

超高効率AI計算基盤の研究開発

実施者	株式会社Preferred Networks、株式会社インターネットイニシアティブ、国立大学法人北陸先端科学技術大学院大学
概要	AI技術を活用した計算可能領域の拡大を実現する超高効率AI計算基盤の研究開発を行う。情報通信システムの一部としての超高効率AI計算基盤に向けて、(1) 高い電力効率を実現する次世代AIアクセラレータ・システムおよびその制御技術の研究開発、(2) 大規模商用サービス高密度データセンター基盤技術の研究開発、(3) AI計算基盤の共同利用における実AIワークロードの効率化に関する研究に取り組む。本研究開発では、開発した技術を組み入れたテストベッドを構築し、実AIワークロードを用いた実証実験により、AI計算基盤としての効率化およびその評価を行う。

【超高効率AIアクセラレータ・システムおよびその制御技術の研究開発】

2023: PFN MN-Core™ 2 (7nm) 2026: 次世代超高電力効率AIアクセラレータ

- AIに最適化した基本アーキテクチャ
- アプリケーション特性に合わせた各種回路への電力最適分配
- システム全体の最適化制御ソフトウェア

⇒ AIアクセラレータ・チップおよびノード・ラックレベルでの高効率システムの実現

【大規模商用サービス展開に向けた高密度データセンター基盤技術の研究開発】

AI計算基盤に求められる**密度**を **直接水冷技術** の応用により実現

- 商用データセンターにおける直接水冷技術利用の**レファレンスモデルと実装**
 - 商用データセンターでの運用に耐える**モジュール化技術**
 - 水冷・空冷ハイブリッド、負荷の差異・変動に対応した水冷仕様等
 - 経済性、継続性、即応性、運用可能性、省エネ性の5要件(*)を充足
- 商用データセンター+AI計算基盤における**省エネ指標**

* 次世代商用データセンターに求められる5要件

- 経済性：IT機器とファシリティ（建物、電気・空調設備）などの**ライフサイクルの異なる機器**の経済的な更改
- 継続性：部品・モジュールの高い**可用性** / 産業としての**継続的に運用可能な事業構造**
- 即応性：需要に応じて**タイムリーに計算キャパシティを増強** / 次世代機器の速やかな導入（新陳代謝促進）
- 運用可能性：IT機器の**負荷変動と消費電力**への**冷却方式**での対応 / 騒音等**労働環境**の低下の防止
- 省エネ性：**データセンター全体**の効率を計測可能な**電力利用効率**の指標

⇒ AI計算基盤を効率良く収容し、運用可能なデータセンターの実現

【超高密度AI計算基盤の共同利用における実AIワークロードの効率化に関する研究】

実AIワークロードとAI計算基盤の協調制御による資源割り当て最適化・効率化

→ 上記次世代AIアクセラレータと高密度データセンター基盤技術の成果と合わせて**圧倒的な高実効スループット**（単位時間）の実現を目指す。

⇒ **ハードウェア・ソフトウェア協調制御によるAI計算基盤全体での最適化と効率化の実現**

