

ディープラーニング(深層学習)を用いた画像認識に最適な SSD を開発 ～ データの「価値」を判断することで、300 倍の長寿命化、26%の高速化に成功

学校法人 中央大学

概 要

中央大学 理工学部 教授 竹内 健のグループは、ディープラーニング(深層学習)を用いた画像認識に最適な記憶デバイス(SSD)を開発しました。開発した SSD は画像データの「価値」を判定し、重要なデータは高信頼なメモリセル、重要性が低いデータは信頼性が低いメモリセルに記憶するように制御します。また、ディープラーニングを用いた画像認識では、認識精度が保てれば計算の精度は低くても影響しないことに着目し、読み出しの高速化と高い画像認識を両立する、メモリのエラーを訂正する誤り訂正回路(ECC)を開発しました。これらの制御技術の開発により、SSD の寿命(データの保持時間)を 300 倍長寿命化し、SSD を 26%高速化することに成功しました。

本研究成果は、科学技術振興機構(JST)戦略的創造研究推進事業(CREST)「素材・デバイス・システム融合による革新的ナノエレクトロニクスの創成」における研究課題「デジタルデータの長期保管を実現する高信頼メモリシステム」において得られました。

本研究成果は、2017年4月30日から5月3日に米国オースチンで開催された「IEEE Custom Integrated Circuits Conference」で発表されました。

【研究者】

竹内 健 中央大学理工学部 教授(電気電子情報通信工学科)

【発表(雑誌・学会)】

本研究成果は、2017年4月30日から5月3日に米国オースチンで開催される「IEEE Custom Integrated Circuits Conference」で発表されました。

論文名: 12x Bit-Error Acceptable, 300x Extended Data-Retention Time, Value-Aware SSD with Vertical 3D-TLC NAND Flash Memories for Image Recognition

【研究内容】

1. 背景

車の自動運転、セキュリティや見守り、ロボットのリアルタイム制御など、ディープラーニング(深層学習)などの AI のアルゴリズム解析による画像認識を用いた様々なサービスが注目されています。AI を用いた認識では、認識自体の精度が高ければ、必ずしも認識に用いるコンピュータやデータを処理する集積回路(LSI)には従来ほど高い計算精度を求められることはありません。この観点から、Google が開発した AI 用プロセッサ TPU(Tensor Processing Unit)のように、LSI の計算精度を下げた高速化・低電力化を図ることが注目されています。このように、従来は AI 向け LSI の研究は主に制御・演算を行うプロセッサなどで行われてきました。

他方、クラウドデータセンタ^{注1)}などの小型、高速、低電力なデータの記憶媒体として SSD(ソリッド・ステート・ドライブ)^{注2)}が使われています。SSD はデータを記憶するフラッシュメモリ^{注3)}と、メモリを制御するコントローラなどで構成されています。今後、リアルタイムのストリーム処理などでは膨大なデータを、深層学習を使って学習、推論を考えると考えられ、大容量なデータを記憶する SSD を使って深層学習を実行することが必要になります。

2. 研究内容と成果

我々は、AI に向けたストレージ・SSD を世界に先駆けて開発しました。

具体的には、通常の立体構造(3次元)フラッシュメモリコントローラの3つのメモリ制御技術を開発し、ディープラーニングに耐えうる SSD の高信頼化、高速化と高い画像認識精度を実現することに成功しました。

開発技術の一例として、画像認識における結果を示します。図1はディープラーニングを用いた画像認識と開発した SSD を示しています。ディープラーニングによって画像は128次元の特徴ベクトルに変換され、SSD に記憶されます。画像認識では、入力した(カメラ等で映された)画像の特徴ベクトルと、SSD に登録された膨大な人の画像の特徴ベクトルを比較することで、入力した画像が誰であるか認識します。

図2に開発したメモリ制御技術1である Value-Aware Data Mapping (VADM)を示します。本研究では、データには「価値」があることを利用しています。例えば32ビットの特徴ベクトルでは、上位ビットは価値が高く、下位ビットは価値が低いということです。従来の TLC(Triple-level cell)フラッシュメモリでは、1つのメモリセルに記憶する3ビットの信頼性が同様になるようにされていました。しかし本提案では、あえて高信頼のビット、中間の信頼性のビット、低い信頼性のビットと、信頼性の価値が異なるようにしました。そして、「価値の高い」データは「信頼性の高い」メモリセル(ビット)に記憶し、「価値の低い」データは「信頼性の低い」メモリセル(ビット)に記憶するように制御しました。これにより従来に比べてメモリのエラーが25倍あっても、高い認識が可能になりました。

図3に開発したメモリ制御技術2である Critical Page Error Reduction (CPEP)を示します。「価値の高い」データを記憶するメモリセルを更に高信頼化するために、「価値の高い」データを記憶するしきい値電圧の状態を、信頼性の低い“D”状態、“E”状態を避けて記憶するように、入力するデータを変調する方式を開発しました。これにより、制御技術1 (VADM) に比べて更にメモリのエラーが19倍あっても、高い認識が可能になりました。

図4に開発したメモリ制御技術である3Accelerated LDPC (A-LDPC)を示します。読み出し中にメモリセルのエラーを訂正する誤り訂正回路(ECC:Error Correcting Code) の復号の繰り返し回数を5回に制限することで、読み出しを26%高速化することに成功しました。現状はわずかに訂正されないエラーが残りますが、画像認識の精度は悪化させることはありません。

図5に開発した Value-Aware SSD を示します。今回開発した3つの制御技術により、画像認識においては従来の SSD に比べて12倍の10%のエラーがあっても高い顔認識が可能になりました(図中シンボル○)。そして、SSD の寿命(データ保持時間)を300倍にすることに成功しました。また、メモリのエラーの訂正にかかる時間を最小にすることで、読み出し時間を26%高速化しました。以上のようにこの SSD を用いることで、高い信頼性、高速化、高い画像認識の両立が可能になります。

3. 今後の展開

今後、膨大に増えていくストリーミングデータの処理などAIで扱うデータ量が増加が見込まれます。クラウドデータセンターやIoTの端末で、今回開発したAIに向けて最適化したSSDやストレージが重要になることを期待しています。

今回は画像認識に適用しましたが、本制御技術は音声認識、文字認識等、ディープラーニングの様々な応用に適用することが可能です。

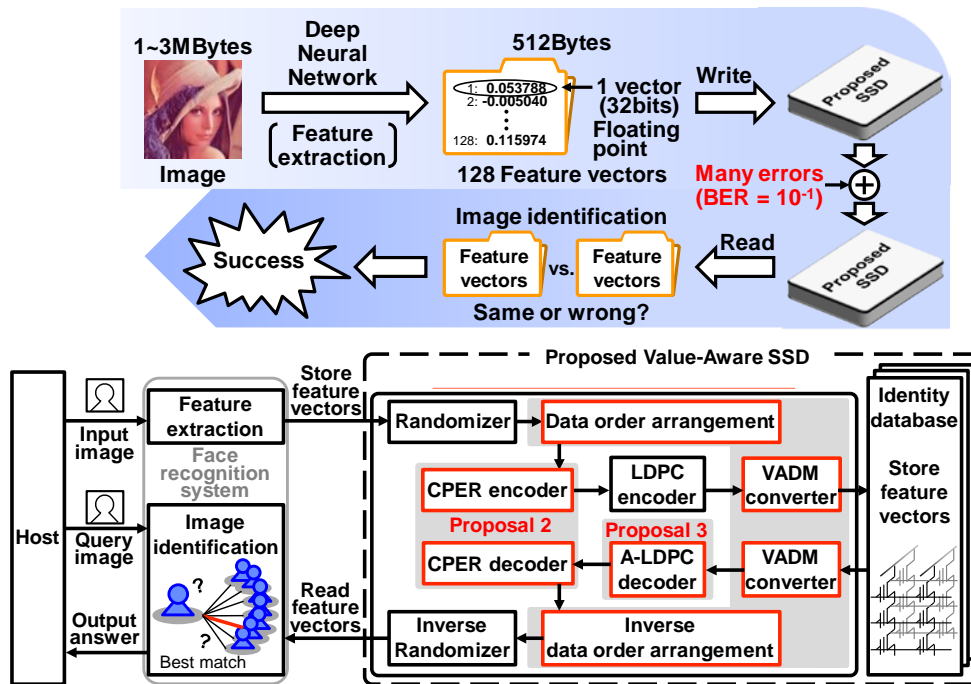


図1. ディープラーニングを用いた画像認識と提案するSSD

ディープラーニングによって画像は128次元の特徴ベクトルに変換され、SSDに記憶される。画像認識では、入力した(カメラ等で映された)画像の特徴ベクトルと、SSDに登録された膨大な人の画像の特徴ベクトルを比較することで、入力した画像が誰であるか認識する。

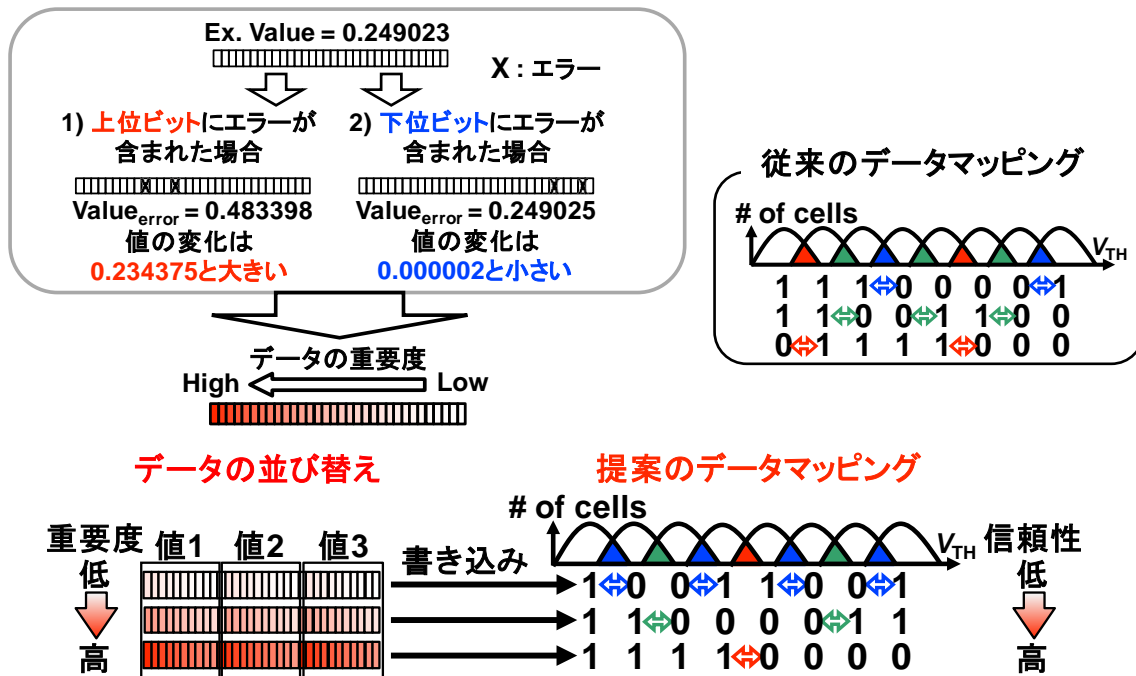


図2. 開発した制御技術1: Value-Aware Data Mapping (VADM)

データには「価値」があることを利用。上位ビットは価値が高く、下位ビットは価値が低い。従来の TLC (Triple-level cell) フラッシュメモリでは、1つのメモリセルに記憶する3ビットの信頼性が同様になるようにされていた。本提案では、あえて高信頼のビット、中間の信頼性のビット、低い信頼性のビットになるようにした。そして、「価値の高い」データは「信頼性の高い」メモリセル(ビット)に記憶し、「価値の低い」データは「信頼性の低い」メモリセル(ビット)に記憶するように制御した。

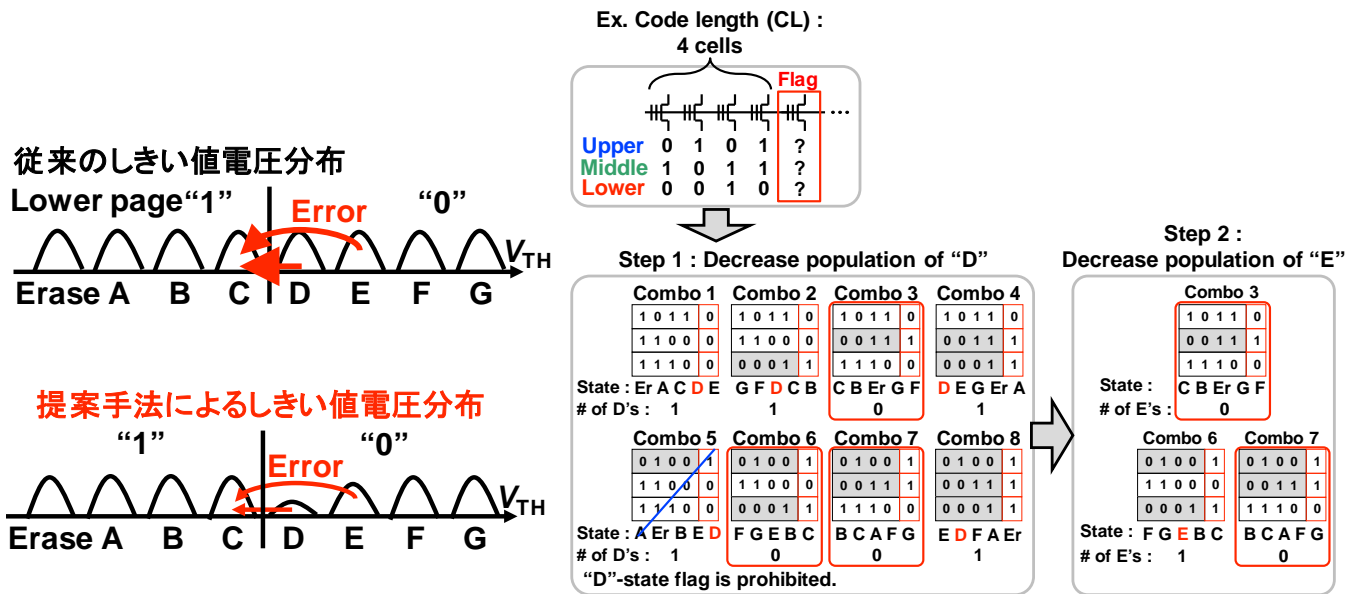


図3. 開発した制御技術2:Critical Page Error Reduction (CPER)

「価値の高い」データを記憶するメモリセルを更に高信頼化するために、「価値の高い」データを記憶するしきい値電圧の状態を、信頼性が低い「D」状態、「E」状態を避けて記憶するように、入力するデータを変調する方式を開発。

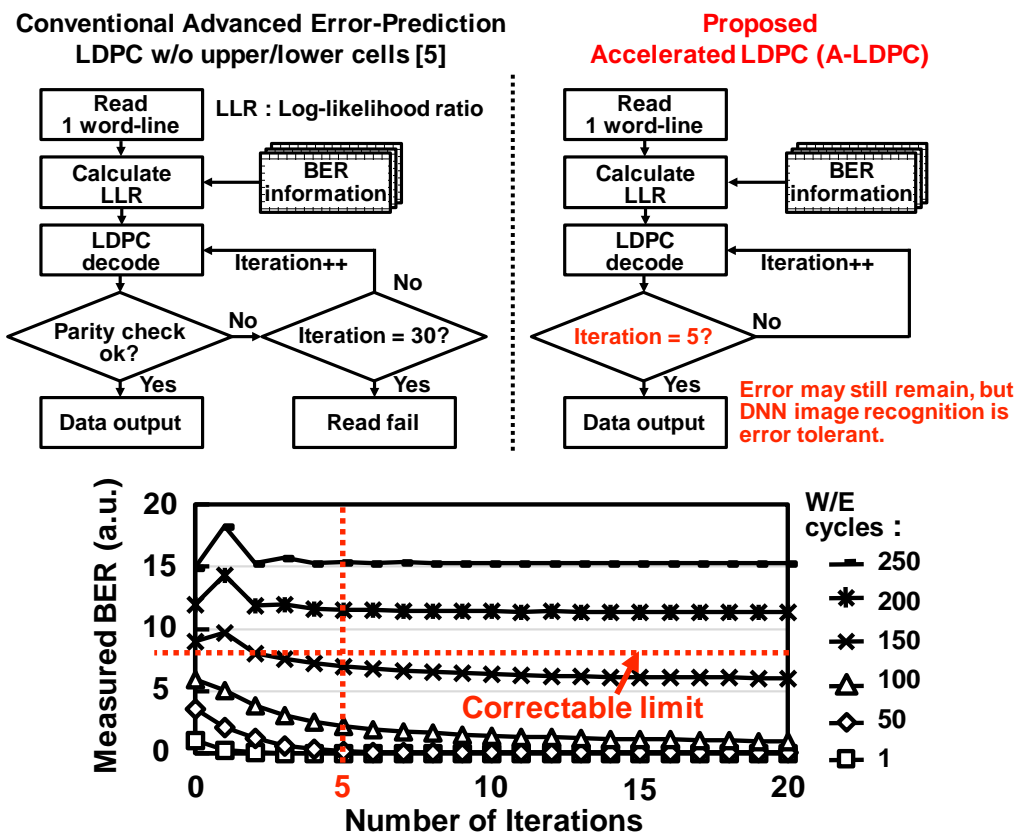


図4. 開発した制御技術3:Accelerated LDPC (A-LDPC)

読み出し中にメモリセルのエラーを訂正するLDPC ECCの復号の繰り返し回数を5回に制約することで、読み出しを26%高速化。わずかに訂正されないエラーが残るが、画像認識の精度を悪化させない。

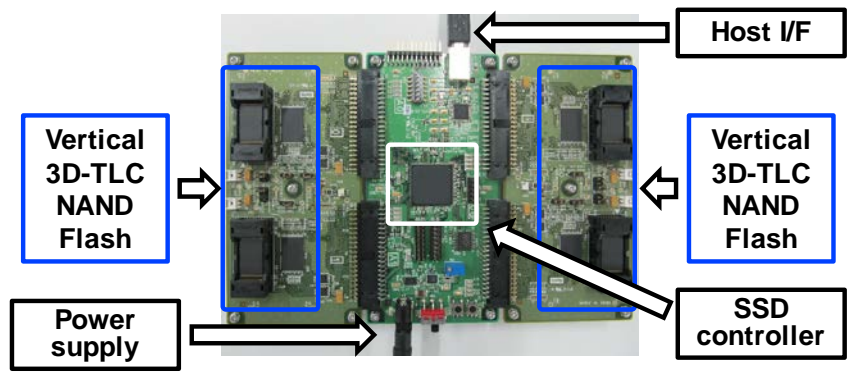
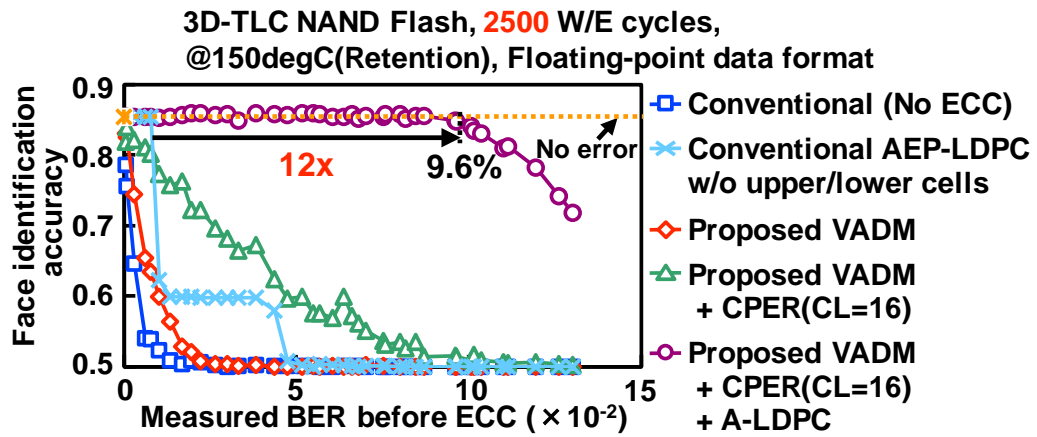


図5. 開発した Value-Aware SSD。従来の SSD に比べて 12 倍の 10%のエラーがあっても高い顔認識が可能になった。これにより SSD の寿命(データ保持時間)を 300 倍にすることができた。また、メモリのエラーの訂正にかかる時間を最小にすることで、読み出しを 26%高速化した。以上のように、高い信頼性、高速化、高い画像認識を両立した。

【お問い合わせ先】

＜研究に関すること＞

竹内 健 （タケウチ ケン）

中央大学理工学部 教授（電気電子情報通信工学科）

TEL： 03-3817-7374

E-mail: takeuchi@takeuchi-lab.org

＜広報に関すること＞

加藤 裕幹 （カトウ ユウキ）

中央大学 研究支援室

TEL 03-3817-1603, FAX 03-3817-1677

E-mail: k-shien@tamajs.chuo-u.ac.jp

【用語解説】

注1)クラウドデータセンタ

SNS やインターネットを使ったサービスを行うためにサーバーやストレージ、ネットワーク機器などの IT 機器を設置・運用する施設。

注2)ソリッド・ステート・ドライブ(SSD)

記憶媒体としてフラッシュメモリを用いるドライブ装置で、ハードディスクの代替としてスマートフォン、パソコンやデータセンタのストレージなどとして広く利用されている。機械的に駆動する部品がないため、高速の読み書きが可能で、消費電力も少なく衝撃にも強い。このため、頻繁にアクセスされるプログラムやデータを保存する用途で、現在幅広く使われている。

注3)フラッシュメモリ

データの一括消去を特徴とする半導体記憶装置。電氣的にデータの読み書きが可能で、電源を切ってもデータが消えない。