10 Quantized Degrees-of-Freedom in a Continuous Signal

We have now encountered several theorems expressing the idea that even though a signal is continuous and dense in time (i.e. the value of the signal is defined at each real-valued moment in time), nevertheless a finite and countable set of discrete numbers suffices to describe it completely, and thus to reconstruct it, provided that its frequency bandwidth is limited.

Such theorems may seem counter-intuitive at first: How could a finite sequence of numbers, at discrete intervals, capture exhaustively the continuous and uncountable stream of numbers that represent all the values taken by a signal over some interval of time?

In general terms, the reason is that bandlimited continuous functions are *not as free* to vary as they might at first seem. Consequently, specifying their values at only certain points, suffices to *determine* their values at all other points.

Three examples that we have already seen are:

- **Nyquist's Sampling Theorem:** If a signal $f(x)$ is strictly bandlimited so that it contains no frequency components higher than W, i.e. its Fourier Transform $F(k)$ satisfies the condition

$$F(k) = 0 \text{ for } |k| > W \tag{1}$$

then $f(x)$ is completely determined just by sampling its values at a rate of at least $2W$. The signal $f(x)$ can be exactly recovered by using each sampled value to fix the amplitude of a $\text{sinc}(x)$ function,

$$\text{sinc}(x) = \frac{\sin(\pi x)}{\pi x} \tag{2}$$

whose width is scaled by the bandwidth parameter W and whose location corresponds to each of the sample points. The continuous signal $f(x)$ can be perfectly recovered from its discrete samples $f_n(\frac{n\pi}{W})$ just by adding all of those displaced $\text{sinc}(x)$ functions together, with their amplitudes equal to the samples taken:

$$f(x) = \sum_n f_n\left(\frac{n\pi}{W}\right)\frac{\sin(Wx - n\pi)}{(Wx - n\pi)} \tag{3}$$

Thus we see that any signal that is limited in its bandwidth to W, during some duration T has at most $2WT$ degrees-of-freedom. It can be completely specified by just $2WT$ real numbers (Nyquist, 1911; R V Hartley, 1928).

- **Logan's Theorem:** If a signal $f(x)$ is strictly bandlimited to one octave or less, so that the highest frequency component it contains is no greater than twice the lowest frequency component it contains

$$k_{max} \leq 2k_{min} \tag{4}$$

i.e. $F(k)$ the Fourier Transform of $f(x)$ obeys

$$F(|k| > k_{max} = 2k_{min}) = 0 \tag{5}$$

and

$$F(|k| < k_{min}) = 0 \tag{6}$$

and if it is also true that the signal $f(x)$ contains no complex zeroes in common with its Hilbert Transform (too complicated to explain here, but this constraint serves to exclude families of signals which are merely amplitude-modulated versions of each other), then the original signal $f(x)$ can be perfectly recovered (up to an amplitude scale constant) merely from knowledge of the set $\{x_i\}$ of zero-crossings of $f(x)$ alone:

$$\{x_i\} \text{ such that } f(x_i) = 0 \tag{7}$$

Comments:
(1) This is a very complicated, surprising, and recent result (W F Logan, 1977).

(2) Only an existence theorem has been proven. There is so far no stable constructive algorithm for actually making this work – i.e. no known procedure that can actually recover $f(x)$ in all cases, within a scale factor, from the mere knowledge of its zero-crossings $f(x) = 0$; only the existence of such algorithms is proven.

(3) The "Hilbert Transform" constraint (where the Hilbert Transform of a signal is obtained by convolving it with a hyperbola, $h(x) = 1/x$, or equivalently by shifting the phase of the positive frequency components of the signal $f(x)$ by $+\pi/2$ and shifting the phase of its negative frequency components by $-\pi/2$), serves to exclude ensembles of signals such as $a(x)\sin(\omega x)$ where $a(x)$ is a purely positive function $a(x) > 0$. Clearly $a(x)$ modulates the amplitudes of such signals, but it could not change any of their zero-crossings, which would always still occur at $x = 0, \frac{\pi}{\omega}, \frac{2\pi}{\omega}, \frac{3\pi}{\omega}, \ldots$, and so such signals could not be uniquely represented by their zero-crossings.

(4) It is very difficult to see how to generalize Logan's Theorem to two-dimensional signals (such as images). In part this is because the zero-crossings of two-dimensional functions are non-denumerable (uncountable): they form continuous "snakes," rather than a discrete and countable set of points. Also, it is not clear whether the one-octave bandlimiting constraint should be isotropic (the same in all directions), in which case the projection of the signal's spectrum onto either frequency axis is really low-pass rather than bandpass; or anisotropic, in which case the projection onto both frequency axes may be strictly bandpass but the different directions are treated differently.

(5) Logan's Theorem has been proposed as a significant part of a "brain theory" by David Marr and Tomaso Poggio, for how the brain's visual cortex processes and interprets retinal image information. The zero-crossings of bandpass-filtered retinal images constitute edge information within the image.

- **The Information Diagram:** The *Similarity Theorem* of Fourier Analysis asserts that if a function becomes narrower in one domain by a factor a, it necessarily becomes broader by the same factor a in the other domain:

$$f(x) \longrightarrow F(k) \tag{8}$$

$$f(ax) \longrightarrow |\frac{1}{a}|F(\frac{k}{a}| \tag{9}$$

The Hungarian Nobel-Laureate Dennis Gabor took this principle further with great insight and with implications that are still revolutionizing the field of signal processing (based upon wavelets), by noting that an *Information Diagram* representation of signals in a plane defined by the axes of time and frequency is fundamentally *quantized*. There is an irreducible, minimal, area that any signal can possibly occupy in this plane. Its uncertainty (or spread) in frequency, times its uncertainty (or duration) in time, has an inescapable lower bound.

11 Gabor-Heisenberg-Weyl Uncertainty Relation. "Logons."

11.1 The Uncertainty Principle

If we define the "effective support" of a function $f(x)$ by its normalized variance, or the normalized second-moment:

$$(\Delta x)^2 = \frac{\int_{-\infty}^{+\infty} f(x)f^*(x)(x-\mu)^2 dx}{\int_{-\infty}^{+\infty} f(x)f^*(x)dx} \tag{10}$$

where μ is the mean value, or normalized first-moment, of the function:

$$\mu = \frac{\int_{-\infty}^{+\infty} xf(x)f^*(x)dx}{\int_{-\infty}^{+\infty} f(x)f^*(x)dx} \tag{11}$$

and if we similarly define the effective support of the Fourier Transform $F(k)$ of the function by its normalized variance in the Fourier domain:

$$(\Delta k)^2 = \frac{\int_{-\infty}^{+\infty} F(k)F^*(k)(k-k_0)^2 dk}{\int_{-\infty}^{+\infty} F(k)F^*(k)dk} \tag{12}$$

where k_0 is the mean value, or normalized first-moment, of the Fourier transform $F(k)$:

$$k_0 = \frac{\int_{-\infty}^{+\infty} kF(k)F^*(k)dk}{\int_{-\infty}^{+\infty} F(k)F^*(k)dk} \tag{13}$$

then it can be proven (by Schwartz Inequality arguments) that there exists a fundamental lower bound on the product of these two "spreads," *regardless* of the function $f(x)$:

$$\boxed{(\Delta x)(\Delta k) \geq \tfrac{1}{4\pi}} \tag{14}$$

This is the famous Gabor-Heisenberg-Weyl Uncertainty Principle. Mathematically it is exactly identical to the uncertainty relation in quantum physics, where (Δx) would be interpreted as the position of an electron or other particle, and (Δk) would be interpreted as its momentum or deBroglie wavelength. We see that this is not just a property of nature, but more abstractly a property of all functions and their Fourier Transforms. It is thus a still further respect in which the information in continuous signals is quantized, since the minimal area they can occupy in the Information Diagram has an irreducible lower bound.

11.2 Gabor "Logons"

Dennis Gabor named such minimal areas "logons" from the Greek word for information, or order: lōgos. He thus established that the Information Diagram for any continuous signal can only contain a fixed number of information "quanta." Each such quantum constitutes an independent datum, and their total number within a region of the Information Diagram represents the number of independent degrees-of-freedom enjoyed by the signal.

The unique family of signals that actually achieve the lower bound in the Gabor-Heisenberg-Weyl Uncertainty Relation are the complex exponentials multiplied by Gaussians. These are sometimes referred to as "Gabor wavelets:"

$$f(x) = e^{-(x-x_0)^2/a^2} e^{-ik_0(x-x_0)} \tag{15}$$

localized at "epoch" x_0, modulated by frequency k_0, and with size or spread constant a. It is noteworthy that such wavelets have Fourier Transforms $F(k)$ with exactly the same functional form, but with their parameters merely interchanged or inverted:

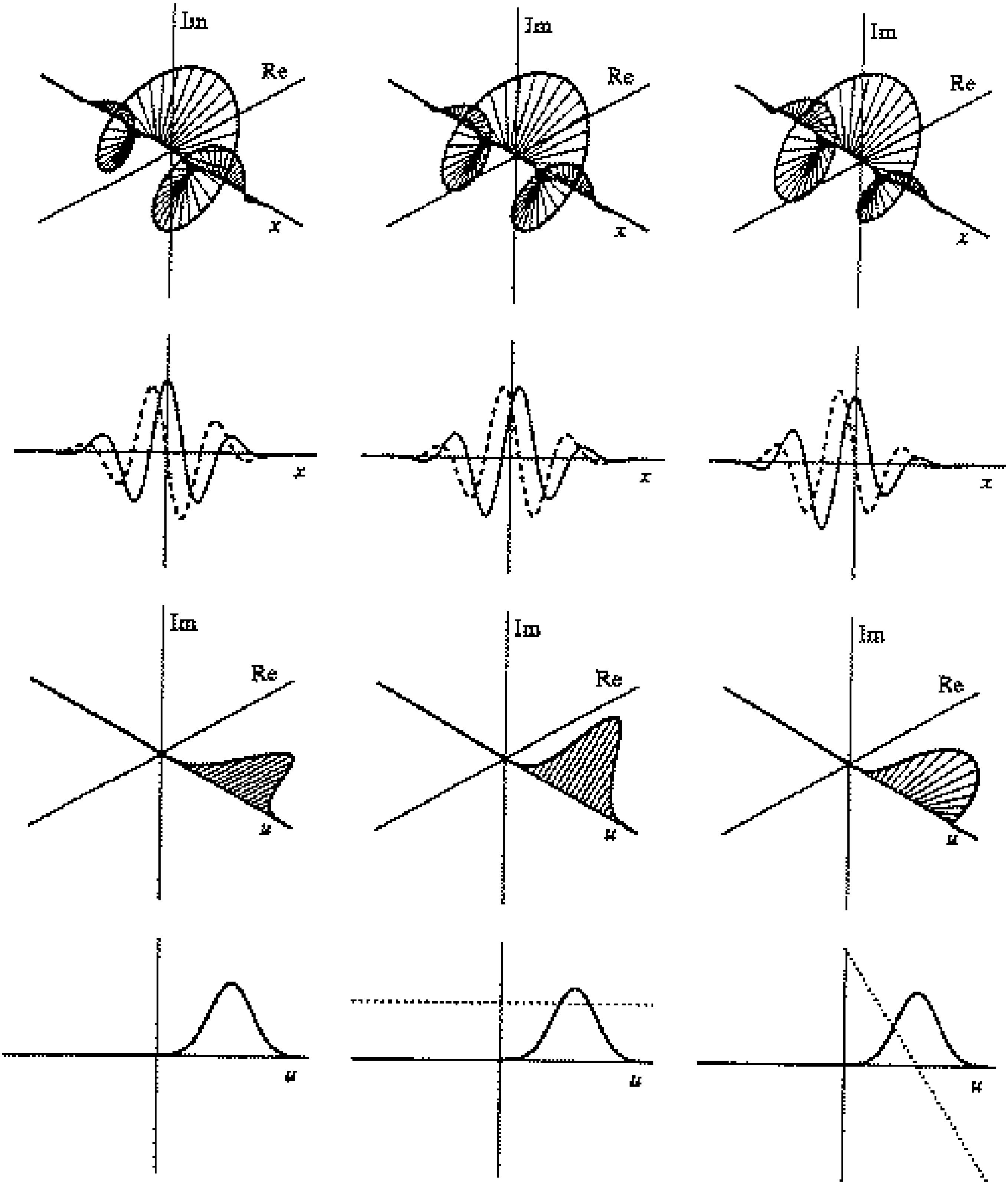
$$F(k) = e^{-(k-k_0)^2 a^2} e^{ix_0(k-k_0)} \tag{16}$$

Note that in the case of a wavelet (or wave-packet) centered on $x_0 = 0$, its Fourier Transform is simply a Gaussian centered at the modulation frequency k_0, and whose size is $1/a$, the reciprocal of the wavelet's space constant.

Because of the optimality of such wavelets under the Uncertainty Principle, Gabor (1946) proposed using them as an expansion basis to represent signals. In particular, he wanted them to be used in broadcast telecommunications for encoding continuous-time information. He called them the "elementary functions" for a signal. Unfortunately, because such functions are mutually <u>non-orthogonal</u>, it is very difficult to obtain the actual coefficients needed as weights on the elementary functions in order to expand a given signal in this basis. The first constructive method for finding such "Gabor coefficients" was developed in 1981 by the Dutch physicist Martin Bastiaans, using a dual basis and a complicated non-local infinite series.

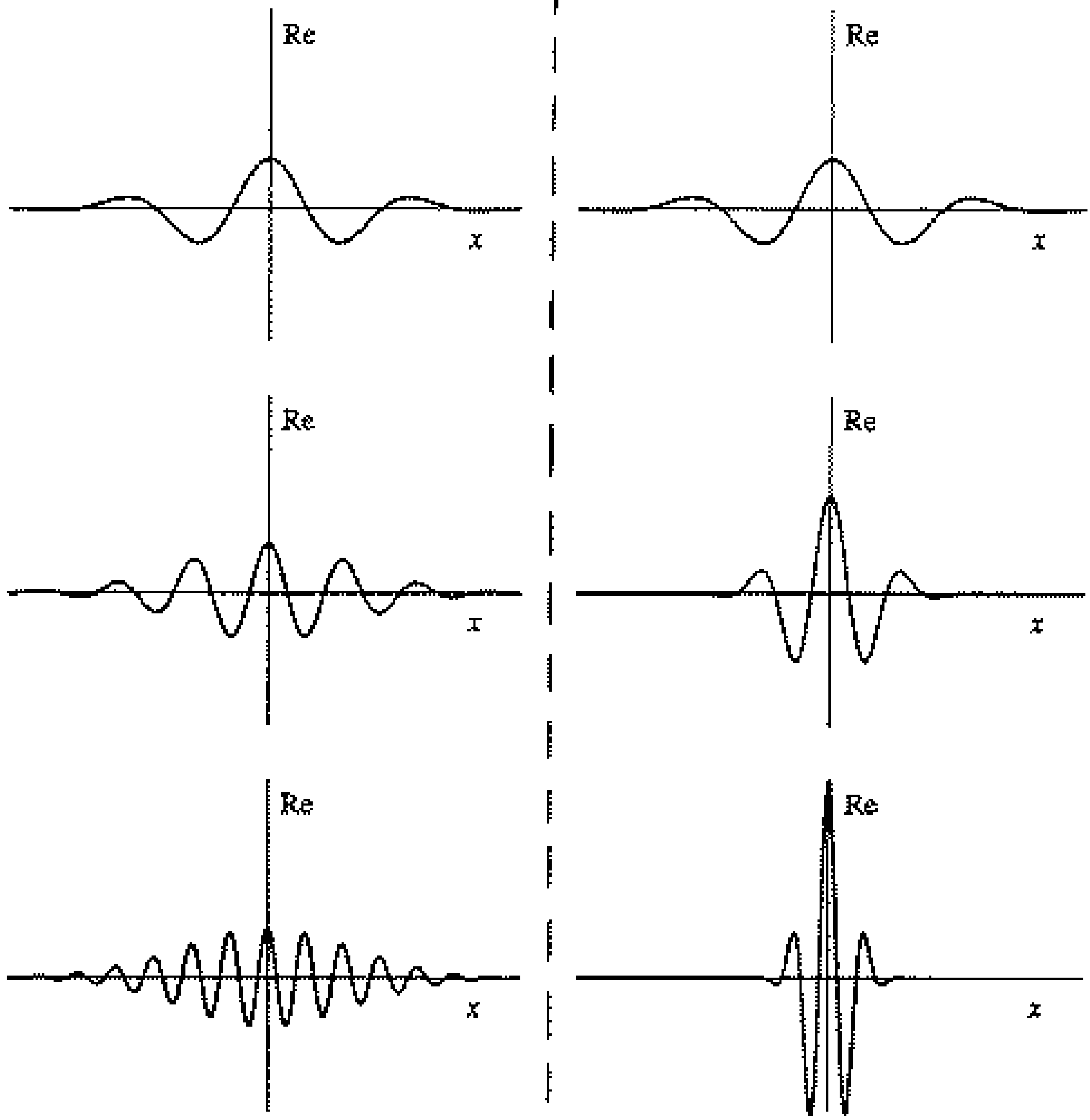
The following diagrams show the behaviour of Gabor elementary functions both as complex wavelets, their separate real and imaginary parts, and their Fourier transforms. When a family of such functions are parameterized to be self-similar, i.e. they are dilates and translates of each other so that they all have a common template ("mother" and "daughter"), then they constitute a (non-orthogonal) *wavelet basis.* Today it is known that an infinite class of wavelets exist which can be used as the expansion basis for signals. Because of the self-similarity property, this amounts to representing or analyzing a signal at different <u>scales</u>. This general field of investigation is called *multi-resolution analysis.*

Complex-valued Gabor wavelets ("phasor logons")

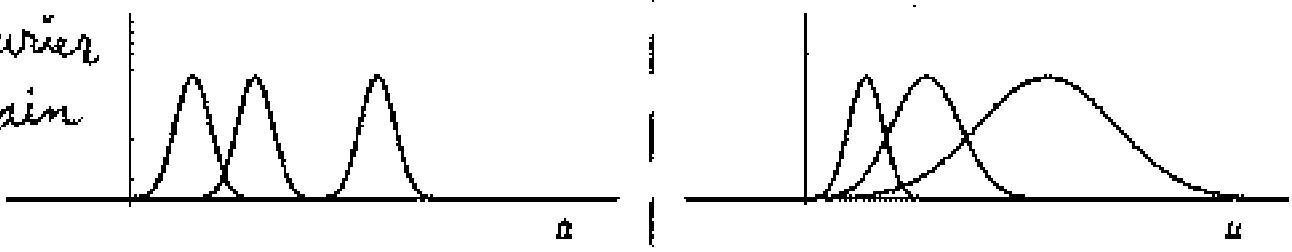


Constant-window Gabor wavelets

Self-similar Gabor wavelets



The Fourier domain



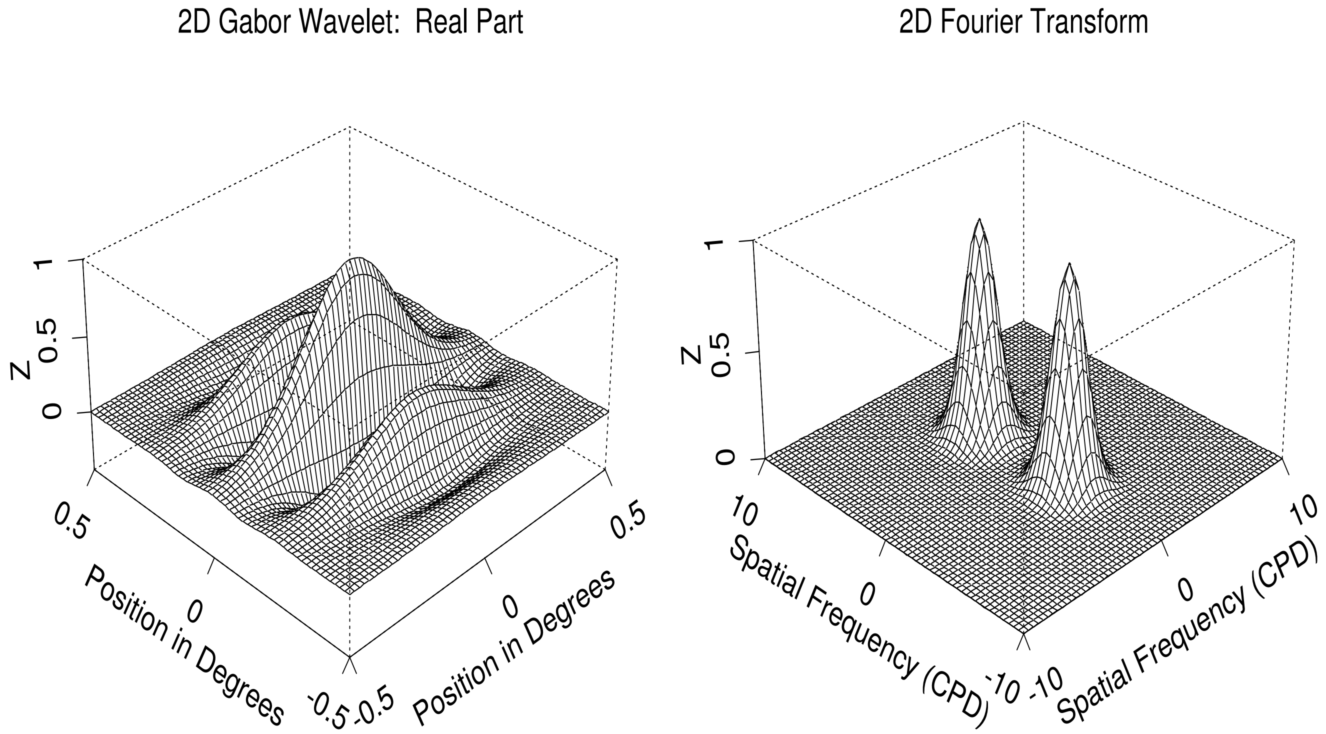


Figure 1: The real part of a 2-D Gabor wavelet, and its 2-D Fourier transform.

11.3 Generalization to Two Dimensional Signals

An effective strategy for extracting both coherent and incoherent image structure is the computation of two-dimensional Gabor coefficients for the image. This family of 2-D filters were originally proposed as a framework for understanding the orientation-selective and spatial-frequency-selective receptive field properties of neurons in the brain's visual cortex, and as useful operators for practical image analysis problems. These 2-D filters are conjointly optimal in providing the maximum possible resolution both for information about the spatial frequency and orientation of image structure (in a sense "what"), simultaneously with information about 2-D position ("where"). The 2-D Gabor filter family uniquely achieves the theoretical lower bound for joint uncertainty over these four variables, as dictated by the inescapable Uncertainty Principle when generalized to four-dimensional space.

These properties are particularly useful for texture analysis because of the 2-D spectral specificity of texture as well as its variation with 2-D spatial position. A rapid method for obtaining the required coefficients on these elementary expansion functions for the purpose of representing any image completely by its "2-D Gabor Transform," despite the non-orthogonality of the expansion basis, is possible through the use of a relaxation neural network. A large and growing literature now exists on the efficient use of this non-orthogonal expansion basis and its applications.

Two-dimensional Gabor filters over the image domain (x, y) have the functional form

$$f(x,y) = e^{-\left[(x-x_0)^2/\alpha^2+(y-y_0)^2/\beta^2\right]} e^{-i[u_0(x-x_0)+v_0(y-y_0)]} \tag{17}$$

where (x_0, y_0) specify position in the image, (α, β) specify effective width and length, and (u_0, v_0) specify modulation, which has spatial frequency $\omega_0 = \sqrt{u_0^2 + v_0^2}$ and direction $\theta_0 = \arctan(v_0/u_0)$. (A further degree-of-freedom not included above is the relative orientation of the elliptic Gaussian envelope, which creates cross-terms in xy.) The 2-D Fourier transform

$F(u,v)$ of a 2-D Gabor filter has exactly the same functional form, with parameters just interchanged or inverted:

$$F(u,v) = e^{-\left[(u-u_0)^2\alpha^2+(v-v_0)^2\beta^2\right]}e^{i[x_0(u-u_0)+y_0(v-v_0)]} \tag{18}$$

The real part of one member of the 2-D Gabor filter family, centered at the origin $(x_0, y_0) = (0,0)$ and with unity aspect ratio $\beta/\alpha = 1$ is shown in the figure, together with its 2-D Fourier transform $F(u,v)$.

2-D Gabor functions can form a complete self-similar 2-D wavelet expansion basis, with the requirements of orthogonality and strictly compact support relaxed, by appropriate parameterization for dilation, rotation, and translation. If we take $\Psi(x,y)$ to be a chosen generic 2-D Gabor wavelet, then we can generate from this one member a complete self-similar family of 2-D wavelets through the generating function:

$$\Psi_{mpq\theta}(x,y) = 2^{-2m}\Psi(x',y') \tag{19}$$

where the substituted variables (x', y') incorporate dilations in size by 2^{-m}, translations in position (p,q), and rotations through orientation θ:

$$x' = 2^{-m}[x\cos(\theta) + y\sin(\theta)] - p \tag{20}$$

$$y' = 2^{-m}[-x\sin(\theta) + y\cos(\theta)] - q \tag{21}$$

It is noteworthy that as consequences of the similarity theorem, shift theorem, and modulation theorem of 2-D Fourier analysis, together with the rotation isomorphism of the 2-D Fourier transform, all of these effects of the generating function applied to a 2-D Gabor mother wavelet $\Psi(x,y) = f(x,y)$ have corresponding identical or reciprocal effects on its 2-D Fourier transform $F(u,v)$. These properties of self-similarity can be exploited when constructing efficient, compact, multi-scale codes for image structure.

11.4 Grand Unification of Domains: an *Entente Cordiale*

Until now we have viewed "the space domain" and "the Fourier domain" as somehow opposite, and incompatible, domains of representation. (Their variables are reciprocals; and the Uncertainty Principle declares that improving the resolution in either domain must reduce it in the other.) But we now can see that the "Gabor domain" of representation actually embraces and unifies both of these other two domains! To compute the representation of a signal or of data in the Gabor domain, we find its expansion in terms of elementary functions having the form

$$f(x) = e^{-ik_0x}e^{-(x-x_0)^2/a^2} \tag{22}$$

The single parameter a (the space-constant in the Gaussian term) actually builds a continuous bridge between the two domains: if the parameter a is made very large, then the second exponential above approaches 1.0, and so in the limit our expansion basis becomes

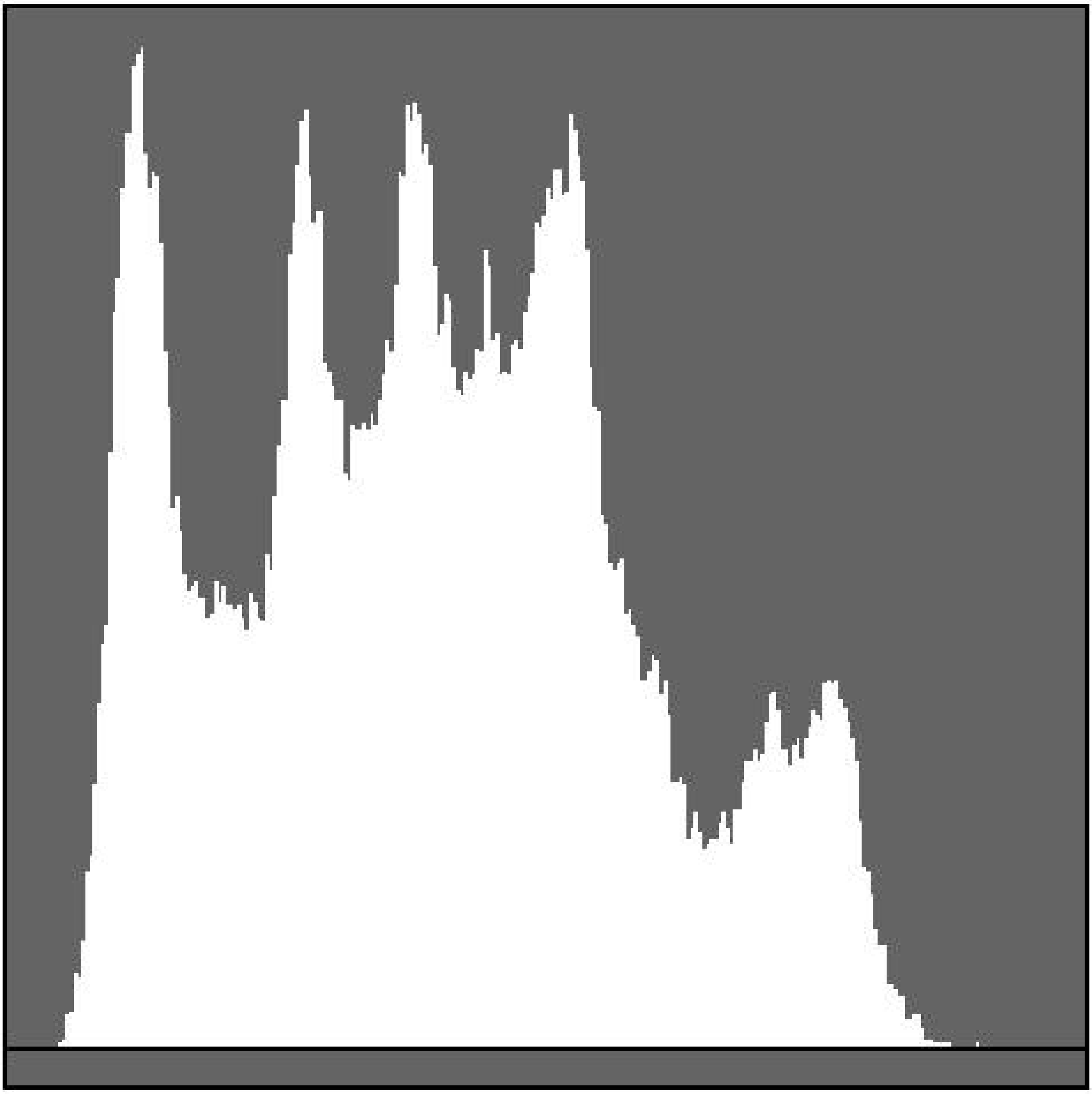
$$\lim_{a\to\infty} f(x) = e^{-ik_0x} \tag{23}$$

the ordinary Fourier basis! If the frequency parameter k_0 and the size parameter a are instead made very small, the Gaussian term becomes the approximation to a delta function at location x_o, and so our expansion basis implements pure space-domain sampling:

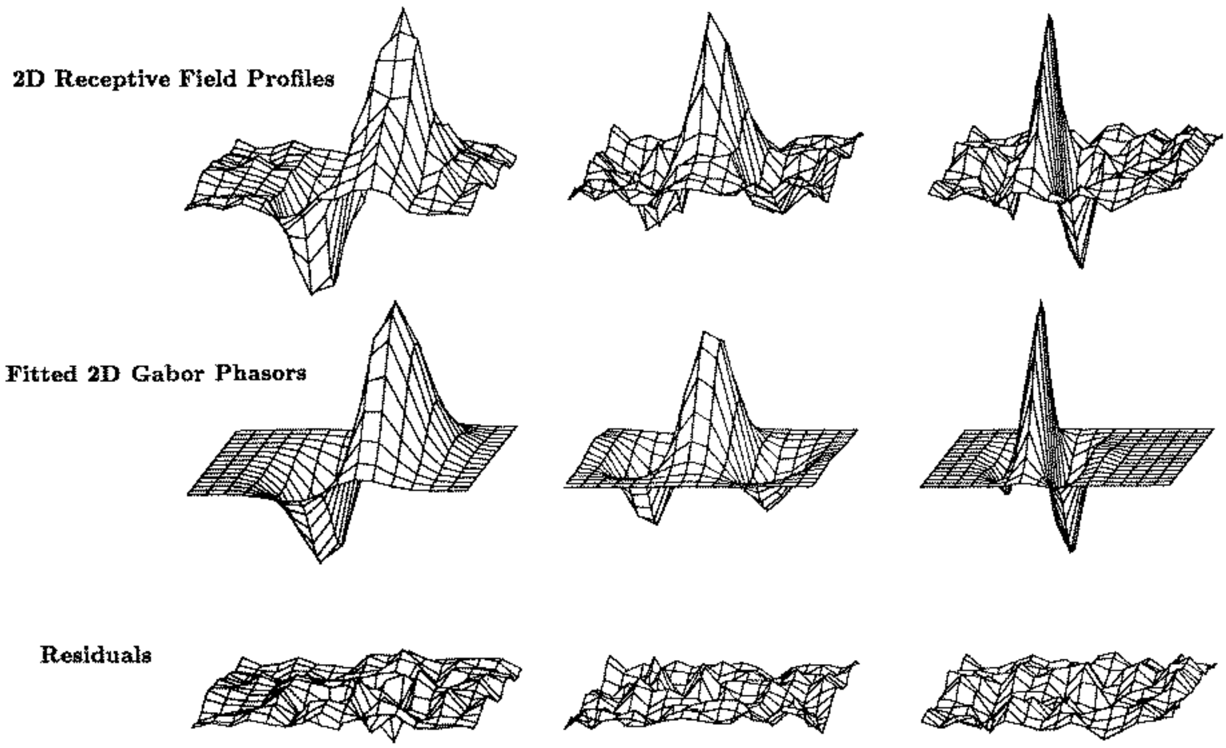
$$\lim_{k_0,a\to 0} f(x) = \delta(x - x_0) \tag{24}$$

Hence the Gabor expansion basis "contains" both domains at once. It allows us to make a continuous deformation that selects a representation lying anywhere on a one-parameter continuum between two domains that were hitherto distinct and mutually unapproachable. A new *Entente Cordiale*, indeed.

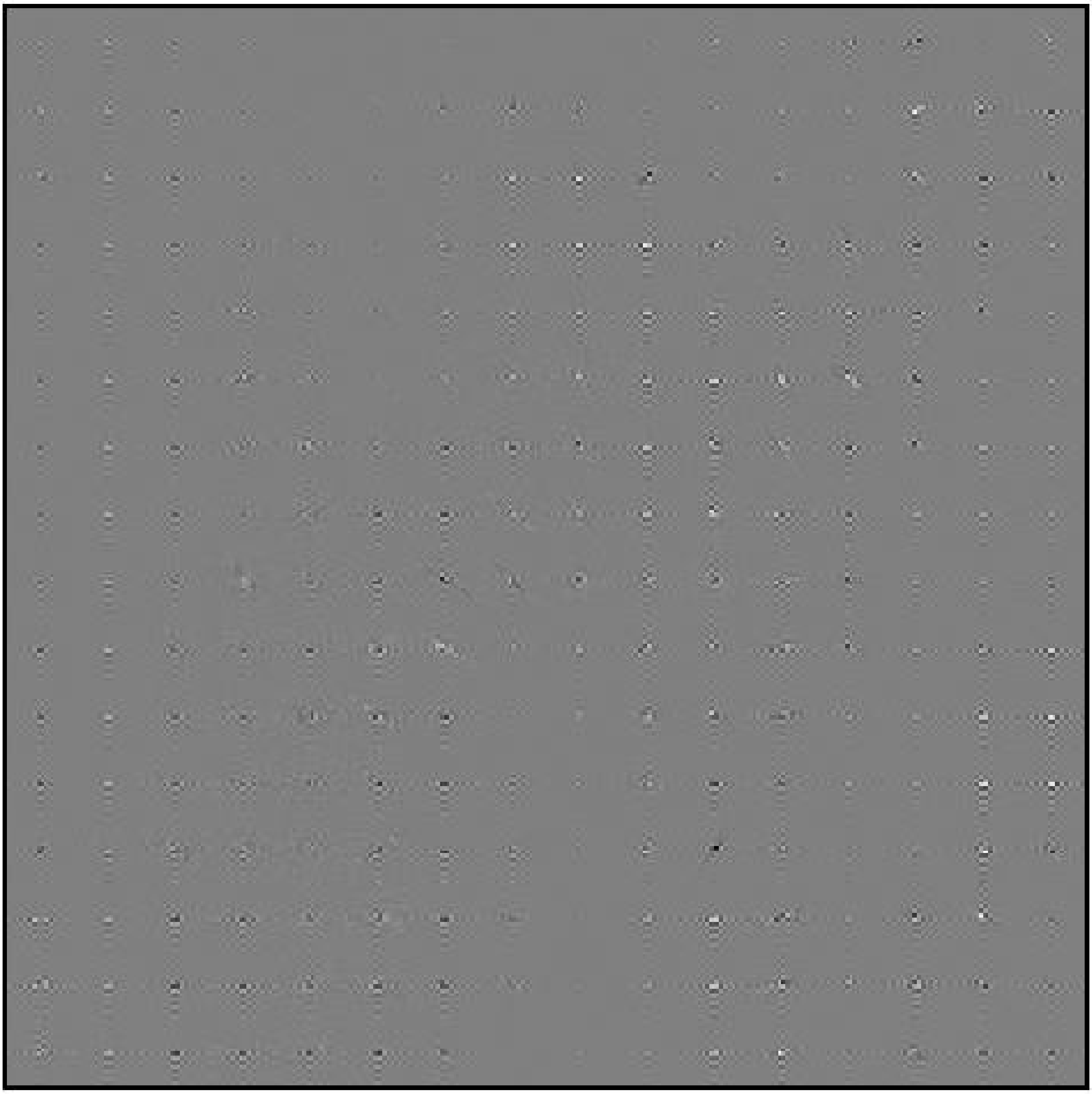
Reconstruction of Lena: 25, 100, 500, and 10,000 Two-Dimensional Gabor Wavelets



Pixel histogram of original Lena image: entropy = 7.57 bits/pixel



Empirical 2D receptive field profiles of neurones in the brain's visual cortex (top row), resembling 2D Gabor wavelets (middle row)



Complete 2D Gabor Transform of the original Lena image
(Gabor coefficients represented as grey value)

Histogram of coefficients in the complete 2D Gabor Transform of Lena: entropy = 2.55 bits/pixel

Reconstruction of Lena from her complete 2D Gabor Transform

12 Kolmogorov Complexity and Minimal Description Length

An idea of fundamental importance is the measure known as Kolmogorov complexity: the complexity of a string of data is defined as the length of the shortest binary program for computing the string. Thus the complexity is the data's "minimal description length."

It is an amazing fact that the Kolmogorov complexity K of a string is approximately equal to the entropy H of the distribution from which the string is a randomly drawn sequence. Thus Kolmogorov descriptive complexity is intimately connected with information theory, and indeed K defines the ultimate data compression. Reducing the data to a program that generates it exactly is obviously a way of compressing it; and running that program is a way of decompressing it. Any set of data can be generated by a computer program, even if (in the worst case) that program simply consists of data statements. The length of such a program defines its algorithmic complexity.

It is important to draw a clear distinction between the notions of *computational complexity* (measured by program execution time), and *algorithmic complexity* (measured by program length). Kolmogorov complexity is concerned with finding descriptions which minimize the latter. Little is known about how (in analogy with the optimal properties of Gabor's elementary logons in the 2D Information Plane) one might try to minimize simultaneously along *both* of these orthogonal axes that form a "Complexity Plane."

Most sequences of length n (where "most" considers all possible permutations of n bits) have Kolmogorov complexity K close to n. The complexity of a truly random binary sequence is as long as the sequence itself. However, it is not clear how to be certain of discovering that a given string has a much lower complexity than its length. It might be clear that the string

```
0101010101010101010101010101010101010101010101010101010101010101
```

has a complexity much less than 32 bits; indeed, its complexity is the length of the program: `Print 32 "01"s`. But consider the string

```
0110101000001001111001100110011111110011101111001100100100001000
```

which looks random and passes most tests for randomness. How could you discover that this sequence is in fact just the binary expansion for the irrational number $\sqrt{2}-1$, and that therefore it can be specified extremely concisely?

Fractals are examples of entities that look very complex but in fact are generated by very simple programs (i.e. iterations of a mapping). Therefore, the Kolmogorov complexity of fractals is nearly zero.

A sequence $x_1, x_2, x_3, ..., x_n$ of length n is said to be *algorithmically random* if its Kolmogorov complexity is at least n (i.e. the shortest possible program that can generate the sequence is a listing of the sequence itself):

$$K(x_1 x_2 x_3 ... x_n | n) \geq n \tag{25}$$

An infinite string is defined to be *incompressible* if its Kolmogorov complexity, in the limit as the string gets arbitrarily long, approaches the length n of the string itself:

$$\lim_{n\to\infty} \frac{K(x_1 x_2 x_3 ... x_n | n)}{n} = 1 \tag{26}$$

An interesting theorem, called the *Strong Law of Large Numbers for Incompressible Sequences,* asserts that the proportions of 0's and 1's in any incompressible string must be nearly equal! Moreover, any incompressible sequence must satisfy all computable statistical tests for randomness. (Otherwise, identifying the statistical test for randomness that the string failed would reduce the descriptive complexity of the string, which contradicts its incompressibility.) Therefore the algorithmic test for randomness is the ultimate test, since it includes within it all other computable tests for randomness.

13 A Short Bibliography

Cover, T M and Thomas, J A (1991) *Elements of Information Theory.* New York: Wiley.

Blahut, R E (1987) *Principles and Practice of Information Theory.* New York: Addison-Wesley.

Bracewell, R (1995) *The Fourier Transform and its Applications.* New York: McGraw-Hill.

Fourier analysis: besides the above, an excellent on-line tutorial is available at:
`http://users.ox.ac.uk/~ball0597/Fourier/`

McEliece, R J (1977) *The Theory of Information and Coding: A Mathematical Framework for Communication.* Addison-Wesley. Reprinted (1984) by Cambridge University Press in *Encyclopedia of Mathematics.*

McMullen, C W (1968) *Communication Theory Principles.* New York: MacMillan.

Shannon, C and Weaver, W (1949) *Mathematical Theory of Communication.* Urbana: University of Illinois Press.

Wiener, N (1948) *Time Series.* Cambridge: MIT Press.