

ReRAM セルの信頼性ばらつきを許容する 機械学習向け高速 Approximate ストレージ

松井千尋¹ 竹内健¹

概要：機械学習アプリケーション向けに、TaO_x ReRAM セルの信頼性ばらつきを許容する Approximate ストレージを提案する。メモリセル間の書き込み回数のばらつきを削減するウェアレベリング技術を用いない場合、アプリケーションワークロードは ReRAM セルの信頼性のばらつきを引き起こす。ストレージシステムと ReRAM デバイス特性を組み合わせることにより、ReRAM セルエラーのばらつきが減少しストレージ性能が7倍向上する。

キーワード：Approximate ストレージ, TaO_x ReRAM, 信頼性

1. はじめに

機械学習アプリケーションの一つである画像認識は、1-10%の TaO_x ReRAM セルエラーを許容し、画像認識精度が低下しないことが分かっている[1]。提案するメモリセルのばらつきを許容するコンピューティングは、ウェアレベリングの余分なデータコピーをなくし、セル数が多くエラーの低いメモリセルを対象とすることで弱い ECC を適用する[2]。ストレージ性能・エネルギー・信頼性のトレードオフを解決するために、提案のメモリセルのばらつきを許容するコンピューティングは TaO_x ReRAM セルのエラーを許容する。

2. ReRAM セルの信頼性ばらつきを許容するストレージ

2.1 ReRAM セルに発生するエラー

データセンタで運用されるアプリケーションワークロードによって、ストレージ内でデータを保存する ReRAM に複雑なエラーが発生する。アプリケーションワークロードで頻りに読み出し・書き込みされるデータをホットデータと呼ぶ。頻りに書きこみが行われると ReRAM セルの書き換え回数が高くなり、導電性フィラメントの酸素欠陥 (V_o) 密度が低下することによって、低抵抗状態 (LRS) にエラーが起こる。また、同一セルの頻りに読み出し動作によって LRS のセル電流 (I_{CELL}) が減少するため、リードディスタブエラーが発生する。読み出し・書き込みともに頻りに起こらない場合、ReRAM セルの V_o が水平方向に拡散することによってデータ保持エラーが発生する[3, 4]。

2.2 システム・回路・デバイスレベルの Approximation

ReRAM に生じるこれらのエラーを補正するために、ReRAM デバイス測定結果と TLM ベースのエミュレータを組み合わせ、システム・回路・デバイスの設計プラットフォームを用いて評価する[2]。たとえばウェアレベリングは、すでにセルに書き込まれたデータを別のセルにコピーすることで、頻りに書き込みによる ReRAM セルの書き換え回数のばらつきを平均化する[5]。つまりシステムレベルの

Approximate 手法として、ウェアレベリングによるデータコピーを削減しストレージ性能を向上させる。

一方でウェアレベリングをしないことで頻りに書き込みによるエラーが増加するが、書き込み頻度の低い ReRAM セルのエラーは低下する。ECC にはエラー訂正強度・復号時間・セル面積のトレードオフがあるため、エラー率の低いセルをターゲットとすることでデコード時間を短縮しチップ面積を小さくできる。

また、LRS のテールエラーが許容できる場合、書き込み時のベリファイ動作を削減および低いセット/リセット電圧の適用によって、ReRAM ストレージの性能向上と消費エネルギーが低減できる。

3. おわりに

システム・回路・デバイスレベルで TaO_x ReRAM の信頼性ばらつきを許容するストレージを提案した。ストレージコントローラがアプリケーション特性に応じて異なる Approximation 手法を用いることで、ReRAM ストレージの性能が7倍向上し、消費エネルギーおよびチップ面積が低減する。

謝辞 この成果は、国立研究開発法人新エネルギー・産業技術総合開発機構 (NEDO) の委託業務の結果得られたものです。

参考文献

- [1] Y. Yamaga et al., "5x Reliability Enhanced 40nm TaO_x Approximate-ReRAM with Domain-Specific Computing for Real-time Image Recognition of IoT Edge Devices," in *Symp. VLSI Tech. Dig. Tech. Papers*, 2018, pp. 109-110.
- [2] C. Matsui et al., "Application-Induced Cell Reliability Variability-Aware Approximate Computing in TaO_x-based ReRAM Data Center Storage for Machine Learning," in *Symp. VLSI Tech. Dig. Tech. Papers*, 2019, pp. 234-235.
- [3] D. Ielmini, "Modeling the Universal Set/Reset Characteristics of Bipolar RRAM by Field- and Temperature-Driven Filament Growth," *IEEE TED*, vol. 58, no. 12, pp. 4309-4317, 2011.
- [4] T. Ninomiya et al., "Improvement of Data Retention During Long-Term Use by Suppressing Conductive Filament Expansion in TaO_x Bipolar-ReRAM," *IEEE EDL*, vol. 34, no. 6, pp. 762-764, 2013.
- [5] T. Onagi et al., "Design Guidelines of All Storage Class Memory (SCM) SSD and Hybrid SCM/NAND Flash SSD to Balance Performance, Power, Endurance and Cost," in *Ext. Abstr. SSDM*, 2014, pp. 106-107.

¹ 東京大学大学院
Graduate School of Engineering, The University of Tokyo