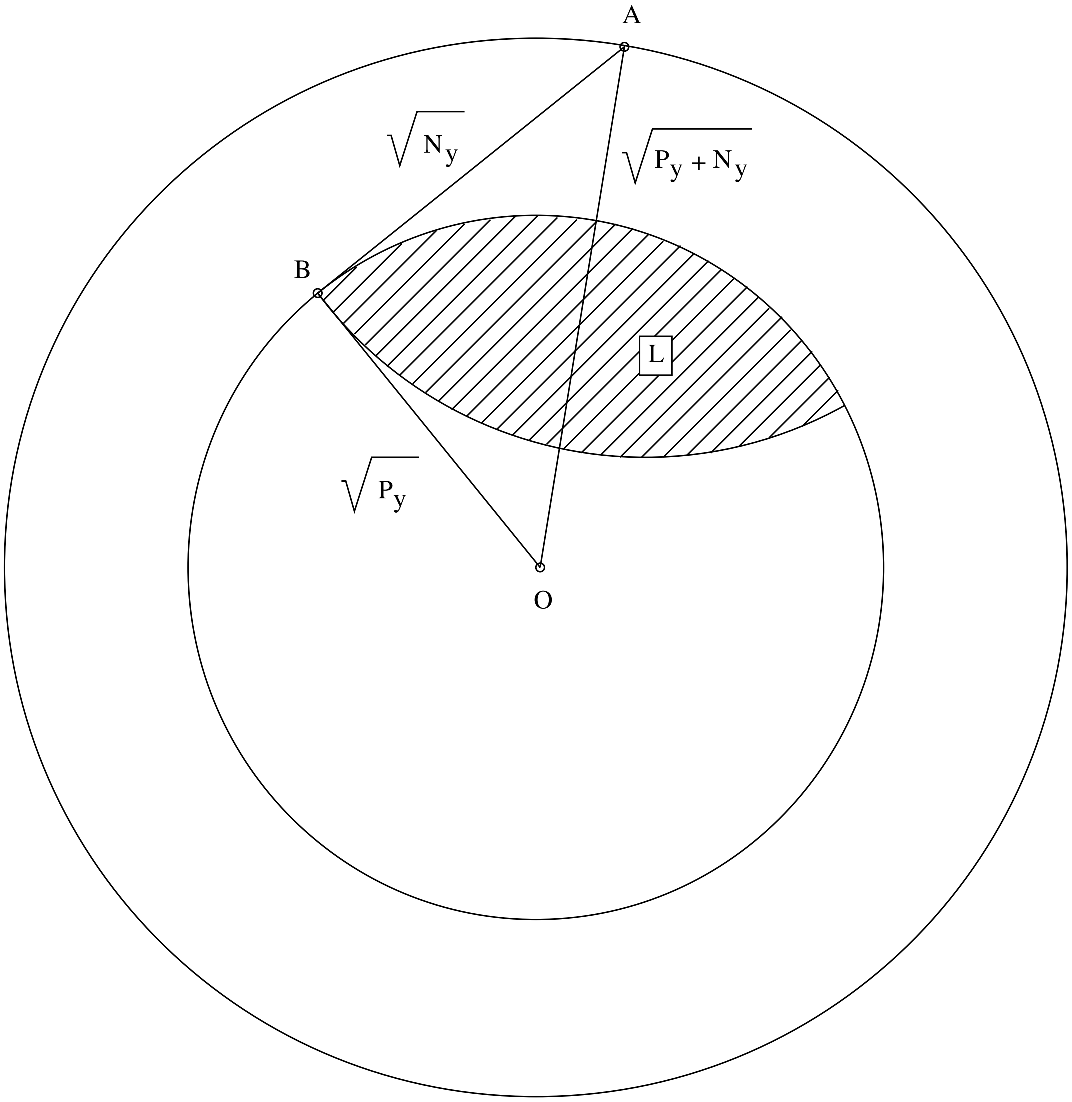

Figure 10: Correspondence Between Communication Theory and Molecular Machine Geometry.
The figure is the same as Figure 5 in [4] except that the distances are given as $\sqrt{P_y}$, $\sqrt{N_y}$ and $\sqrt{P_y + N_y}$ rather than $\sqrt{2tWP}$, $\sqrt{2tWN}$ and $\sqrt{2tW(P+N)}$.

	Channel	Molecular Machine	Molecular Receiver
Coding Space	Shannon	Y	Z
Degrees of Freedom	$d_{time}=2tW$	$2d_{space} \leq 2(3n-6)$	$D=d_{space}d_{time}$
Power	P	P_y	P_z
Noise	$N=Wk_{\mathrm{B}}T$	$N_y=d_{space}k_{\mathrm{B}}T$	$N_z=d_{space}Wk_{\mathrm{B}}T$
Power & Noise units	J / sec	J / op	J / sec
Capacity	$C=W\log_2(P/N+1)$	$C_y=d_{space}\log_2(P_y/N_y+1)$	$C_z=d_{space}W\log_2(P_z/N_z+1)$
Rate	W_sR	R	W_sR
Capacity & Rate units	bits / sec	bits / op	bits / op - sec

Table 1: Information Capacity Theories

The units are J: joules; sec: seconds, op: operation.