公開

研 究 終 了 報 告 書

「アルゴリズム・ソフトウェア・ハードウェアの融合による超低電力ニューラルネットワークの構築」

研究期間： 2019 年 10 月～2023 年 3 月
研 究 者： 陳 オリビア

1. 研究のねらい

ここ数年で深層学習はたくさんの注目を集め、様々なアプリケーションに応用されながら、その多くにおいて劇的な精度向上を果たしている。これら多くは数百万から数十億のパラメータに依存しており、複数の GPU と CPU の組み合わせの超高速計算能力が重要な役割をはたしている。このような計算には大きな消費電力がともない、莫大なコストを必要とする。CMOS 回路に比べ、低消費電力および高速性の点で優れたデバイスとして、超伝導単一磁束量子（Single Flux Quantum: SFQ）回路と超伝導断熱磁束量子パラメトロン（Adiabatic Quantum Flux Parametron: AQFP）がある。超伝導回路は、スイッチングが高速であるのみならず、チップ内およびチップ間でも高速・高スループットな信号伝送が可能である。超伝導回路を用いれば、極めて低消費電力で高性能なプロセッサが実現できる可能性がある。そこで本研究は、超伝導回路を用いることより、アルゴリズム、ソフトウェア、及びハードウェアを統合した超低電力ニューラルネットワークを提案する。提案したニューラルネットワークは、様々なアルゴリズムを学習モデルに導入することで、大量データを軽量化し（最大 1/4000 倍まで）、かつ処理を高速化させ（加算回数の削減と乗算の除去）、従来性能より圧倒的に優れたハードウェアアーキテクチャを確立する。そしてハードウェアに関しては、超伝導素子の有する確率的状態遷移特徴を生かした Stochastic Computing に基づいたニューラルネットワークの演算機構の設計と低温動作実証を行う。最後に、論理合成最適化法、機能検証法、消費電力解析法を含めた大規模な超伝導回路の設計作業に不可欠な自動設計（Electronic Design Automation: EDA）ソフトウェアを開発することで、確立したハードウェアアーキテクチャを超伝導回路で設計と実装をする。研究体表者はこれまでに、超伝導集積回路の開発と自動化設計ツールの開発について、研究成果を創出した実績がある。本研究では、研究代表者がこれまでに培った超伝導集積回路技術を最大限に活用し、アルゴリズム、ソフトウェア、及びハードウェアの連携を通じた、PetaOPS/W（１W で１千兆回演算）級のエネルギー効率を有する新たな超低電力 AI システムタを実現するための基盤技術を確立することを目的とする。

2. 研究成果

(1) 概要

超伝導断熱磁束量子パラメトロン AQFP 回路による超低電力ニューラルネットワークに向けた専用ハードウェアの基盤技術を確立し、将来ペタオプスパーワート級高性能ニューラルネットワークプロセッサが実現可能であることを示すことを目的とし、❶AQFP ニューラルネットワークに向けた数値表現と計算原理の確立、❷学習モデルの圧縮手法の研究、❷AQFP 大規模回路自動化設計技術の開発、❸回路アーキテクチャの確

立と適した演算ユニットの開発を行なった。さらに、❹探索研究として、超伝導フーリエ変換器を用いたニューラルネットワークの加速機構の開発を行なった。

❶AQFP ニューラルネットワークに向けた数値表現と計算原理の確立については、超伝導素子の確率的な状態遷移を着目し、長いビット列の 1 の割合を利用した stochastic 計算法を確立した。当該計算手法は 1 ビットのワード長を採用し、演算機構のハードウェアコストを大幅に削減することにより、高い計算性能を実現する。

❷AQFP 大規模回路自動化設計技術の開発では、多数決論理および超伝導素子の同期性に向けた論理合成法、交互方向乗数法（ADMM）を用いた静的なタイミング分析法、学習ベースの配置配線法、probabilistic 回路電力解析法を含めた設計自動化ツール群を構築した。

❸回路アーキテクチャの確立と適した演算ユニットの開発では、①において確立した stochastic computing に向けた random binary neural network (RBNN)を提案した。RBNN は、メモリ近傍構造、アナログ積和演算と確率活性化関数を用いた。さらに、AQFP デバイスの特徴パラメータを導入した学習の最適化、batch normalization のマッチング、重み整流クランプ方法を含んだアルゴリズムとハードウェアの協調設計を行なった。提案のアーキテクチャに適した 4x4 と 8x8 のクロスバーアレーを開発した。

❹ニューラルネットワークを加速するため，超伝導フーリエ変換器を用いた畳込み演算を提案し，産総研ニオブ 9 層 1μm プロセスによる試作を行い、4.2K の低温下で 48GHz の高速動作実証に成功した．これは今までに，最大規模の超伝導ニューロモーフィックコンピューティング回路となっている。

(2)詳細

研究テーマ A「AQFP ニューラルネットワークに向けた数値表現と計算原理の確立」

AQFP 論理ゲートは２つの Josephson 接合を含む超伝導ループと励起電流印加用のインダクタンスで構成される。AQFP 回路のポテンシャルは励起電流の有無によって１つあるいは２つの極小値を持ち、回路のポテンシャルエネルギーを１極小値の状態から２極小値の状態にゆっくりと断熱的に変化させることで、微小なエネルギーで論理状態の遷移を実現できる。この状態遷移には CMOS 回路の様な非断熱的なエネルギー消費を伴わないため、究極的な低電力ロジックを構成できる。また、ゲートのポテンシャルエネルギーは、熱雑音により確率的に変化させることができるため、極めて容易に確率的値を生成することができる。この特性は、確率値を利用した stochastic computing に極めて有利である。stochastic computing は、乱数列を用いることで複雑な演算を少ないゲートで表現できるため、回路面積の低減、ひいては低電力化を可能とする。例えば、4/8 は 01101010...，6/8 は 10111011...という乱数列（1 の存在確率）で表され、両者を AND

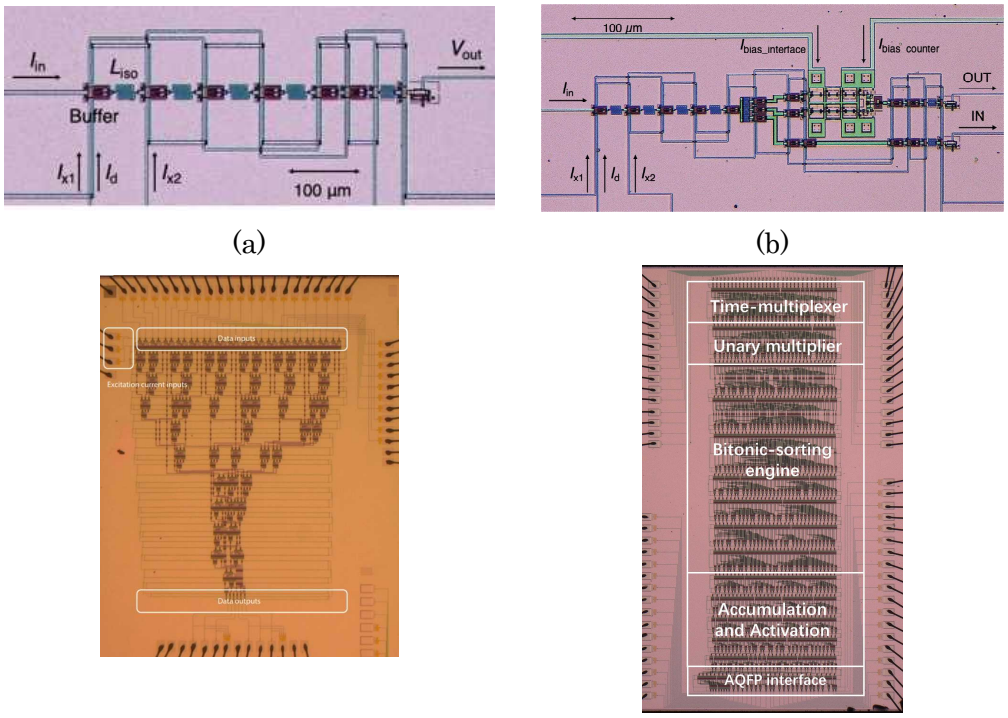


図 A.1 産総研 4 層 1μm プロセスによる試作した (a) stochastic number 生成器(SNG) (b) シグモイド関数生成器(SFG)、 ならびに(c) approximate-parallel-counter (APC)と(d)

ゲートに入力するだけで、乗算結果である 00101010... (=3/8) が得られる。また、画像、音声認識などの深層学習は厳密な計算を必要としないため、近似計算であるストカスティック演算の計算精度で十分に対応できる。本研究テーマにおいては、AQFP ゲートを用いた確率値生成回路を設計と実装し、低温実験により、生成した確率ビット列の数値分析を行い、乱数の質は NIST 乱数検定ツールにより検証された。また、演算の誤り率の調査のため、ケーススタディとして、stochastic computing に基づいた AQFP 非線形関数生成回路と積和演算回路の実装と動作実証を行なった。開発した回路のチップ写真は図 A.1 に示した。さらに、提案手法に基づいた非線形関数生成回路の実験結果を図 A.2 に示すように、十分な精度を得たことが明確された。

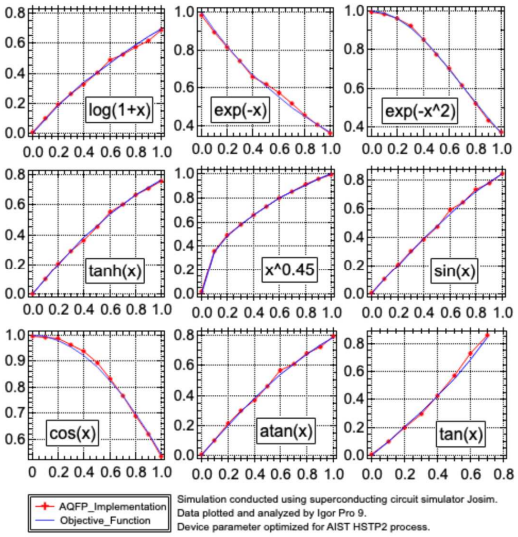


図 A.2 提案手法に基づいた関数の近似結果

研究テーマ B「AQFP 大規模回路自動化設計技術の開発」

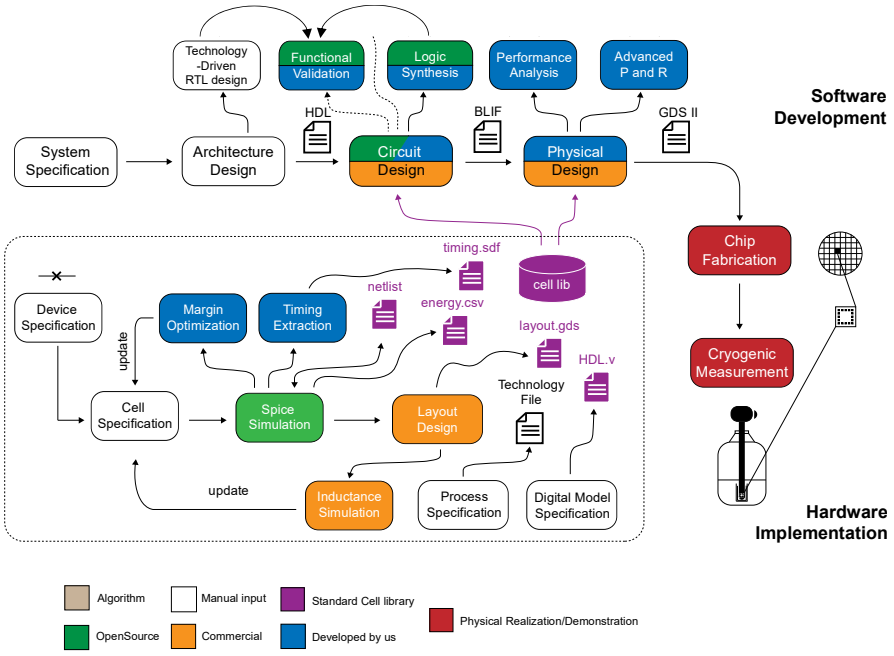


図 B.1 開発した自動化設計流れ

AQFP 回路では、クロックに同期してデータが伝播するため，すべてのデータパスが等しい遅延（クロック位相）を持つことが非常に重要である。また、AQFP ゲートは電流極性を利用した論理表現であることより様々な従来の CMOS と異なる特徴を持つため、大規模回路に向けた専用自動化設計技術を開発した。図 B.1 は設計流れ（ツール群）を示している。その中には、無地で示される工程は手作業であり、水色の 8 工程の中図の上部の 6 工程は本プロジェクトにおいて開発及び最適化することとした。以下で、最も顕著な成果を得た論理合成と配置配線について説明する。

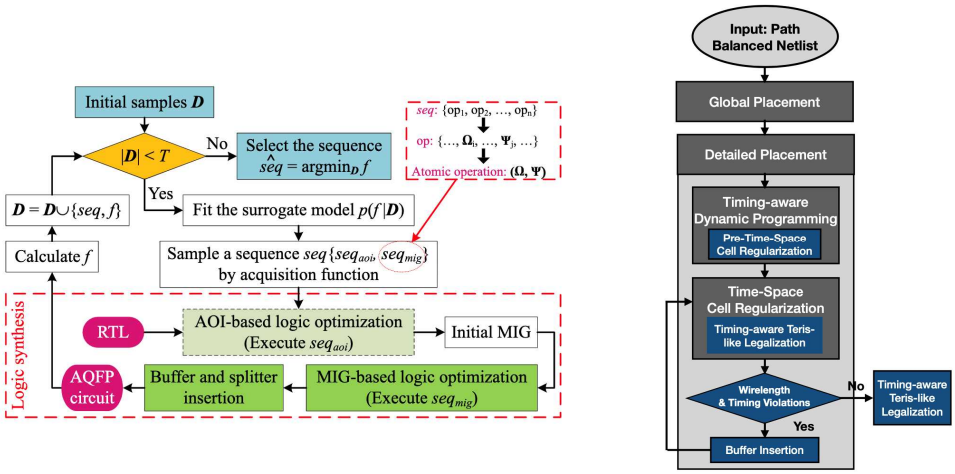
AQFP 論理合成ツール

AQFP 論理合成においては、お主に多数決論理合成、ファンアウトを増加するため分岐素子の挿入と全パスを均一のためのバッファ素子の挿入を行う。既存の手法より良い合成結果を得るため、Bayesian 最適法を導入した新たな論理合成フレームワークを開発した（図 B.2 (a)）。それ以外にも、AQFP ラッチセルに基づいた順序回路の合成のためのツール開発を行なった。

タイミングを考慮した配置配線ツール

研究実施までには、遺伝的アルゴリズム（Genetic Algorithm, 以下 GA と呼ぶ）を用いた大規模断熱型超伝導回路向けの配置配線ツールが既に提案されている。しかしながら、これらの GA を用いた手法はかなりの時間がかかる一方、最適な結果に至っていない。研究では、図 B.2 (b)に示すように、回路動作速度を高めることも目的とし、タイミングと配線距離の制約を同時に最適化できる配置手法を提案し、最大43%の回路量の削減かつ100倍以上の計算時間の縮短を達成した上、回路のタイミング制約および配線長制約の範囲内で，タイミング違反を大幅に削減することに成功した.

回路電力解析ツール



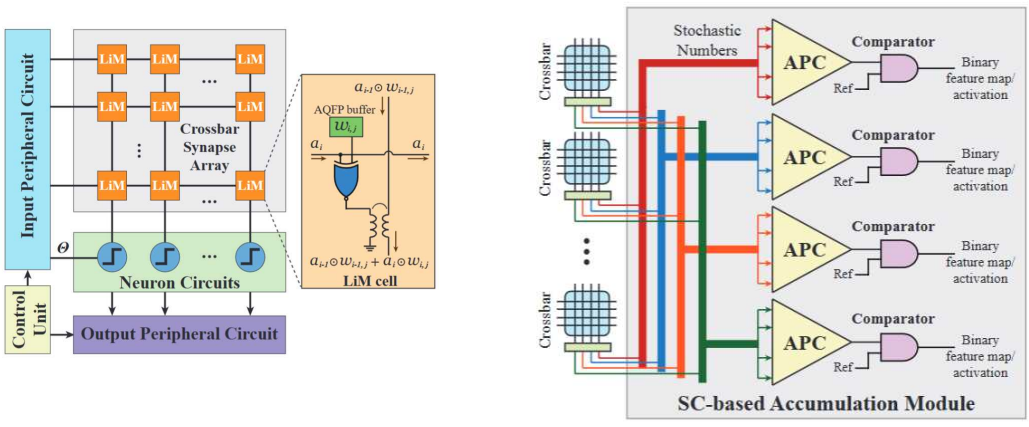
(a) (b)

図 B.2 (a) 開発した論理合成フレームワーク (b) timing-aware 配置配線策

回路電力解析ツール「AQFP-QPA」が開発された．「AQFP-QPA」では，開発された SFQ 電力解析フレームワークをさらに拡張して，AQFP 回路の電力解析をサポートするもので，全ての AQFP 論理セルの事前に計算されたパターン依存の電力値に基づいて，信号確率の計算方法を示している．図 9 に今回開発した「AQFP-QFA」の概要構成を示す．「AQFP-QFA」の主要な入力は，合成された回路ネットリスト，AQFP パワーライブラリ，およびユーザー定義の動作周波数で構成される。

研究テーマ C「回路アーキテクチャの確立と適した演算ユニットの開発」

研究テーマに定めた二値化ニューラルネットワークは、重みと活性化が 1 ビットであるため、メモリ面積が大幅に削減されているにもかかわらず、従来の CPU とメモリが分離したノイマン型コンピュータ方式では、演算回路とデータメモリ間の大規模なデータ



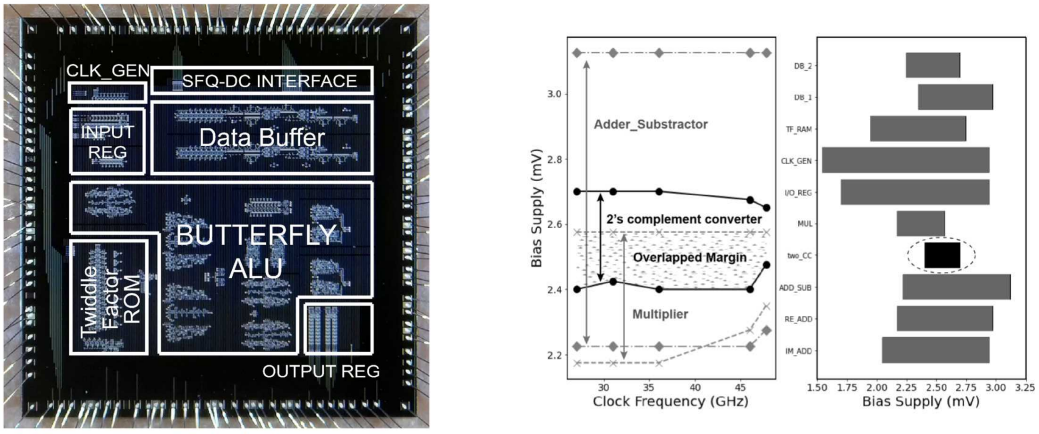
(a) (b)

図 C.1 開発した AQFP-NN のアーキテクチャ (a) ユニットクロスバーの構造 (b) 複数のクロスバー配列の積算モジュール図

の伝搬に課題に残っている。この問題を解消するために、論理演算機能を搭載したロジックインメモリ(LiM)セルを沢山並べたクロスバー構造を用いた BNN を提案する。図 C.1 に、提案回路の構成を示す。二値化された重みは、AQFP LiM セルにあらかじめ格納され、セル内 XNOR マクロで乗算される。従来の BNN におけるホップカウントを用いた積算とは異なり、AQFP の論理“1”と“0”は正と負の電流パルスで表されるため、すべての出力を直接加算するアナログ積算方式を採用している。そして、各列の電流の合計値で表される積算結果は，誤差逆伝播法の過程で生成されるバッチ正規化パラメータを示す閾値電流でさらにオフセットされ, AQFP 電流比較器からなるニューロン回路で二値化される．さらに、図 C.1(b)に示すように、AQFP デバイスの特徴パラメータを導入した学習の最適化、batch normalization のマッチング、重み整流クランプ方法を含んだアルゴリズムとハードウェアの協調設計を行なった。

テーマ D 探索研究：単一磁束量子回路を用いた FFT 回路の開発：

ニューラルネットワークを加速するため，超伝導フーリエ変換器を用いた畳込み演算を提案し，産総研ニオブ 9 層 1μm プロセスによる試作を行い、4.2K の低温下で 48GHz の高速動作実証に成功した．これは今までに，最大規模の超伝導ニューロモーフィックコンピューティング回路（超伝導素子数約 2 万個）となっている。



(a) (b)

図 D　開発した超伝導 FFT 加速回路の（a)チップ写真（b)動作マージン

3. 今後の展開

本研究において創出した基盤技術を最大限に活用し、AI や BC などに向けた次世代情報処理専用計算機の開発を期待できる。さらに、超伝導回路に向けた低熱流入広帯域アクセス技術の推進、室温への出力インタフェースの開発、チップ life cycle 調査等製品化向けの準備を行い、実用化に向けて万進する。また、汎用性の高い CMOS システムと組み合わせたヘテロジニアスアーキテクチャへの拡張による次世代スーパーコンピュータの開発を目指す。他の国プロジェクトにより開発された量子計算機との複合化より、次世代のスーパーコンピュータの実現を目標して、科学的発見を加速し、気候変動予測、医薬品設計、脳型機能マッピング、革新的なスマートマテリアルの開発などの新しい分野の開拓また加速を目指す。さらに、パーソナライズされたアプリケー

ションを推進し、命を救う自然災害の予測、また、計算を通じて、長年の病気の根絶を推進するなどすべて人類の改善への貢献を期待できる。

4. 自己評価

研究代表者は、半導体の微細化問題に対する直接的な挑戦し、「超伝導ロジックを対象としたハードウェア、ソフトウェア、設計技術、アルゴリズム、のシステム階層横断型研究」を行い、その独創性は極めて高いと自負している。さらに、提案者は超伝導素子が有する確率的な動作特性に着眼し、超伝導回路の AI 応用に関するアイディアを独自に見いだし、機械学習アルゴリズム、設計技術、アーキテクチャ／回路技術、といったクロスレイヤーでの最適化技術の探求に次々と成功している。本研究に関する研究成果は、AQFP 論理回路の大規模集積化を可能とする技術として世界から注目を集め、関連する研究は世界多国の研究グループの中で盛んでいる。

5. 主な研究成果リスト

(1) 代表的な論文（原著論文）発表

研究期間累積件数：13件

1. O. Chen, Y. Wang, R. Zhang and N. Yoshikawa, "Design and Implementation of Stochastic Neural Networks Using Superconductor Quantum-Flux-Parametron Devices," *2022 IEEE 35th International System-on-Chip Conference (SOCC)*, 2022, pp. 1-6, doi: 10.1109/SOCC56010.2022.9908075.

Stochastic logic has been proven to be a low-cost hardware solution to accelerate neuromorphic computing thanks to its bit-wise operation feature, which removes the most hardware. On the other hand, the superconductive electronic device exhibiting stochastic behaviour and deep-pipelining nature is a perfect candidate for implementing stochastic logic-based neuromorphic computing. The adiabatic quantum-flux-parametron (AQFP) logic family is demonstrated with the highest energy efficiency among its superconducting cousins. In this work, we introduce the recent designs and implementations of the AQFP-based neural network prototypes utilizing the stochastic computing paradigm. We further report the low-temperature measurement results of two different implementations using approximate-parallel-counter (APC) and bitonic-sorter based design approaches. Thanks to its adiabatic switching nature and zero-static power dissipation, the AQFP-based implementation averages 5-6 orders of energy efficiency compared to its CMOS counterpart.

2. F. Ke, O. Chen, Y. Wang and N. Yoshikawa, "Demonstration of a 47.8 GHz High-Speed FFT Processor Using Single-Flux-Quantum Technology," in IEEE Transactions on Applied Superconductivity, vol. 31, no. 5, pp. 1-5, Aug. 2021, Art no. 1300905, doi: 10.1109/TASC.2021.3059984.

A fast Fourier transform (FFT) is an algorithm that computes the discrete Fourier transform (DFT) of a sequence at high speed. FFT can convert a signal from time domain to frequency domain, and is wildly used in digital signal processing field. In this paper, a high-speed, low-power FFT processor is demonstrated up to 47.8GHz with the measured power

consumption of 5.3mW, using single-flux quantum (SFQ) logic. This is the first complete FFT processor implementation using superconducting technology, performing 8-point 7-bit FFT in a bit-serial computing manner. The test chip fabricated using a 1.0 μm 9-layer process consists of 17 455 Nb/AlO x /Nb Josephson junctions (JJs), rendering itself the largest superconducting digital circuit capable of iterative data computing. The correct operation of the chip has been experimentally confirmed at a maximum operating frequency of 47.8GHz (word speed 8GHz) by conducting on-chip high-speed testing.
3. Luo, W., Chen, O., Yoshikawa, N. *et al.* Scalable true random number generator using adiabatic superconductor logic. *Sci Rep* **12**, 20039 (2022). https://doi.org/10.1038/s41598-022-24230-5
Alternative computing such as stochastic computing and bio-inspired computing holds promise for overcoming the limitations of von Neumann computers. However, one difficulty in the implementation of such alternative computing is the need for a large number of random bits at the same time. To address this issue, we propose a scalable true-random-number generating scheme that we refer to as XORing shift registers (XSR). XSR generates multiple uncorrelated true random bitstreams using only two true random number generators as entropy sources and can thus be implemented by a variety of logic devices. Toward superconducting alternative computing, we implement XSR using an energy-efficient superconductor logic family, adiabatic quantum-flux-parametron (AQFP) logic. Furthermore, to demonstrate its performance, we design and observe an AQFP-based XSR circuit that generates four random bitstreams in parallel. The results of the experiment confirm that the bitstreams generated by the XSR circuit exhibit no autocorrelation and that there is no correlation between the bitstreams.

(2) 特許出願

該当しません。

(3) その他の成果（主要な学会発表、受賞、著作物、プレスリリース等）

- 超伝導回路の深層学習応用（招待講演） 応用物理学会超伝導分科会（日本・東京） 2019 年 11 月
- An AQFP-Based Neural Network Accelerator with On-Chip Stochastic Number Generation 応用超伝導大会（アメリカ・リモート） 2020 年 10 月
- Implementation of an FFT-Based Convolutional Processing Element Using Single-Flux-Quantum Technology 応用超伝導大会（アメリカ・リモート） 2020 年 10 月
- Multi-Output True Random Number Generator Based on Adiabatic Quantum-Flux-Parametron Logic ヨーロピアン応用超伝導大会（リモート）2021 年 9 月
- Non-Linear Function Generator Using Stochastic Superconductive Circuits 応用超伝導大会（アメリカ・ホノルル） 2022 年 10 月