Nonlinear Codes Straighten Up—and Get to Work

"Nonlinearity"—the idea that effects can be more than the sum of their causes—has been all the rage in science for the last 10 years or so. There's a good reason for that, since nonlinearity has been found in phenomena ranging from weather patterns to population dynamics. In one branch of applied mathematics, however, linearity has stubbornly held sway: the error-correcting codes that reliably convert the strings of zeros and ones in digital signal transmissions or on computer disk drives into words, numbers, or images. But now the vogue for nonlinearity may be overtaking coding theory as well.

Linear codes have prevailed until now because their structure, in which the sum of two code words is always another code word, has enabled researchers to develop efficient decoding algorithms, which are essential when digital information has to be interpreted at a rapid clip. Researchers have known all along that nonlinear codes, which are not required to have any particular structure, can often squeeze more words into digital strings of a given length, enabling information to be sent more quickly or stored more compactly. The problem was that there didn't seem to be any efficient way to decode them. As a result, linear codes have predominated in applications, while nonlinear codes remained mathematical curiosities.

Now, however, two separate collaborations—at Hughes Aircraft and the University of Southern California (USC), and at AT&T Bell Laboratories and the CNRS in Valbonne, France—have discovered that some of the most studied nonlinear codes are actually closely related to well-known linear codes. The discovery, to be reported in a joint paper in the *IEEE Transactions in Information Theory*, is raising hopes of putting the more compact nonlinear codes to work in a range of computer and telecommunications products, says Vijay Kumar of USC, among them new high-capacity cellular networks.

Error-correcting codes are born of the need to compensate for the inevitable physical flaws that occasionally—and usually at the worst moment—turn a 1 into a 0 or a 0 into a 1. The ability of these codes to deliver a readable message in spite of such errors is what allows clear images of planets to be received at incredibly low power from distant spacecraft, and music on a compact disk to be enjoyed even when the disk has been scratched. But the cost paid for these abilities is a bulky code, since in order for a code to correct errors, there must be a certain separation, or what theorists call "distance," between the code's words.

Each word is just a string of 0s and 1s of some fixed length, say 5, and the distance between two words is the number of places in which they differ. For example, the distance between 00000 and 11100 is 3. Similarly, 00111 is distance 3 from 00000 and 4 from 11100. And to round things out, 11011 is distance 3 from 11100 and 00111 and distance 4 from 00000.

If each of these four words represents a message—say "Out to lunch," "Gone for the day," "In a meeting," and "On the phone"—then the message can be clearly read even if there's an error in one of the digits. For example, 00011 is closer to 00111 ("On the phone") than it is to any of the other three code words, because it's still two digits away from any other word. In the parlance of coding theory, the set of words {00000, 11100, 00111, 11011} constitutes a single-error-correcting code of length 5.

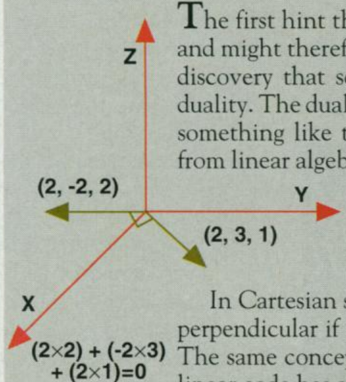
This code is linear: If you "add" any two code words, using the rule 1+1=0, you get another word, as in 11100+11011=00111. Linearity makes it possible to describe a code without having to run through an explicit list of all its code words. A linear code with 1024 words, for example, can be boiled down to a mere 10 words that generate all the rest. That feature and other special algebraic properties of linear codes speed up what might otherwise be awkward, inefficient algorithms for encoding and decoding.

The drawback is that linear codes don't necessarily pack the most code words into strings of a given length. For example, among double-error-correcting codes of length 16, the best linear code contains 128 code words, but a nonlinear code of the same length, known as the Nordstrom-Robinson code, contains 256 words. Nonlinear codes are able to pack in more words while maintaining the same distance between words because they are not required to have any particular algebraic structure. But when it comes to decoding messages, that lack of algebraic structure slows them down, which more than offsets their extra carrying capacity. That handicap has kept the Nordstrom-Robinson and other nonlinear codes out of application.

That could change now. What the two groups have discovered is that many nonlinear codes really are linear codes in disguise. In their joint paper, Kumar of USC and Roger Hammons, Jr. of Hughes, together with Robert Calderbank and Neil Sloane at Bell Labs and Patrick Solé at the CNRS, show how the Nordstrom-Robinson and other nonlinear codes can be described as binary "projections" of linear codes in an algebraic system known as Z_4. (Z_4 consists of the four numbers 0,1,2, and 3, with the rule 2+2= 1+3=0.) Projecting a code in binary digits simply amounts to replacing each 0 with 00, 1 with 01, 2 with 11, and 3 with 10. Thus transformed, a well-known linear code known as the octacode turns out to be equivalent to the nonlinear Nordstrom-Robinson code.

These relations not only offer a shortcut to deciphering nonlinear codes, says Sloane, they also solve a mystery that baffled theorists for decades—namely why some of these codes behave in certain respects as if they were linear. For example, the Nordstrom-Robinson code has properties reminiscent of a "self-dual" code (see box), in spite of the

Getting the Right Angle on Linear Codes



The first hint that nonlinear codes might be related to linear codes—and might therefore be technologically useful (see main text)—was the discovery that some of them display a mathematical property called duality. The duality of codes, technically defined only for linear codes, is something like the perpendicularity of vectors. The analogy emerges from linear algebra, which expresses a linear code as a vector space over a finite field, much as the familiar Cartesian plane is a vector space over the field of real numbers. In that picture, the digits of a code word correspond to the coordinates of an ordinary vector.

In Cartesian space, two vectors—say (x_1, y_1, z_1) and (x_2, y_2, z_2)—are perpendicular if their "dot product," $x_1x_2+y_1y_2+z_1z_2$, is zero (see figure). The same concept turns out to be useful in coding theory, where each linear code has a "dual" code consisting of code words perpendicular to every word in the original code. However, unlike ordinary vectors, a code word can be perpendicular to itself. It's even possible for a linear code to be its own dual.

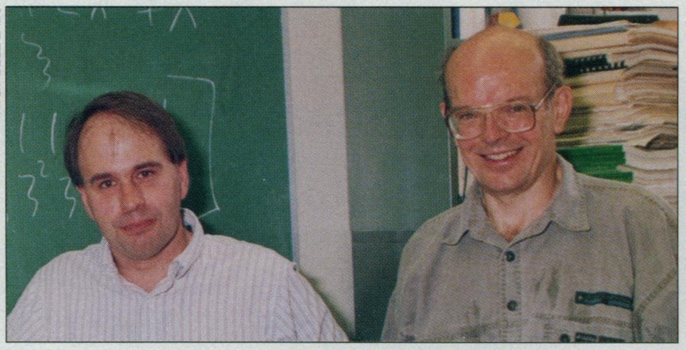
–B.C.

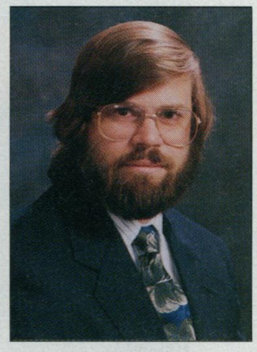
fact that duality is a concept that makes sense only for linear codes. The new results dispel the mystery: The dual-like properties of nonlinear codes are inherited from their linear precursors in Z_4.

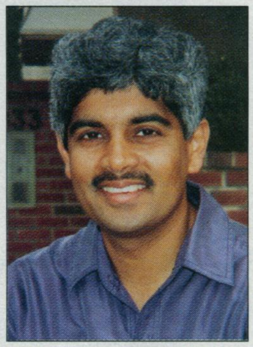
In fact, it was the self-duality of the Nordstrom-Robinson code that finally gave away its relation with the octacode. Sloane, who is one of the leading experts in coding theory, credits David Forney at Motorola Codex and Mitchell Trott at MIT with raising the possibility of a link between the Nordstrom-Robinson code and self-dual linear codes at a conference last October. "I went home and in 2 minutes thinking about it, it became clear that yes indeed, the octacode was really the same thing as the Nordstrom-Robinson code," recalls Sloane.

Calderbank and Solé added several ideas to Sloane's observation, and soon the three theorists had found linear precursors for a number of other nonlinear codes, including generalizations of the Nordstrom-Robinson code known as the Kerdock and Preparata codes. Then they discovered that they weren't alone. Working independently, Hammons and Kumar had been investigating mathematical aspects of a communication technique called code-division multiple access (CDMA), which has been touted as a basis for digital cellular radio. CDMA allows many users to broadcast simultaneously over the same communication channel, keeping their signals straight by assigning a separate code word, or sequence, to each user as an identifying tag. Because more code words means more users can have access to the system, Hammons and Kumar had been searching for ways to apply nonlinear codes. Their efforts paid off in the discovery of Z_4 linearity lurking behind the Kerdock and Preparata codes.

AT&T BELL LABS







Nonlinear thinkers. From left: Bell Labs' Calderbank and Sloane, Hughes' Hammons, and USC's Kumar. A fifth member of the collaboration, Solé of the CNRS, is not shown.

When the two groups discovered the overlap in their results, they decided to join forces. Together, they've gone beyond teasing out the relations between linear and nonlinear codes to devising new decoding schemes that take advantage of those links, including an explicit decoding algorithm for the Preparata code. Meanwhile, the list of no-longer-nonlinear codes is growing almost daily, says Sloane, and it may be only a matter of time before they show up in products like modems and mobile radios. "People were scared of them" because they seemed so complicated, he says. "They won't be so scared anymore."

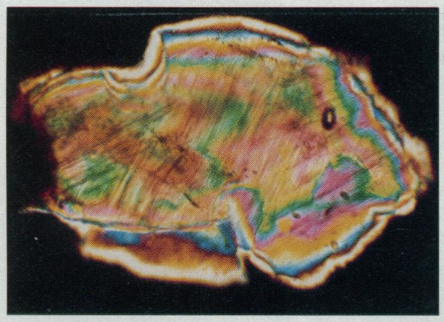
–Barry Cipra

EXTINCTIONS

New Crater Age Undercuts Killer Comets

One well-known Manson is definitely a killer, but another may now have been cleared of a gruesome mass murder charge. In 1989, Michael Kunk of the U.S. Geological Survey (USGS) in Reston, Virginia, and his colleagues published an age for the Manson crater in Iowa of around 65.7 million years. That put Manson tantalizingly close in age to the giant 65-million-year Chicxulub impact in the Yucatan, which may have exterminated the dinosaurs. The match, which Kunk and his colleagues tightened last spring with an additional date, led to speculation that the dinosaurs were pushed to extinction by not just one murderous impact, but two or even a whole swarm, such as a shower of comets (*Science*, 12 March, p. 1543). But the latest look at Manson puts the swarm-of-comets idea back on the shelf.

On page 729, Glen Izett, William Cobban, and John Obradovich of the USGS in Denver, along with Kunk, present two lines of evidence that suggest Manson is nearly 9 million years older than Chicxulub. One comes from grains of the mineral sanidine, extracted from the Manson crater, that yielded a radiometric age of 73.8 ± 0.3 million years. The other comes from sedimentary rocks of roughly the same age found hundreds of kilometers away, which display signs of a powerful but distant impact.



G. IZETT

Against the grain. Millimeter-sized grains of shocked quartz confirm that the Manson crater is 9 million years older than was thought.

The researchers dated the sanidine by measuring the amount of argon formed within the mineral by the steady decay of radioactive potassium. That's just what Kunk did with the earlier samples, but Izett thinks this one is more likely to have yielded an accurate age because of sanidine's straightforward origin. Sanidine only forms from rock that has melted completely, releasing all its argon and thereby resetting the radioactive clock to a zero age. The earlier samples, in contrast, had only been shocked, not melted, during the impact or were a mixture of melted and unmelted minerals.

Izett found evidence supporting the new age by going farther afield. He knew that 73.8 million years ago, the Manson site lay within the seaway that ran up the middle of North America. An impact there would have spewed debris across the seaway and churned out huge waves—effects that should be visible in sedimentary rock formed at that time. To test that idea, Izett examined exposures of sediments of roughly the right age a few hundred kilometers west of the crater.

There, within a 15- to 20-centimeter layer containing sand and centimeter-size shale fragments typical of a wave-scoured sea floor, he found mineral grains scarred by the extreme pressures of an impact. Above and below the disturbed layer, the uniformly fine-grained sedimentary layers suggested the sea was deep and tranquil. The nearest datable rocks above and below the impact layer yielded ages of 73.7 and 72.3 million years, give or take 0.4 million years, which is consistent with the new radiometric age. "The finding of the shocked mineral layer where it was predicted to be really makes the [older] age seem probable," says Kunk. It looks as if Manson the crater has now been exonerated of global mayhem.

–Richard A. Kerr