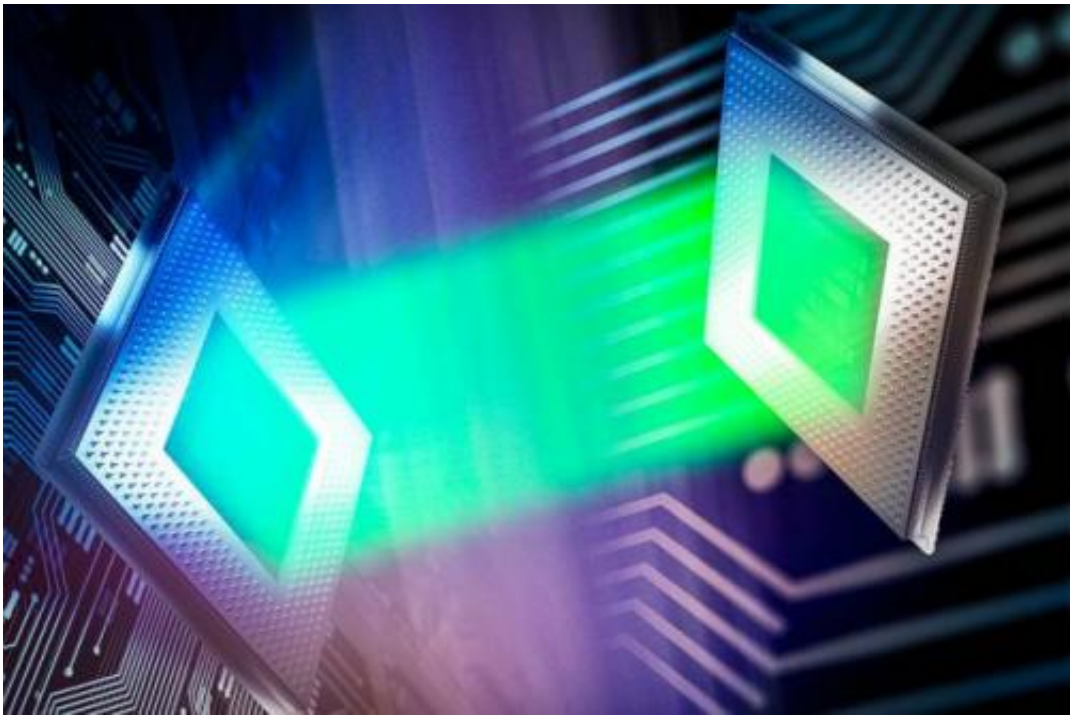
New frontier in error-correcting codes

October 1 2014, by Larry Hardesty



Credit: Jose-Luis Olivares/MIT

Error-correcting codes are one of the glories of the information age: They're what guarantee the flawless transmission of digital information over the airwaves or through copper wire, even in the presence of the corrupting influences that engineers call "noise."

But classical error-correcting codes work best with large chunks of data: The bigger the chunk, the higher the rate at which it can be transmitted error-free. In the Internet age, however, distributed computing is

becoming more and more common, with devices repeatedly exchanging small chunks of data over long periods of time.

So for the last 20 years, researchers have been investigating interactive-coding schemes, which address the problem of long sequences of short exchanges. Like classical error-correcting codes, interactive codes are evaluated according to three criteria: How much noise can they tolerate? What's the maximum transmission rate they afford? And how time-consuming are the encoding and decoding processes?

At the IEEE Symposium on Foundations of Computer Science this month, MIT graduate students past and present will describe the first interactive coding scheme to approach the optimum on all three measures.

"Previous to this work, it was known how to get two out of three of these things to be optimal," says Mohsen Ghaffari, a graduate student in electrical engineering and computer science and one of the paper's two co-authors. "This paper achieves all three of them."

Vicious noise

Moreover, where Claude Shannon's groundbreaking 1948 analysis of error-correcting codes considered the case of random noise, in which every bit of transmitted data has the same chance of being corrupted, Ghaffari and his collaborator—Bernhard Haeupler, who did his graduate work at MIT and is now an assistant professor at Carnegie Mellon University—consider the more stringent case of "adversarial noise," in which an antagonist is trying to interfere with transmission in the most disruptive way possible.

"We don't know what type of random noise will be the one that actually captures reality," Ghaffari explains. "If we knew the best one, we would

just use that. But generally, we don't know. So you try to generate a coding that is as general as possible." A coding scheme that could thwart an active adversary would also thwart any type of random noise.

Error-correcting codes—both classical and interactive—work by adding some extra information to the message to be transmitted. They might, for instance, tack on some bits that describe arithmetic relationships between the message bits. Both the message bits and the extra bits are liable to corruption, so decoding a message—extracting the true sequence of message bits from the sequence that arrives at the receiver—is usually a process of iterating back and forth between the message bits and the extra bits, trying to iron out discrepancies.

In interactive communication, the maximum tolerable error rate is one-fourth: If the adversary can corrupt more than a quarter of the bits sent, perfectly reliable communication is impossible. Some prior interactive-coding schemes, Ghaffari explains, could handle that error rate without requiring too many extra bits. But the decoding process was prohibitively complex.

Making a list

To keep the complexity down, Ghaffari and Haeupler adopted a technique called list decoding. Rather than iterating back and forth between message bits and extra bits until the single most probable interpretation emerges, their algorithm iterates just long enough to create a list of likely candidates. At the end of their mutual computation, each of the interacting devices may have a list with hundreds of entries.

But each device, while it has only imperfect knowledge of the messages sent by the other, has perfect knowledge of the messages it sent. So if, at the computation's end, the devices simply exchange lists, each has enough additional information to zero in on the optimal decoding.

The maximum tolerable error rate for an interactive-coding scheme—one-fourth—is a theoretical result. The minimum length of an encoded message and the minimum decoding complexity, on the other hand, are surmises based on observation.

But Ghaffari and Haeupler's decoding algorithm is nearly linear, meaning that its execution time is roughly proportional to the length of the messages exchanged.

But linear relationships are still defined by constants: y = x is a linear relationship, but so is y = 1,000,000,000x. A linear algorithm that takes an extra second of computation for each additional bit of data it considers isn't as good as a linear algorithm that takes an extra microsecond.

More information: Paper: "Optimal Error Rates for Interactive Coding II: Efficiency and List Decoding." people.csail.mit.edu/ghaffari/… active_protocol2.pdf

Provided by Massachusetts Institute of Technology