

## IGZOと不揮発性メモリを三次元集積した新デバイスの開発に成功 ～ディープラーニングの高効率ハードウェア化へ期待～

### 1. 発表者：

小林 正治（東京大学 生産技術研究所 准教授／

大学院工学系研究科 附属システムデザイン研究センター（d.lab）准教授）

### 2. 発表のポイント：

- ◆ディープラーニングの計算では、プロセッサとメモリの間の大量のデータ移動が性能を律速するため、メモリ自体で演算も行うインメモリコンピューティングが期待されています。
- ◆極薄の酸化物半導体をチャンネルとするトランジスタと不揮発性メモリの三次元集積デバイスを開発し、インメモリコンピューティングの機能の実証に成功しました。
- ◆この技術により、ディープラーニングがクラウドだけでなくエッジデバイスにも実装され、人工知能を用いたより高度で充実した社会サービスの展開が期待されます。

### 3. 発表概要：

東京大学 生産技術研究所の小林 正治 准教授らは、極薄の酸化物半導体 IGZO（注1）を用いたトランジスタと抵抗変化型不揮発性メモリを三次元集積したデバイスの開発に成功しました。

大量のデータを用いるディープラーニングは多層のニューラルネットワークで構成されており、プロセッサとメモリの間の大量のデータ移動が性能を律速するため、メモリ配列に演算機能をもたせたインメモリコンピューティングによるハードウェア実装が期待されています。しかし、通常のメモリ配列は二次元構造であり、ネットワークのモデルが大規模になるにつれて配線長が長くなり、計算速度や消費電力が問題となります。また、同時にアクセスできるメモリ量にも制限があるため、並列計算の効率が上がりません。

本研究では、インメモリコンピューティングのハードウェア実装における二次元メモリ配列での配線の問題を解決し、かつ超並列計算を可能にするため、メモリ配列を三次元積層した三次元ニューラルネットワークの実現に向けて、通常集積回路の配線層プロセスに適用可能な最高温度である 400℃以下のプロセス温度で極薄の酸化物半導体 IGZO を用いたトランジスタと抵抗変化型不揮発性メモリを形成する三次元集積デバイスを開発しました。本デバイスを用いてインメモリコンピューティングの機能を実証し、ディープラーニングの多層ニューラルネットワークを一チップ上に物理的に多層構造で実装することが可能となりました。

本成果はディープラーニングの計算を高いエネルギー効率で計算することを可能にし、クラウドだけでなくエッジデバイス（注2）でも高度な人工知能計算を行うことで、ビッグデータに基づく社会サービスの飛躍的な向上が期待されます。

本研究成果は、2020年6月14日（太平洋夏時間）から「VLSI Technology Symposium 2020」で発表されます。

### 4. 発表内容：

#### <研究の背景と経緯>

ビッグデータを用いた人工知能技術により革新的な社会サービスが実現されてきています。特にディープラーニングを用いた学習・推論システムは強力なアルゴリズムを発揮しますが、

実装には相応のコンピューティング能力が必要とされ、実際、高性能なサーバー・クラウドコンピューティングによって実装されています。大量のデータを用いるディープラーニングを構成する多層ニューラルネットワークを従来のコンピュータで計算する場合、プロセッサとメモリとの間のデータのやりとりが性能を大きく律速することがわかっています。これをフォンノイマンボトルネックと呼びます。そのため、メモリ自体に演算機能をもたせたハードウェアによるインメモリコンピューティングが注目を集めています。

インメモリコンピューティングは、ニューラルネットワークにおける重み係数を二次元のメモリ配列に学習させ、入力ベクトルとして画像などのデータを入力し、メモリ配列の重みに基づいて一括・並列にニューラルネットワークの出力ベクトルを計算できることが特徴です。しかし、ニューラルネットワークのモデルが複雑になるほど二次元メモリ配列のサイズは大きくなり、それに伴ってデータが通る配線の長さが長くなり、計算の遅延や消費電力が増大します。また、同時にアクセスできるメモリ量が制約されるため、並列計算にも制約がおきます。そのため、高性能で低消費電力なインメモリコンピューティングの実現にはデバイスアーキテクチャの刷新が必要であると考えました。

### <研究の内容>

本研究では、インメモリコンピューティングのハードウェア実装における二次元メモリ配列での配線の問題を解決し、かつ超並列計算を可能にするため、メモリ配列を三次元積層した三次元ニューラルネットワークの実現に向けて、通常集積回路の配線層に 400°C 以下のプロセス温度で極薄の酸化物半導体 IGZO をチャンネルとするトランジスタ (IGZO トランジスタ) と抵抗変化型不揮発性メモリ (RRAM) を形成した三次元集積デバイスを提案しました (図 1)。

本三次元集積デバイスの各層は、RRAM と IGZO トランジスタによるメモリセルからなる配列で構成されます (図 1)。本研究では三層を積層したデバイスを試作しました。IGZO トランジスタは 400°C 以下のプロセスで形成されますが、移動度 (注 3) は  $10\text{cm}^2/\text{Vs}$  という高い値を有します。各層の IGZO トランジスタは十分に大きな電流を駆動し、RRAM に適切に書き込むことができます (図 2)。また、各層のメモリセル特性を比較したところ、不揮発性メモリ特性および信頼性はほぼ同じで、積層プロセスによる劣化は見られませんでした (図 3)。この結果はさらなる積層化が可能でありネットワークモデルの拡張に対応できることを示唆しています。

本デバイスの IGZO トランジスタと RRAM によるメモリセルのペアを用いたインメモリコンピューティングとして、ニューラルネットワークの基本計算である XNOR 演算 (注 4) を実証しました。具体的には RRAM にニューラルネットワークの重みを学習させ、IGZO のゲートにつながるワード線に入力データを印加し、その組み合わせによってプリチャージしておいたビット線の電圧が放電され、XNOR 演算結果をビット線の出力電圧として得ることができます (図 4)。この方式では定常電流を発生しないため、従来の方式に比べて 10 分の 1 以下の低消費電力な演算が可能となります。

現在、半導体集積デバイス分野においては、ムーアの法則に従うデバイスの微細化による性能向上が鈍化してきており、三次元集積化が今後の技術の大きな潮流となってきます。本研究で開発した技術もこの流れと人工知能のニーズに沿って発展していく技術であると考えています。

## <今後の展開>

本研究では、極薄の酸化物半導体 IGZO をチャンネルとするトランジスタと RRAM を用いたメモリ配列の三次元集積化を行いました。IGZO の移動度は  $10\text{cm}^2/\text{Vs}$  程度ですが、現在のプロジェクトの研究対象である二次元層状物質（注5）の移動度は IGZO よりも数倍～数十倍高いことがわかっており、 $400^\circ\text{C}$ 以下の温度で二次元層状物質を成膜することができれば、本研究で提案したアーキテクチャを二次元層状物質を用いて実現することができます。より高い移動度のトランジスタを用いることで、より低電圧かつ高速に RRAM の書き込みを行うことができるため、エネルギー効率の高いインメモリコンピューティングの実現に向けて二次元層状物質のプロセスとデバイス技術の研究を鋭意進めていきます。

本成果は、科学技術振興機構（JST）戦略的創造研究推進事業 CREST「二次元機能性原子・分子薄膜の創製と利用に資する基盤技術の創出」研究領域（研究総括：黒部 篤（株）東芝 研究開発センター 首席技監）における研究課題「原子層ヘテロ構造の完全制御成長と超低消費電力・3次元集積デバイスの創出」（代表者：宮田 耕充 東京都立大学 大学院理学研究科 准教授）の支援を受けて得られました。

## 5. 発表雑誌：

学会名：「VLSI Technology Symposium 2020」（オンライン版：6月14日）

論文タイトル：A Monolithic 3D Integration of RRAM Array with Oxide Semiconductor FET for In-memory Computing in Quantized Neural Network AI Applications（「多ビットニューラルネットワーク人工知能応用に向けたインメモリコンピューティングのための酸化物半導体トランジスタを用いた抵抗変化型メモリのモノリシック三次元集積化」）

## 6. 問い合わせ先：

<研究に関すること>

小林 正治（コバヤシ マサハル）

東京大学 生産技術研究所 准教授／

大学院工学系研究科 附属システムデザイン研究センター（d.lab） 准教授

〒153-8505 東京都目黒区駒場4-6-1

Tel：03-5452-6813 Fax：03-5452-6265

E-mail：masa-kobayashi[at]nano.iis.u-tokyo.ac.jp

<JST事業に関すること>

嶋林 ゆう子（シマバヤシ ユウコ）

科学技術振興機構 戦略研究推進部 グリーンイノベーショングループ

〒102-0076 東京都千代田区五番町7 K's 五番町

Tel：03-3512-3531 Fax：03-3222-2066

E-mail：crest[at]jst.go.jp

## 7. 用語解説：

（注1）IGZO

In-Ga-Zn-O の四元素からなる酸化物半導体。現在、パネルディスプレイの駆動トランジスタの材料として使われている。

(注2) エッジデバイス

IoT の枠組みでは大規模なデータセンターで構成されるクラウドに対する言葉。一般的にはネットワークにつながる小型のデバイスを表し、携帯端末やセンサーデバイス、監視カメラなどが当てはまる。

(注3) 移動度

半導体材料の電気的特性の一つ。半導体に電界をかけると電子または正孔が運動する。電界を高くするほど速度は速くなる。移動度は電子または正孔の速度と電界の間の比例係数を指す。

(注4) XNOR 演算

論理演算の一つで、二つの論理変数に関する排他的論理和 (XOR) の否定をとる演算。

(注5) 二次元層状物質

原子間の結合が層内で閉じていて、層間は弱いファンデルワールス力のみで結合した結晶構造を持つ一連の物質を指す。

8. 添付資料 :

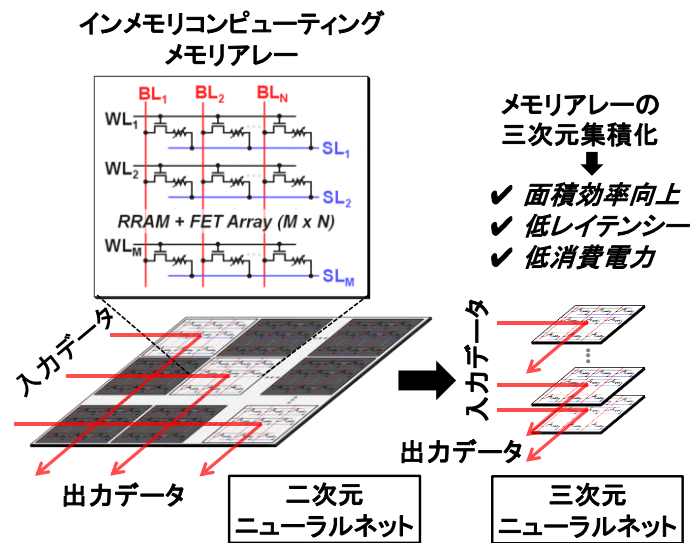


図1

左に従来の二次元メモリアレー、右に本研究で提案する三次元集積メモリアレーによるインメモリコンピューティングの概念図を示す。三次元集積化により面積効率の向上、低レイテンシー（高速動作）、低消費電力を実現することができる。

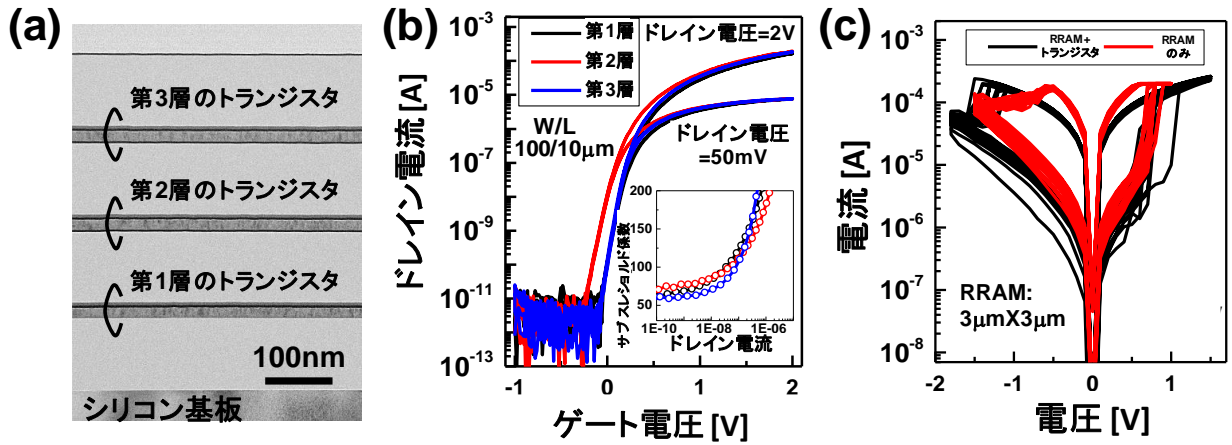


図 2

(a) 三層積層した IGZO トランジスタの断面 TEM 像。(b)三層各層の IGZO トランジスタのドレイン電流-ゲート電圧特性。(c) RRAM のみのメモリ特性 (赤) と IGZO トランジスタで駆動した RRAM のメモリ特性 (黒)。

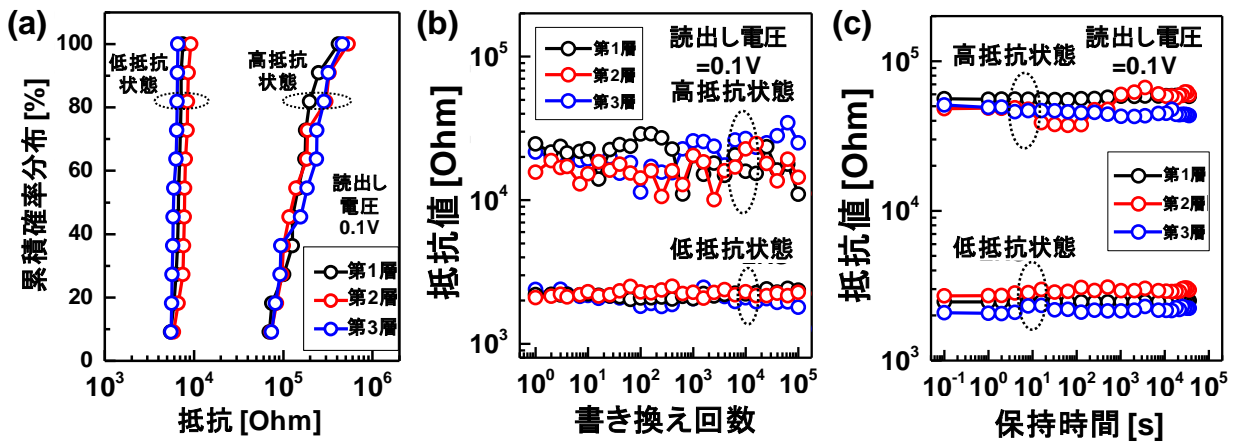


図 3

図 2 と同じく三層積層した RRAM の各層の低抵抗状態と高抵抗状態の(a)累積確率分布、(b)書き換え耐性、(c)保持時間特性。三層ではほぼ同一特性を示し、三次元集積化プロセスによる劣化は見られなかった。

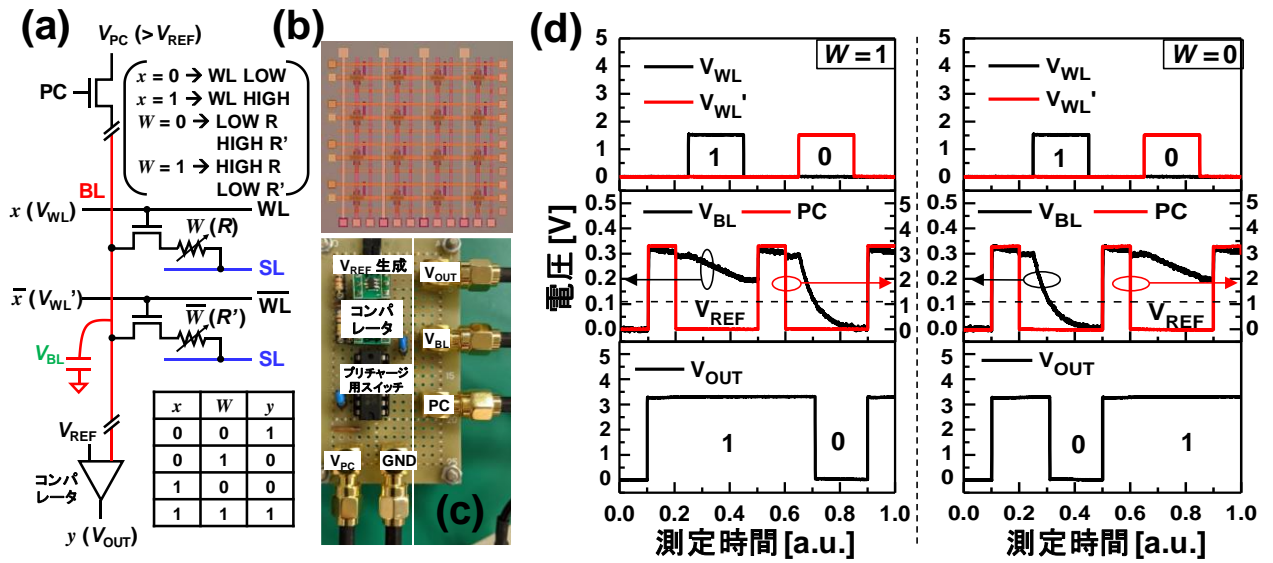


図 4

(a)RRAM と IGZO トランジスタからなるメモリセルのペアによる XNOR 演算の基本ユニット。RRAM の抵抗値( $R$ ,  $R'$ )にニューラルネットワークの重みを学習させ、ワード線の電圧( $V_{WL}$ ,  $V_{WL}'$ )に入力データを印加し、それらの組み合わせによってビット線電圧( $V_{BL}$ )が変化し XNOR 演算の出力データが読みだされる。(b)試作したメモリアレーの写真、(c)外部周辺回路、(d)測定波形。