

# 本村・藤木研究室 リモート研究室紹介

---

□ 3/11 PM1-2:30

- PM1-2: 教員による研究室紹介 & QA
- PM2-: 修士学生による相談会 (教員退席)

# 東工大・情報通信系 本村・藤木研究室 紹介スライド

2024年 3月 11日

東京工業大学 科学技術創成研究院  
AIコンピューティング研究ユニット (ArtIC)  
情報通信系 本村・藤木 研究室



Tokyo Tech



# AIコンピューティング研究ユニット: ArtIC

2019年4月  
に発足  
2020年4月からフル  
メンバで活動中

すずかけ  
台キャン  
パス J3棟  
17F

<http://www.artic.iir.titech.ac.jp>



ARTIC

トップ 新着情報 メンバー 研究活動 発表論文 研究室生活 アクセス リンク 言語

## AIコンピューティング 研究ユニット

情報処理ハードウェアの革新

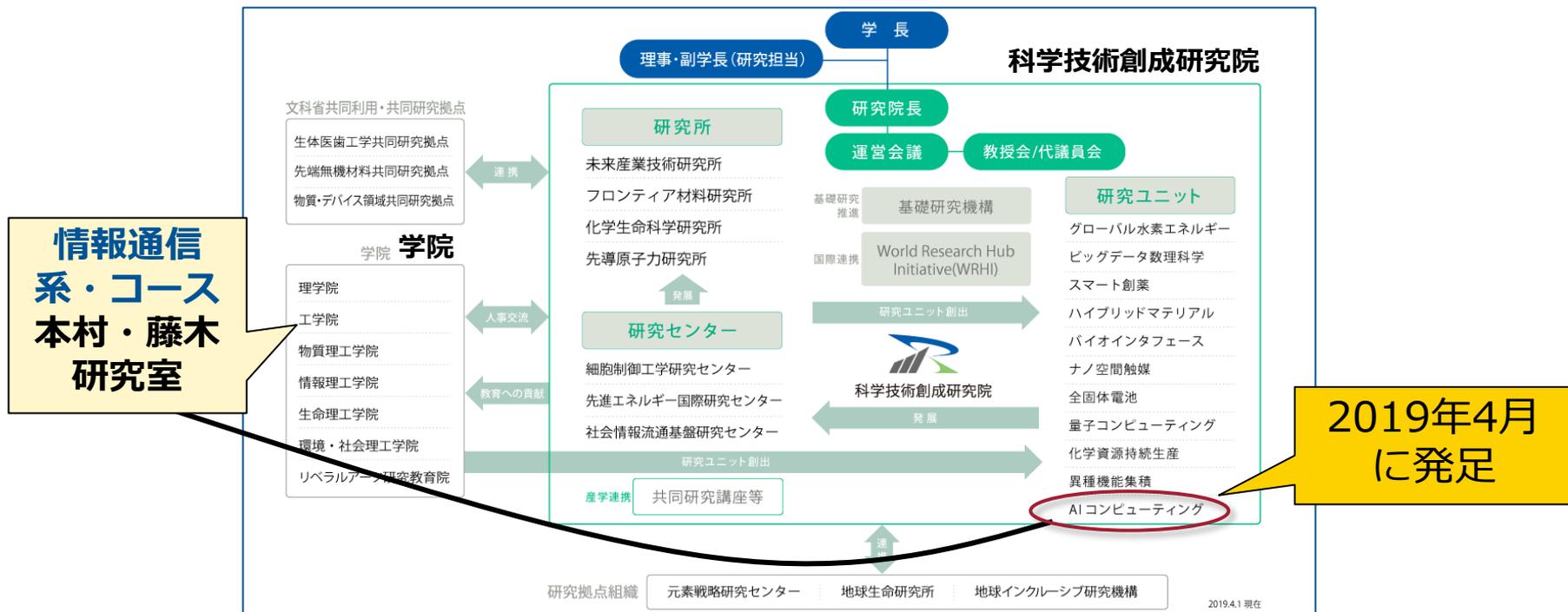
### ArtICのミッション

- 新たな AI プラットフォームの創出**  
Creation of Novel Platform for AI Computing
- 共通基盤コンピューティングアーキテクチャの構築**  
Construction of Common-basis Computing Architecture
- 自律性・安全性・エネルギー効率・コスト効率の高いハードウェア基盤の実現**  
Realization of Hardware Foundation with Good Autonomy, Safety, Energy, and Cost Efficiency



# ArtICの成り立ち

## 工学院・情報通信系 と 科学技術創成研究院



## Artificially Intelligent Computing Research Unit

もう一つの意味: ArtなIC

=> ソフトとハードの協調研究

=> 素敵なハードウェア(HW)



# ArtIC発足の背景

## 急ピッチで進むAI技術の社会応用



自然言語・自動  
対話型AI



スマート  
ロボット



自律航行  
ドローン



スマート社会  
インフラ

## AI処理の急拡大 => エネルギー消費問題の深刻化

より豊かで低環境負荷なSociety5.0を実現したい



「AIコンピューティング」の処理効率を向上する  
情報処理アーキテクチャの技術革新が必要

東工大のアクション：  
AIコンピューティング研究ユニット(ArtIC)の設置 (2019～)

ChatGPT Burns Millions Every Day. Can Computer Scientists Make AI One Million Times More Efficient?

John Koetsier Senior Contributor @  
John Koetsier is a journalist, analyst, author, and  
speaker.

Follow

Feb 10, 2023, 09:00am EST

Listen to article 8 minutes



Running ChatGPT costs millions of dollars a day, which is why OpenAI, the company behind the viral natural-language processing artificial intelligence has started ChatGPT Plus, a \$20/month subscription plan. But our brains are a million times more efficient than the GPUs, CPUs, and memory that make up ChatGPT's cloud hardware. And neuroomorphic computing researchers are working hard to make the miracles that big server farms in the clouds can do today much simpler and cheaper, bringing them down to the small devices in our hands, our homes, our hospitals, and our workplaces.

(Forbes)

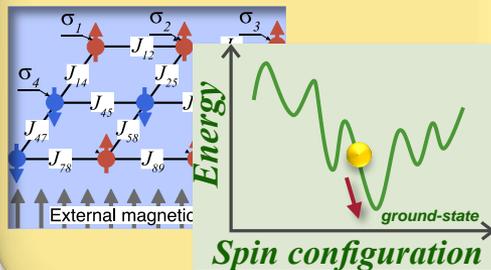
ChatGPTの学習  
には、一般家庭の  
数百年分もの消費  
電力が必要

(Stanford大学報告書)

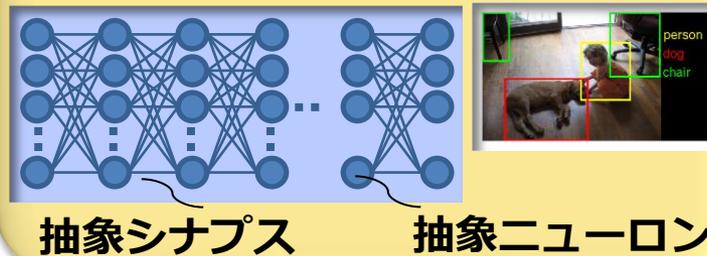
# ArtIC: 研究ターゲット (～2023年度)

人工知能(AI)応用の急速な拡大  
 「**コントロール**駆動から**データ**駆動へ」  
 計算機アーキテクチャの革命

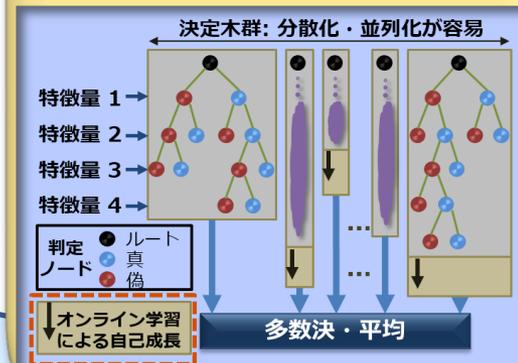
組合せ最適化問題 →  
 スピン格子のエネルギー  
 ギー最小化問題



大量データの学習 →  
 強力な推論・識別・予測能力



説明性・制御性の高さ  
 と低学習負荷の両立



深層ニューラルネット  
 ・ディープラーニング

アニーリング計算機  
 (非量子)

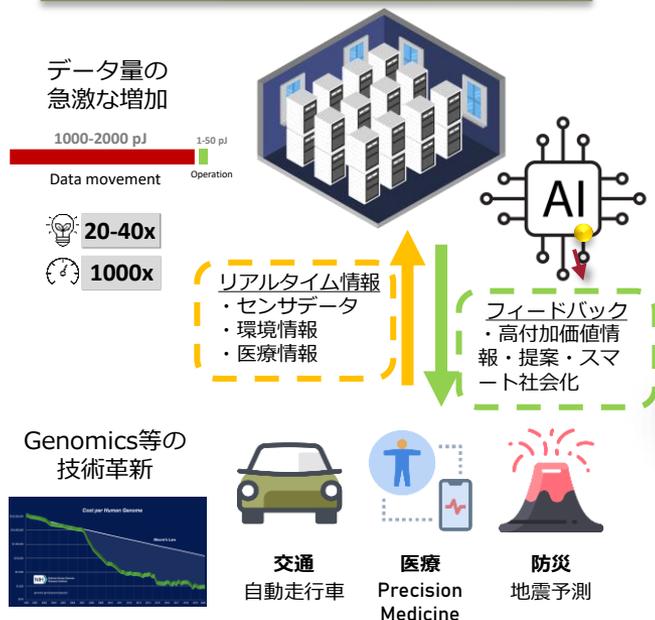
統計的機械学習  
 (アンサンブル学習等)

構造型情報処理アーキテクチャ  
 として共通基盤化

アルゴリズム理解 ⇒ アーキテクチャ研究 ⇒ ハードウェア実現

# ArtIC: 研究ターゲット (2024年度~)

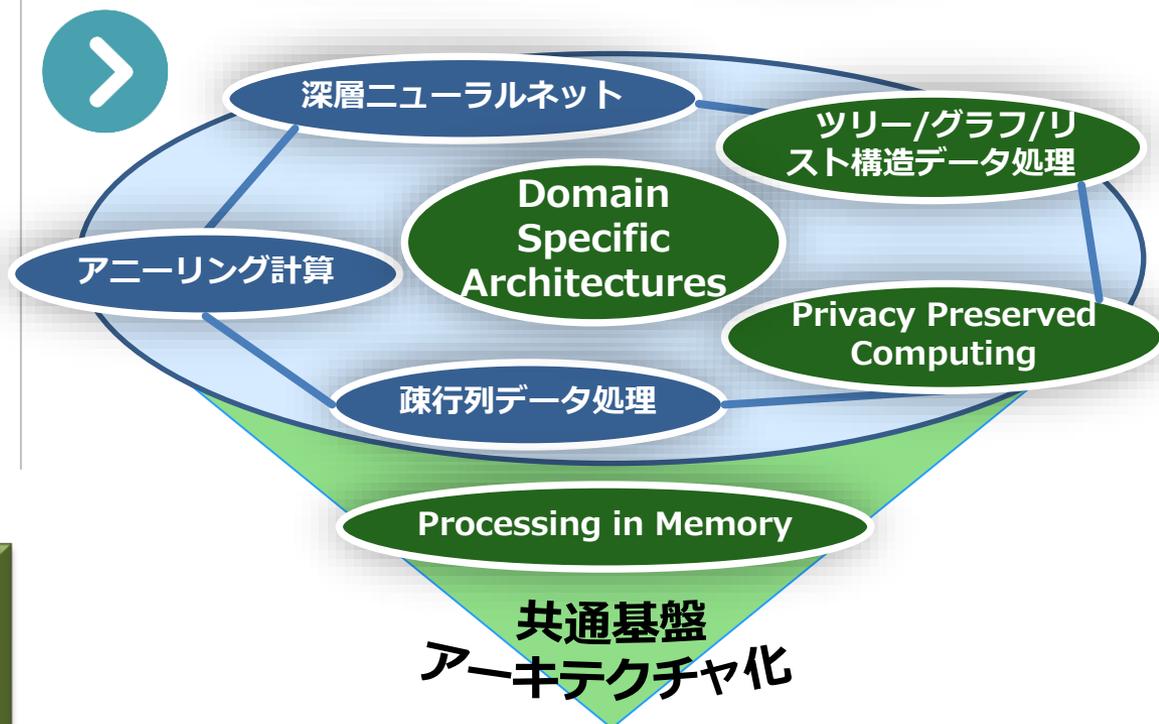
最新の社会課題を  
より強く意識



LLM・生成モデル等の台頭を受け、AIアーキテクチャ分野での闘い方をアップデート

従来テーマを  
ブラッシュアップ

新たな取り組み



アルゴリズム理解 => アーキテクチャ研究  
=> ハードウェア実現

# 2021-22年発表: 機械学習HWアーキテクチャ

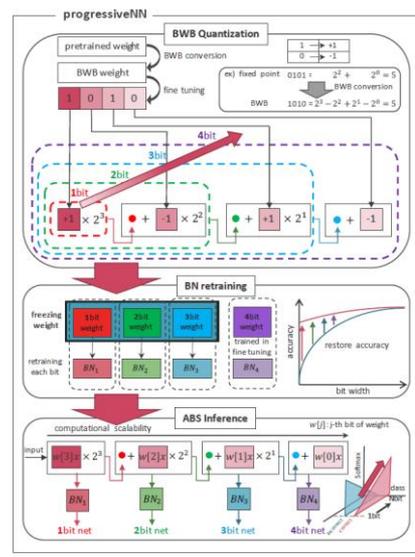
- 深層ニューラルネットワーク・決定木アンサンブルは様々な応用分野で実用
  - 画像認識, 音声認識, 自然言語処理, 意思決定, ...
- 学習・推論にかかる計算量・消費電力の問題
- 高い精度の予測を**低計算量**, **低電力**, **かつ**, **高速**に行う技術への要求

### Hot Chips2021発表: Descartes

東工大ニュース: スマホやロボットなどで高効率なAI処理を行うプロセッサアーキテクチャを開発

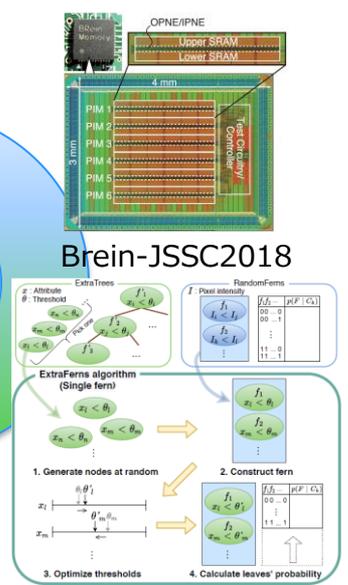
開発した世界トップレベルの集積回路最大26.5 TOPS・消費電力

東工大ニュース: 東工大大学院工学研究科情報学専攻の田中真由美准教授と、共同開発した世界初の「ニューロモック」が、Hot Chips 2021で「最優秀集積回路賞」を受賞した。このチップは、AI処理の効率化を実現する。消費電力は従来のチップの約1/10に抑えられ、処理速度は約10倍に向上した。



スケーラブルネットワーク Progressive-CANDAR2020

HW 協調 Algo



決定木アンサンブル ExtraFerns-CANDAR2020

### ISSCC2022発表: Hiddenite

モデル構築コントローラ

圧縮したスーパーマスク

スーパーマスク展開ユニット

スーパーマスク

スーパーマスク

スーパーマスク

活性値メモリ

活性値

ニューロン演算器

活性値

乱数シード

重み生成ユニット

活性値

圧縮マスク

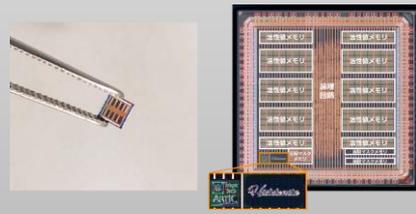
展開

スーパーマスク

スーパーマスク

東工大ニュース: 隠れニューラルネットワーク理論を具現化したAIチップを世界で初めて開発

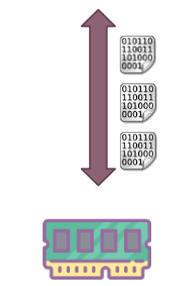
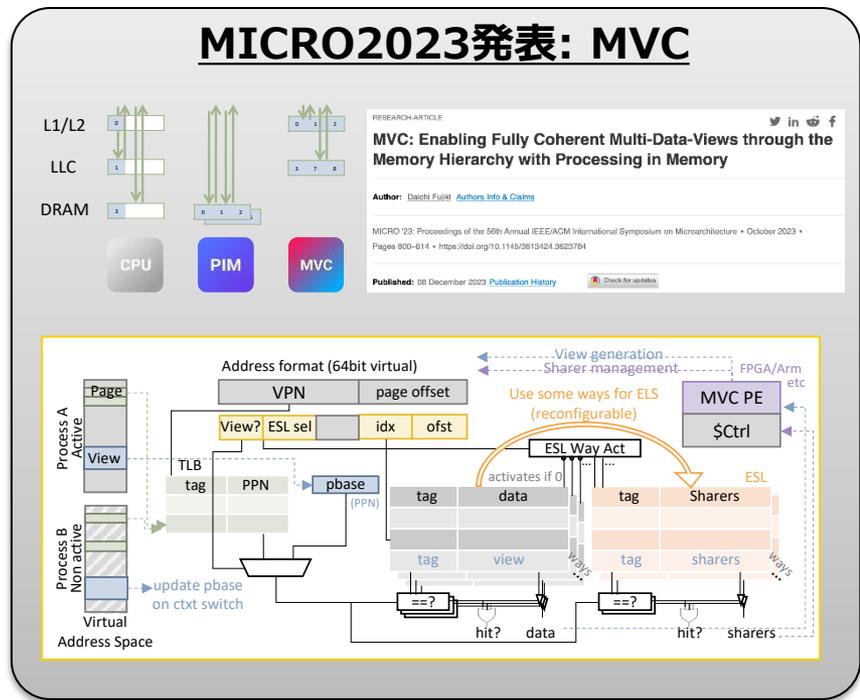
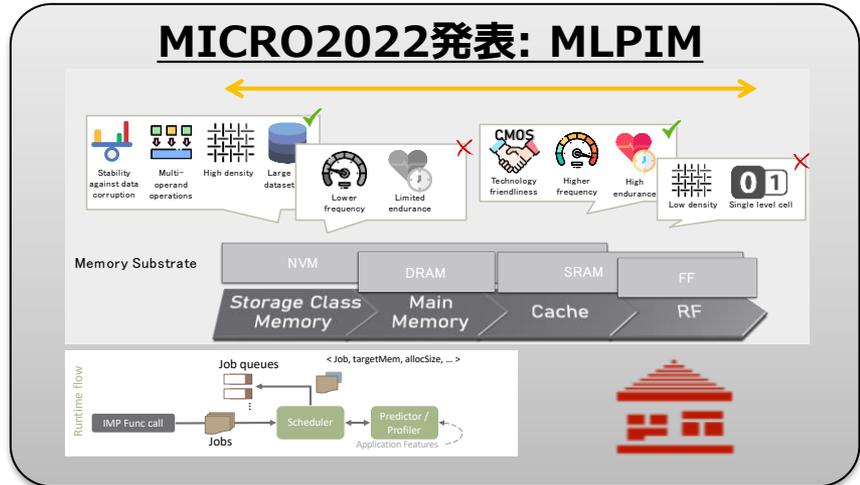
東工大ニュース: 東工大大学院工学研究科情報学専攻の田中真由美准教授と、共同開発した世界初の「ニューロモック」が、Hot Chips 2021で「最優秀集積回路賞」を受賞した。このチップは、AI処理の効率化を実現する。消費電力は従来のチップの約1/10に抑えられ、処理速度は約10倍に向上した。



Hiddenite chip

# 2022-23年発表: インメモリ計算システム

- データ移動コストを低減するため、メモリ内で計算を行う技術
  - 機械学習、データベース処理、ゲノムデータ処理、暗号処理、...
- 計算機は様々なメモリを使用し、特性が大きく異なる
- インメモリ計算という新しい技術をどう計算機に組み込むか? (プログラミング、スケジューリング、CPUとの協調動作)



↔ データ移動の低減  
 ☄ 超並列計算が可能

◇ CPU-Memが狭帯域

● >90%の面積がメモリに割当

🔄 >90%を占める受動的メモリを能動的計算資源に転換

# ArtICの研究レイヤ

## 技術レイヤ

「論理」を相手にする世界



「物理」を相手にする世界

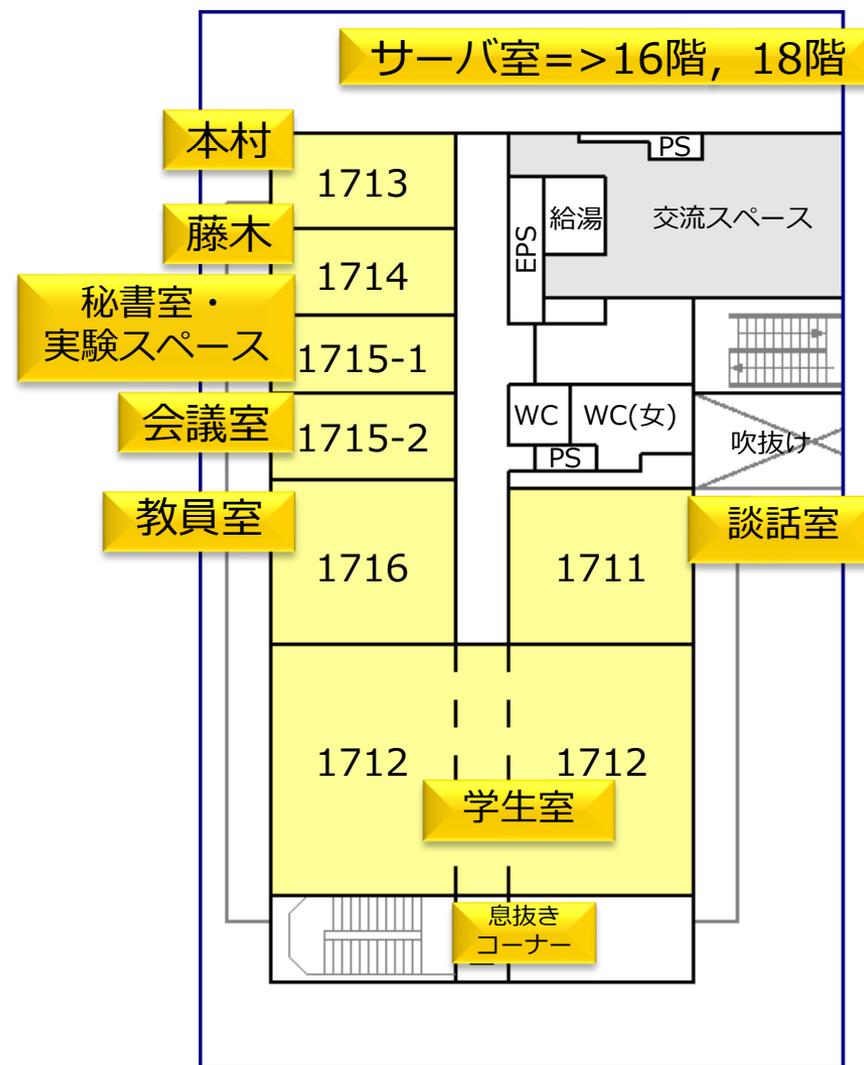
## 身に付くスキル・知識 (例)

- 機械学習 アルゴリズム
- 画像処理 アルゴリズム
- ディープラーニング(人工知能)
- 学習環境構築・GPU活用
- 組み込みシステム設計
- プロセッサアーキテクチャ
- 非線形・近似計算理論・応用
- 柔らかいHW (動的再構成LSI)
- 信号処理
- デジタル回路・Verilog設計
- FPGA設計・利用
- LSI設計

# ArtIC人員・居室構成 (24年4月)

		本村研 藤木研	
教員	教授	本村	
	准教授		藤木
	助教	ティエム	
	特任助教	川村	
スタッフ	技術支援員	(募集中)	
	秘書	2名	
学生	D3	3名	
	D2	3名	
	D1	3名	
	M2	5名	
	M1	0名	
	B4	3名	

ユニット  
一体運営  
(学生17名)



ArtICホームページにフロア紹介  
ビデオ(3年前)があります

# 2021年11月のすずかけ台・ArtICオフィス















← 1711-1716  
1713-1714-1715

715  
HC  
樓上

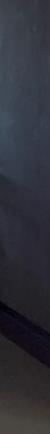
ART C





















# 研究の立ち上げ方・進め方

- 導入教育
  - 輪講: アーキテクチャ, 機械学習
  - 実習: ディープラーニング, FPGA
- 研究テーマ配属
  - 4年生 4月頃, 修士入学 5月頃
  - 本人の希望を聞きながら調整
- 研究の進め方
  - 全体進捗会議: 週1回
  - 4グループ毎(深層学習, 決定木, アニメーション, 疎行列処理)の研究報告: 週1回
  - その他, 適宜個別ミーティング
  - コアタイムはありません
- M1で国際会議発表を目指す

# 修論・卒論 タイトル名 (2022年度)

## □ 修士論文

- 全並列アニーリングの解探索性能を向上させる動的なスピン反転機構の研究、小此木
- 局所解脱出を容易にするアニーリング手法とそのアクセラレータ設計、神保
- 乱数重みニューラルネットワークにおける精度・サイズトレードオフの向上に関する研究、大越
- 高効率な量子化決定森推論アクセラレータのためのモデル最適化手法の研究、北島

## □ 卒業論文

- 強い宝くじ仮説に基づく超軽量物体検出ネットワーク、大塚
- 同変性ネットワークに基づく自律走行向け強化学習手法、塩田
- 表形式データを対象とした決定木とニューラルネットワークの融合型機械学習手法の研究、山倉
- 2スピン同時フリップを並列試行可能なシミュレーテッド アニーリング手法、兵頭
- 組合せ最適化問題のアニーリング解法に関する難易度評価、四元

# 最後に…

- ユーチューブチャンネル (研究室HPからリンクあり/Artic 東工大で検索可能)
  - <https://www.youtube.com/channel/UCJY897-DXhrnfMWC4gYIwFw>
- 計算機アーキテクチャやハードウェア設計に興味を持つ皆さんのArtICへの参加を歓迎します
- 二研究室で共同運営しており, 研究ユニット内に垣根はありません
- 一線級の国際会議で発表できるグループを目指しています
- 近未来の社会ニーズに即した、実戦的な研究活動を主体としています
- 産学連携、大学間連携、国家プロジェクト参画を活発に進めています
- ArtICでは, 博士課程進学もお勧めしています
  - 日本学術振興会(学振)研究員への採択をプッシュ, 修士からRA雇用
  - 国際的には, 先端技術職で勝負するには博士号取得がスタンダード