Approximate and Noisy Computing: Connections to the Information-Theory World

Frederic Sala, Clayton Schoeny, and Lara Dolecek
{fredsala, cschoeny}@ucla.edu, dolecek@ee.ucla.edu
UCLA, Los Angeles, CA 90095

Abstract—**Information and coding-theoretic tools have had a significant impact on communications and data storage systems. Recently, a number of works in this area have begun a study of algorithms and systems operating on faulty or noisy hardware. These efforts have unknowingly paralleled research from other communities in the context of approximate computing. We introduce and discuss a number of relevant papers from the information-theoretic world with the goal of opening a discourse with other communities interested in new computing paradigms.**

I. INTRODUCTION

The information theory (IT) and coding theory communities have been fundamental in establishing the underlying mathematical tools used for communications and data storage systems. Traditionally, these tools have been developed under the assumption that noise occurs only in the transmission channel, not in the peripheral system components. However, due to the relentless shrinking of electronic components, such noise can no longer be ignored and computing under uncertainty has experienced a surge of interest in these communities.

We provide a survey of recent works related to approximate computing, but performed from the information- and coding-theoretic perspectives. Many of these works do not even mention their relevance to approximate computing. Despite this, we believe that a number of ideas introduced in these research efforts can offer a valuable viewpoint to other communities interested in approximate computing. Our goal is to open up an interdisciplinary dialogue; in particular, we wish to ask how the powerful tools developed in information theory can best contribute to novel research in approximate computing.

II. FAULT-TOLERANT SYSTEMS

Most information-theoretic works related to approximate computing study noisy computing, e.g., fault-tolerant systems built from unreliable components. The connection to approximate computing is natural: we relax accuracy or reliability constraints, but allow for operation in an imperfect world.

In this section, we discuss a number of system and algorithm analyses; afterwards we focus on low-density parity-check (LDPC) code decoders. An early work with this flavor is [1], where Hadjicostis and Verghese study fault tolerance for discrete-time dynamic systems, and develop a bound on the computational capacity of linear finite-state machines.

A. Memories

A broad area of interest for coding theorists is memories built from faulty components. When we refer to faulty memories, we not only consider that the memory cells themselves may experience errors (leading to the natural and popular use of error-correcting codes for improved reliability), but also that the encoding and decoding circuitry may itself be faulty as well. Therefore, direct application of error-correcting codes is insufficient, and novel paradigms must be introduced.

Early works include those of Taylor [2] and Kuznetsov [3], introducing mathematical abstractions for faulty components (selecting between, for example, transient or permanent faults, independent or correlated noise, etc...) and deriving the fundamental limits of systems built from these components analogously to the noiseless cases in information theory.

A more explicit application to approximate computing is found in the recent work of Huang, Li, and Dolecek [4]. Practical solutions offered here are based on the idea of introducing an encoding scheme based on the magnitude of the stored data; data of low magnitude can be protected with an error-correcting code by taking advantage of unused bits, while data of large magnitude can be compressed (with the resulting distortion acceptable in the approximate computing scenario) and the resulting saved bits used for code redundancy bits. Huang *et al.* considered machine learning and graph-based inference systems and proved the effectiveness of the adaptive coding framework through image denoising and naïve Bayesian classification examples.

B. Capacity

Another area of interest for information and coding theorists is channel capacity (the maximal data transmission rate allowing for arbitrarily reliable communication.) Traditionally, capacity calculations only consider noise affecting messages transmitted over the channel, and not encoders or decoders. Recently, works have studied capacity in the case of systems with faulty components. In [5], Grover showed that, under certain assumptions, providing equivalently reliable communication in noisy systems (i.e., non-zero Shannon capacity) requires infinite energy. On the other hand, as shown in [6] by Yang *et al.*, if only the encoder is noisy, reliable communication is still possible at non-zero rate. Such analyses are closely related to energy/power considerations. Here too, we note the connection to approximate computing: is it possible to derive similar capacities in approximate computing systems? We observe that attempts in the opposite direction have been made, the goal being to transfer fundamental limits from approximate computing into information and coding theory. Two examples include [7] (in the context of statistical information processing) and [8] (deriving bounds on error magnitudes in different data representation schemes).

III. NOISY DECODERS

A significant portion of the research literature of the last twenty years has focused on low-density parity-check (LDPC) codes. This popularity is a result of the fact that LDPC codes are capacity approaching and their decoders offer manageable complexity [9], [10], [11], [12]. Although message-passing decoders are sub-optimal, they have much lower complexity compared to maximum-likelihood (ML) decoders. The performance of these decoders in the asymptotic regime of very long codes is done through a process known as *density evolution.*

A. Fundamental Performance Limits with Noisy Components

Recently, numerous works have performed a noisy version of density evolution, yielding fundamental performance bounds for decoders built from faulty components. The earliest papers in this topic studied the Gallager B decoder, a message-passing iterative decoder involving a threshold for when the binary message should be modified at variable nodes [13], [14], [15], [16]. These works performed density evolution under various models of component errors. In particular, [13] solves an optimization problem, determining how to best allocate decoding resources in order to optimize the final error probability. In [14], Huang *et al.* derived the approximate density evolution allowing for transient and permanent errors.

There has been substantial research analyzing the performance of LDPC codes through density evolution of noisy min-sum decoders [17], [18]. One surprising result from [18] shows that in certain situations, the noise from the device components can assist the min-sum decoder in escaping certain combinatorial objects that would have resulted in an error during noise-free decoding. Tarighati *et al.* analyzed the effects of noisy message-passing under the sum-product algorithm [19], [20]. In particular, they performed a two-dimensional density evolution analysis and designed LDPC codes based on an EXIT chart analysis [20].

Varshney has conducted comprehensive research on the fundamental performance limits of LDPC codes under a variety of noisy decoders [21], [22]. In [22], Varshney used IT techniques such as entropy dissipation and sphere-packing arguments to calculate upper and lower bounds on storage capacity in memories with unreliable components. These are fundamental, Shannon-type bounds on the code size that are agnostic to the specific code or decoder used.

B. Relevance of Noisy Decoding to Approximate Computing

The research in the previous subsection was performed in reaction to increasing errors in ever-shrinking components. However, by establishing the fundamental limits on decoding with noisy components, the IT and coding researchers were unwittingly also creating mathematical analysis tools for approximate computing. Since LDPC decoders are inherently robust to noise and corruption, they are a prime candidate for approximate computing.

Using the previous ideas, recent work has intentionally introduced noise at the component level-by lowering the voltages-with the goal of saving energy. In [23], Schläfer *et al.* designed a theory-guided unequal error protection technique that achieves up to a 40% energy efficiency increase over currently used techniques. By using dynamic voltage scaling, error probabilities can be controlled on a per iteration basis. A low complexity greedy algorithm solves the optimization problem formulated via the density evolution analysis. The work in [23] is an exact parallel of approximate computing for the IT community: locate a sub-system in which errors can be tolerated, calculate the effect of errors, and establish an energy saving mechanism. An important remaining task is to establish upper bounds on the energy savings that can be attained while still performing reliable computation. Such a task, perhaps, will require the application of ideas from thermodynamics.

Another type of component-error that LDPC decoders are resilient to is timing violations. In [24], Leduc-Primeau *et al.* minimized energy consumption by designing a framework which allows timing faults by taking a quasi-synchronous approach to designing the LDPC decoder processing circuit. By allowing these timing violations within their framework, they were able to reduce energy consumption by up to 28%.

The study of belief-propagation algorithms operating on noisy hardware is not limited to LDPC decoding. A variety of novel algorithms have been created, in conjunction with the inherent error-resiliency in general graph-based decoding, to reduce the energy required for successful decoding. Huang, Li, and Dolecek studied this problem in the general setting of inference in graphical models in [25], with applications to machine learning. In this work, two belief propagation (BP) schemes, called censoring BP and averaging BP, that are robust to computation noise were introduced. Additionally, in [26], Li analyzes the performance of orthogonal matching pursuit on faulty circuits, showing the potential of iterative inference algorithms beyond LDPC decoders.

IV. CONCLUSION AND FUTURE WORK

In this work, we briefly surveyed some of the major themes and research works on approximate computing from the point of view of the information theory community. We predominantly focused on fault-tolerant systems and noisy decoding. We note that other information-theoretic ideas relevant to approximate computing also exist; consider, for example, rate-distortion theory, which can be applied to machine learning tasks such as subspace classification [27], [28].

We wish to bridge the gap between the information-theoretic world and other communities by having the following questions answered:

- What are the right models for approximate computing?
- What are reasonable – and unreasonable – assumptions?
- What kind of results from the IT community would be most useful?
- How can we best integrate theory and practice?

We hope our survey will clarify our field's ideas and approaches towards noisy computation to other communities and will inspire many future collaborative works.

REFERENCES

[1] C. N. Hadjicostis and G. C. Verghese, "Coding approaches to fault tolerance in linear dynamic systems," *IEEE Trans. Inf. Theory*, vol. 51, no. 1, pp. 210–228, Jan. 2005.

[2] M. Taylor, "Reliable information storage in memories designed from unreliable components," *Bell System Technical Journal*, vol. 47, no. 3, pp. 2299–2337, Dec. 1968.

[3] A. V. Kuznetsov, "Information storage in a memory assembled from unreliable components," *Problemy Peredachi Informatsii*, vol. 9, no. 3, pp. 100–114, 1973.

[4] C.-H. Huang, Y. Li, and L. Dolecek, "Acoco: Adaptive coding for approximate computing on faulty memories," *IEEE Trans. Commun.*, vol. 63, no. 12, pp. 4615–4628, Dec. 2015.

[5] P. Grover, "Is Shannon-capacity of noisy computing zero?" in *Proc. IEEE Int. Symp. on Inf. Theory (ISIT)*. IEEE, Honolulu, HI, Jun.-Jul. 2014, pp. 2854–2858.

[6] Y. Yang, P. Grover, and S. Kar, "Can a noisy encoder be used to communicate reliably?" in *Proc. IEEE Allerton Conference on Communication, Control, and Computing*. IEEE, Monticello, IL, Sep.-Oct. 2014, pp. 659–666.

[7] S. Malik, N. Verma, S. Mitra, and N. Shanbhag, "Statistical information processing: Extending the limits of approximate computing," in *Workshop on Approximate Computing (WAX)*, Portland, OR, June 2015.

[8] P. Stanley-Marbell, "Encoding efficiency of digital number representations under deviation constraints," in *IEEE Information Theory Workshop (ITW)*, Volos, Greece, June 2009, pp. 203–207.

[9] D. J. MacKay and R. M. Neal, "Near shannon limit performance of low density parity check codes," *IEEE Electronics Letters*, vol. 32, no. 18, pp. 1645–1646, Aug. 1996.

[10] T. J. Richardson and R. L. Urbanke, "The capacity of low-density parity-check codes under message-passing decoding," *IEEE Trans. Inf. Theory*, vol. 47, no. 2, pp. 599–618, Feb. 2001.

[11] S.-Y. Chung, G. D. Forney, T. J. Richardson, and R. Urbanke, "On the design of low-density parity-check codes within 0.0045 db of the shannon limit," *IEEE Communications Letters*, vol. 5, no. 2, pp. 58–60, Feb. 2001.

[12] M. P. Fossorier, M. Mihaljević, and H. Imai, "Reduced complexity iterative decoding of low-density parity check codes based on belief propagation," *IEEE Trans. Commun.*, vol. 47, no. 5, pp. 673–680, May 1999.

[13] S. Yazdi, C.-H. Huang, and L. Dolecek, "Optimal design of a gallager b noisy decoder for irregular LDPC codes," *IEEE Communications Letters*, vol. 16, no. 12, pp. 2052–2055, Dec. 2012.

[14] C.-H. Huang, Y. Li, and L. Dolecek, "Gallager b LDPC decoder with transient and permanent errors," *IEEE Trans. Commun.*, vol. 62, no. 1, pp. 15–28, 2014.

[15] F. Leduc-Primeau and W. J. Gross, "Faulty gallager-b decoding with optimal message repetition," in *Proc. IEEE Allerton Conference on Communication, Control, and Computing*, Monticello, IL, Oct. 2012, pp. 549–556.

[16] B. Vasic and S. K. Chilappagari, "An information theoretical framework for analysis and design of nanoscale fault-tolerant memories based on low-density parity-check codes," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 54, no. 11, pp. 2438–2446, Nov. 2007.

[17] A. Balatsoukas-Stimming and A. Burg, "Density evolution for min-sum decoding of LDPC codes under unreliable message storage," *IEEE Communications Letters*, vol. 18, no. 5, pp. 849–852, May 2014.

[18] C. Kameni Ngassa, V. Savin, E. Dupraz, and D. Declercq, "Density evolution and functional threshold for the noisy min-sum decoder," *IEEE Trans. Commun.*, vol. 63, no. 5, pp. 1497–1509, May 2015.

[19] A. Tarighati, H. Farhadi, and F. Lahouti, "Performance analysis of noisy message-passing decoding of low-density parity-check codes," in *IEEE Int. Symp. on Turbo Codes & Iterative Information Processing*, Brest, France, Sep. 2010, pp. 389–393.

[20] ——, "Design of LDPC codes robust to noisy message-passing decoding," *arXiv preprint arXiv:1501.02483*, Jun. 2015.

[21] L. R. Varshney, "Performance of LDPC codes under faulty iterative decoding," *IEEE Trans. Inf. Theory*, vol. 57, no. 7, pp. 4427–4444, Jul. 2011.

[22] ——, "Toward limits of constructing reliable memories from unreliable components," in *Proc. IEEE Inf. Theory Workshop (ITW)*, Jeju Island, Korea, Oct. 2015, pp. 114–118.

[23] P. Schläfer, C.-H. Huang, C. Schoeny, C. Weis, Y. Li, N. Wehn, and L. Dolecek, "Error resilience and energy efficiency: An LDPC decoder design study," in *Proc. Design, Automation and Test in Europe (DATE)*, Dresden, Germany, Mar. 2016 (pre-print).

[24] F. Leduc-Primeau, F. R. Kschischang, and W. J. Gross, "Energy optimization of LDPC decoder circuits with timing violations," in *Proc. IEEE Int. Conf. on Commun. (ICC)*, London, UK, Jun. 2015, pp. 412–417.

[25] C.-H. Huang, Y. Li, and L. Dolecek, "Belief propagation algorithms on noisy hardware," *IEEE Trans. Commun.*, vol. 63, no. 1, pp. 11–24, Jan. 2015.

[26] Y. Li, Y. Chi, C.-H. Huang, and L. Dolecek, "Orthogonal matching pursuit on faulty circuits," *IEEE Trans. Commun.*, vol. 63, no. 7, pp. 2541–2554, Jul. 2015.

[27] M. Nokleby, M. Rodrigues, and R. Calderbank, "A rate-distortion framework for supervised learning," in *IEEE Information Workshop on Machine Learning for Signal Processing (MLSP)*, Boston, MA, Sept. 2015.

[28] M. Nokleby, A. Beirami, and R. Calderbank, "Information-theoretic criteria for the design of compressive subspace classifiers," in *IEEE International Conference on Acoustic, Speech and Signal Processing (ICASSP)*, Florence, Italy, May 2014, pp. 3067–3071.