April 17th, 2008

Information Mechanics

John L. Haller Jr.

jlhaller@alumni.princeton.com

Abstract: Advances in science are brought fourth by hypothesizing that the action of a system is a direct measure of the amount of information in that system. We begin to interpret this governing hypothesis by examining its implications to current research. From this investigation, we find four primary conclusions.

1) To properly and completely quantify the amount of information contained within a particle (or system), one must add the self-information of both the wavefunction and its Fourier transform pair.

2) Information in nature is found in packets quantized to an integer number of the natural units.

3) Over a period of time, the energy of a system acts like an information rate and thus the information needed to describe that system for that period of time is equal to the product of the energy and the time divided by the minimum uncertainty.

4) At a given instant in time, the angular momentum, J, of a system is in direct proportion to the amount of information that is contained within or can be transmitted by that system.

Empirical evidence affirming our governing hypothesis is given through twelve examples of systems (ranging from a black hole, to an electric circuit, to an electron). Thus from the very big down to the limits of the Heisenberg uncertainty principal, the conclusions are shown as a self consistent theory, accurately quantifying the amount of information in each given system.

Governing Hypothesis: The amount of information contained within a system in nature (or the amount of information needed to describe the system), when measured in natural units of information, is equal to the number of quantum units of action contained within that system. Or,

$$I = 2\varepsilon\tau/\hbar$$

where I is the information in the system in natural units, $\varepsilon\tau$ is the action of the system in Joule-sec, and $\hbar$ is Planck's constant divided by 2π which is approximately 1.055e-34 Joule-sec. For a derivation of this hypothesis, please refer to Appendix A.

Implied Conclusions: With twelve examples (given later) where our hypothesis is verified, it is useful to assume that the hypothesis is true and ask the question, "what if information equals energy times time?" We will consider the implications of such a relationship on the following research topics, in the hopes of learning how to interpret the equation, $I = 2\varepsilon\tau/\hbar$:

- The Fourier Transform
- The quantization of action
- The Von Neumann entropy
- A "q-bit" from quantum computing

- The hidden variables theorems

The Fourier Transform: Given that the Fourier transform is at the intersection of electrical engineering and physics, it is no coincidence that the Fourier kernel, $\exp(-i2\pi ft)$, is the starting point for the derivation of our governing hypothesis. Thus the Fourier transform will be an integral part of our theory.

Additionally, from the Heisenberg uncertainty principle, we know that both domains simultaneously contain variance [1] and thus both domains carry entropy or self-information [2]. Looking at an example of a cell phone, we know how to use this fact to transmit information using either domain, where we have either: time division multiplexing, frequency division multiplexing, or using both domains simultaneously, code division multiplexing (which alternates frequency bands through time).

In fact, the idea to quantify information in both domains was written about by Hirschmann in 1957. His idea was to add the entropy from both domains in order to completely quantify it [3]. We formalize this phenomenon in our first conclusion.

Conclusion 1) To properly and completely quantify the amount of information contained within a particle (or system), one must add the self-information of both the wavefunction and its Fourier transform pair.

As can be seen in Appendix B, this procedure will be very useful and insightful when we look at our examples.

Quantization of action: Over a century ago, Plank introduced us to the quantization of energy and with it was able to correctly model black body radiation [4]. Bohr then quantized angular momentum and used this to solve for the energy states of the hydrogen atom [5]. Since that time, this quantization has been extended to action. Useful in path integrals, derivation of equations of motion, and other applications (such as geomagnetically trapped radiation), the action of a system has been shown quantized to integer units of the minimum uncertainty [6].

We can further extend this quantization principle to information, vis a vie the governing hypothesis.

Whereby setting: Action = $\varepsilon\tau = n\cdot\hbar/2$, where n is a positive integer

Implies: Information = $2\varepsilon\tau/\hbar = n$, in units of the natural log

Leading us to conclusion 2)

Conclusion 2) Information in nature is found in packets quantized to an integer number of the natural units.

It is worthwhile to note that in this example (and two other examples in the paper), an extra factor of 2π is present in some of the original work. It is shown here as part of Plank's constant where the action is typically quoted as a closed loop path integral of the energy times the differential dt being equal to an integer times Plank's constant (h not ħ). This discrepancy is dealt with in Appendix C. To be clear, the analysis here quantizes the action as an integer number of the minimum uncertainty (which is well known to be $\hbar/2$).

Note that this analysis is in contrast to the basic understanding that the bit, ln(2), is the fundamental building block of nature. Yet as we will see in the examples, an additional ln(e/2) can be found in a number of different places which brings the total information per packet up to one natural unit, ln(e).

The Von Neumann entropy: The Von Neumann entropy, introduced in 1927, is the general expression for the amount of information in a wavefunction, and as we know the Shannon entropy is a special case when the density matrix is diagonal [7,8,9].

Here, we will look at how the governing hypothesis serves as an additional method to measure information, supplementing the Von Neumann Entropy.

The following two examples might be helpful where we look at the Von Neumann entropy after 1) an evolution through time; and 2) a finite rotation.

1) As we know the Von Neumann entropy is invariant through time, that is,

$$S(\rho) = S\left(U(t)\rho U^{\dagger}(t)\right)$$

Where $U(t) = e^{-iHt/\hbar}$, and H is the Hamiltonian [10]

If however we examine why the Von Neumann entropy is invariant we can see that there are two dueling entropy rates of equal magnitude which cancel each other out. Expanding out the time evolution of S(ρ),

$$S\left(U(t)\rho U^{\dagger}(t)\right) = -Tr\left(\left(U(t)\rho U^{\dagger}(t)\right)\ln\left(U(t)\rho U^{\dagger}(t)\right)\right)$$

In the eigenbasis, with all eigenvalues of H being equal to ε,

$$= -Tr\left(\left(\left(e^{-i\varepsilon t/\hbar}I\right)\rho\left(e^{i\varepsilon t/\hbar}I\right)\right)\left(\ln\left(e^{-i\varepsilon t/\hbar}I\right)\rho\left(e^{i\varepsilon t/\hbar}I\right)\right)\right) = -Tr(\rho\ln\rho) - i\varepsilon t/\hbar + i\varepsilon t/\hbar$$

If instead of letting –iεt/ħ cancel itself out with its complex conjugate, let's see what happens if we take the magnitude of both the positive and negative entropy rates add them together. We can see that under this scenario the Von Neumann entropy is not invariant in time, but rather has an additional 2εt/ħ.

$$S\left(U(t)\rho U^{\dagger}(t)\right) = S(\rho) + 2\varepsilon t/\hbar$$

The use of imaginary time might seem "un-real", however space-time becomes Euclidian with this change and thus might be a more "real" description that what we all are used to as real time [11]. In fact, imaginary time is very useful in calculating past histories and is known as analytical continuation [12]. Two further derivations of this same result can be found in appendix A.

This leads us to our third conclusion.

Conclusion 3) Over a period of time, the energy of a system acts like an information rate and thus the information need to describe that system for that period of time is equal to the product of the energy and the time divided by the minimum uncertainty.

This conclusion shows us how to deal with calculating the amount of information needed to digitize a system with a given energy for a given time to the highest precision required by nature. It is simply twice the energy times the time divided by hbar.

In fact, this idea is not new. After the turn of the 20th century, the concept of the information content of a signal was explored by Nyquist and Hartly and shown to be linear in both the time and the frequency of the signal [13,14]. The only missing piece was Einstein's relation from the photoelectric effect where E=hf [15]. This of course makes sense; think of a data port of a computer - the amount of information transmitted is the product of the frequency, or bandwidth, of the signal times how long the signal is sending information.

A more recent discussion of this topic concluded in 1998 with Hawking offering that information is equal to energy times time [16]. Lloyd was also onto the same idea when he stated that the maximum rate at which a physical system can process information is proportional to its energy [17]. As we can see from the results presented here, these conclusions are applicable in a myriad of many different contexts.

2) Our second example will result in a variation of conclusion 3 and begins by showing that S(ρ) is normally invariant under a finite rotation through angle θ

$$S(\rho) = S\left(U\left(R\left(\hat{\theta}\right)\right)\rho U^{\dagger}\left(R\left(\hat{\theta}\right)\right)\right)$$

Where $U\left(R\left(\hat{\theta}\right)\right) = e^{-i\hat{\theta}\cdot\hat{J}/\hbar}$, and J is the angular momentum [18]

Without going through the details, we can again ask the question "what if we separate the positive contribution from the negative contribution of the unitary rotation on the Von Neumann entropy?" The same arguments used above results in an additional factor added onto the initial entropy.

$$S\left(U\left(R\left(\hat{\theta}\right)\right)\rho U^{\dagger}\left(R\left(\hat{\theta}\right)\right)\right) = S(\rho) + 2J\theta/\hbar$$

This then leads us to conclusion 4.

Conclusion 4) At a given instant in time, the angular momentum, J, of a system is in direct proportion to the amount of information that is contained within or can be transmitted by that system.

As one might see, there is another challenge presented here with a factor of 2π. If you rotate the angular momentum by 2π (which should be an identity transformation) you would have an additional 4πJ/ħ in the Von Neumann entropy. However as we will show in the examples the correct proportionality is I=2J/ħ. Appendix C discusses a possible source of this confusion.

Note that conclusion 4 intends that the angular momentum, J, includes both orbital, L, and intrinsic angular momentum, S.

A "q-bit" from quantum computing: Coming off the heals of conclusion 4, it behooves one to further examine how the governing hypothesis deals with particles that have spin. With this interpretation we can see its implications to a "q-bit".

As shown specifically in the examples of the electron, the photon, and the Higgs boson, we see how we should interpret the governing hypothesis by taking each quantum unit of action (including intrinsic angular momentum) as one natural unit of information. Since each q-bit is one quantum unit of action, we can easily calculate from our governing hypothesis that a q-bit contains one natural unit of information, ln(e).

Using the standard Von Neumann approach a "q-bit" obtains its maximal entropy at ln(2), however this approach does not take into account our first conclusion which requires that we include the entropy from both domains. This extra ln(e/2) of information comes about in the dual domain to the "q-bit" wavefunction, or the angular position domain.

In this light, our first conclusion is self consistent with our second and fourth conclusions, whereby adding the self-information from both domains results in an information content for the spin of particles, which is quantized to integer units and proportional to the angular momentum.

These results suggest that current research in quantum information is looking at only ~70% (precisely ln(2)) of the self-information contained in a q-bit. It is possible that a trace of this missing ln(e/2) of information can be found in the experimental ".7" structure found in quantum conductance measurements.

Hidden variables in quantum mechanics: Our governing hypothesis on information presents an interesting challenge to the working conclusion that there are no hidden variables within quantum mechanics; however a full analysis of the implications will require more discussion.

There are many philosophical and scientific papers surrounding the hidden variable theorems and the associated loss of a non-causal nature to quantum mechanics resulting in the "spooky" actions at a distance [19,20,21,22]. However with our governing hypothesis that the quantum of action is one natural unit of information, we can see that the action at a distance is not "spooky" but rather a consequence of one degree of freedom obtaining one natural unit of mutual information with another degree of freedom.

For example, according to our hypothesis a pair of entangled q-bits each contains one natural unit of information. When the two become entangled, the two separate natural units of information overlap and become one natural unit of mutual information. As we will see in the section on electrons, this is enough mutual information to correctly answer (probabilistically speaking) any question you can ask of a q-bit. Thus the particles themselves are the hidden variables we seek.

It is also possible to see how Bohr's correspondence principle fits in. The correspondence principle states that theorems in quantum mechanics should approach the classical theorems when the action of a system is large compared to $\hbar$ [23]. If $\hbar$ were small compared to the action of a system, then the amount of information in that system would be large. In macroscopic systems there are Avogadro's number of natural units of information which overwhelms the quantum uncertainty leaving the classical theorems. However since $\hbar$ is not zero nor the action infinite, the information will be finite and some uncertainty and noise will remain, leaving space for a violation of Bell's theorem.

Examples: Here we look at specific examples where we can independently verify that the amount of information needed to describe those systems is equal to the number of quantum units of action of that system. These twelve examples each give different empirical evidence in support of at least one of our four conclusions. The following table gives a listing detailing which examples supports which conclusions

<table>
<tr><th>“X” implies support for conclusion</th><th>Conclusion 1 (both domains)</th><th>Conclusion 2 (quantized)</th><th>Conclusion 3 ($I = 2ET/\hbar$)</th><th>Conclusion 4 ($I \; \alpha \; 2J/\hbar$)</th></tr>
<tr><td>Black Hole</td><td></td><td></td><td>X</td><td></td></tr>
<tr><td>Electric circuit</td><td></td><td></td><td>X</td><td></td></tr>
<tr><td>Tunneling event</td><td></td><td></td><td>X</td><td></td></tr>
<tr><td>Two slit experiment</td><td></td><td>X</td><td></td><td></td></tr>
<tr><td>Exponential distribution</td><td>X</td><td>X</td><td></td><td></td></tr>
<tr><td>Gaussian distribution</td><td>X</td><td rowspan="2">X</td><td rowspan="2">X</td><td></td></tr>
<tr><td>Sinusoidal distribution</td><td>X</td><td></td></tr>
<tr><td>Signal</td><td>X</td><td>X</td><td>X</td><td></td></tr>
<tr><td>Spin 1 (Photon)</td><td></td><td>X</td><td></td><td>X</td></tr>
<tr><td>Spin ½ (Electron)</td><td>X</td><td>X</td><td></td><td>X</td></tr>
<tr><td>Spin 0 (Higgs boson)</td><td></td><td>X</td><td></td><td></td></tr>
<tr><td>Brownian Motion</td><td></td><td></td><td>X</td><td></td></tr>
</table>

A black hole: Using our governing hypothesis, it is possible to calculate the entropy of a black hole by directly substituting in the energy and time. The geometry is most simple on a non-rotating, charge free black hole, called a Schwarzchild black hole. In a Schwarzchild black hole, the event horizon is a sphere with radius equal to $R_S = 2GM/c^2$, where G is the gravitational constant, c is the speed of light, and M is the mass inside the event horizon [24].

To calculate the information of the Schwarzchild black hole using our governing hypothesis, we take the energy and multiply it by the time. The energy is of course, Mc^2, which we get from Einstein [25]. The time will be the length of the event horizon, or more explicitly the time it takes a photon to travel from one side of the event horizon to the other. In a normal sphere this would be the diameter of the sphere divided by c, however because the equations of general relativity break down inside a black hole, the path of the photon must follow the circumference along the event horizon. Thus the time is equal to π times the radius divided by c. Here we have the following two equations.

$$E = Mc^2$$

$$t = \pi R_S / c = 2\pi GM / c^3$$

Using our equation for information, we can calculate the information as

$$I = 2Et/\hbar = \frac{4\pi GM^2}{\hbar c}$$

We now show that "I" is equal to the thermodynamic entropy of a Schwarzchild black hole. We begin by differentiating I by M

$$dI = \frac{8\pi GM}{\hbar c^3} d\left(Mc^2\right)$$

With assistance from Hawking we can show that the temperature of the black hole, due to black body radiation, is [26]

$$T = \frac{\hbar c^3}{8\pi GMk_B}$$

Where k_B is Boltzmann's constant. Again seeing that $E=Mc^2$, we can rewrite our differential equation as

$$dI = \frac{1}{k_B T} d(E)$$

In this form, we can recognize the equation as the first law of thermodynamics (TdS=dE) [27] and show that the thermodynamic entropy (defined as the log of states) plus a constant is actually Boltzmann's constant times the information we have defined in our governing hypothesis. Assuming a black hole with zero mass has zero entropy, we can integrate to get

$$I = S/k_B \equiv \ln(\Omega)$$

The factor of k_B is a result of the thermodynamic entropy, S, being defined with Boltzmann's constant included. Please note that (similar throughout the paper) I is in natural units.

This example is exciting for two reasons. First the result is simple and holds, even when M is very large. Second, when M is small (about the size of the Planck mass) the amount of information is approximately one natural unit.

As we have shown in the second conclusion on the quantization of information, the number of natural units of information must be an integer greater than or equal to one. Thus when the mass becomes less than approximately the Planck mass ($M \approx M_P$) or when the Schwarzchild radius becomes less than approximately the Planck length ($R_S \approx L_P$) there is not enough information to give physicality to the equations of general relativity.

The black hole is thus our first example of a system where the governing hypothesis is affirmed.

An electric circuit: This example will closely follow the work of Brillouin [28]. Here we assume a noiseless physical RC circuit with time constant RC = γ. If the circuit is excited to voltage V_n, the circuit will decay according to the exponential law

$$V = V_n \cdot \exp[-t/\gamma]$$

The energy stored in the circuit is proportional to the square of the voltage and thus

$$E = E_n \cdot \exp[-2t/\gamma]$$

Now let us assume that we use a code of n equidistant energy levels chosen uniformly

$$0, E_0, 2\cdot E_0, 3\cdot E_0, ..., E_n = (n-1)\cdot E_0$$

Requiring error free transmission leads to the condition that we must wait a time interval between successive pulses such that the energy has dropped below $E_0/2$, whereby we can distinguish the next pulse, which might be either 0 or E_0. Solving for the worst case scenario leads to the condition

$$E_0/2 = (n-1)\cdot E_0 \cdot \exp[-2t/\gamma]$$

$$t = (\gamma/2)\cdot \ln[2\cdot(n-1)]$$

To distinguish between two successive pulses a time period, t, must be used. For a long time T the total number of signals will be T/t. Given that each pulse strength is equally likely, each signal carries ln[n] nats of information and the total information that is transmitted will be the number of signals that can be sent in time T multiplied by the information per pulse.

$$I = (T/t)\cdot \ln[n] = (2T/\gamma)\cdot \ln[n]/\ln[2\cdot(n-1)]$$

The function $f(n) = \ln[n]/\ln[2\cdot(n-1)]$, $n \in Z$, is maximized for $n = 2$, and $n = \infty$ and has a value of $f(2) = f(\infty) = 1$. It is interesting and worthwhile to note that f(n) is maximized for the same values of n that are the allowed maximum occupation numbers for fermions ($n = 2$) and for bosons ($n = \infty$). Plugging $f(n) = 1$ into the equation above leads to

$$I = 2T/\gamma$$

Upon examining these equations, one can see that $1/\gamma$ has the same mathematical form as an imaginary angular frequency ($i\cdot E/\hbar$). Looking at the magnitude of the relationship leaves us again with our hypothesis.

$$I = 2ET/\hbar$$

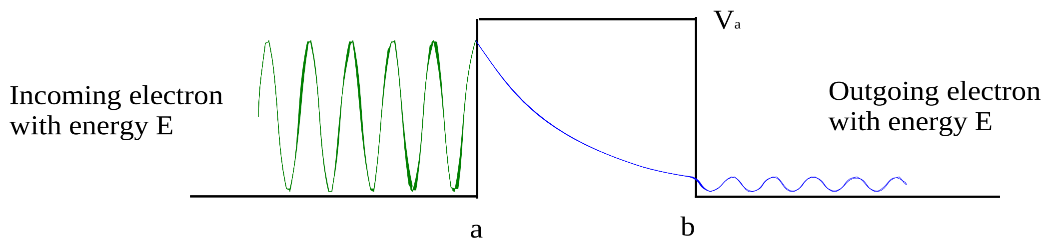
It would be nice to be able to conclude that this relationship is true for all circuits that obey Kirchhoff's laws, but I don't yet possess that proof. However it still holds as an example where our hypothesis is correct.

A tunneling event: The tunneling of an electron across a barrier that is classically forbidden has a finite probability of occurrence, and is a consequence of the requirement that the wave function and its derivative be continuous. Using the WKB method, the Schrödinger equation can be solved with the assumed solutions of the form $\psi(x) = \exp[-iS(x)]$ [29].

The transmission probability across a barrier is defined as the outgoing intensity divided by the incoming intensity, which is proportional to $|\psi(b)|^2/|\psi(a)|^2$. When the wave function tunnels into the classically forbidden area, between a and b, the complex exponential becomes a decaying exponential. The transmission tunneling probability, T, is

$$T = \exp[-2p\cdot x/\hbar]$$

where $x = (b-a)$ and p s.t. $p^2/2m = |V_a - E|$



This result also holds when the barrier is not square, where the transmission T is

$$T = \exp[-2\int_a^b p\cdot dx/\hbar]$$

Thus the transmission probability is exp[-2·action], where the action is defined as the momentum times the distance (instead energy times the time).

If we were to apply the viewpoint of universal probability [30], we can quantify how much information should be associated with each tunneling event by taking the negative natural log, i.e.

$$I = -\ln(T) = 2px/\hbar$$

It is not clear that this is true for all instances of tunneling, but it is clear that for this example our hypothesis is not only true but can be extend into the spatial dimensions as well as the temporal dimension.

The two slit experiment: While the two slit experiment looks like a one bit measurement (upper slit or lower slit with equal probability), it is not.

Assume an experimental setup like below in the figure, with "a" approaching zero and "b" a half integer wavelength of the incoming wave. If a measuring device is asked to locate the particle at one slit, then the probability distribution on a circular measurement screen away from the partition in the two slit experiment is uniform [31]. If a measurement of the particles location at the slits is not attempted, then the particle emerges from both slits with equal amplitude, and produces the cosine squared distribution on the circular measurement screen away from the partition [32].

$$u(x) = 1/(2r) \quad \text{for } x \in [-r, r]$$

$$c(x) = (1/r)\mathrm{Cos}^2(\pi\,(b/\lambda)(x/r)) \quad \text{for } x \in [-r, r],\ b/(2\lambda) \in Z$$

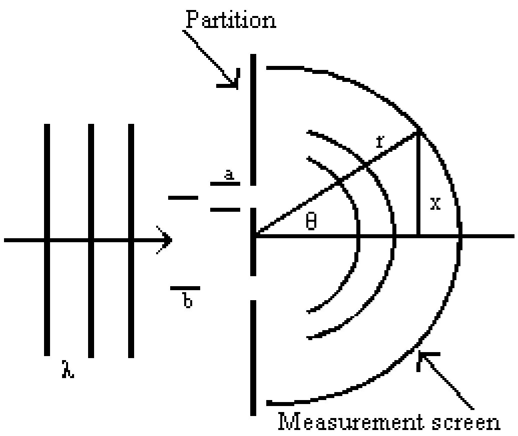
If we have N measurement locations on the measurement screen we can get ln(N) minus the difference in differential entropy between the uniform distribution and the actual distribution. An amazing fact about the sinusoidal distribution is that it has ln(e/2) less entropy than the uniform distribution.

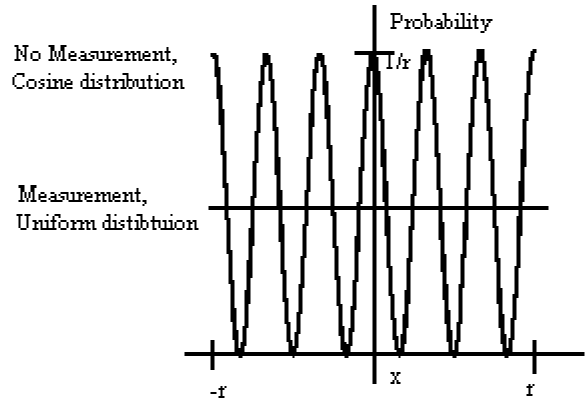
$$h(u) - h(c) = {}_{-r}\int^{r} (1/2\cdot r)\cdot\ln[2\cdot r]\cdot dx + {}_{-r}\int^{r} (1/r)\mathrm{Cos}^2(\pi\cdot(b/\lambda)\cdot x/r)\ln[(1/r)\cdot\mathrm{Cos}^2(\pi\cdot(b/\lambda)\cdot x/r)]dx$$

$$= {}_{-r}\int^{r} (1/r)\mathrm{Cos}^2(\pi\cdot(b/\lambda)\cdot x/r)\ln[2\cdot\mathrm{Cos}^2(\pi\cdot(b/\lambda)\cdot x/r)]dx$$

This integral can be solved analytically for any r and $b/(2\cdot\lambda) \in Z$, and is equal to the natural log of e/2.

Entropy of uniform distribution – entropy of cosine distribution = h(u) - h(c) = ln[e/2]





If a measurement were to occur at the partition, one bit of entropy (ln(2)) could be extracted by knowing which slit the particle when through plus ln(N) from the measurement screen. If a measurement at the partition did not happen the total entropy would be ln(N)-ln(e/2) because in this case the resulting distribution is the cosine distribution. The difference is of course ln(e) = 1.

$$I(\text{two slit}) = I(\text{measurement at partition}) - I(\text{no measurement at partition})$$

$$= \ln(2) + \ln(N) - (\ln(N) - \ln(e/2)) = 1$$

Thus we can see that a simple left/right (ln(2)) measurement spills over to where the resulting distribution allows an additional ln(e/2) to be extracted.

An exponential distribution: The exponential distribution is a perfect example of how both our first and second conclusions work together. The self-information, $h = -\int g(t)\ln[g(t)]dt$, of an exponential distribution is dependent on its mean parameter, however we will use our first conclusion and the result from Hirshmann, in 1957, where he explained that one should take the entropy of both domains and add them together [3]. In this case the addition of the differential entropies from both domains is equal to a constant, ln(e).

Abbreviated here (but explicit in Appendix B), we start with an exponential wavefunction and its Fourier transform pair. After we take the magnitude squared of each, we arrive at the following two distributions

$$\psi^*(t)\psi(t) = a\cdot\exp[-a\cdot t] \text{ for } t \geq 0 \qquad \text{exponential distribution}$$

$$\Psi^*(s)\Psi(s) = (4/a)/[1 + (4\pi\cdot s/a)^2] \qquad \text{Cauchy distribution}$$

The differential entropy of the exponential distribution is ln(e/a) and the differential entropy of the Cauchy distribution is ln(a) [33]. Thus the self-information in the exponential-Cauchy pair is one natural unit.

$$I = \ln(e/a) + \ln(a) = 1$$

Using our first conclusion to add the differential entropy in both domains we arrive at a specific verification of our second conclusion that information in nature comes in packets of natural units.

The Gaussian: The Gaussian wavefunction is the function that minimizes the product of the variances of two conjugate variables and by doing so hits the limit of the

Heisenberg uncertainty principal [1]. Taking the magnitude squared of the Gaussian wavefunction results in the Gaussian distribution, which is the distribution that maximizes the self-information of a random variable for a given variance [2]. A pair of Gaussian wavefunctions and their self-information is covered in the end of Appendix B.

As shown in the appendix, the self-information, $h = -\int g(t)\ln[g(t)]dt$, of a Gaussian is dependent on its variance, however like we used in the example with the exponential distribution we will use our first conclusion and add the differential entropy of both domains together.

Abbreviated here, the differential entropy of a pair of Gaussian distributions, derived from the time domain Gaussian wavefunction, $\psi(t)$, and the frequency domain wavefunction, $\Psi(f)$, is

$$h(\psi^*(t)\psi(t)) = (½)\ln[2\pi e(\Delta t)^2]$$

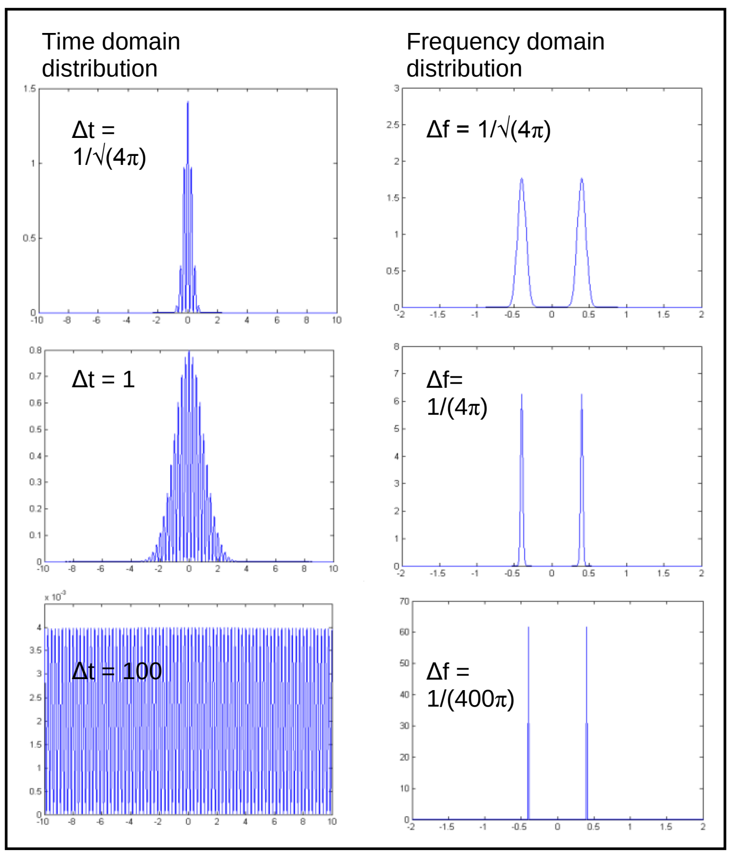
$$h(\Psi^*(f)\Psi(f)) = (½)\ln[2\pi e(\Delta f)^2]$$

Adding the two differential entropies and substituting $1/(4\pi)$ for $(\Delta f)\cdot(\Delta t)$ leads to

$$h(\psi^*(t)\psi(t)) + h(\Psi^*(f)\Psi(f)) = (½)\ln[(2\pi e)^2(\Delta t)^2(\Delta f)^2] = \ln[e/2]$$

This result is true for all Gaussians, regardless of the value of (Δt) and (Δf). Yet this is not the whole picture for as we know through conclusion 2 that information is quantized to integer units. As we will see in the next example on sinusoidal distributions, the missing bit of the Gaussian distribution comes from the sinusoidal distribution.

A sine wave: As we know from Fourier Theory, an infinite sine wave transforms to an impulse pair (or two opposed delta functions) [34]. Here, through numerical methods, we will calculate the self-information of both domains of the sine wave and show in Appendix D that the sum is equal to one bit, ln(2).



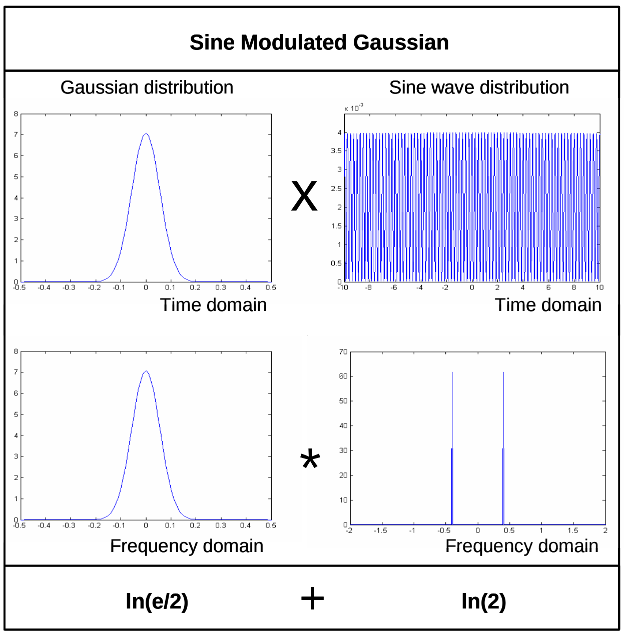
The differential self information of an infinite sine wave is infinite. Oppositely the differential self information of an impulse pair is negatively infinite. However using the Gaussian wavefunction as the generating function for the delta function, we can keep the differential entropy in either domain, finite and notice that the sum of the two differential entropies approaches the constant ln(2) as the numerical integration gets more accurate.

To generate a Dirac delta function, (ie. the impulse function), we will start with a pure sine wave and convolve it with the Gaussian wavefunction. We then take the magnitude squared of both domains to get the probability distribution and lastly take the limit as the width of the Gaussian goes to zero or as the width of its dual domain goes to infinity. Like we have seen in the example on Gaussians, we can calculate the amount of information in a pair of Gaussian spread sine waves, for any value of (Δt) and (Δf).

The included graph above shows three pairs of Gaussian spread sine waves with different values of (Δt) and (Δf). As you can see, as (Δt) gets bigger, the time domain gets closer to an infinite sine wave, while the frequency domain approaches the impulse pair. However, the take away is that the sum of the differential entropy of both domains approaches ln(2) as the numerical integration gets more precise for any value of (Δt) and (Δf). See Appendix D for more detail.

This is great because it gives us the exact missing piece we needed from the analysis of the Gaussian distribution (which was missing ln(2) of self information to make one discrete natural unit). From this we can see that the Gaussian distribution and the sinusoidal distribution are inextricably linked.

From these two distributions, we can build what I am calling a sine modulated Gaussian.



Putting a Gaussian distribution together with a sinusoidal distribution has a number of nice qualities. To start with it gets rid of the problem that a theoretical sine wave goes on forever in the time domain and has an infinite value in the frequency domain at its pure frequency, both of which are undesirable. However if we were to convolve a sine wave with a Gaussian in the frequency domain, we would end up with a random variable that has much more physicality to it, ie a pure frequency with noise. Similarly, looking from the other side, one would not expect a Gaussian all on its own without any sinusoidal modulation or else the Gaussian would be at absolute zero energy which has its own problems with re-normalization. There are also problems when the center frequency of the sine wave is less than the width of the Gaussian since in that case there would be significant overlap between the two Gaussians.

Still, the majority of the time these problems do not arise and we can use a pure sine wave convolved with a Gaussian in the frequency domain as a fundamental distribution. We can calculate the total self-information of the sine modulated Gaussian by assuming that the sinusoidal distribution and Gaussian distributions have zero correlation between them and are thus independent. In this case we simply add the self-information from one distribution to the self information from the other distribution.

$$I = I(\text{Gaussian}) + I(\text{sine wave}) = \ln(e/2) + \ln(2) = 1 \text{ natural unit}$$

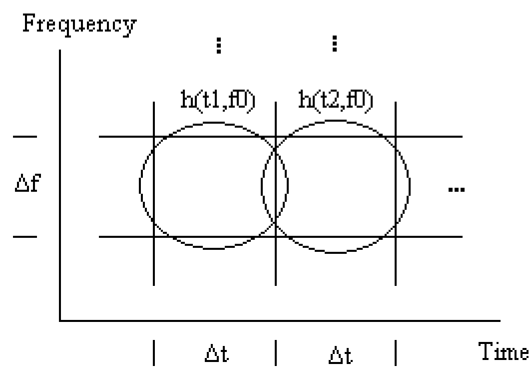
This result fits in beautifully with our conclusion that information comes in integer units of natural information. Thus, for this example where we started with a Gaussian distribution that strictly satisfies the Heisenberg uncertainty principle, and convolved it with a sinusoidal distribution we arrive at a fundamental distribution that has one natural unit of information.

As we will see in the next example, this analysis also supports conclusion 3 where we take a signal and break it into a basis of shifted and modulated Gaussian wave packets to reproduce a signal of energy, E, for time, t, with $2Et/\hbar$ natural units of information.

A signal: Armed with our results from the exponential distribution, the Gaussian distribution and the sinusoidal distribution, we move onto our next example, showing how much information can be encoded into a signal that is mostly bound in frequency to f and in time to t. We will rely on Gabor's work from 1946 [35], which was later supported by Slepian, Pollak and Landau in 1961 and 1962 [36,37,38]. Gabor's insight was to partition a signal into a sine modulated Gaussian basis and assign one quantized unit of information to each pair of sine modulated Gaussian distributions.

Given the direct proportion between the area in the time frequency plane that a signal occupies and the amount of information encoded into that signal, you can double the area, by either doubling the bandwidth, or doubling the duration and thus double the amount of information encoded in the signal (and if you double both the bandwidth and the duration, the information quadruples). Again this is what Nyuist and Hartley saw at the turn of the 20th century [13,14], and Lloyd saw after the turn of the 21st century [17].

To help visualize this picture, think of a signal when it is sampled. The result is a random variable with statistical properties. We know two ways we can visualize this picture. First a Venn diagram which denotes how much information is encoded into that random variable. Second the time-frequency plane which denotes the density of power that the signal occupies. Think of these two pictures (the Venn diagram and the time-frequency plane) as the same theoretical object.



This suggests a way one can encode this information is to produce independent pairs of sine modulated Gaussian distributions that are shifted in time by units of Δt (one standard deviation of time) and modulated in frequency by units of Δf (one standard deviation of frequency) with each sine modulated Gaussian being encoded with one natural unit of information. When properly weighted and mixed together the resulting signal will statistically reproduce any measurable quantity of the original signal to within the noise floor.

If the signal is mostly limited to frequency f and duration t [36,37,38] we can tile that area of the time-frequency plane of size $f\cdot t$ with sine modulated Gaussians that each occupy an area of size $\Delta f \Delta t$. The number of sine modulated Gaussian pairs is therefore $f\cdot t/(\Delta f \Delta t) = 4\pi f\cdot t$ (when we substitute in the Heisenberg Uncertainty). With the final substitution that energy is equal to Plank's constant times the frequency, $E= 2\pi\hbar f$ [15], we can state that the number of independent sine modulated Gaussian pairs is equal $2Et/\hbar$.

As we stated in the previous examples each sine modulated Gaussian pair carries with it one natural unit of information and we can verify that for a signal

$$I=2Et/\hbar$$

To decode this information one would simply apply filters to the signal, separating out each sine modulated Gaussian packet and decoding the one natural unit of information for each.

If one were to look at the original (and brilliant) paper by Gabor, they would find yet another extra factor of 2π from definition of the density of the Gaussian basis. Again this factor is explained in Appendix C.

Note that the basis decomposition could also use the exponential distributions to arrive at the same result. This is because each exponential distribution has a width in the time domain that is inversely proportional to its width in the frequency domain (even though Δf is not defined) and carries with it ln(e) of information from both domains.

Spin 1 particle (the photon): The photon is a spin one particle and thus has two quantum units of action. Using our governing hypothesis, this implies that each photon carries with it $2\cdot\ln(e)$ of information. This is precisely the capacity that Yamamato and Haus calculated in 1986 for a photon channel when prepared in a squeezed state [39].

In a squeezed state both the electric and magnetic fields have the form of a sine modulated Gaussian. Thus if a measurement were to happen, say of the electric field, the results would be distributed like a sine modulated Gaussian and would be independent of a measurement of the magnetic field. From our previous analysis, the independent measurements of two modulated Gaussian variables should produce one natural unit of information for each modulated Gaussian or two natural units per photon.

Adding the next example of the electron, (which will show that one quantum unit of angular momentum contains one natural unit of information) to this example (where we have shown that two quantum units of angular momentum contain two natural units of information), we can conclude that angular momentum is in direct proportion to the amount of information in a system; a direct affirmation of conclusion 4. [The even stronger result that $I = 2J/\hbar$ can also be concluded from these examples].

Spin ½ particle (the electron): A lot of work has been done on quantum information of an electron [40,41,42], but instead of possibly sidetracking the emphasis here, I will jump right into showing that the amount of information (or entropy) in a measurement of an electron's intrinsic angular momentum is ln(e). Showing this result will support the conclusion that twice the spin quantum number is equal to the number of natural units of information that the particle contains.

The procedure to show that a spin ½ particle contains ln(e) of information will be mathematically identical to the procedure used to show that a measurement of the two slit experiment necessarily extracts ln(e) of information from the system. Yet I will re-state the main points again for clarity, to simply show that abiding by conclusion 1 (that you need to sum the information from both domains) you necessarily extract ln(e) more of information from the particle than if you don't make any measurement at all.

To begin with take a mixed state where the particle has a 50/50 chance of being spin up / spin down along a chosen direction. In this state the particle's angular momentum wavefunction is equally divided between being in the spin up direction and the spin down direction. In the dual domain, the particle's angular position wavefunction looks like the cosine distribution.

$$\Psi(\theta)^*\Psi(\theta) = (1/\pi)\cos^2(\theta/2) \text{ for } \theta \in [-\pi,\pi]$$

Since the angular momentum domain is quantized to only the up or down state we can not measure more accurately than ln(2). However in the angular position domain we can measure at an arbitrarily small angle ($d\theta = 2\pi/N$) and thus we can have N different possible outcomes. If a measurement were to occur of the angular position domain, the amount of information that could be extracted from the measurement would approach ln(N) minus the difference in differential entropy between the uniform distribution and the cosine distribution as N approaches infinity. As we have shown in the previous example of the two-slit experiment, the cosine distribution has ln(e/2) less self-information than the uniform distribution. Thus a measurement of the angular position domain would extract ln(2N/e) of information as N goes to infinity or as $d\theta$ goes to zero.

If however we first measure the angular momentum, we would extract ln(2) from a sample of the 50/50 distribution; and then an additional ln(N) from the angular position domain. The ln(N) is because a measurement of the angular momentum domain places the angular momentum into an eigenstate and thus the angular position domain will be uniform, $1/(2\pi)$ for $\theta \in [-\pi,\pi]$. With a uniform distribution on N possible outcomes, the amount of information that can be extracted is ln(N).

We are now in a position to show that a measurement of the angular momentum domain necessarily extracts ln(e) more out of the system than if we were to only measure the angular position domain. If we were to keep the particle in a mixed state and measure its angular position we could get ln(2N/e) out of the system. If we measure the angular momentum domain and then the angular position domain we get out ln(2) + ln(N). The difference is ln(e).

$$I(\text{measurement of electron}) = \ln(2) + \ln(N) - \ln(2N/e) = \ln(e) = 1$$

From this it is clear that a measurement of the angular momentum extracts ln(e) more information (or entropy) than without a measurement if the angular position domain is taken into consideration. Again this analysis on the electron is inconsistent with the current understanding. However by looking at all the places where information can be stored (i.e. both domains) we can see that the electron has ln(e) of information.

Spin 0 particle (Higgs boson): In the previous two examples, we showed that the amount of information contained within the intrinsic angular momentum of a particle is twice the quantum spin number of that particle. In this example we will examine how that result deals with particles that have zero spin.

Using our rule, that you double the quantum spin number, for the Higgs boson, we would get zero information. This does not make sense since the existence of the particle tells us something. To solve this dilemma, preliminary research indicates that one additional natural unit of information should be added to total magnitude of intrinsic angular momentum that comes from the x, y and z directions. Similar to how other spin operators are summed, this extra degree of freedom will add in quadrature [18]. One can think of this extra degree of freedom being associated with the information relating to the existence the particle and its phase relative to its anti-particle.

Using the stronger preliminary result, that information from the x, y, and z directions is $2S/\hbar$ and if we add this in quadrature to the one additional natural unit of information relating to the existence of the particle, the total information will take on the simple form below.

$$I^2 = \left(\tfrac{2}{\hbar} S\right)^2 + (1)^2$$

We know that the magnitude of spin, S, is equal to $\hbar\sqrt{(s(s+1))}$ where s is the quantum spin number.

$$I^2 = 4s(s+1) + 1 = (2s+1)^2$$

$$I = 2s + 1$$

Where: I is the information (in natural units). Thus with s being either zero or a half integer, the total information is still a positive integer.

With this understanding of the relationship between the quantum spin number and the amount of information associated with spin, we can see that the existence of a particle necessarily requires one natural unit of information even when the quantum spin number is zero.

It is also interesting to note that given angular momentum is a conserved quantity we can say that information is conserved as well.

Brownian Motion: As a quick note on the system described by Brownian motion, it is possible to show that a diffusing free particle generates entropy at a rate equal to twice its temperature times Boltzmann's constant divided by ħ. This implies that over a period of time t, it requires $I = 2k_BTt/\hbar$ natural units of information to tract a diffusing particle. It is interesting to note that while most of the systems being examined here are time reversible, the equations of Brownian motion are not. Details can be found in the work by Haller in 2007 [43].

Conclusion - Nature of particles / information: It would not be a discovery to say that particles (or atoms) store information; however the intent of this paper is to convince the community that particles don't just store information, they are information (with the dual being that information fundamentally consists of discrete particles). Hopefully the governing hypothesis and its four accompanying conclusions give the insight needed to examine new directions of research that can open up a whole new way to look at (and analyze) more than just the 12 examples shown here.

Acknowledgements: The author would like to thank the many people who have helped in the development of this work; notably my family for their continuous support, the electrical engineering departments at Princeton and Stanford for a base from which I could conduct my research and Professor Hawking for his inspiration and direction.

Appendix A (Derivation of governing hypothesis): The basic calculation to arrive at the governing hypothesis, begins with finding the probability of a system at energy ε evolving for a time τ, and then taking the logarithm to find the information needed to describe that evolution.

The time evolution of a wavefunction is the complex exponential, $\exp(-i\varepsilon\tau/\hbar)$, also known as the Fourier kernel. Using a Minkowski transformation the complex exponential is turned into a decaying exponential, $\exp(-\varepsilon\tau/\hbar)$. The order of when to apply the Minkowski transformation is important because if the magnitude squared of the wavefunction is calculated before the transformation is used, the probability will be one and the information is thus zero.

However by applying the transformation before the operation of magnitude squared, the phase information is extracted and the result is a real probability of $\exp(-2\varepsilon\tau/\hbar)$ with an associated information of $2\varepsilon\tau/\hbar$. [Note that since the natural logarithm is used, the units of information are in natural units]

The time dependent Schrödinger equation is solved for a system with eigenvalue ε.

$$i\hbar d/dt\ \psi_k(t) = H_k\psi_k(t)$$
$$H_k\psi_k(t) = \varepsilon_k\psi_k(t)$$
$$\psi_k(t) = \exp[-i\varepsilon_k(t-t_k)/\hbar]$$

Now a system of n independent and symmetric particles is assumed.

$$H^n = \sum^n_{k=1} H_k$$
$$\Rightarrow$$
$$\Psi_n\ (t) = \prod^n_{k=1} \psi_k\ (t)$$

Assuming further that each particle has the same eigenvalue, $H_k = H$, and that each particle has the same phase, $t_k = t_0 = 0$, the overall wave function, $\Psi_n\ (t)$, takes on a simple form.

$$\Psi_n\ (t) = \prod^n_{k=1} \psi(t) = \exp[-in\varepsilon t/\hbar]$$
$$\Psi_n\ (t) = \exp[-n\varepsilon it/\hbar]$$

Making the Minkowski transformation ($i\cdot t \rightarrow \tau$) allows the phase information inside the Fourier kernel to be analyzed. One might be familiar with the Minkowski transformation as it was made popular in its use on the invariant distance in four space within special relativity. The transformation has also been used much more recently with path action integrals and is known as analytic continuation [11,12].

$$i\cdot t \rightarrow \tau$$
$$\Psi_n\ (\tau) = \exp[-n\varepsilon\tau/\hbar]$$

The wave function is now a pure real number, and thus the magnitude squared of the wave function is just the square of it.

$$\Psi^*_n\ (\tau)\Psi_n\ (\tau) = \exp[-2n\varepsilon\tau/\hbar]$$

Taking the natural log of both sides, multiplying by -1/n, and substituting $\prod^n_{k=1} \psi(\tau)$ for $\Psi_n\ (\tau)$ allows the weak law of large numbers to be evoked.

$$\ln[\Psi^*_n\ (\tau)\Psi_n\ (\tau)] = -n2\varepsilon\tau/\hbar$$
$$-1/n\ \ln[\Psi^*_n\ (\tau)\Psi_n\ (\tau)] = 2\varepsilon\tau/\hbar$$
$$-1/n\ \ln[\prod^n_{k=1} \psi^*(\tau)\psi(\tau)] = -1/n \sum^n_{k=1} \ln[\psi^*(\tau)\psi(\tau)] = 2\varepsilon\tau/\hbar$$

As $n \rightarrow \infty$, the sum approaches the expected value.

$$n \rightarrow \infty$$
$$-1/n \sum \ln[\psi^*(\tau)\psi(\tau)] \rightarrow -E[\ln[\psi^*(\tau)\psi(\tau)]]$$
with probability one by the weak law of large numbers

Since the average approaches the expected value, we can show that information (the negative expected log probability) is equal to twice energy times time divided by hbar.

$$I = 2\varepsilon\tau/\hbar$$

Setting $\hbar = 1$ emphasizes the idea that information is unitless (or more properly, measured in units of the natural logarithm), and doing so results in the compact answer below.

$$I = 2\varepsilon\tau$$

This argument states that by applying the Minkowski transformation at the right time, $\psi^*(\tau)\psi(\tau)$ is still a probability and equal to a real decaying exponential. One could thus say that the probability a system with an energy ε evolves for the time τ is $\exp[-2\varepsilon\tau/\hbar]$. Taken on average, this results in information that is equal to twice energy times time.

An alternate way to derive the governing hypothesis is to assume that information is linear in both time and energy (which we get from Nyquest and Hartley [13,14]) and then measure the energy in units of the smallest possible standard deviation of the energy for that system, and likewise for the time. Thus $I = (\varepsilon/\Delta\varepsilon)(\tau/\Delta\tau)$. Plugging in the Heisenberg uncertainty principle, $\hbar/2$, for $\Delta\varepsilon\Delta\tau$ (as the smallest possible uncertainty in a system), we arrive at our hypothesis, $I=2\varepsilon\tau/\hbar$.

Appendix B (Self information in a Fourier transform pair): These examples will use conclusion 1 (adding the entropy of both domains to completely quantify the information in a system) to show examples of conclusion 2 (that information comes in packets of one natural unit. Here we have explicitly,

$$I(\text{wavefunction}) = h(\psi^*(t)\psi(t)) + h(\Psi^*(f)\Psi(f))$$

$$\text{Where } h(g(t)) = -\int g(t)\ln[g(t)]dt$$

$$\text{And } \Psi(f) = \int\psi(t)\exp[-i2\pi ft]dt$$

An issue might arise when applying this conclusion, as to the integrity of using the differential entropy to compute the self-information, as it is well known that the differential entropy is not a true measure (since it scales and can be negative). However adding the differential entropy from both domains solves this problem.

Hirschmann [3] showed that the sum of the differential entropy from both domains is always greater than ln(e/2) (and thus positive); and secondly, shown below we can see that the sum is invariant under a scale transformation, for any a > 0.

$$\sqrt{a}\psi(at) \rightarrow \sqrt{(1/a)}\Psi(f/a)$$

$$-\int a\psi(at)^*\psi(at)\ln\left(a\psi(at)^*\psi(at)\right)dt = h\left(\psi(t)^*\psi(t)\right) - \ln(a)$$

$$-\int (1/a)\Psi(f/a)^*\Psi(f/a)\ln\left((1/a)\Psi(f/a)^*\Psi(f/a)\right)df = h\left(\Psi(f)^*\Psi(f)\right) + \ln(a)$$

$$h\left(a\Psi(at)^*\Psi(at)\right) + h\left((1/a)\Psi(f/a)^*\Psi(f/a)\right) = h\left(\Psi(f)^*\Psi(f)\right) + h\left(\psi(t)^*\psi(t)\right)$$

Thus the total self-information can be used as a proper metric as shown specifically in the next examples.

The exponential distribution: Abiding by our first conclusion, the differential self-information contained within an exponential distribution of any width is equal to exactly one natural unit when the differential self-information from the Cauchy distribution (the exponential's Fourier transform pair) is included. We can see this below

$$\psi(t) = a^{1/2}\cdot\exp[-a\cdot t/2] \text{ for } t \geq 0$$

$$\Psi(s) = \int\psi(t)\exp[-i2\pi st]dt = (2/a^{1/2})\cdot[1 - i\cdot 4\pi\cdot s/a]/[1 + (4\pi\cdot s/a)^2]$$

$\psi^*(t)\psi(t) = a\cdot\exp[-a\cdot t]$ for $t \geq 0$ exponential distribution

$\Psi^*(s)\Psi(s) = (4/a)/[1 + (4\pi\cdot s/a)^2]$ Cauchy distribution

$$h(\psi^*(t)\psi(t)) = -\int\psi^*(t)\psi(t)\ \ln[\psi^*(t)\psi(t)]dt$$

$$= \int_0^\infty a\cdot\exp[-a\cdot t]\cdot[a\cdot t + \ln[1/a]]\cdot dt$$

$$= \ln[e/a]$$

$$h(\Psi^*(s)\Psi(s)) = -\int\Psi^*(s)\Psi(s)\ \ln[\Psi^*(s)\Psi(s)]ds$$

$$= \int_{-\infty}^{\infty}(4/a)/\left(1+(4\pi s/a)^2\right)\ln\left[(a/4)\left(1+(4\pi s/a)^2\right)\right]ds$$

$$= \ln[a]$$

$$h(\psi^*(t)\psi(t)) + h(\Psi^*(s)\Psi(s)) = \ln[e/a] + \ln[a] = 1 \text{ nat}$$

The Gaussian distribution: The Gaussian distribution has a differential self-information that is proportional to Δt, but when added to its dual domain (which has a self-information inversely proportional to Δt), the two cancel each other out and the constant ln(e/2) remains as seen below.

The Gaussian wave function

$$\psi(t) = \sqrt{(2a/\sqrt{(2)})}\exp[-a^2\pi t^2]$$
$$\Psi(s) = \int\psi(t)\exp[-i2\pi st]dt = (1/a)\sqrt{(2a/\sqrt{(2)})}\exp[-\pi s^2/a^2]$$
$$\psi^*(t)\psi(t) = (2a/\sqrt{(2)})\exp[-2a^2\pi t^2]$$
$$\Psi^*(s)\Psi(s) = (\sqrt{(2)}/a)\exp[-2\pi s^2/a^2]$$

$$(\Delta t)^2 = \int t^2\ \psi^*(t)\psi(t)dt = 1/(4a^2\pi)$$
$$(\Delta s)^2 = \int s^2\ \Psi^*(s)\Psi(s)ds = a^2/(4\pi)$$
$$(\Delta s)^2(\Delta t)^2 = 1/(4\pi)^2$$

$$h(\psi^*(t)\psi(t)) = -\int\psi^*(t)\psi(t)\ \ln[\psi^*(t)\psi(t)]dt$$
$$= \int(2a/\sqrt{(2)})\exp[-2a^2\pi t^2](2a^2\pi t^2)dt - \int(2a/\sqrt{(2)})\exp[-2a^2\pi t^2]\ln[(2a/\sqrt{(2)})]dt$$
$$= 2a^2\pi/(4a^2\pi) - (½)\ln[2a^2]$$
$$= (½)\ln[e] + (½)\ln[2\pi(\Delta t)^2]$$
$$= (½)\ln[2\pi e(\Delta t)^2]$$

$$h(\Psi^*(s)\Psi(s)) = -\int\Psi^*(s)\Psi(s)\ \ln[\Psi^*(s)\Psi(s)]ds$$
$$= (½)\ln[2\pi e(\Delta s)^2]$$

$$h(\psi^*(t)\psi(t)) + h(\Psi^*(s)\Psi(s)) = (½)\ln[(2\pi e)^2(\Delta s)^2(\Delta t)^2]$$
$$= (½)\ln[(e/2)^2] = \ln[e/2] \text{ nats}$$

Appendix C (Factors of 2π): In the original papers on quantum uncertainty, it was unclear what the exact proportion is for the product of the variance of two conjugate variables. It was not uncommon to see the uncertainty principle quoted as $\Delta E\Delta t\sim h$ [44]. Yet with a century of research, we know the minimum uncertainty is precisely $\Delta E\Delta t \geq \hbar/2$ with an exact bound when the underlying wavefunction is Gaussian [45].

The three examples mentioned above where an there is an extra 2π in the literature are:

1) Quantization of action given as, $\oint E dt = n \cdot h$, [6].

2) 2π rotation of the angular momentum leads to an extra term in the Von Neumann entropy of magnitude $4\pi J/\hbar = 2J/h$.

3) The density of Gabor's Gaussian covering space is 2π natural units per Gaussian

I will discuss example 3), and mention that the first two examples are due to integration over θ from 0 to 2π.

As for the third example, Gabor used Gaussian wave packets to tile the time frequency plane and associated one quantized "logon" of information to each degree of freedom of a signal [34]. He defined each degree of freedom by tiling the time frequency plane with shifted and modulated Gaussian wave packets. Each wave packet, that was separated by $(2\pi)^{1/2}\Delta t$ in the time domain and by $(2\pi)^{1/2}\Delta f$ in the frequency domain, would have one real number associated with it. Thus each area of size $(2\pi)^{1/2}\Delta t(2\pi)^{1/2}\Delta f = 1/2$ would have one real number to cover that space, or two real numbers per unit area of phase space. This coincides with the more rigorous development of the number of degrees of freedom of a signal, known as the analysis of the prolate spheroidal wave functions where the frequency is bound upto W and where the time is bound to a finite duration T [36,37,38]. One of the main conclusions from the work of Slepian, Pollak, and Landau is that the number of degrees of freedom of a signal, bound in frequency to W and of time duration T, is roughly 2WT + 1. Again we see the result that (excluding the +1 DOF) there are twice as many degrees of freedom as the area in phase space that the signal occupies.

If we attempt to reconcile this result with the equation $I = 2\varepsilon\tau/\hbar$, we can find the density of information per degree of freedom. For an area of size WT Gabor and Slepian, Pollak, and Landau tells us that there should be 2WT degrees of freedom. Moreover, $I = 2\varepsilon\tau/\hbar$ tells us there should be 4πWT natural units of information. It appears that the degree of freedom that Gabor defined has 2π natural units of information per degree. This is precisely what happens as we will see below.

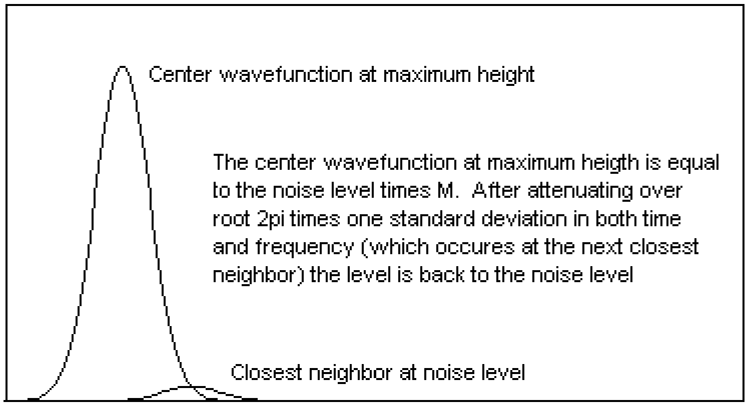
We can explain the 2π by describing a model of how information is stored in the Gaussian basis. The model simply covers phasespace by offset and overlapping Gaussian wavepackets. Assume each Gaussian has a magnitude that is independent of the other wavepackets and takes on a value that is a certain amount greater than the background noise level. We will call the maximum magnitude M and solve for it below.

Our argument will be similar to the argument that Brillouin [28] made when deciding how long to wait for a signal to decay before determining if the signal was from the current time period or from a transient left over from the previous time period. In this case we have overlapping Gaussians that are separated by $(2\pi)^{1/2}\Delta t$ in the time domain and by $(2\pi)^{1/2}\Delta f$ in the frequency domain. If we require that the maximum strength of the signal at its neighbor's site must not be greater than the height of the smallest local signal (the background noise) then we can get the range of what the magnitude of each Gaussian wavepacket is.

The two dimensional probability distribution that represents the conjugate pair of Gaussian wave packets, centered around t=0, f=0 is P(t,f).

$$P(t,f) = \psi^*(t)\psi(t)\cdot\Psi^*(f)\Psi(f)$$

$$P(t,f) = 2\cdot\exp[-t^2/2(\Delta t)^2 - f^2/2(\Delta f)^2]$$



We will assume that the magnitude of each wavepacket has to be at least one. Thus a neighboring site that can take on the highest possible value, call it M, must decay fast enough so that M times the wavepacket, that decays over the distance between neighbors, must be no greater than one.

Thus to find M, we say that the conjugate wavepacket at t=0, f=0 has a height of M times P(0,0). After P(t,f) decays over the range $t=(2\pi)^{½}\Delta t$ and $f=(2\pi)^{½}\Delta f$ (the distance between two neighbors) the total height can not be greater than the height of the smallest possible value of the next conjugate probability distribution, which is then centered around $t=(2\pi)^{½}\Delta t$ and $f=(2\pi)^{½}\Delta f$.

$$M\cdot P((2\pi)^{½}\Delta t,(2\pi)^{½}\Delta f) = P(0,0)$$

$$M\cdot 2\cdot\exp[-2\pi] = 2$$

$$M=\exp[2\pi]$$

We can now see where the 2π natural units per degree of freedom come in. If each Gaussian wavepacket has a range of magnitude equal to M, and say that all values of M are equally likely, we can get the information content by taking the ln of M. Thus the information in each degree of freedom associated with the height of Gaussian wave packets, that are separated by $t=(2\pi)^{½}\Delta t$ and $f=(2\pi)^{½}\Delta f$, is equal to 2π natural units.

If however we were to cover the space with Gaussians that were separated by $t=\Delta t$ and $f=\Delta f$, using the same analysis, we would get M=exp(1). This would imply an information content for each degree of freedom equal to 1 natural unit, just as our analysis of the sine modulated Gaussian describes. Thus the information content in a degree of freedom will be dependent upon the density of the covering space used to map each degree of freedom, however the information density per Gaussian times the quantity of Gaussians will be constant for any given area and we are left with $I = 2\varepsilon\tau/\hbar$.

Appendix D (Numerical evaluation of the self information in a sine wave): Matlab was used to numerically integrate the self information in a Gaussian spread impulse pair. We find that the results are not sensitive to the width of the impulse pair and thus we conclude that in the limit when (Δf) goes to zero (ie the impulse pair) and when (Δt) goes to infinity (ie the infinite sine wave) the sum of the differential entropy of both domains will be equal to ln(2).

Sum of differential entropy from both domains minus ln(2)		Size of differential (sec,Hz)			
		dt,df=.01	dt,df=.001	dt,df=.0001	dt,df=.00001
Angle of sine wave (radians)	Π	-1.5 e-4	-1.5 e-7	-1.5 e-10	-1.5 e-12
	π/2	1.2 e-4	-1.5 e-7	-1.5 e-10	-0.65 e-12
	π/3	0.69 e-4	0.68 e-7	-0.68 e-10	-0.38 e-12
	π/4	0.14 e-4	1.1 e-7	-1.5 e-10	-0.49 e-12
	π/5	-1.5 e-4	-1.5 e-7	-1.5 e-10	-0.70 e-12
	π/6	-0.51 e-4	0.67 e-7	-0.68 e-10	-0.70 e-12
	π/7	1.1 e-4	0.40 e-7	-0.71 e-10	-0.67 e-12
	π/8	-0.85 e-4	0.14 e-7	1.2 e-10	-0.92 e-12
	π/9	-0.07 e-4	-0.07 e-7	-0.07 e-10	-0.92 e-12
	0	-1.5 e-4	-1.5 e-7	-1.5 e-10	-1.5 e-12

Sum of differential entropy from both domains minus ln(2)		Size of differential (sec,Hz)		
		dt,df=.01	dt,df=.001	dt,df=.0001
Width (Δt) (sec)	1/sqrt(4π)	-1.5 e-4	-0.75 e-7	-2.3 e-10
	1	1.2 e-4	-1.5 e-7	-1.5 e-10
	10	0.021 e-4	-1.5 e-7	-1.6 e-10
	25	-1.5*	-1.5 e-7	-1.6 e-10
* high error comes from width Δf=Hz/(100π) being less than differential, df = Hz/100				

Above are two tables which present the calculated value of the differential entropy for different values of the differential, dt, and 1) the offset angle of the sine wave, and 2) the width of the wavepacket (Δt).

As you can see the numerical evaluation of the sum of the differential entropy of both domains approaches ln(2) as the differential approaches zero. Notice that for any given size of differential, the error away from ln(2) is insensitive to either the angle of the sine wave, or the width of the wavepacket. The angle tells us that the result is true for either a sine wave or a cosine wave (or any combination). The width tells us two things. First by a scale factor one can show that since the result is insensitive to the width of the wavepacket, it is also insensitive to the frequency of the sine wave. Second, extrapolating the insensitivity to width to any value of Δt, even extremely large values, we can claim that the sum of the self information in both domains is ln(2) for an infinite value of Δt, or for a pure sine wave.

It should be noted that there is a difference in the self-information between a Gaussian spread sine wave (self-information equals ln(2)) and a sine modulated Gaussian (self-information equals 1). The first was generated by convolving a pure sine wavefunction with a Gaussian wave function and then each domain was magnitude squared to arrive at the Gaussian spread sine wave distributions. Because the convolution happened in the wave function level, we need to calculate the self-information using numerical methods of the resulting distributions. A sine modulated Gaussian on the other hand was generated by convolving an infinite sine wave distribution with a Gaussian distribution. In this case since both are independent and have their own self-information (self-information of pure sine distribution equals ln(2) and self-information of Gaussian distribution equals ln(e/2)), and the convolution of the two results in a self-information that is the sum of both.

References:

1. Ronald N. Bracewell, The Fourier Transform and Its Applications 2nd ed. McGraw-Hill Inc., New York, NY, 1986 Ch 8
2. Thomas M. Cover & Joy A. Thomas, Elements of Information Theory, John Wiley & Sons Inc., New York, NY, 1991 Ch 10
3. Hirshman, I. I., "A Note on Entropy," *American J. of Mathematics*, January 1957, **79**, No 1, p. 152
4. Feynman, Lectures on Physics, Addison-Wesley Publishing, Reading Massachusetts, 1965 Vol. I Ch 41
5. ibid, Vol. III Ch 19
6. Albert T. Fromhold, Jr., Quantum Mechanics for Applied Physics and Engineering, Dover Publishing Inc., New York, NY, 1981 Ch. 1
7. Sarah Nichols and William Wooters, "Between Entropy and Subentropy," *Quantum Information and Computation*, Vol. **3**, No 1, p. 1
8. Other von neuman
9. C. E. Shannon, The Mathematical Theory of Communication, The University of Illinois Press, Urbana Il., 1964
10. John Von Neumann, Mathematical Foundations of Quantum Mechanics, Princeton University Press, Princeton, NJ, 1955 Ch V.3
11. Stephen W. Hawking, The Theory of Everything: The origin and fare of the universe, new edition, Phoenix Books, Beverly Hills, CA 2007, Ch 5
12. R. Shankar, Principles of Quantum Mechanics 2nd ed, Plenum Press, New York, NY 1994 Ch 21
13. H. Nyquist, "Certain factors affecting Telegraph speed," *Bell Systems Tech. J.*, 1924, **3**, pp. 324.
14. R. V. L. Hartley, "Transmission of Information," *Bell Systems Tech. J.*, 1928, **7**, pp.535.
15. Feynman, Lectures on Physics, Addison-Wesley Publishing, Reading Massachusetts, 1965 Vol. III Ch 2
16. John Haller & Stephen Hawking, private conversation, Stanford University, Stanford, CA 1998
17. Seth Lloyd, Programming the Universe, Knopf, New York, NY, 2006, ch. 3
18. R. Shankar, Principles of Quantum Mechanics 2nd ed, Plenum Press, New York, NY, 1994 Ch 14
19. http://en.wikipedia.org/wiki/Interpretation_of_quantum_mechanics
20. David Bohm, Quantum Theory, Dover Publications, New York, 1951, 1989
21. J. S. Bell, "On the Einstein Podolsky Rosen Paradox," Physics ,1964, **1**, 195
22. A. Einstein, B. Podolsky, and N. Rosen, "Can quantum-mechanical description of physical reality be considered complete?" Phys. Rev., 1935, **47** 777
23. Niels Bohr, The philosophical Writings of Niels Bohr, "Discussion with Einstein on Epistemological Problems in Atomic Physics," Ox Bow Press, Woodbridge Connecticut, 1987
24. Bernard Schutz, A first course in general realativity, Cambridge University Press, Cambridge England, 1985 Ch 11
25. Albert Einstein, The Meaning of Relativity, 5th Edition, Princeton University Press, Princeton NJ 1956
26. J. B. Hartle & S. W. Hawking, "Path-integral derivation of black-hole radiance," Phys. Rev. D, 1976, **13** 2188

27. Reif, Fundamentals of Statistical and Thermal Physics, Mc Graw Hill, Boston, MA 1965 Ch 5
28. Leon Brillouin, Science and Information Theory, 2nd Edition, Academic Press Inc., 1962, pp. 75,76
29. R. Shankar, Principles of Quantum Mechanics 2nd ed, Plenum Press, New York, NY 1994 Ch 16
30. Thomas M. Cover & Joy A. Thomas, Elements of Information Theory, John Wiley & Sons, Inc., New York, NY, 1991, pp. 160
31. William C. Elmore & Mark A. Heald, Physics of Waves, McGraw-Hill, Inc. 1969, pp. 351-354
32. ibid, pp. 365-367
33. Thomas M. Cover & Joy A. Thomas, Elements of Information Theory, John Wiley & Sons Inc., New York, NY, 1991, pp. 486
34. Ronald N. Bracewell, The Fourier Transform and Its Applications 2nd ed. McGraw-Hill Inc., New York, NY, 1986, Ch 5
35. D. Gabor, "Theory of Communication," *J. Inst. Elec. Engrs.*, 1946, **93**, III p.429
36. D. Slepian and H. O. Pollak, "Prolate spheroidal wave functions, Fourier analysis and uncertainty: Part I" *Bell Sys. Tech. Journal*, 40: 43-64, 1961
37. H. J. Landau and H. O. Pollak, "Prolate spheroidal wave functions, Fourier analysis and uncertainty: Part II" *Bell Sys. Tech. Journal*, 40: 65-84, 1961
38. H. J. Landau and H. O. Pollak, "Prolate spheroidal wave functions, Fourier analysis and uncertainty: Part III" *Bell Sys. Tech. Journal*, 41: 1295-1336, 1962
39. Y. Yamamoto and H. A. Haus, "Preparation, measurement and information capacity of optical quantum states," Review of Modern Physics, October 1986, Vol. 58, No. 4
40. C. Soo and C. C. Y. Lin, "Wigner rotations, Bell states, and Lorentz invariance of entanglement and von Neumann entropy," *Int. J. Quantum Info*. 2, 2004 183-200
41. C. Moustakidis and S.E. Massend, "The dependence of information entropy of uniform Fermi systems on correlations and thermal effects," *Phys. Rev. B*71 2005, 045102
42. R. P. Sagar and N. L. Guevara, "Shannon entropies of atomic structure factors, off-diagonal order and electron correlation," *Mol. Phys*. 105 (2007) 1359-1368
43. John Haller, "The Entropy Rate of Thermal Diffusion", 2007 http://arxiv.org/abs/math/0106215
44. Niels Bohr, "Discussion with Einstein on Epistemological Problems in Atomic Physics," Albert Einstein: Philospher-Scientist Library of Living Philosophers, Inc., Tudor Publishing, New York, NY 1951 pg 201-241
45. R. Shankar, Principles of Quantum Mechanics 2nd ed, Plenum Press, New York, NY 1994 Ch 9