Modulation and Information Hiding in Images

Published in Proceedings of the First Information Hiding Workshop, Isaac Newton Institute, Cambridge, U.K., May 1996. Springer-Verlag Lecture Notes in Computer Science Volume 1174.

Joshua R. Smith and Barrett O. Comiskey

{jrs, elwood}@media.mit.edu
Physics and Media Group
MIT Media Lab
20 Ames Street
Cambridge, MA 02139
USA

Abstract. We use concepts from communication theory to characterize information hiding schemes: the amount of information that can be hidden, its perceptibility, and its robustness to removal can be modeled using the quantities channel capacity, signal-to-noise ratio, and jamming margin. We then introduce new information hiding schemes whose parameters can easily be adjusted to trade off capacity, imperceptibility, and robustness as required in the application. The theory indicates the most aggressive feasible parameter settings. We also introduce a technique called *predistortion* for increasing resistance to JPEG compression. Analogous tactics are presumably possible whenever a model of anticipated distortion is available.

1 Introduction

In this paper, we discuss schemes for imperceptibly encoding extra information in an image by making small modifications to large numbers of its pixels. Potential applications include copyright protection, embedded or "in-band" captioning and indexing, and secret communication.

Ideally, one would like to find a representation that satisfies the conflicting goals of not being perceivable, and being difficult to remove, accidentally or otherwise. But because these goals *do* conflict, because it is *not* possible to simultaneously maximize robustness and imperceptibility, we will introduce a framework for quantifying the tradeoffs among three conflicting figures of merit useful for characterizing information hiding schemes: (1) capacity (the number of bits that may be hidden and then recovered) (2) robustness to accidental removal, and (3) imperceptibility. We will then present new information hiding schemes that can be tailored to trade off these figures of merit as needed in the particular application. For example, capacity may be more important in a

captioning application, robustness may be most desired for copyright protection schemes, and imperceptibility might be favored in a secret communication scenario.

1.1 Information theoretic view of the problem

We view an image in which extra information has been embedded as an approximately continuous (in amplitude), two-dimensional, band-limited channel with large average noise power. The noise is the original unmodified image, which we will refer to as the *cover image,* and the signal is the set of small modifications introduced by the hider. The modifications encode the *embedded message*. We will refer to the modified, distribution image as the *stego-image,* following the convention suggested at the Information Hiding Workshop. From this point of view, any scheme for communicating over a continuous channel—that is, any modulation scheme—is a potential information hiding scheme, and concepts used to analyze these schemes, such as channel capacity, ratio of signal power to noise power, and jamming margin can be invoked to quantify the trade-offs between the amount of information that can be hidden, the visibility of that information, and its robustness to removal.

1.2 Relationship to other approaches

In our framework, it becomes obvious why *cover image escrow* hiding schemes such as those presented in [CKLS] and [BOD95] have high robustness to distortion. In cover image escrow schemes, the extractor is required to have the original unmodified cover image, so that the original cover image can be subtracted from the stego-image before extraction of the embedded message. Because the cover image is subtracted off before decoding, there is no noise due to the cover image itself; the only noise that must be resisted is the noise introduced by distortion such as compression, printing, and scanning. While the image escrow hiding schemes must respect the same information theoretic limits as ours, the noise in their case is very small, since it arises solely from distortions to the stego-image.

In our view, image escrow schemes are of limited interest because of their narrow range of practical applications. Since the embedded message can only be extracted by one who possesses the original, the embedded information cannot be accessed by the user. For example, it would not be possible for a user's web browser to extract and display a caption or "property of" warning embedded in a downloaded image. The need to identify the original image before extraction also precludes oblivious, batch extraction. One might desire a web crawler or search engine to automatically find all illegal copies of any one of the many images belonging to, say, a particular photo archive, or all images with a certain embedded caption, but this is not possible with cover image escrow schemes (at least not without invoking computer vision). Finally, even assuming that the cover image has been identified and subtracted out, the proof value of such a watermark is questionable at best, since an "original" can always be constructed a posteriori to make any image appear to contain any watermark. The only

practical application of cover image escrow schemes we have been able to identify is fingerprinting or traitor tracing[Pfi], in which many apparently identical copies of the cover image are distributed, but the owner wants to be able distinguish among them in order to identify users who have been giving away illegal copies.

The hiding methods presented in this paper are *oblivious*, meaning that the message can be read with no prior knowledge of the cover image. Other oblivious schemes have been proposed [BGM91, Cor95], but the information-theoretic limits on the problem have not been explicitly considered. We make comparisons between our hiding schemes and these other oblivious schemes later in the paper.

In the next section, we will estimate the amount of information that can be hidden (with minimal robustness) in an image as a function of signal-to-noise ratio. The bulk of the paper is a description of some new hiding schemes that fall short but are within a small constant factor of the theoretical hiding capacity. In the implementations of these schemes presented in this paper, we have chosen capacity over robustness, but we could have done otherwise. In the conclusion, we return to the discussion of modeling the trade offs between hiding capacity, perceptibility, and robustness using the quantities channel capacity, signal-to-noise, and process gain.

2 Channel Capacity

By Nyquist's theorem, the highest frequency that can be represented in our cover image is $\frac{1}{2}\frac{cycle}{pixel}$. The band of frequencies that may be represented in the image ranges from $-\frac{1}{2}\frac{cycle}{pixel}$ to $+\frac{1}{2}\frac{cycle}{pixel}$, and therefore the bandwidth W available for information hiding is $2 \times \frac{1}{2}\frac{cycle}{pixel} = 1\frac{cycle}{pixel}$.

For a channel subject to Gaussian noise, the channel capacity, which is an upper bound on the rate at which communication can reliably occur, is given by [SW49]

$$C = W \log_2(1 + \frac{S}{N})$$

Since the bandwidth W is given in units of pixel^{-1} and the base of the logarithm is 2, the channel capacity has units of bits per pixel. For some applications (particularly print) it might be desirable to specify the bandwidth in units of millimeters^{-1}, in which case the channel capacity would have units of bits per millimeter.

This formula can be rewritten to find a lower bound on the $\frac{S}{N}$ required to achieve a communication rate C given bandwidth W. Shannon proved that this lower bound is in principle tight, in the sense that there exist ideal systems capable of achieving communications rate C using only bandwidth W and signal-to-noise $\frac{S}{N}$. However, for practical systems, there is a tighter, empirically determined lower bound: given a desired communication rate C and an available bandwidth W, a message can be successfully received if the signal-to-noise ratio is at least some small *headroom factor* α above the Shannon lower bound. The

headroom α is greater than 1 and typically around 3. [She95]

$$\frac{S}{N} \geq \alpha \left(2^{\frac{C}{W}} - 1 \right)$$

In information hiding, $\frac{S}{N} < 1$, so $\log_2(1 + \frac{S}{N})$ may be approximated as $\frac{S/N}{\ln 2}$ or about $1.44\frac{S}{N}$.[She95] Thus $\frac{S}{N} \geq \frac{\alpha}{1.44}\frac{C}{W}$. So in the low signal-to-noise regime relevant to information hiding, channel capacity goes linearly with signal-to-noise.

The average noise power of our example cover image was measured to be 902 (in units of squared amplitude). For signal powers 1, 4, and 9 (amplitude^2), the channel capacity figures are 1.6×10^{-3} bits per pixel, 6.4×10^{-3} bits per pixel, and 1.4×10^{-2} bits per pixel. In an image of size 320×320, the upper bound on the number of bits that can be hidden and reliably recovered is then $320^2 C$. In our cover image of this size, then, using gain factors of 1, 2, and 3 (units of amplitude), the Shannon bound is 160 bits, 650 bits, and 1460 bits. With a headroom factor of $\alpha = 3$, we might realistically expect to hide 50, 210 or 490 bits using these signal levels.

3 Modulation Schemes

In the modulation schemes we discuss in this paper, each bit b_i is represented by some basis function ϕ_i multiplied by either positive or negative one, depending on the value of the bit. The modulated message $S(x, y)$ is added pixel-wise to the cover image $N(x, y)$ to create the stego-image $D(x, y) = S(x, y) + N(x, y)$. The modulated signal is given by

$$S(x, y) = \sum_i b_i \phi_i(x, y)$$

Our basis functions will always be chosen to be orthogonal to each other, so that embedded bits do not equivocate:

$$< \phi_i, \phi_j >= \sum_{x,y} \phi_i(x, y)\phi_j(x, y) = nG^2\delta_{ij}$$

where n is the number of pixels and G^2 is the average power per pixel of the carrier.

In the ideal case, the basis functions are also uncorrelated with (orthogonal to) the cover image N. In reality, they are not completely orthogonal to N; if they were, we could hide our signal using arbitrarily little energy, and still recover it later.

$$< \phi_i, N >= \sum_{x,y} \phi_i(x, y)N(x, y) \approx 0$$

For information hiding, basis functions that are orthogonal to typical images are needed; image coding has the opposite requirement: the ideal is a small set of basis functions that approximately spans image space. These requirements come

in to conflict when an image holding hidden information is compressed: the ideal compression scheme would not be able to represent the carriers (bases) used for hiding at all.

The basis functions used in the various schemes may be organized and compared according to properties such as total power, degree of spatial spreading (or localization), and degree of spatial frequency spreading (or localization). We will now explain and compare several new image information hiding schemes, by describing the modulation functions ϕ_i used.

3.1 Spread Spectrum Techniques

In the spectrum-spreading techniques used in RF communications[Dix94, SOSL94], signal-to-noise is traded for bandwidth: the signal energy is spread over a wide frequency band at low SNR so that it is difficult to detect, intercept, or jam. Though the total signal power may be large, the signal to noise ratio in any band is small; this makes the signal whose spectrum has been spread difficult to detect in RF communications, and, in the context of information hiding, difficult for a human to perceive. It is the fact that the signal energy resides in all frequency bands that makes spread RF signals difficult to jam, and embedded information difficult to remove from a cover image. Compression and other degradation may remove signal energy from certain parts of the spectrum, but since the energy has been distributed everywhere, some of the signal should remain. Finally, if the key used to generate the carrier is kept secret, then in the context of either ordinary communications or data hiding, it is difficult for eavesdroppers to decode the message.

Three schemes are commonly used for spectrum spreading in RF communications: direct sequence, frequency hopping, and chirp. In the first, the signal is modulated by a function that alternates pseudo-randomly between $+G$ and $-G$, at multiples of a time constant called the chiprate. In our application, the chiprate is the pixel spacing. This pseudo-random carrier contains components of all frequencies, which is why it spreads the modulated signal's energy over a large frequency band. In frequency hopping spread spectrum, the transmitter rapidly hops from one frequency to another. The pseudo-random "key" in this case is the sequence of frequencies. As we will see, this technique can also be generalized to the spatial domain. In chirp spreading, the signal is modulated by a chirp, a function whose frequency changes with time. This technique could also be used in the spatial domain, though we have not yet implemented it.

3.2 Direct-Sequence Spread Spectrum

In these schemes, the modulation function consists of a constant, integral-valued gain factor G multiplied by a pseudo-random block ϕ_i of $+1$ and -1 values. Each block ϕ_i has a distinct location in the (x, y) plane. In both versions of direct sequence spread spectrum we have considered, the blocks ϕ_i are non-overlapping (and therefore trivially orthogonal); they tile the (x, y) plane without gaps. Because distinct basis functions ϕ_i do not overlap in the x and y coordinates, we

do not need to worry about interference and can write the total power

$$P \equiv \sum_{x,y}^{X,Y} (\sum_i G b_i \phi_i(x,y))^2 = \sum_i \sum_{x,y}^{X,Y} (G b_i \phi_i(x,y))^2 = G^2 XY = nG^2$$

The definition holds in general, but the first equation only holds if the ϕ_i tile the (x,y) plane without overlaps. Non-integral values of power can be implemented by "dithering": choosing step values

$$g \in (-G), (-G+1), \ldots, (-1), (0), (1), \ldots, (G-1), (G)$$

with probabilities $p(g)$ such that the average power $G^2 = \sum_g p(g) g^2$.

The embedded image is recovered by demodulating with the original modulating function. A TRUE $(+1)$ bit appears as a positive correlation value; a FALSE (-1) bit is indicated by a negative correlation value. We have found the median of the maximum and minimum correlation values to be an effective decision threshold, though it may not be optimal. For this scheme to work, at least one value of the embedded image must be TRUE and one FALSE. In the version of direct sequence data hiding presented in [Cor95], a similar problem is avoided by including 0101 at the beginning of each line.

A more sophisticated scheme would be to use a "dual-rail" representation in which each ϕ_i is broken in two pieces and modulated with $(-1)(1)$ to represent FALSE and $(1)(-1)$ to represent TRUE. Then to recover the message, each bit can be demodulated twice, once with $(-1)(1)$ and once with $(1)(-1)$. Whichever correlation value is higher gives the bit's value. This dual rail scheme also has advantages for carrier recovery.

Bender et al.'s Patchwork algorithm[BGM91] for data hiding in images can be viewed as a form of spread spectrum in which the pseudo-random carrier is sparse (is mostly 0s) and with the constraint that its integrated amplitude be zero enforced by explicit construction, rather than enforced statistically as in ordinary spread spectrum schemes.

In the Patchwork algorithm, a sequence of random pairs of pixels is chosen. The brightness value of one member of the pair is increased, and the other decreased by the same amount, G in our terminology. This leaves the total amplitude of the image (and therefore the average amplitude) unchanged. To demodulate, they find the sum $S = \sum_{i=1}^{n} a_i - b_i$, where a_i is the first pixel of pair i, and b_i is the second pixel of pair i. Notice that because addition is commutative, the order in which the pixel pairs were chosen is irrelevant. Thus the set of pixels at which single changes are made can be viewed as the non-zero entries in a single two-dimensional carrier $\phi(x,y)$. Bender et al. always modulate this carrier with a coefficient $b = 1$, but $b = -1$ could also be used. In this case, the recovered value of s would be negative. If the same pixel is chosen twice in the original formulation of the Patchwork algorithm, the result is still a carrier $\phi(x,y)$ with definite power and bandwidth. Thus Patchwork can be viewed as a special form of spread spectrum (with extra constraints on the carrier), and evaluated quantitatively in our information-theoretic framework.

Fully Spread Version We have implemented a "fully spread" version of direct sequence spread spectrum by chosing a different pseudo-random ϕ_i for each value of i. This fully spreads the spectrum, as the second figure in the second column of Figure 2 shows. The figure shows both space and spatial frequency representations of the cover image, the modulated pseudo-random carrier, and the sum of the two, the stego-image.

To extract the embedded message (to demodulate), we must first recover the carrier phase. If the image has only been cropped and translated, this can be accomplished by a two dimensional search, which is simple but effective. The point at which the cross-correlation of the stego-image and the carrier is maximized gives the relative carrier phase. We have implemented this brute force carrier phase recovery scheme, and found it to be effective. Rotation or scaling could also be overcome with more general searches.

Once the carrier has been recovered, we project the stego-image onto each basis vector ϕ_i:

$$o_i =< D, \phi_i >= \sum_{x,y} D(x,y)\phi_i(x,y)$$

and then threshold the o_i values. We have used the median of the maximum and minimum o_i value as the threshold value. Note that for this to work, there must be at least one $b_i = -1$ and one $b_i = +1$. Above we discussed more sophisticated schemes that avoid this problem. Figure 2 shows the original input to be embedded, the demodulated signal recovered from the stego-image, the threshold value, and the recovered original input.

Tiled Version This scheme is identical to the "fully spread" scheme, except that the same pseudo-random sequence is used for each ϕ_i. The ϕ_i differ from one another only in their location in the (x, y) plane. Unlike the fully spread version, which is effectively a one-time pad, some information about the embedded icon is recoverable from the modulated carrier alone, without a priori knowledge of the unmodulated carrier. This information appears as the inhomogeneities in the spatial frequency plane of the modulated carrier visible in Figure 3. If a different icon were hidden, the inhomogeneity would look different. One advantage of the tiled scheme is that carrier recovery requires less computation, since the scale of the search is just the size of one of the ϕ_i tiles, instead of the entire (x, y) plane. Given identical transmit power, this scheme seems to be slightly more robust than the "fully spread" scheme.

These two spread spectrum techniques are resistant to JPEGing, if the modulated carrier is given enough power (or more generally, as long as the jamming margin is made high enough). With carrier recovery, the two direct sequence schemes are resistant to translation and some cropping. However, unlike the frequency hopping scheme that we will describe below, the direct sequence basis functions are fairly localized in space, so it is possible to lose some bits to cropping.

Predistortion In addition to simply increasing the signal to improve compression immunity, Figure 4 illustrates a trick, called *predistortion,* for increasing the robustness of the embedded information when it is known that the image will be, for example, JPEG compressed. We generate the pseudo-random carrier, then JPEG compress the carrier by itself (before it has been modulated by the embedded information and added to the cover image), and uncompress it before modulating. The idea is to use the compression routine to filter out in advance all the power that would otherwise be lost later in the course of compression.[1] Then the gain can be increased if necessary to compensate for the power lost to compression. The once JPEGed carrier is invariant to further JPEGing using the same quality factor (except for small numerical artifacts).[2] Figure 4 shows both the space and spatial frequency representation of the JPEG compressed carrier. Note the suppression of high spatial frequencies. Using the same power levels, we achieved error-free decoding with this scheme, but had several errors using the usual fully spread scheme without the pre-distortion of the carrier. Tricks analogous to this are probably possible whenever the information hider has a model of the type of distortion that will be applied. Note that this version of predistortion cannot be applied to our next scheme, or to the version of direct sequence spread spectrum in [Cor95], because in these schemes carriers overlap in space and therefore interfere.

3.3 Frequency Hopping Spread Spectrum

This scheme produces perceptually nice results because it does not create hard edges in the space domain. However, its computational complexity, for both encoding and decoding, is higher than that of the direct sequence schemes.

Each bit is encoded in a particular spatial frequency; which bit of the embedded message is represented by which frequency is specified by the pseudo-random key. In our trial implementation of frequency hopping spread spectrum, however, we have skipped the pseudo random key, and instead chosen a fixed block of 10 by 10 spatial frequencies, one spatial frequency for each bit. One advantage of the frequency hopping scheme over the direct sequence techniques is that each bit is fully spread spatially: the bits are not spatially localized at all. This means that the scheme is robust to cropping and translation, which only induce phase shifts.

An apparent disadvantage of the frequency hopping scheme is that because the functions overlap in the space domain, the time to compute the modulated carrier appears to be kXY, where k is the number of bits, instead of just XY,

[1] By compressing the carrier separately from the image, we are treating the JPEG algorithm as an operator that obeys a superposition principle, which it does in an approximate sense defined in the Appendix.

[2] It should be apparent from the description of JPEG compression in the Appendix that the output of the JPEG operator (or more precisely, the operator consisting of JPEG followed by inverse JPEG, which maps an image to an image) is an eigenfunction and in fact a fixed point of that operator, ignoring small numerical artifacts.

the time required for the direct sequence schemes. However, the Fast Fourier Transform (more precisely, a Fast Discrete Cosine Transform) can be used to implement this scheme, reducing the time to $XY \log_2 XY$. This is a savings if $\log_2 XY < k$. In our example, $\log_2 320 \times 320 = 16.6$ and $k = 100$, so the FFT is indeed the faster implementation.

Figure 5 illustrates the frequency hopping modulation scheme. The results, shown in figure 6, are superior to the direct sequence schemes both perceptually and in terms of robustness to accidental removal. There is little need to threshold the output of the demodulator in this case. However, encoding and decoding require significantly more computation time.

This scheme survived gentle JPEGing[3] with no predistortion, as illustrated in figure 7.[4]

A disadvantage of this scheme for some purposes is that it would be relatively easy to intentionally remove the embedded message, by applying a spatial filter of the appropriate frequency. A more secure implementation of the scheme would disperse the frequencies from one another, to make this sort of filtering operation more difficult. The main disadvantage of this scheme relative to the direct sequence schemes is that, even using the FFT, its computational complexity for encoding and decoding is greater ($XY \log XY$ rather than XY).

4 Discussion

We have suggested that information and communication theory are useful tools both for analyzing information hiding, and for creating new information hiding schemes. We showed how to estimate the signal-to-noise needed to hide a certain number of bits given bandwidth W. A shortcoming of our channel capacity estimate is that we used the capacity formula for a Gaussian channel, which is not the best model of the "noise" in a single image, as a glance at any of the frequency domain plots in the figures will reveal. The Gaussian channel has the same power at each frequency, but clearly these images do not, especially after compression. A more refined theory would use a better statistical model of the image channel, and would therefore be able to make better estimates of the signal-to-noise needed to hide a certain number of bits. This would also lead to better hiding schemes, since the signal energy could be distributed more effectively.

[3] All the JPEG compression reported here was done in Photoshop using the "high quality" setting.

[4] In fact, it is not possible to predistort in the frequency hopping scheme: because the basis functions overlap, the resulting interference pattern depends strongly on the particular values of the bits being encoded. There is no single pattern onto which we can project the stego-image to recover the embedded data; we must (naively) project it onto a sequence of vectors, or (more sophisticated) use the FFT. In either case the idea of predistortion does not apply, at least not in the same way it did in the non-overlapping direct sequence schemes.

The scheme we have called "frequency hopping" is superior perceptually, and in terms of robustness to accidental removal, to the direct sequence schemes with which we experimented. Direct sequence may be less vulnerable to intentional removal, and wins in terms of computational complexity.

Assuming that the Gaussian channel approximation discussed above is not too misleading, our capacity estimates suggest that there exist significantly better schemes than we have presented, capable of hiding several hundred bits in an image in which we hid one hundred. Hybrid modulation/coding schemes such as trellis coding are a promising route toward higher hiding densities. But better models of channel noise (the noise due to cover images themselves, plus distortion) would lead immediately to better capacity estimates, and better hiding schemes.

In all the practical examples in this paper, we have tried to hide as much information as possible using a given signal-to-noise. However, keeping signal-to-noise and bandwidth fixed, communication rate can instead be traded for robustness to jamming. The quantities known as jamming margin and processing gain in spread spectrum communication theory are helpful in capturing this notion of robustness.

Processing gain is the ratio $\frac{W}{M}$ of available bandwidth W to the bandwidth M actually needed to represent the message. Jamming margin, the useful measure of robustness, is the product of signal-to-noise and processing gain. If the actual signal-to-noise ratio is $\frac{S}{N}$, then the jamming margin or effective signal-to-noise ratio $\frac{E}{J}$ after demodulation is given by $\frac{E}{J} = \frac{W}{M}\frac{S}{N}$. So robustness may be increased either by increasing signal-to-noise (at the cost of perceptibility, as we will explain in more detail below), or by decreasing the size of the embedded message (the capacity), which increases the processing gain. For example, in the case of our direct sequence schemes, the processing gain increases when we hide fewer bits because each bit can be represented by a larger block. The Patchwork hiding scheme referred to earlier sacrifices communication rate entirely (hiding just one bit) in order to buy as much robustness as possible.

Signal-to-noise ratio provides a rough estimate of perceptibility, because, all other things being equal, the higher the signal-to-noise, the more visible the modulated carrier will be. However, keeping signal-to-noise constant, some carriers—particularly those with mid-range spatial frequencies, our experience so far suggests—will be more more perceptible than others. So the crudest model of perceptibilty is simply signal-to-noise ratio; a plausible refinement might be the integral over all spatial frequencies of the signal-to-noise as a function of frequency weighted by a model of the frequency response of the human visual system. Methods for quantifying visibility to humans might be a new theoretical avenue to explore, and developing systematic methods for minimizing the visibility of hidden signals is certainly a challenge to information hiding practice. The pre-distortion technique demonstrated in this paper can be viewed as a first step in this direction, in the sense that successful compression schemes comprise implicit, algorithmic models of the human visual system (the ideal compression scheme would encompass a complete model of the human visual system). It

will be interesting to watch the development of information hiding schemes and their co-evolutionary "arms race" with compression methods in the challenging environment of the human visual system.

A Approximate superposition property for JPEG operator

An operator O obeys superposition if $O\{f+g\} - (O\{f\} + O\{g\}) = 0$. Each coefficient generated by the JPEG operator J satisfies $-1 \leq J\{f+g\} - (J\{f\} + J\{g\}) \leq 1$. In other words, JPEGing a pair of images separately and then adding them yields a set of coefficients each of which differs by no more than one quantization level from the corresponding coefficient found by adding the images first and then JPEGing them (using the same compression parameters in both cases).

The proof is simple. For a gray scale image, the unquantized JPEG coefficients S_{ij} are found by expanding each 8×8 block in a cosine basis. The final quantized coefficients a_{ij} are found by dividing each S_{ij} by a quantization factor q_{ij} (where each q_{ij} is greater than one, since the purpose of the JPEG representation is to decrease the file size), and rounding toward zero[BH93]:

$$a_{ij} = \lfloor \frac{S_{ij}}{q_{ij}} \rfloor$$

The cosine expansion is a linear operation, and therefore obeys superposition, so (as long as $q_{ij} > 1$) we need only show that for any real numbers f and g, $-1 \leq \lfloor f+g \rfloor - \lfloor f \rfloor - \lfloor g \rfloor \leq 1$. Without loss of generality, we may take f and g to be non-negative and less than one, since the integer parts F and G of f and g satisfy $\lfloor F+G \rfloor - \lfloor F \rfloor - \lfloor G \rfloor = 0$. So, for such an f and g, $0 \leq f+g < 2$. There are now two cases to consider. If $0 \leq f+g < 1$, then $\lfloor f+g \rfloor - \lfloor f \rfloor - \lfloor g \rfloor = 0 - 0 - 0 = 0$. If $1 \leq f+g < 2$ then $\lfloor f+g \rfloor - \lfloor f \rfloor - \lfloor g \rfloor = 1 - 0 - 0 = 1$. Since $f+g < 2$, these are the only two cases. The case of f and g negative is analogous, yielding a discrepancy of either -1 or 0. The discrepancy in the case that f and g have opposite sign is less than in the same sign case. Therefore each a_{ij} coefficient produced by the JPEG operator satisfies our approximate superposition principle, $-1 \leq J\{f+g\} - (J\{f\} + J\{g\}) \leq 1$. Since each a_{ij} coefficient has a discrepancy of $+1$, 0, or -1, each S_{ij} has a discrepancy of $+q_{ij}$, 0, or $-q_{ij}$. Thus the total power of the deviation from superposition (in either the spatial frequency or pixel representation, by Parseval's theorem) is bounded above by $\sum_{ij} q_{ij}^2$. This explains why JPEGing the carrier separately from the cover image is a reasonable predistortion tactic.

Note that the more aggressive the compression (the larger the q_{ij} values), the larger the discrepancies, or deviations from superposition.

Acknowledgments

This research was performed in the laboratory of Neil Gershenfeld. The authors thank him for his advice and support. The second author thanks Joe Jacobson for his support. We thank Walter Bender, Dan Gruhl, and the News in the Future Consortium for introducing us to the problem of data hiding. We acknowledge the other members of the Physics and Media group, especially Joe Paradiso and Tom Zimmerman, for helpful conversations about modulation techniques. Maggie Orth made useful suggestions about the proof of the approximate superposition principle.

This work was supported in part by the MIT Media Lab's News in the Future Consortium, a Motorola Fellowship, the Hewlett-Packard Corporation, Festo Corporation, Microsoft, Compaq Computer Corporation, and the MIT Media Lab's Things That Think consortium.

References

[BGM91] W. Bender, D. Gruhl, and N. Morimoto. Techniques for data hiding. In *Proceedings of the SPIE*, pages 2420–2440, San Jose, CA, February 1991.

[BH93] M.F. Barnsley and L.P. Hurd. *Fractal Image Compression.* AK Peters, Ltd., Wellesley, Massachusetts, 1993.

[BOD95] F.M. Boland, J.J.K. O'Ruanaidh, and C Dautzenberg. Watermarking digital images for copyright protection. In *Proceedings, IEE International Conference on Image Processing and its Application*, Edinburgh, 1995.

[CKLS] I. Cox, J. Kilian, T. Leighton, and T. Shamoon. A secure, robust watermark for multimedia. *This volume.*

[Cor95] Digimarc Corporation. Identification/authentication coding method and apparatus. *U.S. Patent Application*, June 1995.

[Dix94] R.C. Dixon. *Spread Spectrum Systems with Commercial Applications.* John Wiley and Sons, New York, 1994.

[Pfi] B. Pfitzmann. Trials of traced traitors. *This volume.*

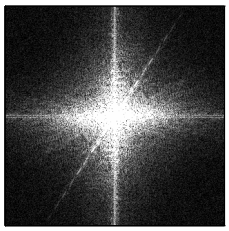
[She95] T.J. Shepard. *Decentralized Channel Management in Scalable Multihop Spread-Spectrum Packet Radio Networks.* PhD thesis, Massachusetts Institute of Technology, July 1995.

[SOSL94] M.K. Simon, J.K. Omura, R.A. Scholtz, and B.K. Levitt. *The Spread Spectrum Communications Handbook.* McGraw-Hill, New York, 1994.

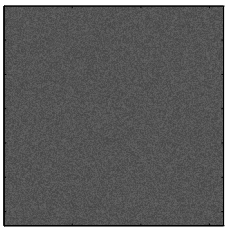
[SW49] C.E. Shannon and W.W. Weaver. *The Mathematical Theory of Communication.* The University of Illinois Press, Urbana, Illinois, 1949.

N(x,y)

Abs(FFT(N(x,y)))



c+S(x,y)



N(x,y) + S(x,y)