Efficient Constrained Codes That Enable Page Separation in Modern Flash Memories

Presenter: ***Simeng Zheng***, Graduate Student, UCSD ECE

Researcher: ***Simeng Zheng***, Graduate Student, UCSD ECE

Collaborators: ***Ahmed Hareedy***, Professor, METU

Robert Calderbank, Professor, Duke

Advisor: ***Paul H. Siegel,*** Professor, UCSD ECE

The pivotal storage density win achieved by solid-state devices over magnetic devices in 2015 is a result of multiple innovations in physics, architecture, and signal processing. One of the most important innovations in that regard is enabling the storage of more than one bit per cell in the Flash device, i.e., having more than two charge levels per cell.

Constrained coding is used in Flash devices to increase reliability via mitigating inter-cell interference that stems from charge propagation among cells. Recently, capacity-achieving constrained codes were introduced to serve that purpose in modern Flash devices which have more than two levels per cell. While these codes result in minimal redundancy via exploiting the underlying physics, they result in non-negligible complexity increase and access speed limitation since pages cannot be read separately.

In this work, we suggest new constrained coding schemes that have low-complexity and preserve the desirable high access speed in modern Flash devices. The idea is to eliminate error-prone patterns by coding data either only on the left-most page (binary coding) [1] or only on the two left-most pages (4-ary coding) while leaving data on all the remaining pages uncoded. Our coding schemes work for any number of levels $q \geq 4$ per cell, offer systematic encoding and decoding, and are capacity-approaching. Since the proposed schemes enable the separation of pages, except the two left-most pages in the case of 4-ary coding, we refer to them as *read-and-run* (RR) constrained coding schemes. The 4-ary RR coding scheme reduces the rate loss incurred by the binary RR coding schemes, and we show that our 4-ary RR coding scheme is also competitive when it comes to complexity and error propagation. We analyze the new RR coding schemes and discuss their impact on the probability of occurrence of different charge levels.

We empirically evaluated the performance of the proposed RR codes on a triple-level cell (TLC) Flash memory using our Flash characterization platform. Experimental results reveal significant program/erase (P/E)-cycle lifetime gains using both binary and 4-ary constrained schemes.

[1] A. Hareedy, S. Zheng, P. H. Siegel, and R. Calderbank, "Read-and-run constrained coding for modern Flash devices," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Seoul, South Korea, May 2022, pp. 1-6.

Error Correction for Distributed Neural Networks using Deep Reinforcement Learning

Presenter: ***Wenyu Peng***, Graduate Student, ECE

Researcher: ***Wenyu Peng***, Graduate Student, ECE

Advisor: ***Paul H. Siegel,*** Professor, CMRR

Collaborators: Simeng Zheng, Graduate Student, ECE

When implementing neural networks (NeuralNets) in hardware, it is crucial to store their weights in memory devices. However, the accumulation of noise in the stored weights can lead to a degradation in the NeuralNet's performance [1]. Unlike traditional error correction techniques employed in data storage, the optimization objective here is to enhance the NeuralNet's performance after error correction, rather than solely minimizing the Uncorrectable Bit Error Rate in the protected bits. A significant hurdle arises due to the deep nature of NeuralNets, which often contain millions to hundreds of millions of weights, resulting in a substantial redundancy overhead for error correction codes (ECCs)[2]. Moreover, the relationship between the weights and the NeuralNet's performance can be highly intricate. This challenge is further exacerbated when the weights are distributed across numerous devices in a datacenter[3], each device having a distinct bit error rate. To tackle this challenge, we propose employing deep reinforcement learning to determine which bits are essential to protect, achieving an optimized tradeoff between ECC redundancy and NeuralNet performance. The experiments results show that we can find the critical layer and protect it by using the Selective Protection Scheme in a distributed scenario.

[1] P. Upadhyaya P, X. Yu, J. Mink, et al., “Error correction for hardware-implemented deep neural networks,” *Proc. Non-Volatile Memories Workshop*, March 2019, pp. 1-2.

[2] K. Huang, P.H. Siegel, A. Jiang, “Functional error correction for robust neural networks,” *IEEE Journal on Selected Areas in Information Theory*, vol. 1, no. 1, pp. 267-276, May 2020.

[3] K. Hazelwood, S. Bird, D. Brooks, et al., “Applied machine learning at facebook: A datacenter infrastructure perspective,” *Proc. IEEE International Symposium on High Performance Computer Architecture (HPCA)*, 2018, pp. 620-629.

Coding for DNA Synthesis

Presenter: ***Andrew Tan,*** Graduate Student, ECE

Researcher: ***Andrew Tan,*** Graduate Student, ECE

Advisor: ***Paul Siegel***, Professor, CMRR

Collaborator: ***Yi Liu,*** CMRR (PhD 2020)

Abstract: A significant obstacle in making DNA data storage a feasible technology is the exorbitant cost of DNA synthesis, which is the process of generating DNA sequences to represent sequences of bits [1]. Therefore, it is of importance to improve cost efficiency by maximizing the amount of encoded information for a given cost. Since DNA synthesis involves an array-based, parallel construction of strands, then cost can be quantified by the amount of time (i.e. iterations or cycles) it takes to produce the strands, and a graphical model can be obtained and analyzed using methods outlined in [2] to determine a maximum information rate and an optimal encoding scheme. Here, we will review an analytical characterization of DNA synthesis with costly constrained coding, and we will provide a code construction [3] that yields asymptotically optimal codebooks.

References:

[1] A. Lenz, Y. Liu, C. Rashtchian, P. H. Siegel, A. Wachter-Zeh, and E. Yaakobi, "Coding for Efficient DNA Synthesis," IEEE ISIT, May 2020.

[2] A. Khandekar, R. J. McEliece, and E.R. Rodemich, "The Discrete Noiseless Channel Revisited," Coding, Communications, and Broadcasting. Badlock: Research Studies Series Ltd., UK, pp. 115 – 137, 2022.

[3] Y. Liu, Y. Li, P. Huang, and P. H. Siegel, "Rate-Constrained Shaping Codes for Finite-State Channels with Cost," IEEE ISIT, pp. 1354 -1359, 2022.

Temperature-dependent magnetic properties of Heat Assisted Magnetic Recording media and their implications on potential methods for sanitization of the recording medium

Sergio A. Montoya, Frederick. E. Spada
Center for Memory and Recording Research
University of California San Diego

Degaussing is commonly employed for sanitization of magnetic recording storage products containing conventional perpendicular recording media at product end-of-life. This talk describes experimental work detailing the temperature- and field-dependent magnetic properties of Heat Assisted Magnetic Recording (HAMR) media [1-3] comprised of a recording layer with high perpendicular magnetic anisotropy and a soft underlayer. The experiments were motivated by the need to develop suitable sanitization methods and technologies for Energy Assisted Magnetic Recording (EAMR) storage products containing HAMR or Microwave Assisted Magnetic Recording (MAMR) media. We first discuss perpendicular magnetic hysteresis measurements conducted at different temperatures, which show that both the saturation field H_{sat} and coercivity field H_c rapidly drop with increasing temperature ($e.g., H_{sat}$ = 5.34 *Tesla* and H_c = 3.07 *Tesla* at *300K*, decreasing *to* H_{sat} = 1.07 *Tesla* and H_c = 0.05 *Tesla* at *600K)*. We then focus on media remanent DC demagnetization behavior at various temperatures which reveals how temperature influences the switching field distributions of HAMR media. Finally, we show that, even without the application of external magnetic fields, HAMR media can be thermally demagnetized at modest temperatures via thermal cycling under different heating conditions. Overall, our results suggest potential pathways for modifying the current degaussing approach to achieve sanitization of EAMR storage products.

1. A. Chernyshow, D. Treves, T. Lee, C. Papusoi, H. Yuan, A. Ajan, and R. Acharya. *Measurement of magnetic properties relevant to heat assisted magnetic recording.* IEEE Trans. Magnt. 49, 7 (2013).
2. D. Weller, G. Parker, O. Mosendz, A. Lyberatos, D. Mittin, N. Y. Sadonova, and M. Albrecht. *FePt heat assisted magnetic recording media*. J. Vac. Sci. Technol. B 34, 060801 (2016).
3. M. Chapline, C. Papusoi, A. Ajan, P. Dorsey, M. Desai, and R. Acharya. *Characterization of magnetic properties of HAMR media using temperature dependence of the AC susceptibility*. IEEE Trans. Magnt. 52, 2 (2016).

Critical slowing of the spin and charge density wave order in thin film Cr following photoexcitation

Presenter: Sheena Patel, Graduate Student Researcher, CMRR
Advisor: Eric Fullerton, Director, CMRR
Collaborators: Andrej Singer, Assistant Professor, Cornell University
Oleg Shpyrko, Professor and Physics Department Chair, UCSD

In strongly correlated systems, the coupling between spin, charge, and lattice degrees of freedom leads to the emergence of complex order such as antiferromagnetism and density wave systems. One such system is single-crystal chromium, the archetypal itinerant antiferromagnet, which exhibits a spin density wave (SDW) below its Néel temperature of 311 K that is incommensurate with the lattice. This results in a periodic lattice distortion or strain wave and a charge density wave (CDW) which are coupled to and a second harmonic of the SDW ordering and can be measured by X-ray diffraction as satellite peaks around the (002) Bragg peak [1,2,3]. Confinement of the geometry to a thin film results in the pinning of order at the interfaces [2,4] and the quantization of the ordering wavevector with hysteretic switches in wavevector as a function of temperature [2,5]. Here we conduct a time-resolved X-ray diffraction experiment at the X-ray free-electron laser at LCLS of a thin film of Cr to measure the dynamics of the periodic lattice distortion following photoexcitation from different temperatures. The evolution of the system back to the ground state occurs on different time scales at different temperatures, with a slowing down by more than two orders of magnitude as we approach the hysteretic regions, indicative of an increasing activation energy for reorientation of the spins. We extend existing phenomenological Landau modeling of the mean-field free energy of incommensurate SDW/CDW systems to include pinning of available wavevectors and reproduce the observed temperature dependence of the wavevector and energy barriers. Thus, we use confinement of the geometry and the unique capabilities of X-ray free electron lasers to assess the energy of system the through the time domain.

Work at UCSD was supported by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences, under Contracts No. DE-SC0018237 and DE-SC0001805.

[1] E. Fawcett, "Spin-density-wave antiferromagnetism in chromium," *Rev. Mod. Phys.*, **60** (1988).
[2] A. Singer, S. K. K. Patel, et al., "Phase coexistence and pinning of charge density waves by interfaces in chromium," *Phys. Rev. B*, **94** (2016).
[3] A. Singer, S. K. K. Patel, et al., "Photoinduced enhancement of the charge density wave amplitude," *Phys. Rev. Lett.*, **117** (2016).
[4] H. Zabel, "Magnetism of chromium at surfaces, at interfaces and in thin films," *J. Phys. Condens. Matter*, **11** (1999).
[5] R. K. Kummamuru and Y.-A. Soh, "Electrical effects of spin density wave quantization and magnetic domain walls in chromium," *Nature*, **452** (2008).

Spin Torque effects during magnetic phase transition of epitaxial Ho(*0001*) thin films.

Biswajit Sahoo, Yuxuan Xiao, Eric E Fullerton

The operation of numerous spintronic devices are enabled through spin currents and thus, it is important to study how the generation, polarization and transmission of spin currents can be controlled by the local magnetic/electronic environment. Holmium is a rare earth metal which undergoes a paramagnetic to anti-ferromagnetic (AFM) transition at T_N=131K and an AFM to ferromagnetic (FM) transition at T_c=20K. The AFM phase is characterized by an incommensurate spin helix in which the average moments are ferromagnetically aligned within the basal planes but rotate from plane to plane along the *c* axis with an average turn angle varying from 50^0/layer at T_N to to 30^0/layer at T_C . Ho is further expected to have a good charge to spin conversion efficiency owing to its high atomic number and initial results of ploy-crystalline Ho films at room temperature show a spin Hall Angle (SHA) of 0.14[1]. Here, we explore the SHA of epitaxial Ho thin films in the paramagnetic, AFM and FM phases to observe the effect of magnetic ordering on the generation of spin current. We deposited epitaxial Ho (*0001*) using Al_2O_3(*0001*) as a substrate with a 3 nm W(*110*) as a seed layer and confirm the different magnetic phases via magnetometry and transport measurements. We create the following thin-film stack: Al_2O_3(0001)/W(3)/Ho(10)/W(1)/Py(7)/W(1.5) (where the thicknesses are in nm) and study the SHA of Ho via spin torque ferromagnetic resonance experiments. We systematically vary the temperature down to 70K to study the charge to spin conversion efficiency. By relating various magnetic parameters such as effective magnetization, gilbert damping and SHA at various temperatures, we explore how the magnetic ordering affects the spin dynamics in this system.

References:

[1] N. Reynolds *et al.* Phys. Rev. B **95**, 064412 (2017)

Acknowledgement:

This work was supported as part of Quantum Materials for Energy Efficient Neuromorphic Computing (Q-MEEN-C), an Energy Frontier Research Centre funded by the U.S. Department of Energy (DOE), Office of Science, Basic Energy Sciences (BES), under Award # DE-SC0019273.

THE USE OF AN ARTIFICIAL CORNEA FOR VALIDATION OF A NOVEL INTRAOCULAR PRESSURE MEASUREMENT DEVICE

Presenter: *Avinash Laha, Graduate Student, CMRR & MAE*

Advisor: *Dr. Frank Talke, CMRR & MAE*

Abstract

Today, glaucoma is the leading cause of irreversible blindness in the world. Its highest risk factor is related to an elevated intraocular pressure (IOP) within the eye globe that can cause stress on the optic nerve and deteriorate the visual field in patients. Regular monitoring of the IOP can be crucial in detecting early onset of glaucoma, especially since therapeutic and surgical cures might not exist in the later stages. To improve the prospect of self-examining ophthalmic devices, we have previously developed a handheld 3-in-1 device consisting of 3 commonly used glaucoma screening instruments. This includes a compact slit lamp biomicroscope, visual acuity tester, and a non-contact tonometer. The aim of this device is to enable easy self-examination of the eye more frequently, especially for patients physically unable to travel to in-person clinic appointments to perform them.

However, it is essential to ensure these measurements accurately capture a patient's optical health, and are standardized in relation to current commercially available devices. To validate the tonometer measurement principle, we have recently developed an artificial cornea using different silicone materials to study the deflection of the cornea as a function of the intraocular pressure inside the human eye. Samples were fabricated using 2-part 3D printed molds that accurately capture the geometry of the human cornea. Young's moduli for a range of selected silicones were evaluated and compared with that of human cornea as well. Eye pressure was simulated using a water-filled manometer channel and varied between 0 and 40 mmHg. Readings were obtained from both a commercial and the 3-in-1 device's tonometer to test the designed artificial corneas at these pressures. Results from both devices validate the use of such a phantom eye apparatus to test similar ophthalmic instrumentation before clinical testing on humans.

Presenter: Darin Tsui
Lab Group: Frank Talke

Optical Surgical Navigation: A Promising Low-cost Alternative.

Computer-assisted surgical navigation has become a popular solution in difficult operations where a high amount of precision is required. Current state-of-the-art methods of surgical navigation involve tracking reflective 3D marker spheres using IR stereoscopic cameras. However, the cost of implementing such systems may not be affordable for smaller healthcare systems. Here, we propose that fully optical navigation has the potential to be a viable alternative to state-of-the-art reflective marker navigation. We use fiducial ArUco markers to facilitate the tracking of real-time position. Using two inexpensive cameras, we design and calibrate a stereoscopic camera to record the 3D position of an ArUco marker moving through space along a positioning platform. Additionally, we explore the possibility of using different color spaces and physical marker colors to improve the detection percentage and accuracy of markers. We identified that black-and-white ArUco markers using the Hue, Saturation, and Lightness (HSL) color space gave a mean error of 5.38 mm. Using the Red, Green, and Blue (RGB) color space gave the highest detection percentage for the same ArUco markers. In the future, the mean error can be reduced by increasing camera quality and by using a multi-stereoscopic camera setup.

Presenter: Brian Li
Lab Group: Frank Talke

3D Printed Gyroid Elastomer and Silicone Composite for Controlled Anisotropy Simulating Human Tissue

This research proposes a novel approach for designing a composite material that can mimic the strain hardening and anisotropic mechanical characteristics of vaginal tissue. The composite consists of a soft elastomeric silicone matrix and a thermoplastic polyurethane 3D scaffold. The strain hardening effect of vaginal tissue was achieved through the altering of the scaffold's gyroid structure. As the composite deforms, the mechanical response changes from silicone-dominated to polyurethane-dominated. The composite's anisotropic properties were achieved by introducing anisotropy in the gyroid structure to replicate the effects of collagen fiber orientation in vaginal tissue. The change in mechanical characteristics was investigated using uniaxial tensile testing. Three types of distorted gyroids were investigated, x-axis lengthened, y-axis lengthened, and x and y-axis lengthened unit cells.

Presenter: Suyash Mahar
Lab Group: Steve Swanson

Puddles: Application-Independent Recovery and Location-Independent Data for Persistent Memory

ABSTRACT

With Puddles, we argue that current work has failed to provide a comprehensive and maintainable in-memory representation for persistent memory.

PM data should be easily mappable into a process address space, shareable across processes, shippable between machines, consistent after a crash, and accessible to legacy code with fast, efficient pointers as first-class abstractions.

While existing systems have provided niceties like mmap()-based load/store access, they have not been able to support all these necessary properties due to conflicting requirements.

We propose Puddles, a new persistent memory abstraction, to solve these problems. Puddles provide application-independent recovery after a power outage; they make recovery from a system failure a system-level property of the stored data rather than the responsibility of the programs that access it. Puddles use native pointers, so they are compatible with existing code. Finally, Puddles implement support for sharing and shipping of PM data between processes and systems without expensive serialization and deserialization.

Compared to existing systems, Puddles are at least as fast as and up to 1.34× faster than PMDK across YCSB workloads. Finally, we showcase Puddles' ability to relocate data using a data-aggregation workload that results in a 2.6× speedup over PMDK.

Presenters: Keli Wang & Jeffrey Liu
Lab Group: Gert Cauwenberghs

Accelerating Spiking Neural Network Emulation on an Efficient Reconfigurable Hardware Platform

In this work, we introduce a set of principles for the optimization of hardware-aware software for compiling general spiking neural networks to run in real-time on a large-scale, reconfigurable event-driven neuromorphic computing platform. Our hardware-software co-design framework is optimized for run-time, massively parallel processing, and hierarchical address-event routing (HiAER) of spikes while promoting memory-efficient network storage and execution. Implemented on a high-end Field Programmable Gate Array (FPGA), the HiAER-Spike platform leverages storage of synaptic look up tables in High Bandwidth Memory coupled with a hierarchical interconnect system to allow for real time simulation of large scale spiking neuronal networks, on the scale of 160 million neurons and 40 billion synapses. This architecture is coupled with a Python programming interface, agnostic to hardware-level details in the implementation, that shields the user from complexity in the configuration and execution of general spiking neural networks on neuromorphic hardware while allowing for interoperability with already widely used SNN simulation packages such as SpikingJelly and SNNTorch.

Efficient Approach for Solving Periodic Unit Cell Problems in Micro-Magnetic Simulations

Presenter: ***Fangzhou Ai***, Ph.D. student, Material Science and Engineering Department

Advisor: ***Prof. Vitaliy Lomakin***, Professor, Electrical and Computer Engineering Department

Abstract: Periodic structures commonly arise in micro-magnetic modeling, presenting unique challenges for achieving accurate solutions. The periodic Green's function (PGF) is a well-established approach for addressing such problems. However, in static micro-magnetic modeling scenarios where the free space wavenumber vanishes and there is no phase shift between adjacent unit cells, the direct application of the PGF becomes infeasible due to its divergence. This study introduces a fast and effective technique for estimating the periodic potential by subtracting the near field contributions from neighboring cells and subsequently interpolating the remaining field. We demonstrate that despite the divergence of the PGF itself, in a neutral system where the total "charges" (in our case, magnetic monopoles) sum to zero, the potential remains convergent and computable. By isolating the near field, we achieve a considerably smoother distribution of the far field, enabling the computation of the far field potential on a sparser grid that is independent of the original source distribution. This advancement greatly accelerates the overall procedure. To control the error, we can adjust the density of the sparse grid, the number of subtracted near-zone cells, and the interpolation order (Fig. 1). In theory, our method allows interpolation of the far field to arbitrary precision. However, in practice, we strike a balance between performance and precision by selecting a moderate error threshold. Importantly, our method is highly versatile and applicable to a broad range of neutral systems, extending beyond micro-magnetic issues.

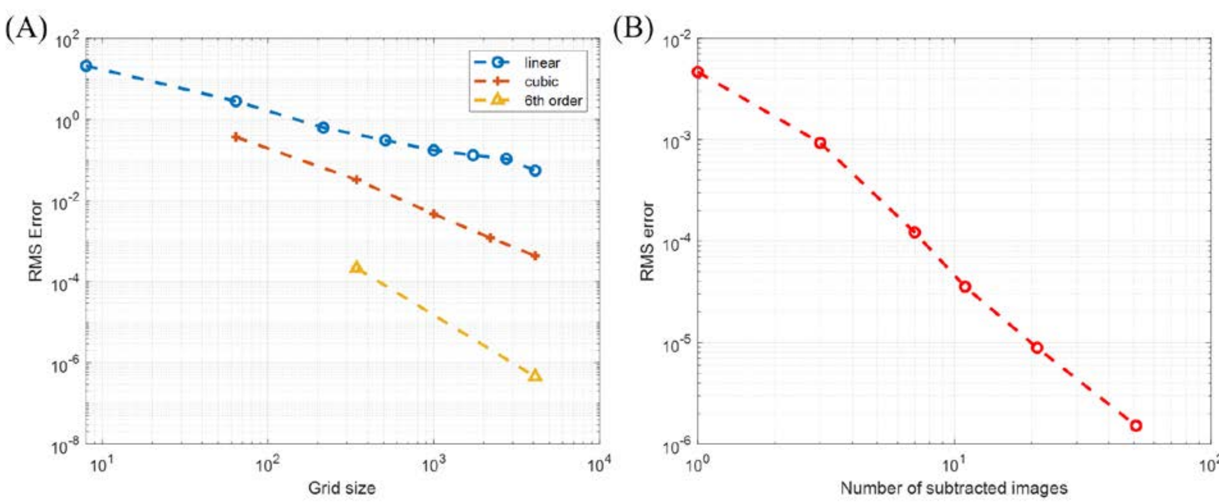

Fig 1. Illustration of the Error between Interpolated Far Field and Theoretical Value in 1D Periodicity along the x-axis. (A) Error comparison for 1st, 3rd, and 6th order interpolation with varying sparse grid sizes, considering the subtraction of only 0th and 1st unit cells. (B) Error analysis of cubic interpolation with 1000 sparse grid points and different numbers of subtracted near-zone unit cells.

Stochasticity hiding behind synchronization.

Erbin QIU, Yuanhang Zhang, Max Di Ventra and Ivan Schuller

We investigated the synchronization of spiking nano-oscillators emerging from thermal interactions due to the close physical proximity of the devices. Based on the thermally coupled spiking oscillators, I discovered an unexpected behavior behind the collective modes. The two spiking oscillators are well synchronized both in phase and frequency from a large time scale (μs). However, from nanosecond scale level, the spike from either oscillator is randomly triggered ahead of that of the other if we zoomed the time window into the two 'synchronized' spikes. The randomness hiding behind the synchronization can be used to directly implement a compact random number generator and other interesting neuromorphic functionalities.

This work was funded by CMRR and the Air Force Office of Scientific Research under award number FA9550-22-1-0135.