Explained: Gallager codes

January 21 2010, by Larry Hardesty

Professor Emeritus Robert Gallager

In the 1948 paper that created the field of information theory, MIT grad (and future professor) Claude Shannon threw down the gauntlet to future generations of researchers. In those predigital days, communications channels — such as phone lines or radio bands — were particularly susceptible to the electrical or electromagnetic disruptions known as "noise."

Shannon proved the counterintuitive result that no matter how noisy a channel, information could be sent over it error free. All you needed was a way to add enough redundancy to the information so that errors could be corrected. He also demonstrated that there was a hard limit on how efficient those error-correcting codes could be — a minimum amount of

extra information that would guarantee near-zero error. Since longer codes take longer to send, a minimum code length implied a maximum transmission rate — the Shannon limit. Finally, Shannon proved that codes approaching that limit must exist. But he didn't show how to find them.

For the next 45 years, researchers sought those codes. Along the way, there were improvements of the kind that helped increase modem speeds from 9.6 kilobits per second to 14.4 kilobits per second in the early 1980s. But according to Muriel Médard, a professor of computer science and electrical engineering at MIT, proposed codes tended to run up against a limit called the computational cutoff rate. That rate varied according to the transmission power and noise characteristics of a channel, but in practical communications systems, it might be only halfway to the Shannon limit.

Then, in 1993, at the Institute of Electrical and Electronics Engineers' International Communications Conference, Alain Glavieux and Claude Berrou of the École Nationale Supérieure des Télécommunications de Bretagne presented a new set of codes that they claimed came very close to the Shannon limit. "People almost laughed them out of the room," Médard says, "especially because they were not coming from the coding side; they were coming from the electronics side." The researchers had developed their codes — dubbed "turbo codes" — largely through trial and error and had no elegant formal explanation for why they worked so well. Nonetheless, subsequent investigation quickly confirmed their results.

Turbo codes are so-called iterative codes, which means that the decoder makes a series of guesses about what the encoded message is supposed to be. Each successive guess is fed back into the decoder, and the result is a more refined guess. Ideally, repeating the process over and over will get the error rate as low as you want.

The startling performance of turbo codes mobilized researchers to try to explain why they worked so well. Within a few years, investigation of iterative coding schemes had yielded a perhaps even more surprising result: a set of codes that worked at least as well as turbo codes had been around since 1960, when they were introduced in the MIT doctoral thesis of Robert Gallager.

Quiet revolution

The power of Gallager's codes went unappreciated for so long because the decoding process he proposed was simply too complicated for 1960s-era technology. Which is ironic, since simplifying the decoding process was his motive in creating the codes. "The crux of the whole thing was, How do you design a good decoding algorithm?" says Gallager, who taught at MIT from 1960 to 2001 and still supervises graduate students as a professor emeritus. "And then given that idea for how to do that, how do you generate codes that you can actually decode in this way?" At the time, however, research on new coding schemes frequently depended on statistical claims about the performance of hypothetical ideal decoders. For researchers like Gallager, who were trying to develop codes that approached the Shannon limit, specifying a concrete decoding algorithm at all was already an uncommon step in the direction of practical deployment.

Gallager's codes use so-called parity bits — extra bits that contain information about message bits. One parity bit might indicate, say, whether the sum of message bits 1, 2, and 4 is even or odd; the next parity bit might do the same for message bits 3, 4, and 6; and so on, through successive triplets of bits. Reliable information about any two bits in a triplet conveys reliable information about the third. "Iterative techniques involve making a first guess of what a received bit might be and giving it a weight according to how reliable it was," says David Forney, an adjunct professor in MIT's Laboratory for Information and

Decision Systems. “Then maybe you get more information about it because it’s involved in parity checks with other bits, and so that gives you an improved estimate of its reliability — might go the same way, might go the opposite way — and through a series of computations like this, hopefully the thing will converge to where all the bits are known highly reliably.” Problems arise, Forney says, “if you begin to confuse yourself because you’re just feeding back reliabilities that you’ve already used in the same computation, so you get a false positive increase in reliability. It’s like a rumor mill. If you keep hearing the same rumor from the same people again and again, you can all begin to think it’s true, when it’s really just a closed circuit.” The trick to the design of Gallager’s codes, Forney says, was to minimize the likelihood of such closed loops. “It should take a long time for the telephone chain to go all around the world before it gets back to you again,” he says.

To date, Gallager’s codes have enabled the closest approaches to the Shannon limit for a given communications channel — closer even than turbo codes. They’ve been integrated into standards for wireless data transmissions, and computer chips dedicated to decoding Gallager’s codes can be found in commercial cell phones. During their long eclipse, did Gallager have any inkling of how good they were?

“I had a little bit of an inkling, but I also had a suspicion that they well might not be,” Gallager says. “And I spent a long time trying to resolve whether they were or weren’t.” His conclusion was equivocal: “What I showed is that with different classes of these codes, you could achieve positive [transmission] rates. As you change the class to make it more complicated, the rate would continue to increase. If you made it complicated enough, you could reach capacity — but you would probably never decode it. What’s happened since is that people have found ways of somewhat streamlining the way you choose the codes to make them better codes.”

This is the second part of a two-part Explained about information theory. The first part, on the Shannon limit, appeared on Tuesday.

Provided by Massachusetts Institute of Technology