# Codes for computationally bounded channels:
## Bridging Shannon and Hamming via computational thinking

VENKATESAN GURUSWAMI

**Preamble.** This is a proposal to the MSR-CMU Center for Computational Thinking on a new research direction in the subject of error-correction. This research endeavor raises new questions of both foundational and practical interest through a computational perspective to bridge between two prevalent and dichotomous schools of thought in information and coding theory. The solutions to these problems will likely call for a novel blend of computational ideas (eg., from cryptography and pseudorandomness) and coding-theoretic methods. Further, the research has the potential for broader significance beyond error-correction, for instance by initiating a study of the algorithmic complexity of hard problems when instances are not picked by an all powerful adversary but are themselves generated by a computationally bounded process (like CAPTCHAs, for instance). Such a theory, if successful, would be a new chapter in computational complexity bridging between worst-case and average-case analyses of algorithmic efficiency.

**Project Description.** Coding theory has had two divergent schools of thought, dating back to its origins, based on the underlying model of the noisy channel. Shannon's theory modeled the channel as a benign stochastic process with a known probability law. Hamming's work suggested a worst-case model, where the channel is subject only to a limit on the number of errors it may cause. These two approaches share several common tools, however in terms of quantitative results, the results in the harsher Hamming model are much weaker. For example, when transmitting bits at a positive rate, error recovery from a fraction $1/4$ of worst-case errors is not possible, whereas even close to a fraction $1/2$ of random errors can be corrected.

Codes that tolerate adversarial errors are attractive because they can transmit reliably over a large range of channels whose true behavior is unknown and hard to model. In contrast, codes tailored to a particular model tend to fail when the model changes — eg., concatenated codes, which can communicate reliably at Shannon capacity with random errors, fail miserably in the presence of burst errors.

This proposal puts forward a line of research aimed at bridging between these models, by constructing codes of rates close to Shannon capacity that work for more general channels, namely *computationally bounded channels*. These are channels that can introduce an arbitrary set of errors as long as (a) the total fraction of errors is bounded by $p$ (the error parameter) and (b) the process which adds the errors is sufficiently "simple" computationally. Such channel models are well-motivated since physical noise processes may be mercurial, but are not computationally intensive. Also, as with codes for worst-case errors, codes for such channels can handle errors whose true behavior is unknown or varying over time.

The above model is in fact not new, and was suggested by Lipton already back in 1994. However, until recently the only solutions in this model required a random secret shared between the sender and receiver (or other setup assumptions like a public key). Such assumptions are not practical in the point-to-point communication settings where codes are used. In a paper with A. Smith to be presented at FOCS 2010, we give the first solution that requires *no* setup assumptions and enables efficient and reliable communication at rates approaching Shannon capacity against channels with natural computational restrictions. In particular, our codes work against oblivious channels (whose error is independent of the codeword) and channels that use only logarithmic space to decide which bits to flip. Our approach is to send the "shared secret" over the channel in a manner that enables the decoder to retrieve it (and the message) despite the adverse effects of the space limited channel. But this must be done with great care (we employ an array of tools from cryptography and pseudorandomness) and indeed it is quite surprising that such coding can be done at all.

This work opens a plethora of exciting questions that I hope to investigate under this proposal. Our

result for logspace channels returns a small list including the correct message. While this is necessary for large error rates ($p > 1/4$), can one achieve Shannon capacity for low error-rates without this restriction? Can we at least go beyond the current bounds for adversarial errors? Can we improve the efficiency of the construction? Could these new codes suffice in any of the many powerful applications of codes list-decodable from adversarial errors?

What about online (aka causal) channels which have no space restriction but whose decision on whether to flip bit $x_i$ or not only depends on bits $x_j$ for $j \leq i$ and not on future bits? Understanding the "capacity" of these and related channels calls for a whole new investigation of classical results, and development of novel ways to integrate computational considerations with combinatorial arguments. The challenge is to quantify the impact the computational power of the adversary has on the limitations it imposes on codes, and finding codes that operate at rates close to the corresponding information-theoretic limit.

**Broader context.** The proposed project is of direct interest to the information theory and the broader communications communities. The best paper award at SIGCOMM 2010 went to a paper "Efficient error estimating coding: feasibility and applications," which gave codes that used shared randomness. The assumption of computationally bounded channels is perfectly viable in their application, so an adaptation of our techniques to eliminate shared randomness while retaining practicality would be very interesting.

Beyond coding theory, this paradigm of "computationally bounded adversaries" could have significantly broader relevance in understanding algorithmic efficiency. There is a rich body of work on the complexity of optimization problems on instances generated by some well-specified random process (eg., models of the web graph, random 3SAT instances, etc.) But what can one say about the complexity of optimization problems when the instances are neither adversarial nor random, but are promised to be generated, *with knowledge of their solution*, by some low complexity process? Various cryptographic primitives rely on the *hardness* of such problems (eg. factoring random Blum integers, solving CAPTCHAs. etc.). But can we design useful *algorithms* in this model that are much better than what is known in the worst-case?

**Collaboration with MSR.** I have visited Principal Researcher Madhu Sudan at MSR New England multiple times, and also hosted him for a talk in computational thinking seminar series at CMU in February 2010. Parikshit Gopalan at MSR Silicon Valley was my postdoctoral fellow from 2007-08; we have co-authored several papers, and continue to interact on research on various coding-theoretic questions (Parikshit is scheduled to speak in CMU's theory seminar in spring 2011). I have shared many research discussions with Sergey Yekhanin at MSR SVC who also works on many aspects of coding theory of interest to me. I have some short visits to MSR SVC planned in 2011. Thus I anticipate plenty of interest in the proposed project amongst multiple researchers at MSR.

**Budget information.** My funding request is support for one Ph.D. student and half-a-month summer salary per year for two years. Assuming that this grant would not incur overhead, but including fringe benefits, I estimate this funding request to be \$75K per year for two years, for a total requested budget of \$150,000 (spread over two years). If awarded, I'd plan to use the funds to support two Ph.D. students, Srivatsan Narayanan and Carol Wang, for one semester each per academic year, and for one summer each over the two-year period. They are currently funded on an NSF CDI grant which will expire in August 2011. Rather than tailoring their research to the objectives of the CDI proposal, I'd very much like to explore aspects of this project with them, and funding from the MSR-CMU center will make this possible.

Depending on how this project takes off, I might plan on submitting a proposal to the CIF program of NSF CCF in 2011-12. As I haven't submitted a proposal to this program before (having only applied to AF and CDI in the past), the opportunity to undertake this project with funding from MSR-CMU center will be invaluable to me, both in pushing forward immediately on some exciting but risky research directions, and in leveraging the research progress over the next couple of years towards external NSF funding.