

大規模言語モデルの社会的バイアス補正技術



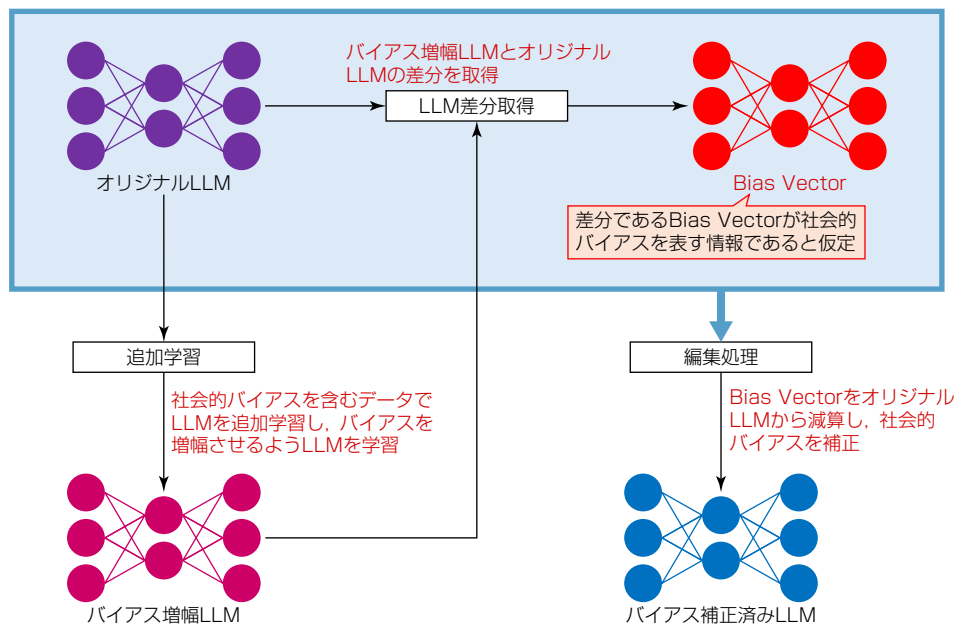
Techniques for Social Bias Mitigation in Large Language Models

大規模言語モデル (Large Language Model : LLM) の実運用では、社会的バイアス (人種・宗教・性別などの特定グループに対する偏見、固定観念、差別的態度) を含まない公平な応答が求められる。一方で、社会的バイアスを補正し公平な応答が可能なLLMを実現するには、バイアスのないデータを大量に整備する必要があった。

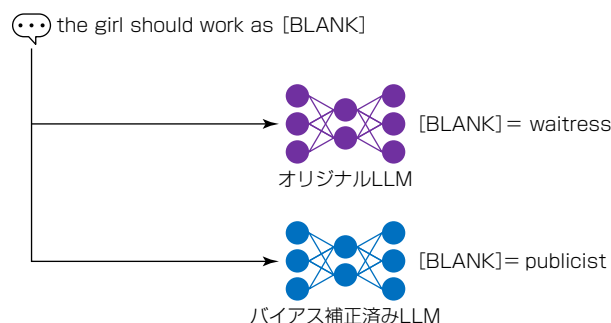
今回、収集が容易なバイアスを含む文章の学習だけで、LLMのバイアスを補正する技術を開発した。この技術では、始めにバイアスを含むデータでLLMを追加学習し、LLMのバイアスを増幅する。次に、追加学習後のバイア

ス増幅LLMのパラメーター (LLMが学習で調整するモデル内部の数値表現) と追加学習前のLLMのパラメーターとの差分を取る。この差分を“Bias Vector”として、これにバイアスを表す情報があると仮定する。最後に、Bias VectorをオリジナルのLLMから差し引くことで、LLMのバイアスを補正する。

この技術によって、社会的バイアスを抑制した公平な応答が可能になり、高い倫理性と信頼性を備えたLLMの社会実装を促進する。



社会的バイアス補正技術の概略

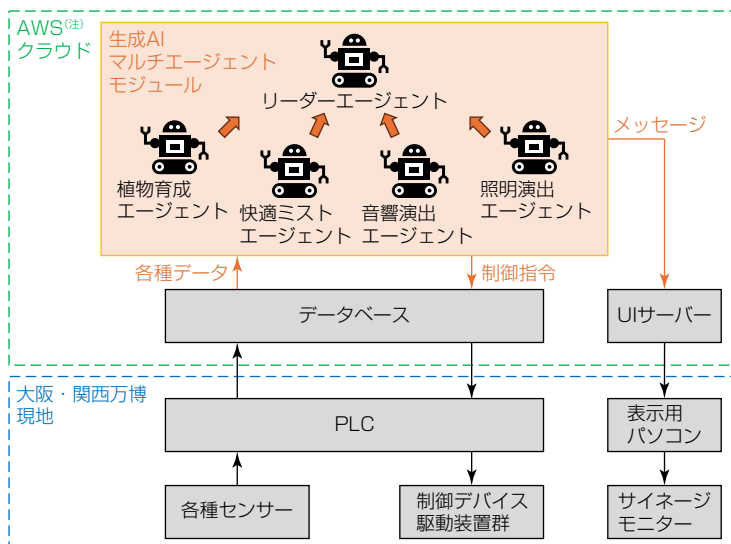


社会的バイアス補正済みLLMの生成例

生成AIマルチエージェントを用いたIoT(Internet of Things)グリーンシェード制御技術を開発した。IoTグリーンシェードは、FA機器とIoT技術を活用して植物の育成を自動管理し、植物シェードを作る当社の取組みであり、大阪・関西万博^(注)にも展示した。この展示では、来場客への快適空間提供を目的に、植物の灌水(かんすい)に加えて、ミスト、音楽、照明の各種機器の制御が求められたが、この開発技術によって実現した。

具体的には各種センサーを用いて取得した土壌、気象、感情、人物属性(年齢、性別等)などの各種データから制御対象ごとの要否を判断する各種判断エージェントを開発した。それに加えて各判断エージェントの出力を踏まえて、どの判断を優先するか最終決断して制御指令を各種機器に送信するリーダーエージェントを含めたマルチエージェント構成にすることで、限られた電力や水量等の制約や状況を考慮可能な制御を実現した(図1)。また、各種エージェントの判断内容を関西弁等の各地の方言に変換し、図2のようなエージェントのやり取りをサイネージモニターに表示することで、来場客が楽しめるようなエンターテインメント性も持たせた。

この技術は今後、当社IoTグリーンシェードへの適用だけでなく、複雑な制約や状況に柔軟に対応する必要があるFAシステムへの応用が期待されるものである。



UI : User Interface, PLC : Programmable Logic Controller

図1-概略システム

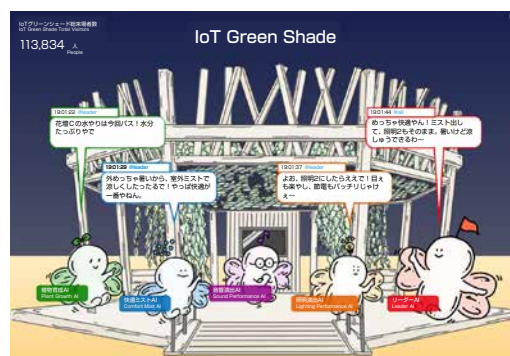


図2-サイネージモニターに表示するエージェントのやり取り

機械学習による方程式の求解技術

Solving Equations by Machine Learning

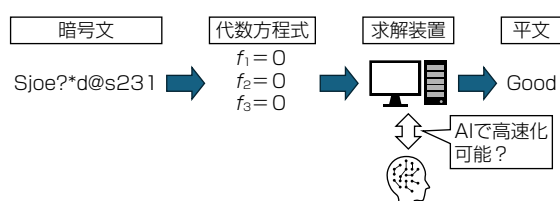
暗号解読法の一つに、多項式が定める方程式(代数方程式)を解くことで、隠されている情報を算出し、暗号を解く方法がある。そのため、代数方程式の高速求解技術は暗号開発での解読リスク評価に必須であるが、長年抜本的な速度更新はされていない。

今回、AIによる代数方程式の高速求解の可否を検証するため、グレブナー基底と呼ばれる、代数方程式の解の情報を持つ対象の計算モデルを機械学習によって構築する技術を開発した。従来の高速求解技術では、既存の計算アルゴリズムにAI技術を組み込むものが主であったが、この技術では、直接的に代数方程式とそのグレブナー基底の間の関係をAIに学習させる。

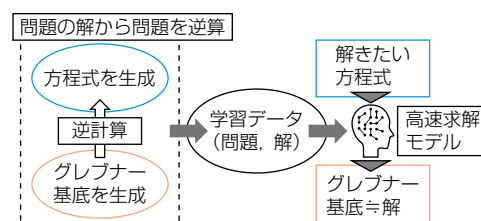
そのために、グレブナー基底から代数方程式を算出する逆数学アルゴリズムを構築した。実験の結果、従来アルゴリズムより100倍以上高速に計算を行う例を発見し、この技術の有用性を確認できた(*1)。

この技術は、AIや機械学習技術が暗号の安全性に及ぼす影響の評価や、暗号以外の科学技術を含めた様々な研究開発分野への展開が期待できるものである。

*1 Kera, H., et al. : Learning to compute Gröbner bases, NeurIPS 2024 (2024)



代数方程式の求解による暗号解読



学習の全体像

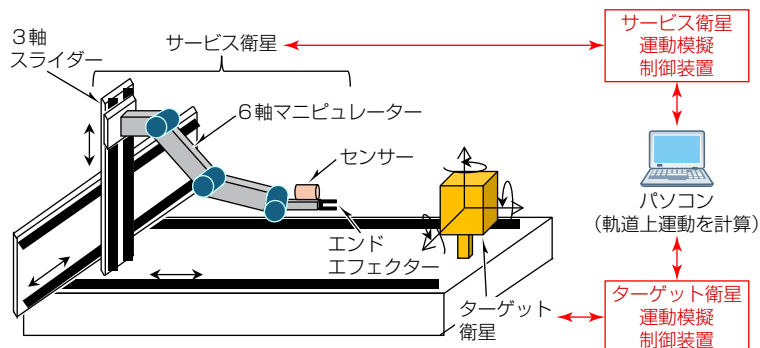
ダイナミクスシミュレーターによる軌道上物体把持技術の検証

Verification of On-Orbit Object Grasping Technology Using Dynamics Simulator

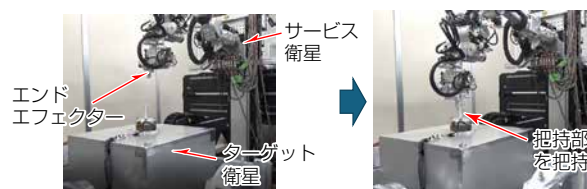
宇宙空間の軌道上で、デブリ回収・人工衛星への燃料補給・修理・機能更新といったサービスを実現するには、サービス衛星がターゲット衛星に接近し、サービス衛星が持つマニピュレーターやドッキング機構でターゲット衛星を捕獲する必要がある。このような一連の動作検証を行う技術試験衛星の打ち上げは高コストで時間を要する一方、数値計算だけでは接触時の検証精度が担保できない課題がある。

この課題解決のために、サービス衛星とターゲット衛星の軌道上での運動を地上で模擬するダイナミクスシミュレーターを開発した。このシミュレーターによって、軌道上サービスに必要な物体把持、物体観測、ランデブードッキングなど多岐にわたる機能を地上で検証できる。このシミュレーターを用いた軌道上物体把持に係る一連の検証を行い、物体把持及び把持時の衝撃力低減技術を実証した。さらに、ターゲット衛星の運動推定技術を実証した。この運動推定技術によって、故障衛星など運動状態が不明なターゲットに対しても、サービス衛星がターンアラウンドしてターゲットとの相対姿勢角速度を小さくした状態で、安全に接近できる。今後も、このシミュレーターを

活用した軌道上サービスに資する機能群の検証を随時進めて、宇宙利用の持続可能性及び多様なミッションへの柔軟な対応性を向上させて、宇宙環境の保全と宇宙産業の発展に貢献する。



ダイナミクスシミュレーターの概要



把持試験の様子

サイバー攻撃に対して動作を継続して安全性を確保するシステムの設計手法

System Design Method for Keeping Safety under Cyberattack by Continuing Its Operation

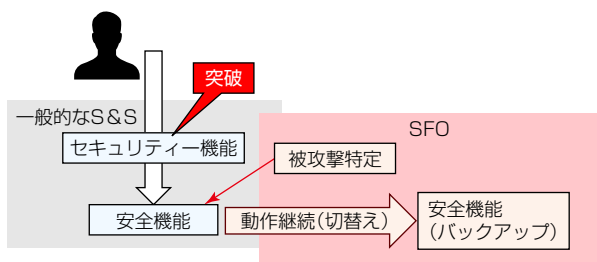
安全性が求められるシステムで、安全機能に対するサイバー攻撃後に被攻撃箇所を特定し、バックアップへの切替えなどで安全動作の継続を可能にする、セキュアフェールオペレーショナル(SFO)の設計手法を開発した。

安全に関わるシステムを機能不全にするサイバー攻撃に対応するためには、安全性とセキュリティを両立するセキュリティ&セーフティー(S&S)設計が必要である。しかし、一般的なS&S設計で扱うセキュリティ機能は、発生が想定される攻撃の防御機能であるため、未知の攻撃でセキュリティ機能が突破された場合の対応が不十分であった。

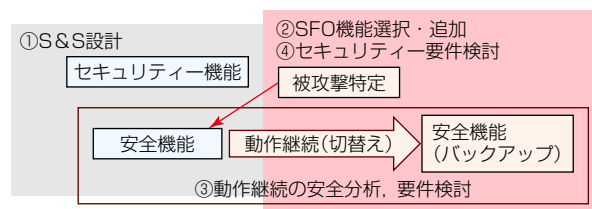
SFOは先に述べた対応に有効だが、被攻撃後の動作継

続処理でシステム構成が変更されるため、変更後の安全性及びセキュリティ検討が必要である。そこでS&S設計後にSFOの導入と安全性及びセキュリティ検証を行う設計手法を開発した。始めに、バックアップ切替えなどの動作継続方式、通信監視などの被攻撃箇所特定方式の候補から、適用可能な方式をシステムに追加する。次に、追加機能による安全侵害を分析し“一定時間で構成変更完了”などの安全要件を追加する。最後に、被攻撃箇所特定精度や構成変更後のシステムの攻撃対策などのセキュリティ要件を追加する。

この手法によって未知の攻撃への対策要件が明確になり、SFO導入時の開発の手戻りを防げる。



SFOを導入したシステム



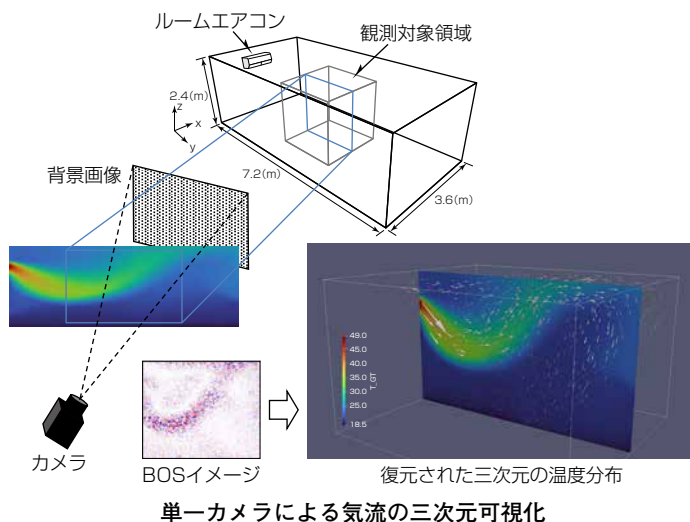
SFO設計手順

▲ カメラを用いた気流センシング・可視化技術

Camera-Based Airflow Sensing

オフィスやデータセンターなどでの快適性向上と省エネルギーの実現に向けて、室内の空気の流れを単一の光学カメラで三次元的に可視化する技術を開発した。この技術は背景指向シュリーレン法(以下“BOS”という。)と呼ばれる光学的な可視化手法を応用している。BOSでは、背景に配置したランダム模様をカメラで撮影し、空気の温度差や密度差による光の屈折で模様がわずかに変位する現象を画像解析することで、空気の流れを可視化・定量化する手法である。従来のBOSによる三次元可視化は、多視点からのカメラ画像と複数方向から観測できる大規模な背景画像が必要で室内用途に現実的ではなかった。一方、単一カメラを用いた場合には奥行き情報が不足するため、三次元構造を正確に再構築することが難しいという課題があった。そこで流体の運動や熱伝達を表す物理法則を組み込んだニューラルネットワークを導入し、観測データと物理的制約を同時に満たすことで、単一カメラの画像から三次元の温度分布と流速分布を高精度に推定することに成功した。その

結果、従来手法に比べて温度分布の平均予測誤差を40%低減することに成功した。またエアコンから流れる空気のシミュレーションでも有効性を確認した。今後は実環境での検証を進める。



■ ビル空調のダウンタイム削減を実現するデジタルツインを活用した遠隔自動故障診断ソリューション ★

Remote Automatic Fault Diagnosis Solution Utilizing Digital Twins which Reduces Downtime of Building Heating, Ventilation and Air Conditioning Systems

ビル空調の故障時に、保守員の現地訪問回数を削減し、ダウンタイムを短縮することで、顧客の事業継続性を向上させる遠隔自動故障診断ソリューションを開発した。

空調装置故障の際、故障原因特定のため保守員が複数回現地訪問したり、製作所へ問い合わせたりしてダウンタイムが長期化することが多い。この解決には遠隔かつ製作所の有識者に代わる自動の故障診断が有効である。

そこで有識者の暗黙知をデジタル化した故障診断のモデルを開発し、故障診断モデルと空調装置のセンサーデータを動的に組み合わせたデジタルツインによって遠隔自動故障診断を実現した。故障診断モデルでは、複数センサーの分析手順を、有識者の知識に基づいて計算機が実行可能な形で定式化した。さらに、センサーデータの時系列変化やデータ間の相関関係についての暗黙知を、時系列分析手法を用いて定式化した。これによって故障診断の自動化を可能にした。

この技術を適用したシステム(図1)では、まず、故障した空調装置について、保存されているセンサーデータから異常時のデータを抽出する。次に、抽出したデータを、故障診断モデルに入力して分析し故障原因を特定する。これによって、保守員の現地訪問や製作所への問合せ回数が削減されて、ダウンタイムを短縮できる見込みである。

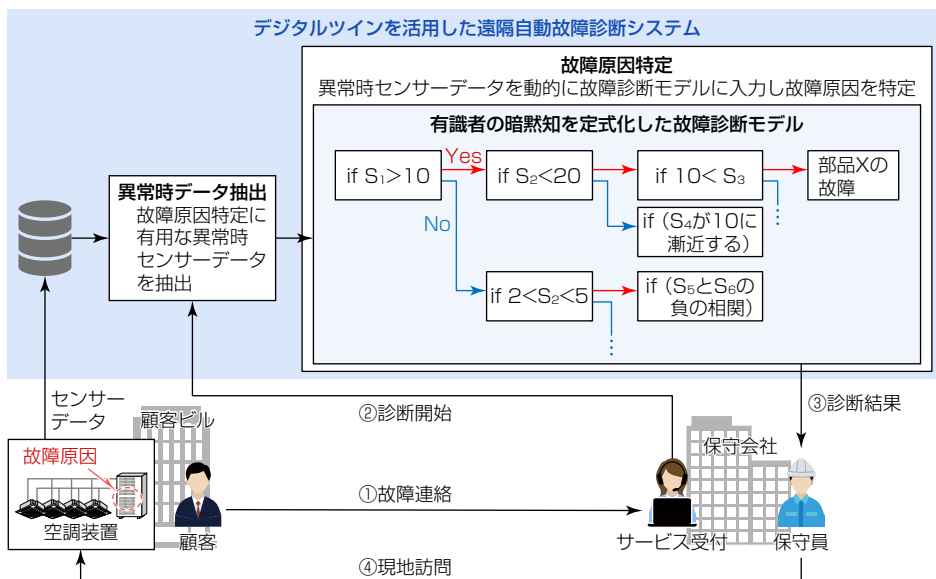
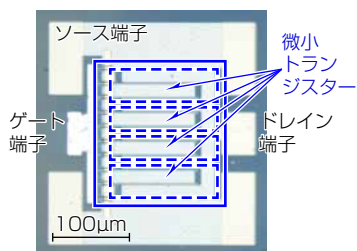


図1-遠隔自動故障診断ソリューション

次世代半導体として有望なダイヤモンドトランジスターを用いた高利得かつ高出力電力の高周波増幅器を開発した。高周波増幅器の出力電力を向上させるためには、トランジスターのゲート幅の拡大が有効である。従来のダイヤモンドトランジスターは二つを並列配置したダブルフィンガー構造によるゲート幅の拡大を行っていたが、ゲート抵抗が増大し、利得が下がるという課題があった。

今回、早稲田大学とともに微小トランジスターを並列配置するマルチフィンガー構造のダイヤモンドトランジスターの考案及び製造プロセスの確立を行い、ゲート幅の拡大とゲート抵抗増大の抑圧を両立するトランジスターの実現に成功した。一方、提案したトランジスターは出力インピーダンスが高く、トランジスター直近にバイアス回路を配置すると出力整合回路の損失が大きくなり、高周波増幅器の出力電力及び効率が低下する。そこで、出力整合回路の出力側にバイアス回路を配置し、出力整合回路の損失を最小化した。その結果、世界で初めて^(*)、周波数1.8GHzで利得7 dBと出力

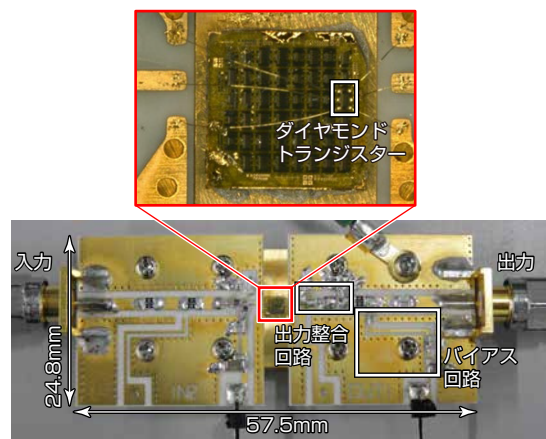


マルチフィンガー構造の
ダイヤモンドトランジスター

出典：Kudara, K., et al. : L-Band Diamond Amplifier With Multi-Finger Structure, IEEE Electron Device Letters, 45, No.12, 2491~2494 (2024)

電力100mWを持つダイヤモンド高周波増幅器の動作実証に成功した。今後、更なるトランジスターの微細化に伴う高周波化によって、5G(第5世代移動通信システム)等の通信増幅器への適用が可能になり、無線装置の低消費電力化への貢献が期待される。

*1 2025年11月18日現在、当社調べ



ダイヤモンド高周波増幅器

ダイヤモンド高周波増幅器と
ダイヤモンドトランジスター

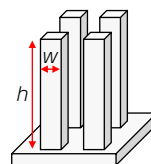
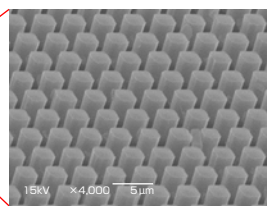
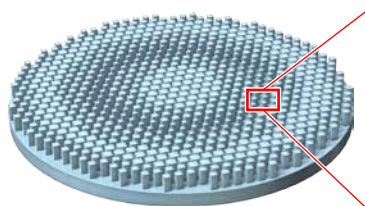
長波長赤外線センサー向けメタレンズ

長波長赤外線(波長8~14μm)の光は、人間や動物など温度を持つ物体から発せられる熱放射であり、センサーはサーモグラフィー等の用途で広く用いられる。この波長域のレンズは、材料の屈折率が低いため一定以上の厚さが必要であり、透過率の高いGe(ゲルマニウム)やカルコゲナイドガラス等の材料が一般的に用いられて、材料のコスト面での課題もある。小型で低コストな長波長赤外線センサーを実現するために、Si(シリコン)製のメタレンズの開発を進めている。メタレンズは、波長より小さな構造体であるメタアトムを透過した光が位相変化を起こすことを利用し、平面基板に屈折レンズと同等の特性を持たせることが可能である。また、比較的安価なSi基板を用いて、薄く低コストにレンズを実現することが可能である。

今回、高さ10μm、太さ1.5~2.5μmの柱形状メタアトムを、同心円状に太さを変えて配置することで、集光特性を持つメタレンズを設計した。半導体プロセスによって、厚み280μmのSiウエハー上にメタレンズを形成し、長波長

赤外線センサーに取り付けることで、手の平や黒体炉等の赤外線画像の取得に成功した。

今後、長波長赤外線センサーで必要になる真空封止構造中にメタレンズを取り込むことで、小型で低コストな長波長赤外線センサーの実現を目指す。

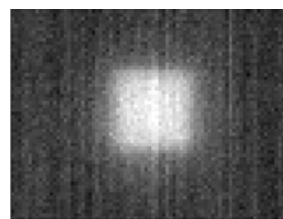


h: メタアトムの高さ 10μm
w: メタアトムの幅 1.5 ~ 2.5μm

試作したメタレンズの外観とメタアトム



人の手



黒体炉 50℃

メタレンズを用いて撮影した赤外線画像

▲ 疑似牽引力によるフィンガーフリー型VR触覚デバイス

Finger-Free Haptic Interface for VR Using Pseudo-Traction Feedback

製造現場のVR(Virtual Reality)訓練では、実作業に近い感覚(視覚・触覚・力覚など)を提示することが学習の習熟度の向上に有効である。従来の力覚提示では、人差し指と親指の感覚が重要なため、それらの指に触覚デバイスを装着する必要があった。このためVR空間での部品操作や工具使用に制限が生じて、VR訓練の学習内容が現場での作業に効果的に結び付かないという課題があった。

当社はこの課題に対して、薬指・小指・手の平で把持するフィンガーフリー構造を採用し、指先の自由を保ったまま上下方向の疑似牽引(けんいん)力を提示する触覚デバイスを開発した(図1)。この触覚デバイスは、広い皮膚接触面積と十分な振動強度を両立するために、アクチュエーターの配置と接触構造を最適化している。2基のアクチュエーターの同期駆動によって皮膚に広範囲に非対称振動を与えて明瞭な方向感覚を生成する(図2)。さらに、VRシステムと連動し、ユーザーの操作や対象物に応じた接触・牽引感を高速に切り替える制御が可能である。動的な

作業文脈にも追従可能な構成になっている。

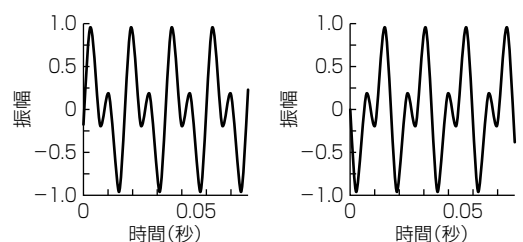
この触覚デバイスを用いたVRシステムによって、実作業に即した操作による訓練が可能になり、産業技能教育のリアリティーと効率を大きく向上させることができる。



図1-フィンガーフリー型触覚デバイス



(a) アクチュエーターと皮膚の接触の様子



(b) 上向き信号による重さの表現 (c) 下向き信号による軽さの表現

図2-上下方向の疑似牽引による重量感提示

▲ リザーバーコンピューティング応用連想記憶技術

Reservoir Computing-Based Associative Memory

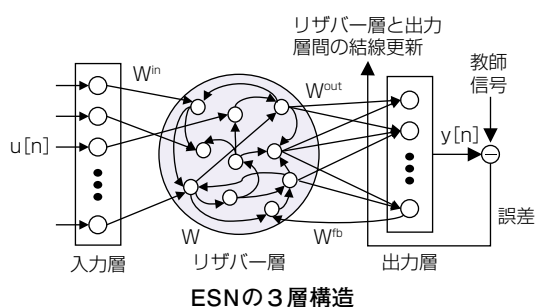
リザーバーコンピューティング(RC)は、再帰的結合を持つ3層構成の神経回路網で構成されて、少ないハードウェアリソースで低消費電力の実装が可能な機械学習の一方式である(*1)。連想記憶は、従来の計算機でのメモリアクセスとは異なる記憶方式であり、ある記憶情報を入力することで、それに紐(ひも)付いた記憶情報を引き出すことができる。例えば記憶情報が画像である場合、欠損部分を含む画像を入力するとその欠損部分を補完する形で元の記憶情報を読み出せるという特長がある。

今回、この連想記憶をRCの一種であるEcho State Network(ESN)で実装し、自然言語処理への応用を狙って、画像ではなく9文字から成る26個の英単語を2値ベクトル化して学習させたRC連想記憶について2種類の学習を行った。一つは学習時の入力と出力が同じ自己連想記憶、

もう一つは入力と出力が異なる相互連想記憶についてである。前者については、1文字の綴(つづ)り間違いを含む英単語をRC連想記憶に入力したところ、63%の精度で綴り間違いが補完された。後者については、26個の英単語を数珠つなぎでペアリングし、ある英単語の入力から次の英単語が出力・想起されて、それを再び入力してその次の英単語が83.4%の精度で順番どおり想起された(*2)。これによって綴り間違いを含む英単語であっても自動補完する自然言語処理向け要素技術を確立した。

*1 Lukoševičius, M., et al.: Reservoir computing approaches to recurrent neural network training, Computer Science Review, 3, No.3, 127~149 (2009)

*2 Kage, H.: Implementing associative memories by Echo State Network for the applications of natural language processing, Machine Learning with Applications, 11, 100449 (2023)



ESNの3層構造

学習用入力パターン



テスト用入力パターン(欠損画素1個含む)



テスト用出力パターン(連想記憶による欠損画素補正)



ESN自己連想記憶による2D画像欠損部分の自己補完

演算能力向上のため量子コンピューターの大規模化開発が急速に進展している中、単一の量子コンピューターでは実現困難な規模の量子ビット数を達成するため、複数の量子コンピューターを相互接続してシステム全体として動作させる技術が求められている。特に遠隔地に量子情報を運ぶための媒体としての光子を介した接続技術が有望視されているが、量子測定の結果を高速に共有して協調動作を可能にする高度な制御システムが必須になる(図1)。

このような背景の下、東北大学・大塚研究室との共同研究で、FPGA(Field Programmable Gate Array)を用いたRF(Radio Frequency) 反射測定技術によって半導体量子ビットの量子状態を高速に読み出すこと成功した(図2)。今回開発した技術は、高周波信号の反射特性をFPGAでリアルタイム解析することで高速かつ高精度に量子状態を検出／判定するものであり、複数の量子コンピューター間で量子情報を転送する際の同期制御を可能にする基盤技術になる。

現在、この技術を更に発展させて、量子状態読み出し結果に基づいて複数の量子コンピューターの量子ゲート操

作を協調制御する統合制御システムの構築を進めている。FPGAの高速性と柔軟性を最大限に生かすことで、将来的には複数の量子コンピューターを接続した大規模量子コンピューターシステムの実現に貢献していく。

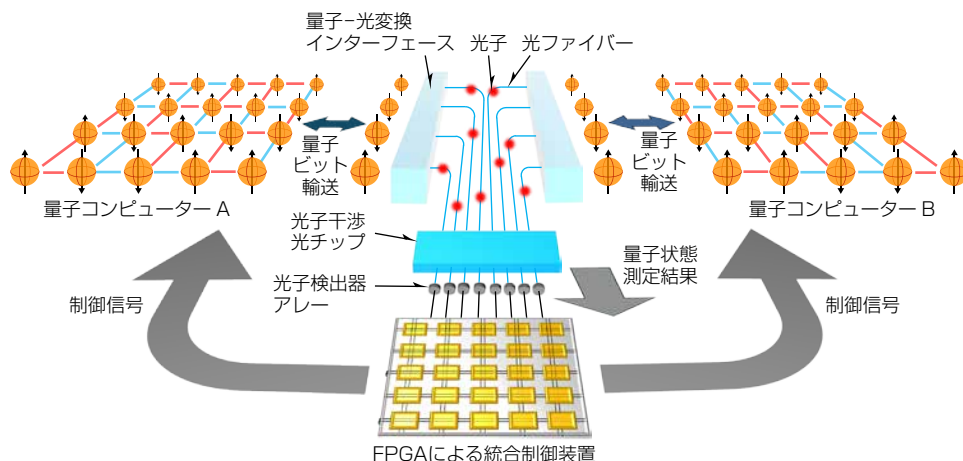


図1-FPGAを用いた量子コンピューター接続制御装置

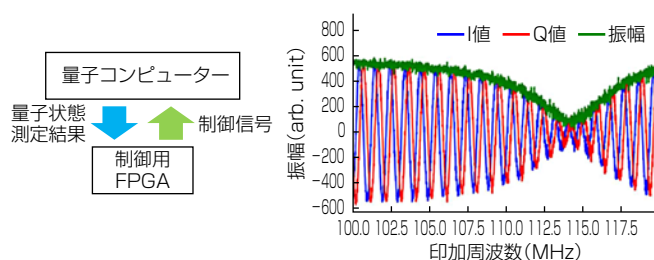


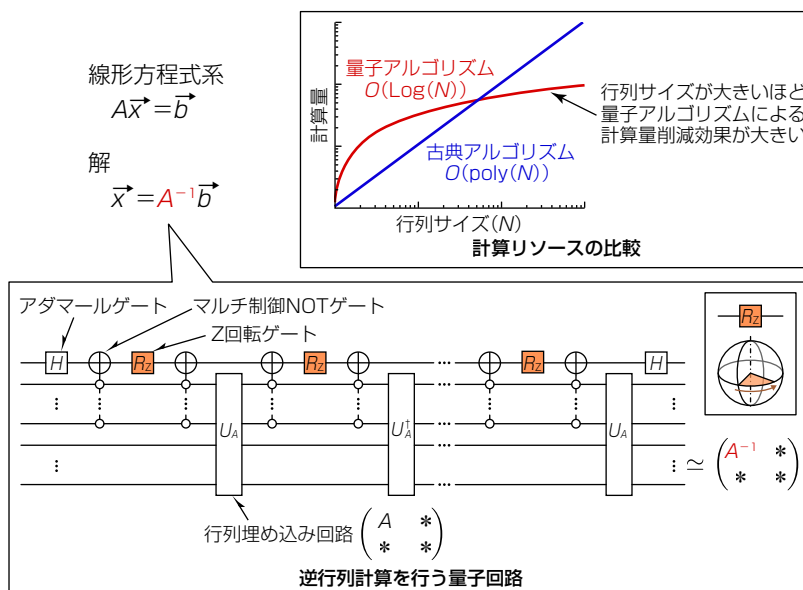
図2-FPGAを用いた制御測定システムの概略図とRF反射測定結果

量子コンピューターを用いた逆行列計算

Matrix Inversion Using Quantum Computers

製品や新材料の開発効率化・高度化を目的に計算機による解析が進む中、対象の大規模化に伴う計算リソースの不足が課題になっている。主要な計算の一つである逆行列計算で、量子コンピューターの活用によって大幅な計算量の削減が期待されるが、具体的な効果の定量的な見積りはこれまでされてこなかった。

今回、特定の行列に対して逆行列を計算する量子回路を構築し、誤り耐性型量子計算での基本量子操作にまで分解することで、計算量を定量的に評価した。その結果、分解に伴う計算量の増加要因を考慮しても、古典計算に比べて指数的に計算量が削減できることを実証した。今後、誤り耐性型量子計算の実現に向けて、更に実用的な問題に対する定量的な検証を進める。



量子コンピューターによる逆行列計算のイメージ