Optics and Information

Of all physical phenomena, optics allows by far the highest capacity, greatest bandwidth and most effective means of information transfer. An optical fiber can transfer information at bandwidths 3-4 orders of magnitude beyond the fastest copper wire and optical imaging systems surpass the bandwidths and parallelism of ultrasonic and radio wave imaging systems by even greater margins. Given these vast information transfer capabilities, analysis of information capacity and flow often lies at the heart of optical system analysis. This note briefly reviews the quantitative concept of information and briefly describes how it arises in optical analysis. Sampling theorems, which lie at the heart of information analysis in analog systems, are discussed at the conclusion of the note.

Quantitative Measurement of Information

Information is stored in systems, transferred over channels and detected by measurements. A system might be "my crazy old aunt Beatrice, sitting on a sofa in the living room." Let's call this system S. Information from S might vary from details such as "Aunt Bea's knees are dirty" to "there is a nanogram of plutonium on the third gray hair back from Aunt Bea's left earlobe.". Leaving aside the philosophical question as to whether or not system S stores an infinite amount of information, the quantity of information about S which can be transferred over a channel or detected by a measurement is always finite. Given that this is the case, we often consider the communication channel or the measurement device to be part of system S and define a quantitative measure for the information capacity of S. A suitable definition is:

The information capacity, H, of a system S is the minimum number of bits necessary to completely specify the state of S.

The number of yes/no questions that could be answered about S is an upper bound on H. For example the answer to the question "Are Aunt Bea's knees dirty?" could convey one bit of information. The number of yes/no questions is an upper bound on H because the answers to some questions may be correlated or implicit in the nature of the system. For example, if Aunt Bea's knees are always dirty, the answer to the question may convey no information.

The definition of the information capacity makes comparisons between systems possible and makes information a universal concept. For example, if we can fully specify the state of S using one megabit, than the information capacity of S is equivalent to a one megabit computer memory and the systems themselves are in some sense equivalent. It is important to stress, however, that there is a difference between specifying the state of a system and describing the state of a system. The number of bits needed to fully describe Aunt Bea sitting on the sofa may greatly exceed the number needed to specify the state, particularly if describing Aunt Bea requires the use of poetry. To further see the distinction, consider the system "the book I checked out of the library." If the library contains 32 books, this system contains 5 bits of information, which is the number of bits needed to specify which book I checked out. On the other hand the system "a 200 page book" may contain between one and several hundred megabytes of information, depending on whether or not the book is all text or contains figures.

The library book example leads to a simple means of determining the information capacity of a system: If a system S exists with equal probability in N different states, than the information capacity, H, of S is the base 2 logarithm of N. As an example of the use of this definition to calculate information, consider the system "images in the movie Citizen Kane." Citizen Kane runs 119 minutes at 24 frames per second, so this system contains log(119*24*60) or 17.4 bits. In contrast, the system "a binary image on black and white film," contains a million bits if we assume that there are a million resolvable spots in on the film.
Information in Analog Systems

In digital systems, the quantity information is equal to the number of bits available and is thus quite precise. In analog systems, such as an imaging system, information is more difficult to quantify. Noise and measurement resolution must be considered in analog systems and play key roles in determining information capacity. In analyzing the information capacity of imaging systems, it is often desirable to separate the impact of noise and measurement resolution, which are determined by the illumination source and the detector, from the impact of diffraction and spatial apertures. Analysis of field propagation (diffraction) and boundary conditions (apertures) determines the number of electromagnetic modes which propagate through an imaging system. The temporal bandwidth and noise characteristics of the illumination source and the detector and digitization resolution on detection determine the quantity of information each mode can carry.

In analyzing the field in imaging systems, one focuses on the number of degrees of freedom. Each degree of freedom is a variable on which information can be encoded. For example, the magnitude and phase of a particular electromagnetic mode might constitute a degree of freedom and the number of available modes might constitute the number of degrees of freedom. Discussions of degrees of freedom allow one to estimate the potential information capacity of an analog system without fully considering the impact of noise and measurement. A system with N degrees of freedom in which each degree of freedom could be independently placed and measured in r different states has an information capacity of

\[ H = \log_c r \cdot N \log r \]

Of course, it is critically important that the noise and measurement processes be such that r is greater than 2. In general, r is rarely large enough to play a major role in increasing the information capacity. Since the information is linear in N and linear in the log of r, it is almost always more useful to increase N rather than r.

As mentioned above, analysis of the degrees of freedom of an optical system may be considered from an electromagnetic perspective in terms of the modes of the system. A mode is a field distribution which satisfies both the Maxwell equations and the boundary conditions for a particular geometry. Usually, one chooses to work with a set of modes which are orthogonal, which means that their inner product is zero, and which are complete, which means that any field satisfying the Maxwell equations in the geometry of interest can be described by a superposition of modes drawn from the set. In geometries where the field is confined to a limited range of space or spatial frequencies, a set of orthogonal and complete modes is generally discrete. In this case, the spatially continuous field distribution can be represented by a discrete (although potentially infinite) list of mode amplitudes.

The Sampling Theorem

Similar methods for representing continuous functions using discrete lists arise in linear systems analysis in the context of sampling theorems. Consider as an example a square integrable bandlimited function \( f(x,y) \). We focus on 2D functions because our main focus will be on fields measured across 2D planes, such as image planes. By bandlimited, we mean that the Fourier transform of \( f \) is vanishes outside a certain range. We might express \( f \) as

\[ f(x,y) = \sum_{k=-B_x}^{B_x} \sum_{l=-B_y}^{B_y} a(k,l) \exp[j2\pi(kx + ly)] \]

The Fourier transform of \( f \), \( F(k,l) \), is limited to a square region of area \( 4B_xB_y \) in the Fourier plane. A density plot of an example form of \( F(k,l) \) is sketched below:
Consider a function $F_{u,v}^p$, which is a periodic tiling of the Fourier plane with copies of $F_{u,v}$, as sketched below.

Since $F_{u,v}^p$ is a periodic function, it can be expressed as the Fourier series
Over the range of \( F_u, v, F_{\mu v} = F_p, v \). This means that

\[
f_{x,y} = \sum_{n=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} a_{nm} \exp \left[ j2\pi \left( x + vy \right) \right] \]

where

\[
sinc(x) = \frac{\sin(\pi x)}{\pi x}
\]

Considering the case \( x = \frac{n}{2B_x}, y = \frac{m}{2B_y} \), we find

\[
f_{x,y} = \sum_{n=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} a_{nm} \exp \left[ j2\pi \left( x + vy \right) \right] = a_{nm},
\]

which yields the Whittaker-Shannon sampling theorem.

As discussed above, the sampling theorem means that the continuous function \( f_{x,y} \) can be fully represented by the discrete set of samples

\[
f_{x,y} = \sum_{n=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} a_{nm} \exp \left[ j2\pi \left( x + vy \right) \right]
\]

In cases of particular interest, we focus on the value of \( f_{x,y} \) only over some finite aperture. If we leave \( f_{x,y} \) unconstrained outside this aperture, only a finite number of samples is needed to characterize the function. If the extent of the aperture of interest is \((-X,X)\) and \((-Y,Y)\), the number of samples needed is \(16B_x B_y XY\). This number of samples corresponds to the number of degrees of freedom in the system \( f_{x,y} \) over the aperture of interest.” As discussed above, the information capacity of this function is proportional to this number of degrees of freedom.