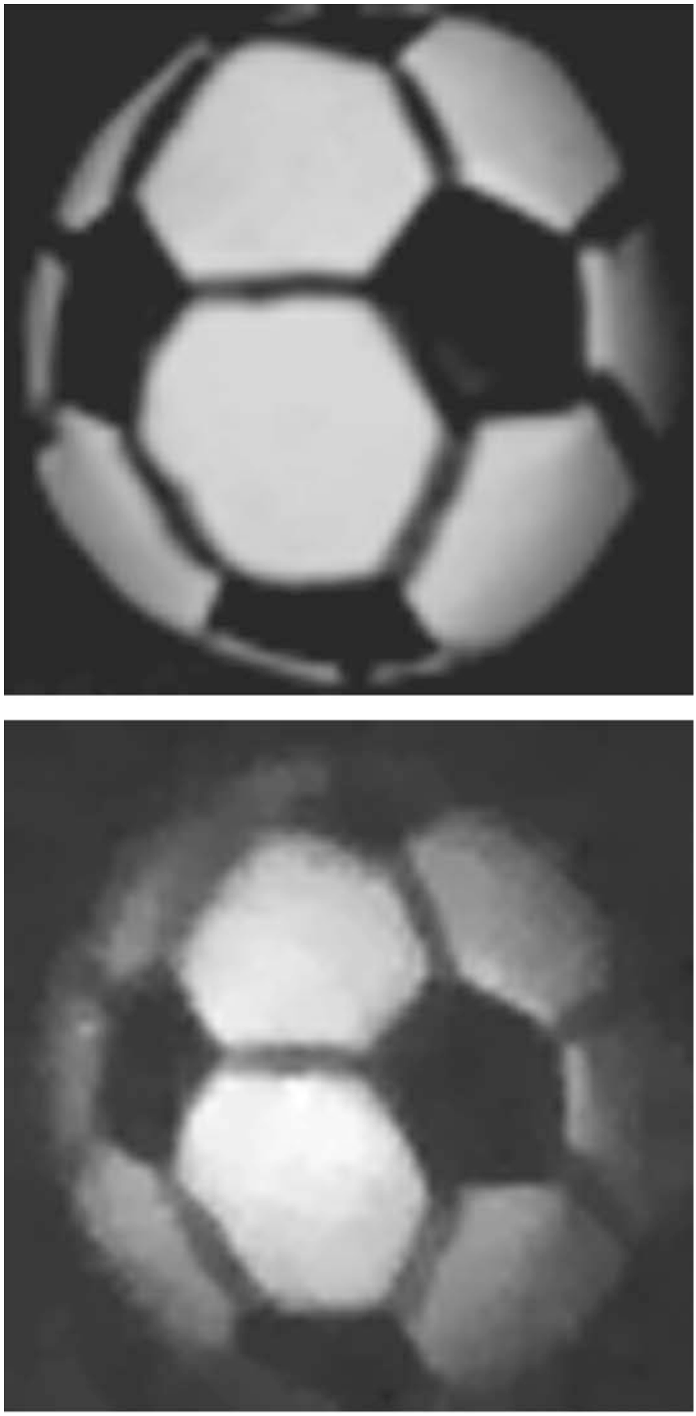


One is Enough. *A photograph taken by the "single-pixel camera" built by Richard Baraniuk and Kevin Kelly of Rice University. (a) A photograph of a soccer ball, taken by a conventional digital camera at* 64×64 *resolution. (b) The same soccer ball, photographed by a single-pixel camera. The image is derived mathematically from 1600 separate, randomly selected measurements, using a method called compressed sensing. (Photos courtesy of R. G. Baraniuk, Compressive Sensing [Lecture Notes],* Signal Processing Magazine, *July 2007. © 2007 IEEE.)*

Compressed Sensing Makes Every Pixel Count

Emmanuel Candes. *(Photo courtesy of Emmanuel Candes.)*

TRASH AND COMPUTER FILES HAVE ONE THING in common: compact is beautiful. But if you've ever shopped for a digital camera, you might have noticed that camera manufacturers haven't gotten the message. A few years ago, electronic stores were full of 1- or 2-megapixel cameras. Then along came cameras with 3-megapixel chips, 10 megapixels, and even 60 megapixels.

Unfortunately, these multi-megapixel cameras create enormous computer files. So the first thing most people do, if they plan to send a photo by e-mail or post it on the Web, is to compact it to a more manageable size. Usually it is impossible to discern the difference between the compressed photo and the original with the naked eye (see Figure 1, next page). Thus, a strange dynamic has evolved, in which camera engineers cram more and more data onto a chip, while software engineers design cleverer and cleverer ways to get rid of it.

In 2004, mathematicians discovered a way to bring this "arms race" to a halt. Why make 10 million measurements, they asked, when you might need only 10 thousand to adequately describe your image? Wouldn't it be better if you could just acquire the 10 thousand most relevant pieces of information at the outset? Thanks to Emmanuel Candes of Caltech, Terence Tao of the University of California at Los Angeles, Justin Romberg of Georgia Tech, and David Donoho of Stanford University, a powerful mathematical technique can reduce the data a thousandfold *before* it is acquired. Their technique, called *compressed sensing*, has become a new buzzword in engineering, but its mathematical roots are decades old.

As a proof of concept, Richard Baraniuk and Kevin Kelly of Rice University even developed a *single-pixel* camera. However, don't expect it to show up next to the 10-megapixel cameras at your local Wal-Mart because megapixel camera chips have a built-in economic advantage. "The fact that we can so cheaply build them is due to a very fortunate coincidence, that the wavelengths of light that our eyes respond to are the same ones that silicon responds to," says Baraniuk. "This has allowed camera makers to jump on the Moore's Law bandwagon"—in other words, to double the number of pixels every couple of years.

Thus, the true market for compressed sensing lies in nonvisible wavelengths. Sensors in these wavelengths are not so cheap to build, and they have many applications. For example, cell phones detect encoded signals from a broad spectrum of radio frequencies. Detectors of terahertz radiation[1] could be used to spot contraband or concealed weapons under clothing.

[1]This is a part of the electromagnetic spectrum that could either be described as ultra-ultra high frequency radio or infra-infrared light, depending on your point of view.

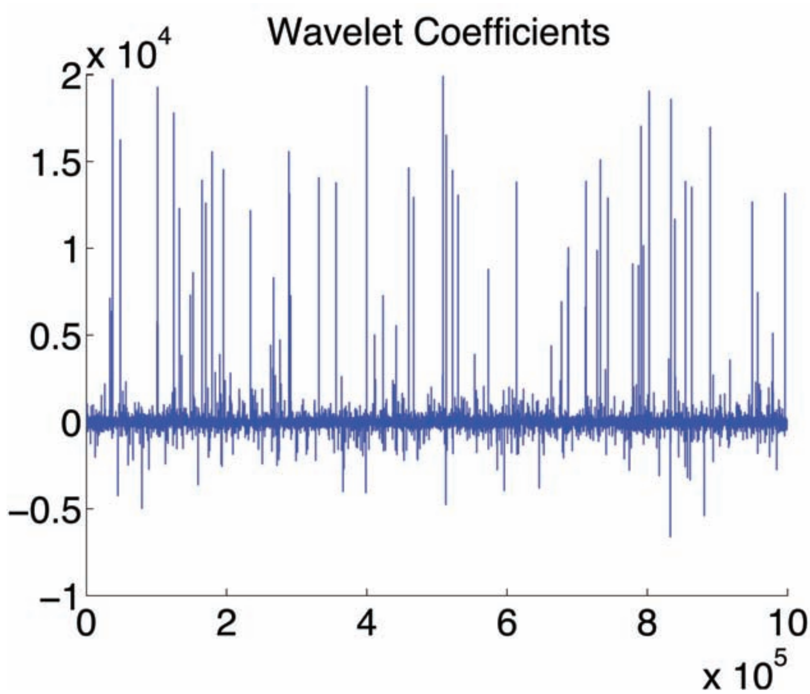
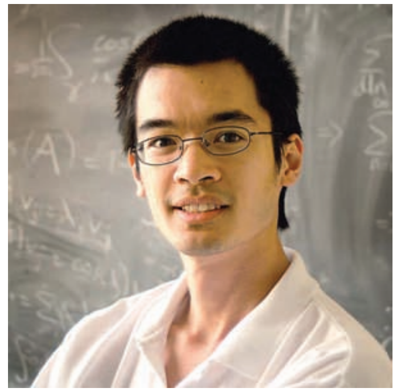


Figure 1. *Normal scenes from everyday life are compressible with respect to a basis of wavelets. (left) A test image. (top) One standard compression procedure is to represent the image as a sum of wavelets. Here, the coefficients of the wavelets are plotted, with large coefficients identifying wavelets that make a significant contribution to the image (such as identifying an edge or a texture). (right) When the wavelets with small coefficients are discarded and the image is reconstructed from only the remaining wavelets, it is nearly indistinguishable from the original. (Photos and figure courtesy of Emmanuel Candes.)*

Even conventional infrared light is expensive to image. "When you move outside the range where silicon is sensitive, your $100 camera becomes a $100,000 camera," says Baraniuk. In some applications, such as spacecraft, there may not be enough room for a lot of sensors. For applications like these, it makes sense to think seriously about how to make every pixel count.



Terence Tao. *(Photo courtesy of Reed Hutchinson/UCLA.)*

The Old Conventional Wisdom

The story of compressed sensing begins with Claude Shannon, the pioneer of information theory. In 1949, Shannon proved that a time-varying signal with no frequencies higher than N hertz can be perfectly reconstructed by sampling the signal at regular intervals of $1/2N$ seconds. But it is the converse theorem that became gospel to generations of signal processors: a signal with frequencies higher than N hertz *cannot* be reconstructed uniquely; there is always a possibility of aliasing (two different signals that have the same samples).

In the digital imaging world, a "signal" is an image, and a "sample" of the image is typically a pixel, in other words a measurement of light intensity (perhaps coupled with color information) at a particular point. Shannon's theorem (also called the Shannon-Nyquist sampling theorem) then says that the resolution of an image is proportional to the number of measurements. If you want to double the resolution, you'd better double the number of pixels. This is exactly the world as seen by digital-camera salesmen.

Candes, Tao, Romberg, and Donoho have turned that world upside down. In the compressed-sensing view of the world, the achievable resolution is controlled primarily by the *information content of the image*. An image with low information content can be reconstructed perfectly from a small number of measurements. Once you have made the requisite number of measurements, it doesn't help you to add more. If such images were rare or unusual, this news might not be very exciting. But in fact, *virtually all real-world images have low information content* (as shown in Figure 1).

This point may seem extremely counterintuitive because the mathematical meaning of "information" is nearly the opposite of the common-sense meaning. An example of an image with high information content is a picture of random static on a TV screen. Most laymen would probably consider such a signal to contain no information at all! But to a mathematician, it has high information content precisely because it has no pattern; in order to describe the image or distinguish between two such images, you literally have to specify every pixel. By contrast, any real-world scene has low information content because it is possible to convey the content of the image with a small number of descriptors. A few lines are sufficient to convey the idea of a face, and a skilled artist can create a recognizable likeness of any face with a relatively small number of brush strokes.[2]

[2]The modern-day version of the "skilled artist" is an image compression algorithm, such as the JPEG-2000 standard, which reconstructs a copy of the original image from a small number of components called *wavelets*. (See "Parlez-vous Wavelets?" in *What's Happening in the Mathematical Sciences*, Volume 2.)

Justin Romberg. *(Photo courtesy of Justin Romberg.)*

The idea of compressed sensing is to use the low information content of most real-life images to circumvent the Shannon-Nyquist sampling theorem. If you have no information at all about the signal or image you are trying to reconstruct, then Shannon's theorem correctly limits the resolution that you can achieve. But if you know that the image is sparse or compressible, then Shannon's limits do not apply.

Long before "compressed sensing" became a buzzword, there had been hints of this fact. In the late 1970s, seismic engineers started to discover that "the so-called fundamental limits weren't fundamental," says Donoho. Seismologists gather information about underground rock formations by bouncing seismic waves off the discontinuities between strata. (Any abrupt change in the rock's state or composition, such as a layer of oil-bearing rock, will reflect a vibrational wave back to the surface.) In theory the reflected waves did not contain enough information to reconstruct the rock layers uniquely. Nevertheless, seismologists were able to acquire better images than they had a right to expect. The ability to "see underground" made oil prospecting into less of a hit-or-miss proposition. The seismologists explained their good fortune with the "sparse spike train hypothesis," Donoho says. The hypothesis is that underground rock structures are fairly simple. At most depths, the rock is homogeneous, and so an incoming seismic wave sees nothing at all. Intermittently, the seismic waves encounter a discontinuity in the rock, and they return a sharp spike to the sender. Thus, the signal is a sparse sequence of spikes with long gaps between them.

In this circumstance, it is possible to beat the constraints of Shannon's theorem. It may be easier to think of the dual situation: a sparse *wave train* that is the superposition of just a few sinusoidal waves, whose frequency does not exceed N hertz. If there are K frequency spikes in a signal with maximal frequency N, Shannon's theorem would tell you to collect N equally spaced samples. But the sparse wave train hypothesis lets you get by with only $3K$ samples, or even sometimes just $2K$. The trick is to sample at random intervals, not at regular intervals (see Figures 2 and 3). If $K << N$ (which is the meaning of a "sparse" signal), then random sampling is much more efficient.

In other fields, such as magnetic resonance imaging, researchers also found that they could "undersample" the data and still get good results. At scientific meetings, Donoho says, they always encountered skepticism because they were trying to do something that was supposed to be impossible. In retrospect, he says that they needed a sort of mathematical "certificate," a stamp of approval that would guarantee when random sampling works.

The New Certificate

Emmanuel Candes, a former student of Donoho, faced the same skepticism in 2004, while working with a team of radiologists on magnetic resonance imaging. In trial runs with a "phantom image" (in other words, not a real patient), he was able to reconstruct the image perfectly from undersampled data. "There was no discrepancy at all between the original and the reconstruction," Candes says. "I actually got into a bit of trouble, because they thought I was fudging."

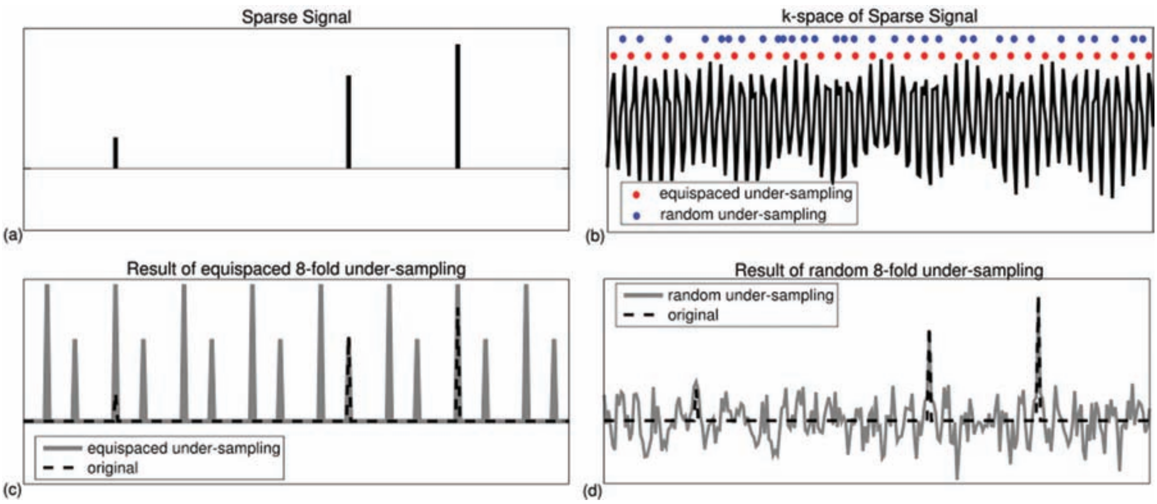


Figure 2. *Reconstructing a sparse wave train. (a) The frequency spectrum of a 3-sparse signal. (b) The signal itself, with two sampling strategies: regular sampling (red dots) and random sampling (blue dots). (c) When the spectrum is reconstructed from the regular samples, severe "aliasing" results because the number of samples is 8 times less than the Shannon-Nyquist limit. It is impossible to tell which frequencies are genuine and which are impostors. (d) With random samples, the two highest spikes can easily be picked out from the background. (Figure courtesy of M. Lustig, D. Donoho, J.Santos and J. Pauly, Compressed Sensing MRI,* Signal Processing Magazine, *March 2008. © 2008 IEEE.)*

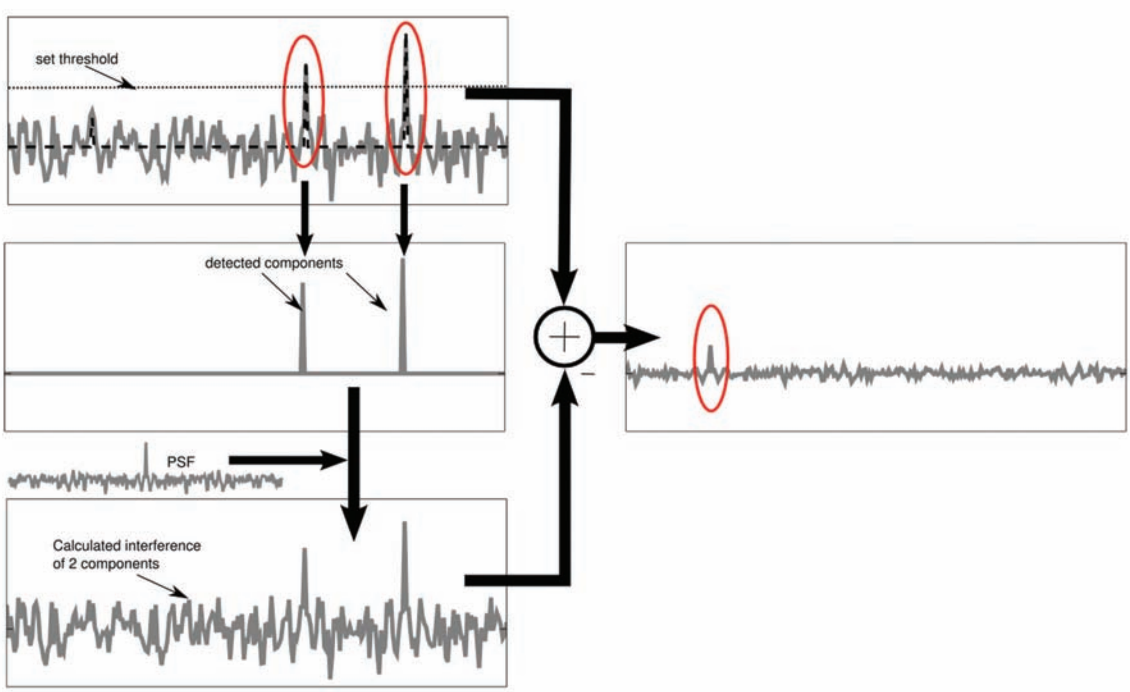


Figure 3. *In the situation of Figure 2, the third frequency spike can be recovered by an iterative thresholding procedure. If the signal was known to be 3-sparse to begin with, then the signal can be reconstructed perfectly, in spite of the 8-fold undersampling. In short, sparsity plus random sampling enables perfect (or near-perfect) reconstruction. (Figure courtesy of M. Lustig, D. Donoho, J.Santos and J. Pauly, Compressed Sensing MRI,* Signal Processing Magazine, *March 2008. © 2008 IEEE.)*

Candes and Tao found a shortcut that not only runs faster on a computer, but also explains why random sampling works so much better than regular sampling.

At this point, Candes did a fortuitous thing: he talked with Terry Tao, a 2006 Fields medalist. The two mathematicians happened to have children at the same pre-school. While they were dropping them off one day, Candes told Tao about the too-good-to-be-true reconstructions. “I had begun looking for an explanation and made some headway, but I was stuck at a particular point,” Candes says.

“Terry reacted like a mathematician,” Candes continues. “He said, ‘I’m going to find a counterexample, showing that what you have in mind cannot be true.”’ But a strange thing happened. None of the counterexamples seemed to work, and Tao started listening more closely to Candes’ reasoning. “After a while, he looked at me and said, ‘Maybe you’re right,”’ Candes says. With the speed for which Tao is legendary, within a few days he had helped Candes overcome his obstacle and the two of them began to sketch out the first truly general theory of compressed sensing.

In the Candes-Romberg-Tao framework, a signal or an image is represented as a vector $\mathbf{x}$, a string of N real numbers. This vector is assumed to be K-sparse, which means that in some prescribed basis it is known to have at most K nonzero coefficients. (K is assumed to be much less than N.) For example, if the basis elements are standard coordinate vectors in $\mathbf{R}^N$, then $\mathbf{x}$ literally consists of mostly zeroes. This is exactly the situation of the sparse spike train hypothesis.

However, compressed sensing does not require a particular basis. Photographs, for example, are not at all sparse with respect to the standard basis; they have many nonzero coefficients (i.e., non-black pixels). JPEG compression has proven that photographs are almost always approximately sparse with respect to a different basis—the basis of wavelets. If Ψ represents the $N \times N$ matrix of basis vectors, then a K-sparse signal with respect to that basis is one that can be written in the form $\Psi\mathbf{x}$, where $\mathbf{x}$ has at most K nonzero coefficients.

A sample $\mathbf{y}$ of the signal $\mathbf{x}$, in the Candes-Romberg-Tao framework, is a linear function of $\mathbf{x}$: that is, $\mathbf{y} = \Phi\mathbf{x}$. The number of measurements in the sample is assumed to be smaller than the signal, so Φ is an $M \times N$ matrix with $M << N$. By elementary linear algebra, there are infinitely many other vectors $\mathbf{x}^*$ such that $\Phi\mathbf{x}^* = \mathbf{y}$. However, provided that $M \geq 2K$, it will normally be the case that none of the other solutions to the equation $\Phi\mathbf{x}^* = \mathbf{y}$ are sparse. Thus, if $\mathbf{x}$ is known in advance to be sparse, it can in theory be reconstructed exactly from M measurements.

Knowing that a unique solution exists is not the same thing as being able to find it. The problem is that there is no way to know in advance which K coordinates of $\mathbf{x}$ are nonzero. The naive approach is to try all the possibilities until you hit on the right one, but this turns out to be a hopelessly slow algorithm. However, Candes and Tao found a shortcut that not only runs faster on a computer, but also explains why random sampling works so much better than regular sampling.

If your image consists of a few sparse dots or a few sharp lines, the *worst* way to sample it is by capturing individual pixels (the way a regular camera works!). The *best* way to sample the image is to compare it with widely spread-out noise functions. One could draw an analogy with the game of "20 questions." If you have to find a number between 1 and N, the worst way to proceed is to guess individual numbers (the analog of measuring individual pixels). On average, it will take you $N/2$

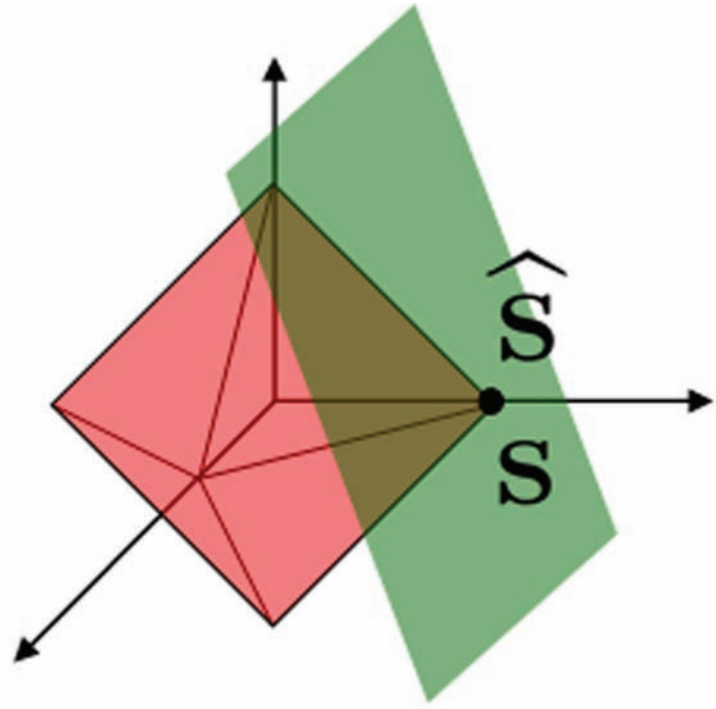


Figure 4. *A random measurement of a sparse signal, S, generates a subspace of possible signals (green) that could have produced that measurement. Within that green subspace, the vector of smallest l_1-norm ($\hat{S}$) is usually equal to S. (Figure courtesy of R. G. Baraniuk, Compressive Sensing [Lecture Notes],* Signal Processing Magazine, *July 2007. © 2007 IEEE.)*

guesses. By contrast, if you ask questions like, "Is the number less than $N/2$?" and then "Is the number less than $N/4$?" and so on, you can find the concealed number with at most $\log_2 N$ questions. If N is a large number, this is an enormous speed-up.

Notice that the "20 questions" strategy is adaptive: you are allowed to adapt your questions in light of the previous answers. To be practically relevant, Candes and Tao needed to make the measurement process nonadaptive, yet with the same guaranteed performance as the adaptive strategy just described. In other words, they needed to find out ahead of time what would be the most informative questions about the signal **x**. That this can be done effectively is one of the great surprises of the new theory. The idea of their approach is called l_1-minimization.

The l_0-norm of a vector is simply the number of nonzero entries in the vector, which can be somewhat informally written as follows:

$$\|(x_1, x_2, ..., x_N)\|_0 = \sum |x_i|^0.$$

(This formula uses the convention that $0^0 = 0$.) The l_1-norm is obtained by replacing the 0's in this equation by 1's:

$$\|(x_1, x_2, ..., x_N)\|_1 = \sum |x_i|^1.$$

In this language, the signal $\mathbf{x}$ is the unique solution to $\Phi\mathbf{x}^* = \mathbf{y}$ with the smallest l_0-norm. But in many cases, Candes and Tao proved, it is *also* the unique solution with the smallest l_1-norm. This was a critical insight because l_1-minimization is a linear programming problem, which can be solved by known, efficient computer algorithms. (See "Smooth(ed) Moves," *What's Happening in the Mathematical Sciences*, Volume 6.)

A random plane that passes through a vertex is *virtually certain* to miss the interior of the cross polytope. Thanks to this "miracle of high-dimensional geometry," as Candes calls it, the l_1-minimizer will almost always be the correct signal, x.

Figure 4 illustrates why the l_1-minimizer is often the same as the l_0-minimizer. In 3-dimensional space, the set of unit vectors in the l_1-norm is an octahedron. Think of the sparse vector $\mathbf{x}$ as lying on a coordinate axis (because it has lots of zero coordinates). Therefore it is at one of the vertices of the octahedron. The set of vectors $\mathbf{x}^*$ such that $\Phi\mathbf{x}^* = \mathbf{y}$ is a plane passing through the point $\mathbf{x}$. Most planes that pass through $\mathbf{x}$ intersect the octahedron *only* at the point $\mathbf{x}$; in other words, $\mathbf{x}$ is the unique point on the plane with the minimum l_1-norm. So if you simply pick the measurement Φ "at random," you have a very good chance of reconstructing $\mathbf{x}$ uniquely.

Unfortunately, picking Φ at random won't always work. You might get unlucky and choose a plane through $\mathbf{x}$ that passes through the interior of the octahedron. If so, the l_1-minimizer will not be the same as the l_0-minimizer. The algorithm will produce an erroneous signal, $\mathbf{x}^*$. But the three-dimensional picture in Figure 4 (page 121) is somewhat misleading because the image vectors typically lie in a space with thousands or millions of dimensions. The analog of the octahedron in million-dimensional space is called the cross polytope; and in million-dimensional space the cross polytope is very, very, very pointy. A random plane that passes through a vertex is *virtually certain* to miss the interior of the cross polytope. Thanks to this "miracle of high-dimensional geometry," as Candes calls it, the l_1-minimizer will almost always be the correct signal, $\mathbf{x}$.

In summary, this is what the theory of compressed sensing says:

- For many $M \times N$ matrices Φ, the unique K-sparse solution, $\mathbf{x}$, to the equation $\Phi\mathbf{x}^* = \mathbf{y}$, can be recovered *exactly*.
- N must be much larger than K. However, M (the number of measurements) need only be a little larger than K. Specifically, M must be roughly $K \log(N/K)$. Notice that the dependence on N is logarithmic, so the "20 questions" speed-up has been achieved.
- The K-sparse solution is found by l_1-minimization, which can be proved to be equivalent to l_0-minimization under certain assumptions on the measurement matrix, Φ.
- Random matrices Φ almost always satisfy those assumptions.

The whole story remains essentially unchanged if the signal is sparse with respect to a basis Ψ that is not the standard basis of coordinate vectors (e.g., the wavelet basis). The only modification required is that the constraint, $\Phi \mathbf{x}^* = \mathbf{y}$, is replaced by the constraint $\Phi\Psi\mathbf{x}^* = \mathbf{y}$. In this context, the randomness of the measurement matrix Φ serves a double purpose. First, it provides the easiest set of circumstances under which l_1-minimization is provably equivalent to l_0-minimization. Secondly, and independently, it ensures that the set of measurement vectors (the rows of Φ) are as dissimilar to the image basis (the columns of Ψ) as possible. If the image basis consists of spikes, the measurement basis should consist of spread-out random noise. If the image basis consists of wavelets, then the measurement basis should consist of a complementary type of signal called "noiselets."

"Our paper showed something really unexpected," says Candes. "It showed that using randomness as a sensing mechanism is extremely powerful. That's claim number one. Claim number two is that it is amenable to rigorous analysis.

"What mathematicians liked [about the paper] was the way it merged analysis and probability theory. A lot of people in my field, analysis, did not think about probability theory as being useful or worthy of attention. At the very intellectual level, it changed the mindset of those people and caused them to engage this field."

Richard Baraniuk. *(Photo courtesy of Richard Baraniuk.)*

Recent Developments

Tao and Candes' preprint appeared in 2004, as did a paper by Donoho announcing similar results. By the time that Tao and Candes' paper actually appeared in print, in 2006, it had been cited more than 100 times. Since then, there have been many advances, both from the theoretical and the practical side.

One question left unanswered by the original paper was how well compressed sensing would hold up if the measurements contained some random error (an inevitable problem of real-world devices), or if the images themselves were not exactly sparse. In photography, for instance, the assumption of a sparse signal is not literally true. It is more realistic to assume the signal is *compressible*, which means that the vast majority of the information in the signal is contained in a few coefficients. The remaining coefficients are not literally zero, but they are small. Under these circumstances, even the l_0-minimizer does not match the signal exactly, so there is no hope for the l_1-minimizer to be exactly correct.

In 2005, Candes, Romberg, and Tao showed that even with noisy measurements and compressible (but not sparse) signals, compressed sensing works well. The error in the reconstructed signal will not be much larger than the error in the measurements, and the error due to using the l_1-minimizer will not be much greater than the penalty already incurred by the l_0-minimizer. That is, the l_1-minimizer accurately recovers the most important pieces of information, the largest components of the signal. Figure 5 (see next page) shows an example of the performance of compressed sensing on a simulated image with added noise.

Mathematicians have also been working on new algorithms that run even faster than the standard linear programming techniques that solve the l_1-minimization problem. Instead of finding the largest K coefficients of $\mathbf{x}$ all at once, they find them iteratively: first the largest nonzero coefficient, then the

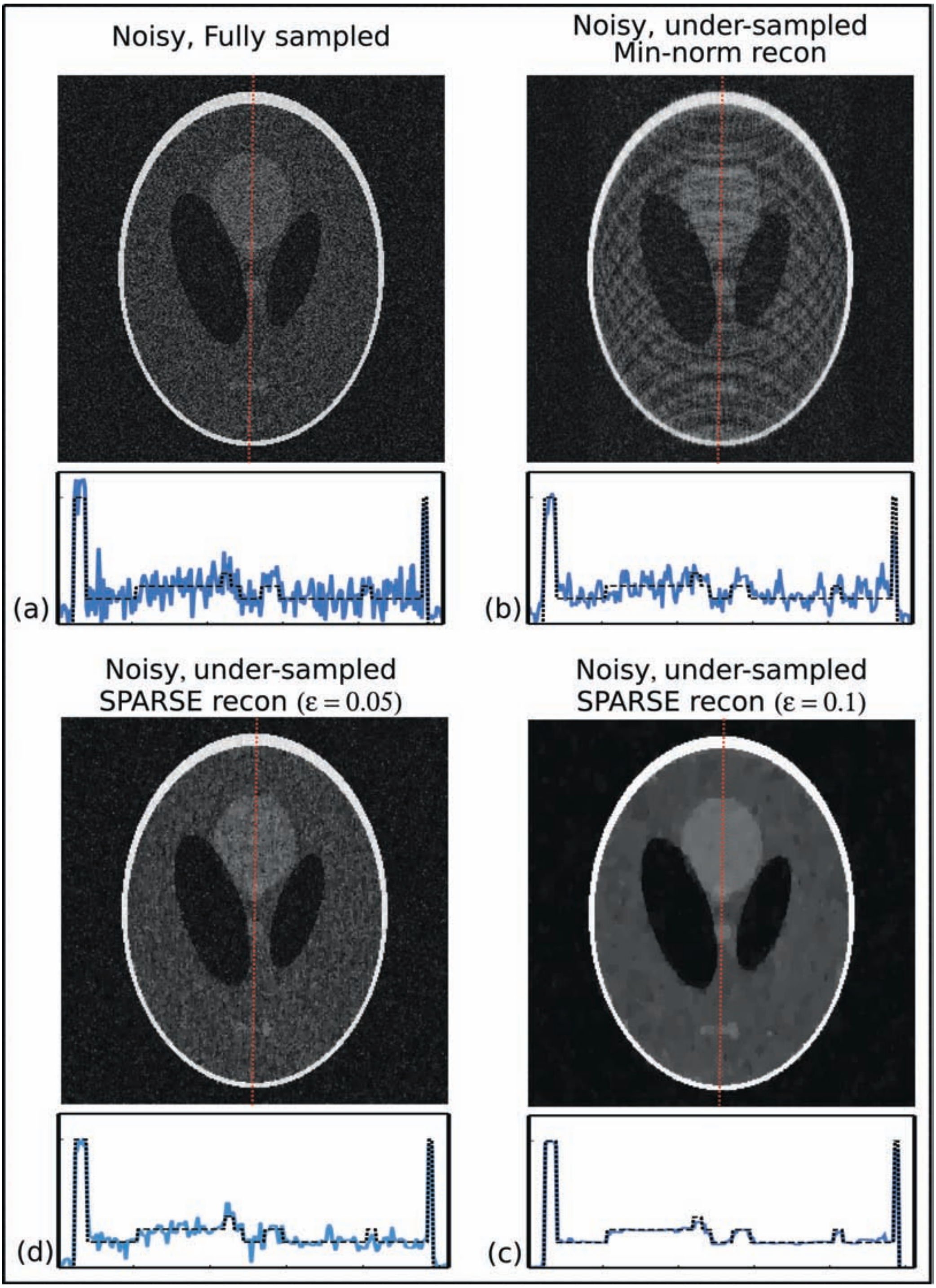


Figure 5. *Compressed sensing with noisy data. (a) An image with added noise. (b) The image, undersampled and reconstructed using the Shannon-Nyquist approach. As in Figure 2, artifacts appear in the reconstructed image. (d) The same image, undersampled randomly and reconstructed with a "too optimistic" noise model. Although there are no artifacts, some of the noise has been misinterpreted as real variation. (c) The same image, reconstructed from a random sample with a more tolerant noise model. The noise is suppressed and there are no artifacts. (Figure courtesy of Michael Lustig.)*

second-largest one, and so on. The first such algorithm, called Orthogonal Matching Pursuit (OMP), did not offer the same guarantees of accuracy that l_1-minimization did. However, there is now a variety of colorfully named variations, such as Regularized OMP (ROMP) and Stagewise OMP (StOMP), which successfully combine the accuracy of l_1-minimization with the speed of OMP. These algorithms have the advantage of being somewhat more intuitive than the "high-dimensional miracle" of l_1-minimization; Figure 3 shows an example.

Meanwhile, researchers in several different fields are exploring practical applications of compressed sensing. Baraniuk and Kelly's single-pixel camera, built in 2006, uses an array of bacteria-sized mirrors to acquire a random sample of the incoming light. (See Figure 6.) Each mirror can be tilted in one of two ways, either to reflect the light toward the single sensor or away from it. Thus the light that the sensor receives is a weighted average of many different pixels, all combined into one pixel. By taking $K \log(N/K)$ snapshots, with a different random selection of pixels each time, the single-pixel camera was able to acquire a recognizable picture with a resolution comparable to N pixels. (See figure **"One Is Enough,"** page 114.)

Baraniuk and Kelly's team is now working on "hyperspectral cameras," which would reconstruct a complete spectrum at each point of the image. "A conventional digital image has red, blue and green pixels," Baraniuk says. "It's great for making a picture that fools the human eye, but it doesn't capture the essence of the wavelengths given off by different materials. What you'd really like would be a spectrum of thousands of colors instead of just three. This would allow you to tell the difference between green paint on a car and a green leaf on a bush." But with thousands of colors at each of millions of pixels, data compression becomes a serious issue.

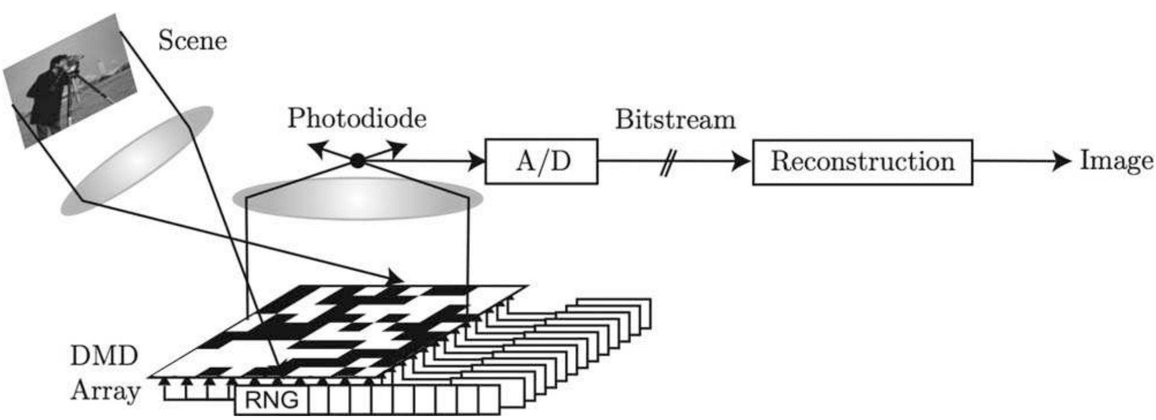


Figure 6. *A schematic diagram of the "one-pixel camera." The "DMD" is the grid of micro-mirrors that reflect some parts of the incoming light beam toward the sensor, which is a single photodiode. Other parts of the image (the black squares) are diverted away. Each measurement made by the photodiode is a random combination of many pixels. In **"One is Enough"** (p.114), 1600 random measurements suffice to create an image comparable to a 4096-pixel camera. (Figure courtesy of Richard Baraniuk.)*

With so many questions and so many choices, it is impossible at present to say what the most successful application of compressed sensing will be. However, one thing is clear: Engineers are finally thinking outside the box of Shannon's theorem.

Baraniuk and his former student Michael Wakin, now at the University of Michigan, have also worked on a problem of object detection. For many applications, producing an actual photograph may not be as important as recognizing quickly what is there. For example, a security system may have to identify a face or a vehicle. For example, Baraniuk says, you could teach it to recognize the difference between a Corolla and a Porsche. The computer will have images of Corollas and Porsches stored in it, but the vehicle in front of the camera may be rotated in a way that does not precisely match the photos. In this application, the image vector has a different kind of sparse structure. Instead of lying on a coordinate K-plane, the vector will lie on a curved K-dimensional manifold in N-dimensional space. (In this case, K would be equal to 3.) In this context, Wakin showed that on the order of $K \log N$ measurements still suffice to make the call.

Some applications of compressed sensing may lie completely outside the realm of imaging. One such example is "analog to digital conversion," a fundamental aspect of wireless communications. For example, the CDMA cell phone standard takes a voice message, which contains sound frequencies up to 4096 hertz, and spreads it out over a radio spectrum that spans hundreds of thousands of hertz. The signal is sparse because it still contains only the information that was squeezed inside those 4096 hertz. So a detector that performs compressed sensing should be able to recover the signal more rapidly than a detector based on Shannon's theorem.

In digital photography, Moore's law lets you pack twice as many detectors on a chip every two years. But in the world of analog to digital conversion, Baraniuk says, "the equivalent figure of merit doubles every 6 to 8 years." So instead of waiting decades for a hardware solution, it really makes sense to solve the problem with software based on compressed sensing.

Finally, compressed sensing may find some medical applications—which would be only natural because the theory was directly inspired by a problem in magnetic resonance imaging. MRI scanners have traditionally been limited to imaging static structures over a short period of time, and the patient has been instructed to hold his or her breath. But now, by treating the image as a sparse signal in space and time, MRI scanners have begun to overcome these limitations and produce images, for example, of a beating heart. Figure 7 shows how a sparse reconstruction algorithm can provide a sharp image of the arteries in a patient's leg even with as many as 20 times less data than a conventional angiogram.

One hurdle that compressed sensing may have to overcome is how to develop practical "incoherent sensors." A single measurement, in compressed sensing, is an inner product of the incoming compressible signal with a random, noisy test signal. Baraniuk's single-pixel camera accomplishes the inner product by using mirrors to deflect certain parts of the light beam toward the sensor, while deflecting other parts away. In real applications, if the hardware that performs the incoherent measurements is more expensive than the array of sensors that it is designed to replace, then the economic case for compressed sensing will disappear.

With so many questions and so many choices, it is impossible at present to say what the most successful application of compressed sensing will be. However, one thing is clear: Engineers are finally thinking outside the box of Shannon's theorem.

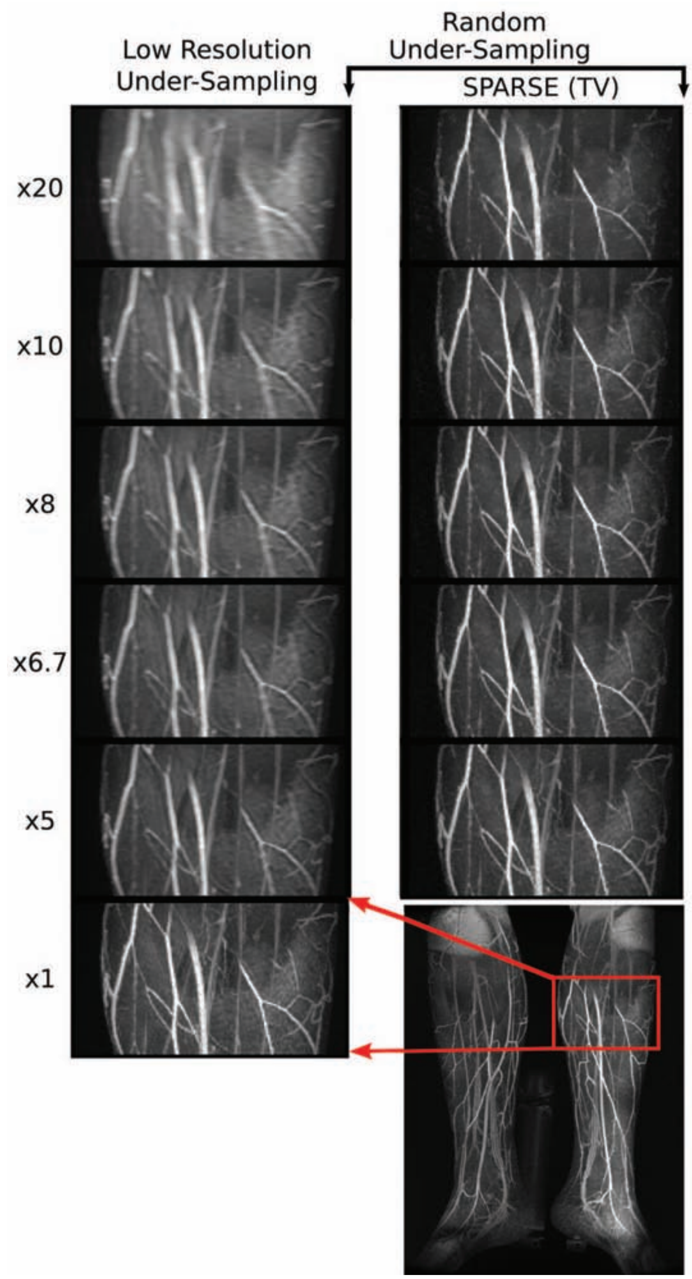


Figure 7. *An angiogram. From bottom to top, the angiogram is progressively undersampled by larger and larger factors. With a Shannon-Nyquist sampling strategy, the image degrades as the degree of undersampling increases. With compressed sensing, the image remains very crisp even at 20-fold undersampling. The approach used here and in Figure 5 is not l_1-minimization but l_1-minimization of the spatial gradient. (Figure courtesy of Michael Lustig.)*