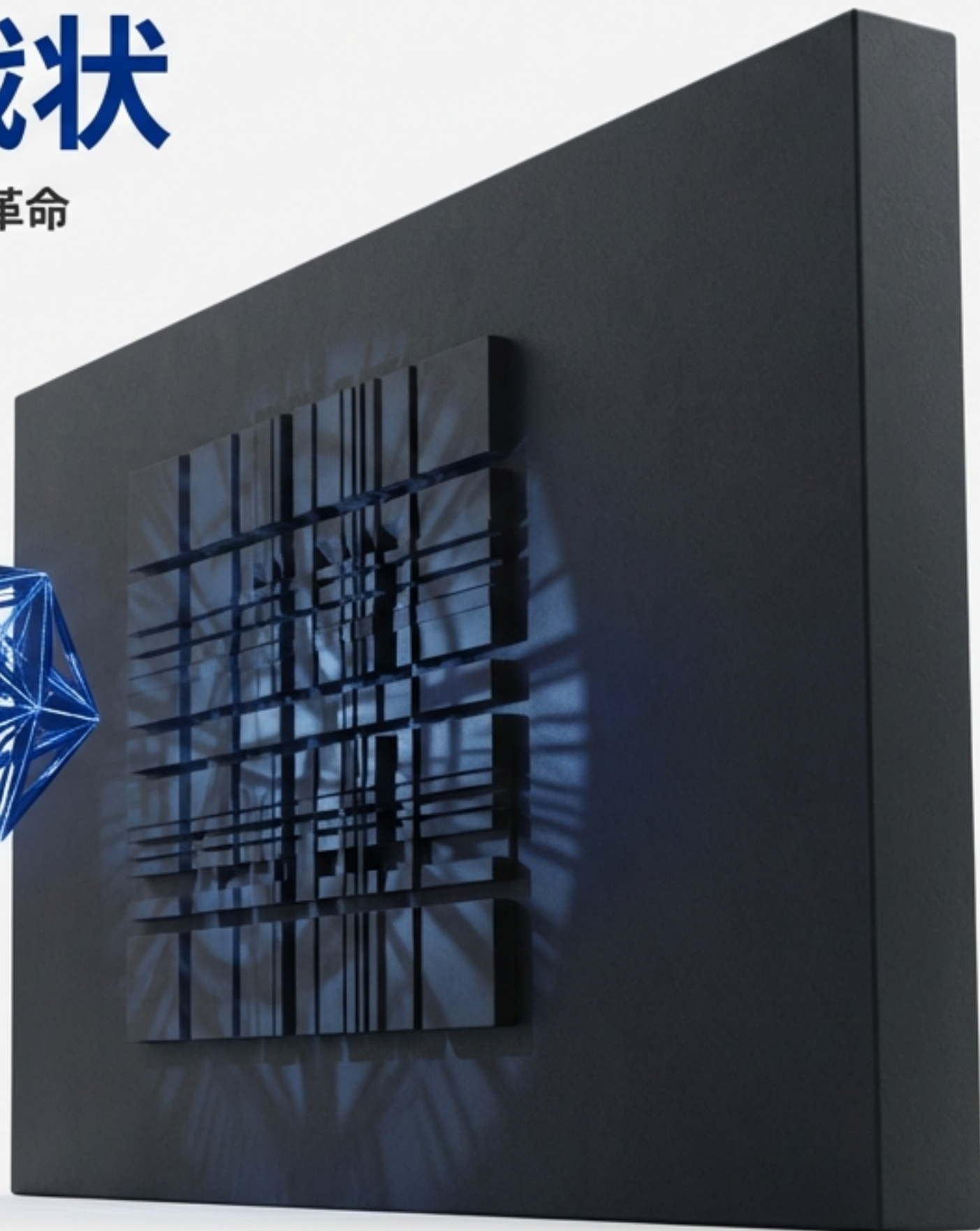


AI—強時代への挑戦状

元PlayStation開発陣が仕掛ける、AI半導体のアーキテクチャ革命



NVIDIAが築いた巨大な市場。
しかし、その牙城を崩す鍵は、30年来の
アーキテクチャに潜む根本的な欠陥にあった。



AIを動かす絶対王者、NVIDIA

AIの学習から実行まで、あらゆる場面でNVIDIAのGPUは不可欠な存在となっている。その市場支配力は絶対的であり、代替不可能なインフラとして君臨している。



600兆円

時価総額

伝説のチーム、再び集結

日本発のAI半導体スタートアップ。



創業者 藤原氏

元ソニー。PlayStation 2, 3向け
半導体開発を主導。



共同創業者 中島氏

元富士通。スーパーコンピュータ
向け半導体開発に従事。



開発チーム

元東芝。同じくPlayStationの半導体
開発に従事した技術者たちが集結。



現代計算機に潜む、 30年来の根本的欠陥

フォン・ノイマン・ボトルネック

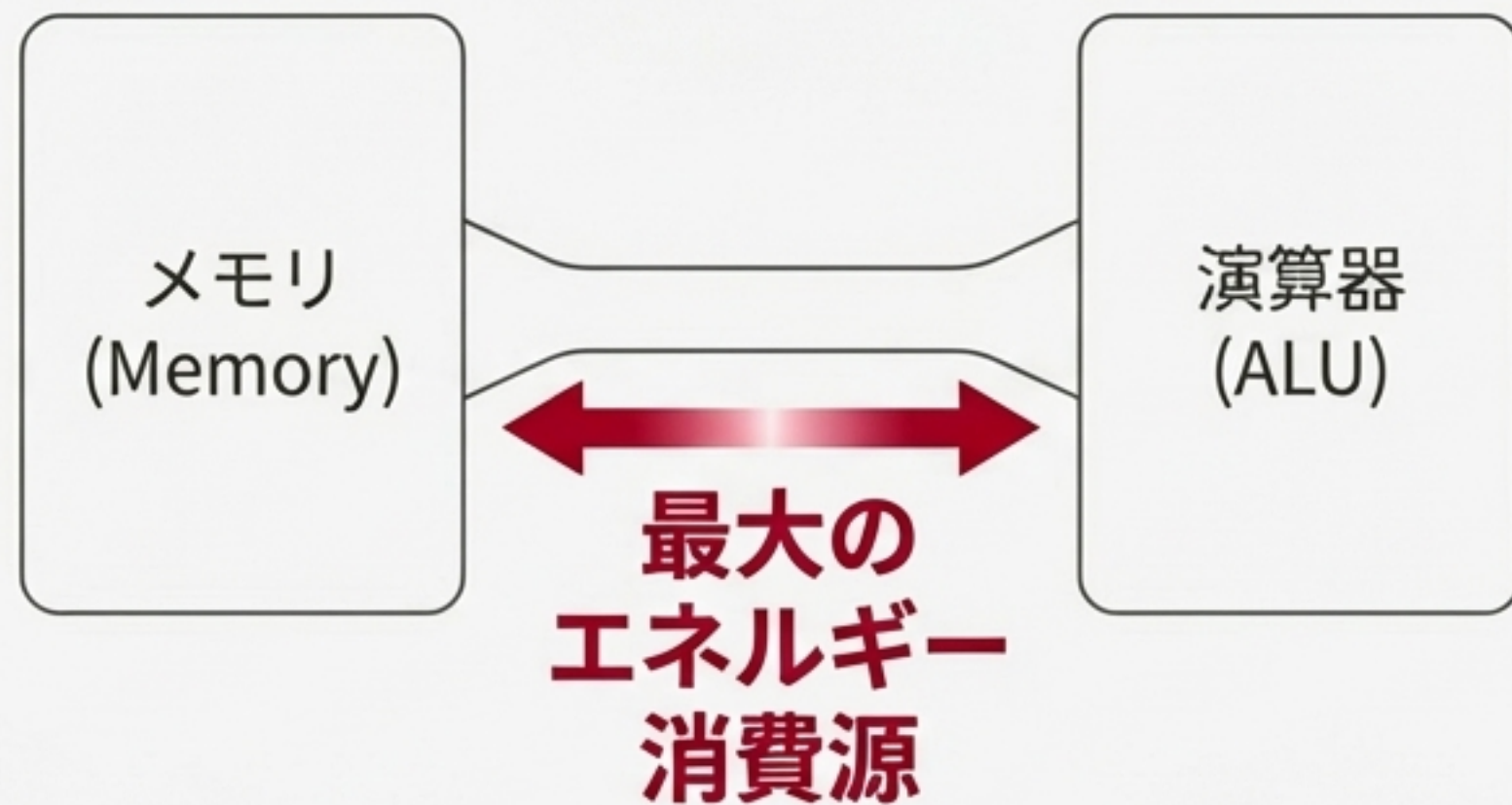
問題の本質：計算そのものより、データの「移動」がエネルギーを消費する。

NVIDIA研究者による試算：

計算1回に必要なエネルギー：**20ピコジュール**

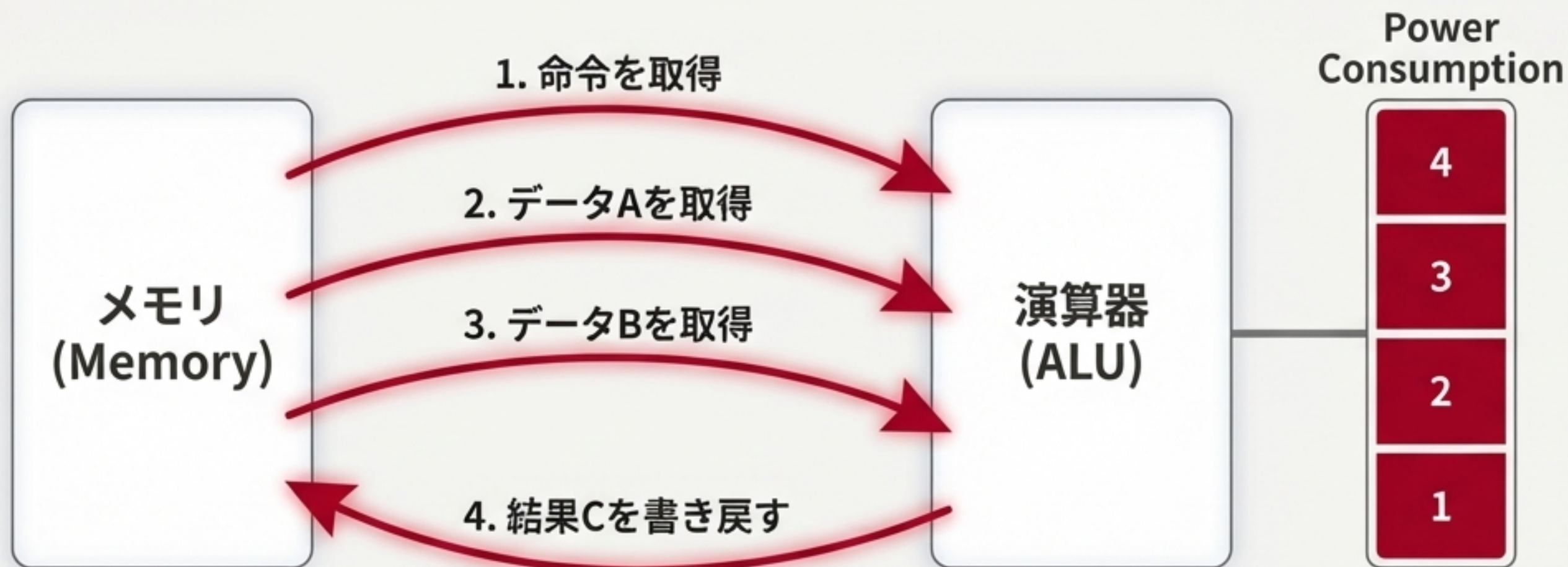
データを1mm移動するのに必要なエネルギー：

26ピコジュール



なぜGPUはこれほど電力を消費するのか？

フォン・ノイマン型（CPU/GPUが採用）



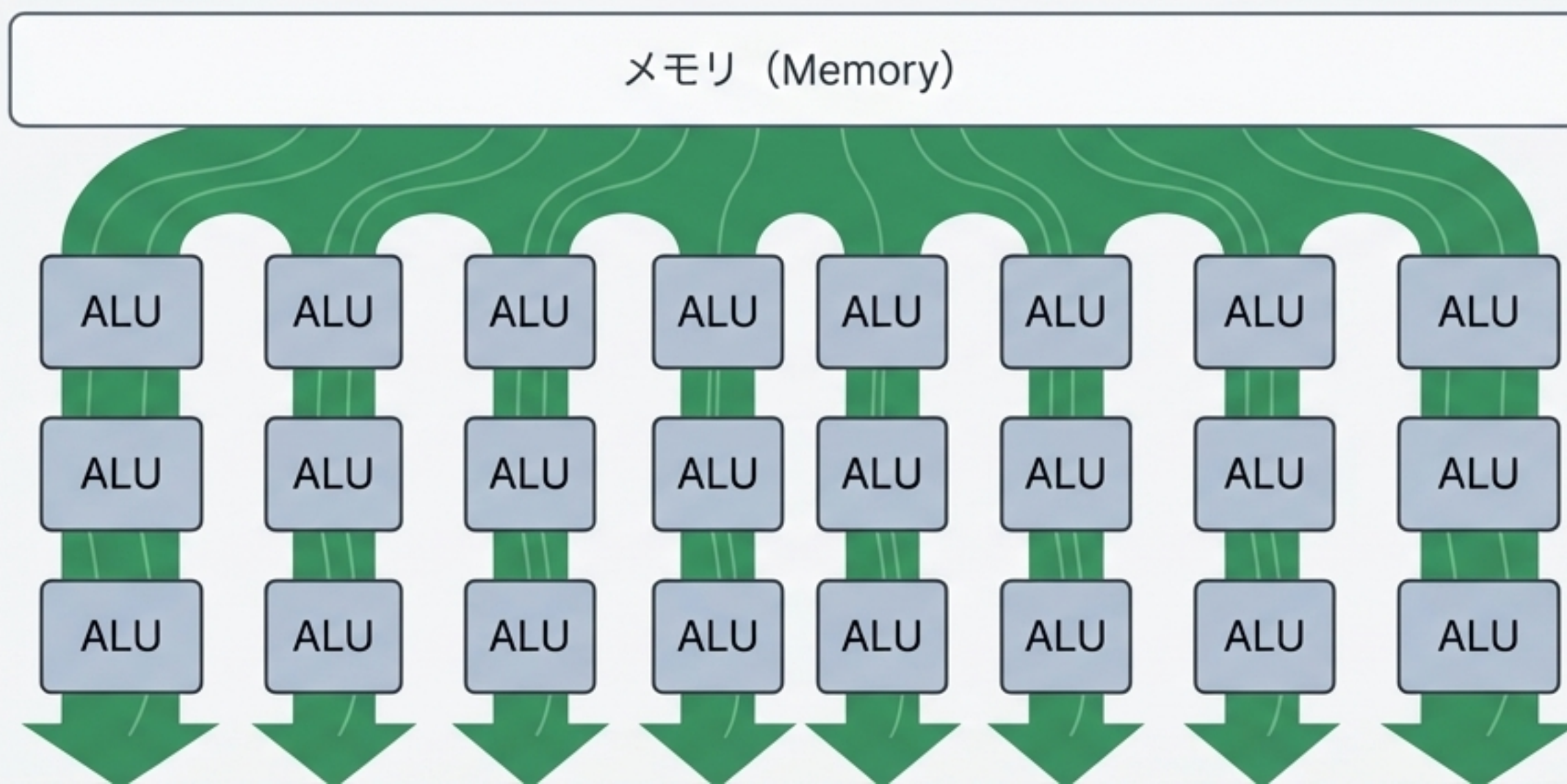
1回の計算サイクルで、最低でも**4回**のメモリアクセスが発生。
この頻繁なデータの往復が、電力消費と性能のボトルネックとなる。
AIサーバーの消費電力の半分以上は、このデータ移動だけで消費されている。

Googleの答え：データフロー型アーキテクチャ「TPU」

データフロー型（TPUが採用）

演算器がデータを取りに行くのではなく、メモリ側から計算に必要なデータを一方通行で送り込み続ける（プッシュ型）。データの往復をなくすことで、メモリアクセスの要求をゼロにし、電力効率を劇的に向上させる。

技術名: シストリック・アレイ (Systolic Array)

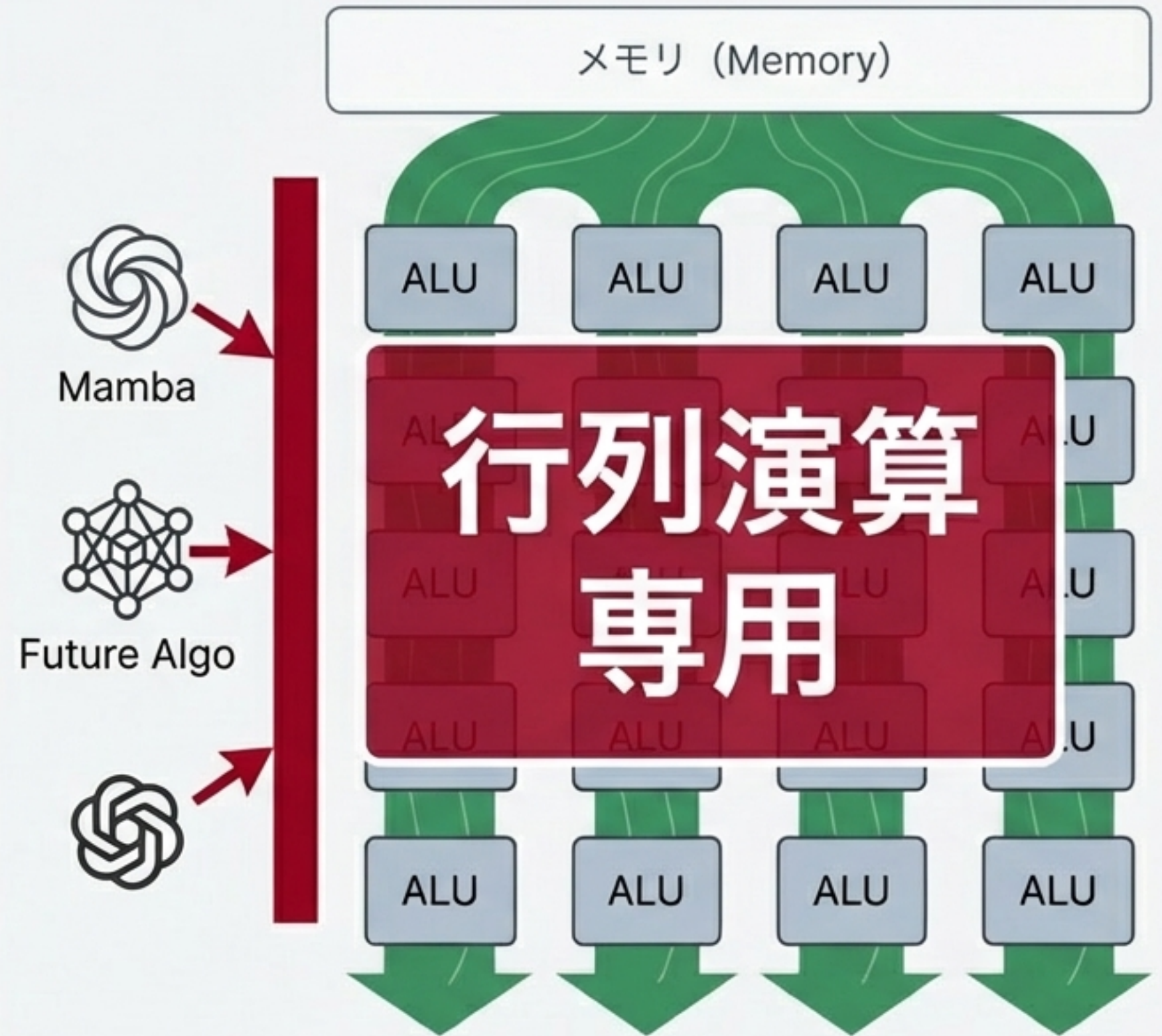


効率性と引き換えにした 「不自由さ」

得意分野: 行列演算 (Tensor Calculation)。これはTransformerモデルの計算に最適化されている。

課題: AIの世界はTransformer以降も進化を続けている（例: Mamba）。主流のアルゴリズムが変化した場合、ハードウェアそのものが最適ではなくなる可能性がある。

結論: TPUは高い電力効率を持つが、汎用性が低い。



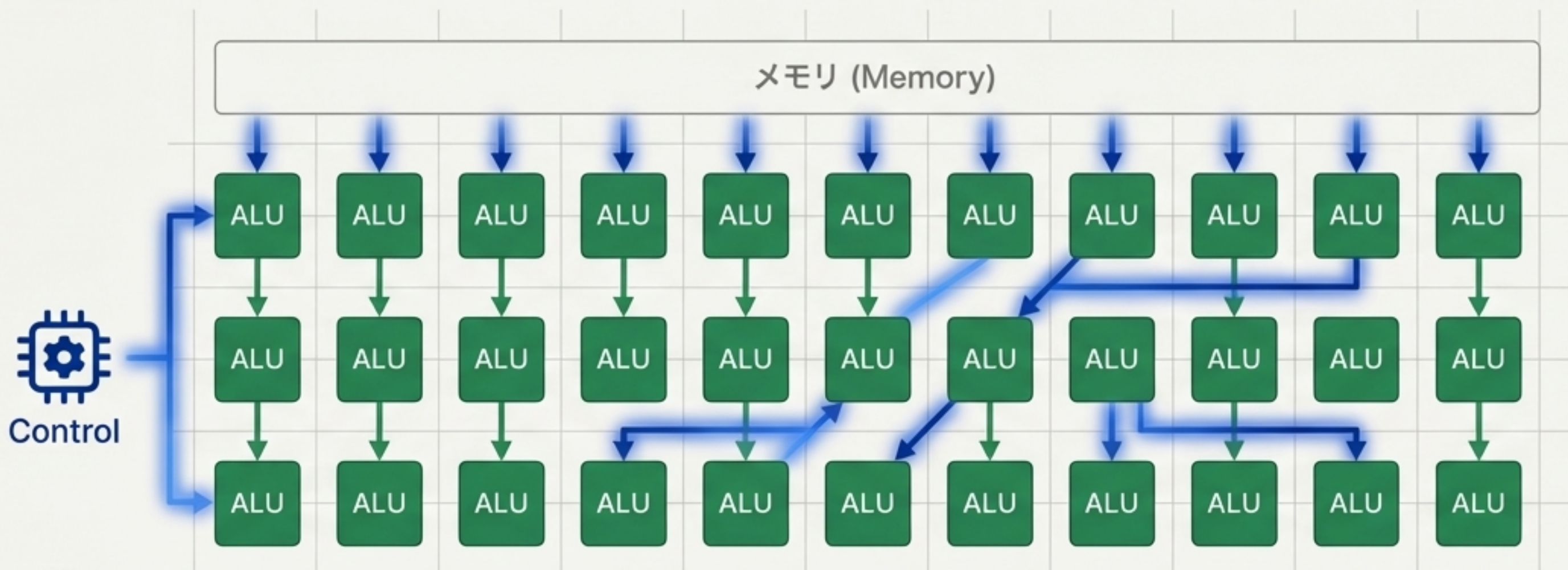
Renzuの妙手：第三のアーキテクチャ「CGLA」

Coarse-Grained Reconfigurable Architecture

データフロー型を基本とし、高い電力効率を実現。

最大の特徴：演算器間のデータ経路を、後から自由に再構成（リコンフィギュア）可能。

結果：行列演算だけでなく、将来登場するであろう新しいアルゴリズムにも柔軟に対応できる。

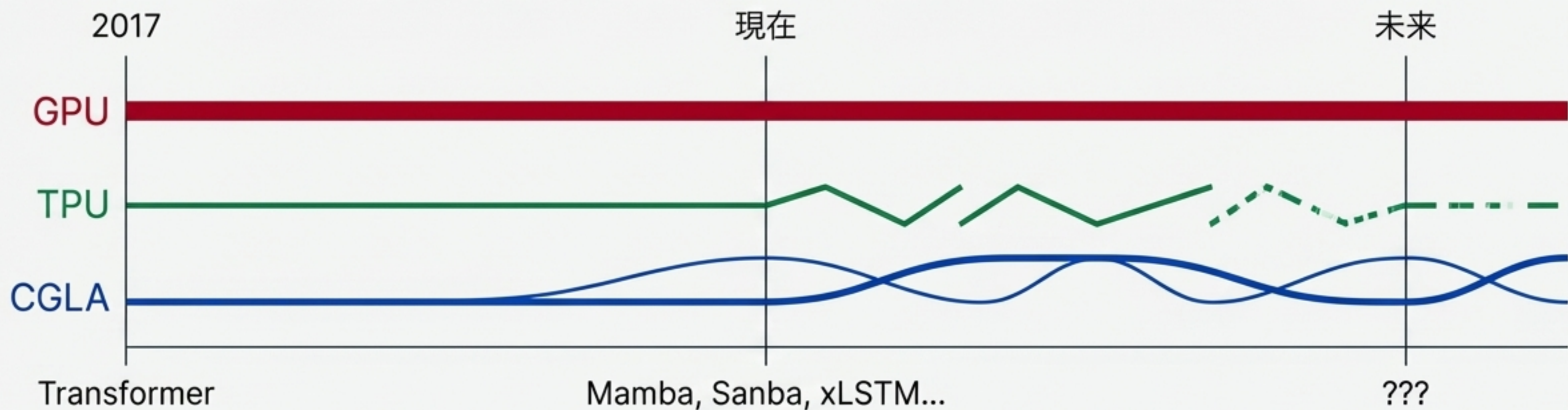


効率と汎用性、二律背反の克服

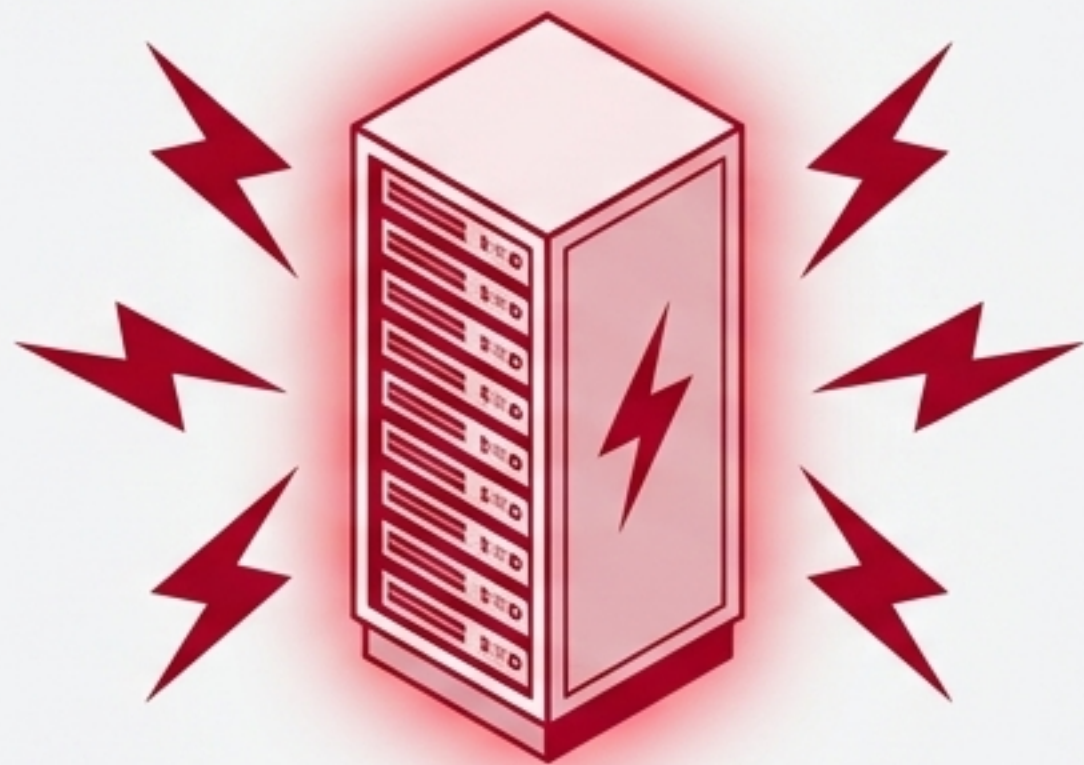


AIアルゴリズムの急速な進化に追従する唯一の選択肢

- **AIモデルの変遷:** 2017年のTransformer（行列演算中心）から、現在はMamba, Sanba, xLSTMなど、多様なアプローチが次々と登場。
- **CGLAの優位性:** 新しいアルゴリズムが登場しても、**ソフトウェアでデータ経路を再構成するだけで最適化が可能**。ハードウェアを作り変える必要のない「**将来性 (Future-Proof)**」を実現。



消費電力「最大9割削減」がもたらす事業価値



GPU搭載サーバー

- データセンターの運用コスト（電気代）の**大幅な削減**
- 冷却コストの削減
- 環境負荷（CO2排出量）の低減



CGLA搭載サーバー

- 電力制約で不可能だった場所へのAI導入（エッジコンピューティング）
- より大規模で複雑なAIモデルの運用が可能に

最後の壁：最強の堀「CUDA」エコシステム

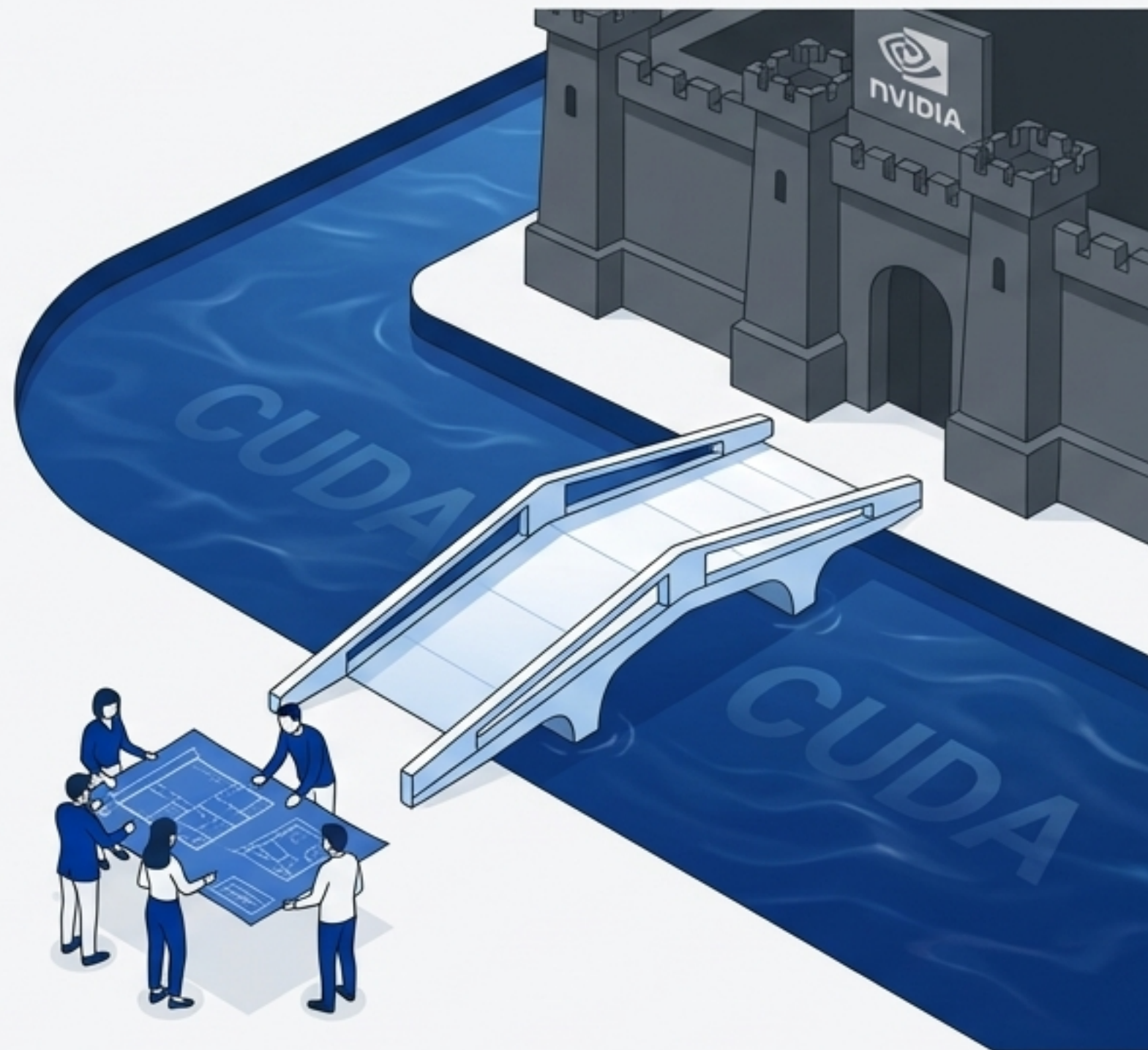
CUDAとは：NVIDIAが提供する独自のソフトウェア開発環境。

課題：世界中のAIソフトウェアはCUDA上で動くことを前提に作られている。既存のAI資産を活かしづらい。

```
device='cuda'
```

この一行が、NVIDIAの優位性を支えている。

Renzuの認識：ハードウェアの性能だけでは勝てない。このソフトウェアの壁を越えることが最大の挑戦。



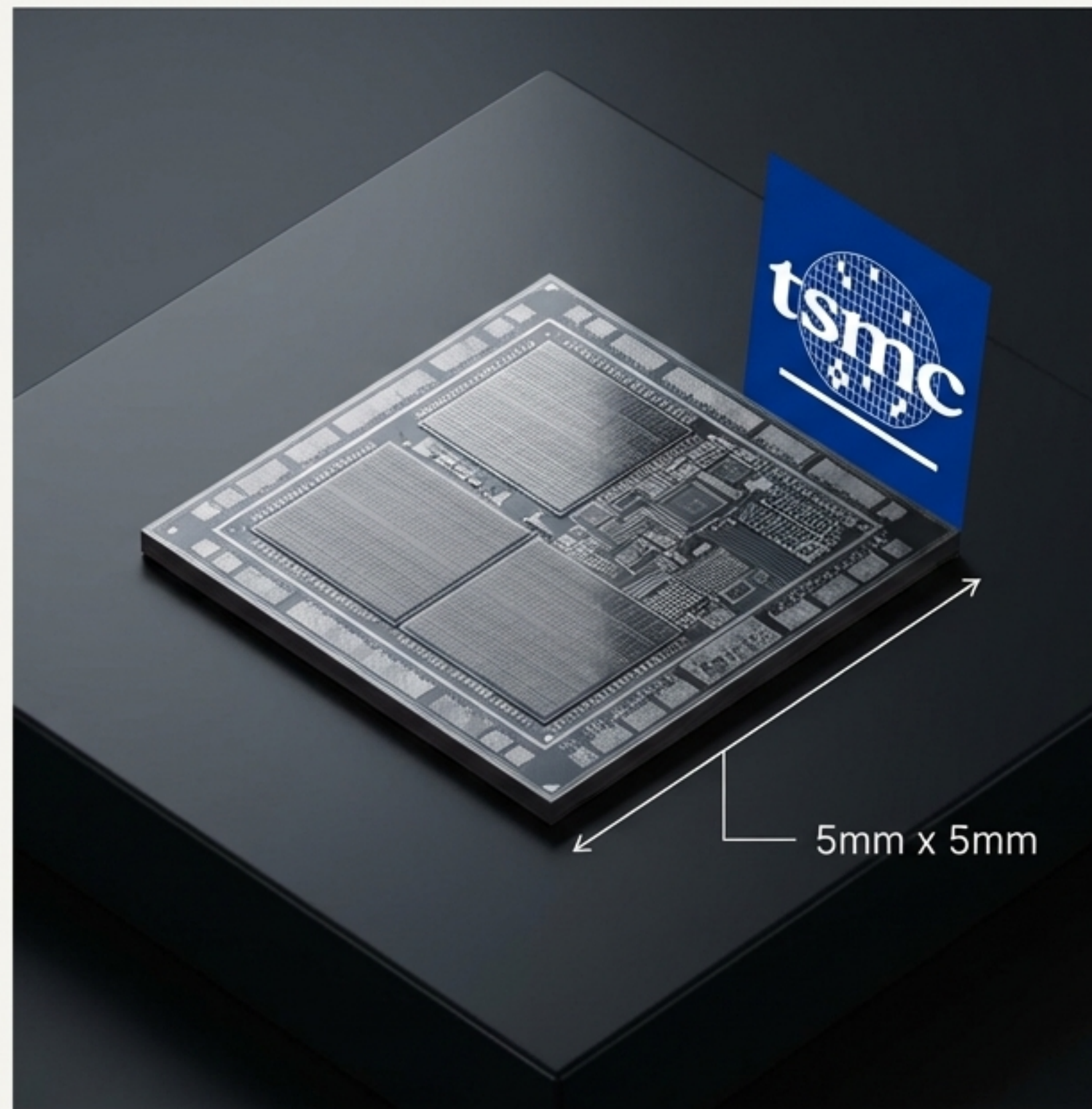
設計から製造へ： 構想は現実になる

- アーキテクチャ設計は完了。
- 今年春、TSMCにて製造委託。
- **5mm角**の半導体チップとして完成予定。

ラピダスとの違い

Renzu: ファブレス。半導体の「設計」に特化。

Rapidus: ファウンドリ。半導体の「製造」を担う。



市場に生まれる、新たな選択肢

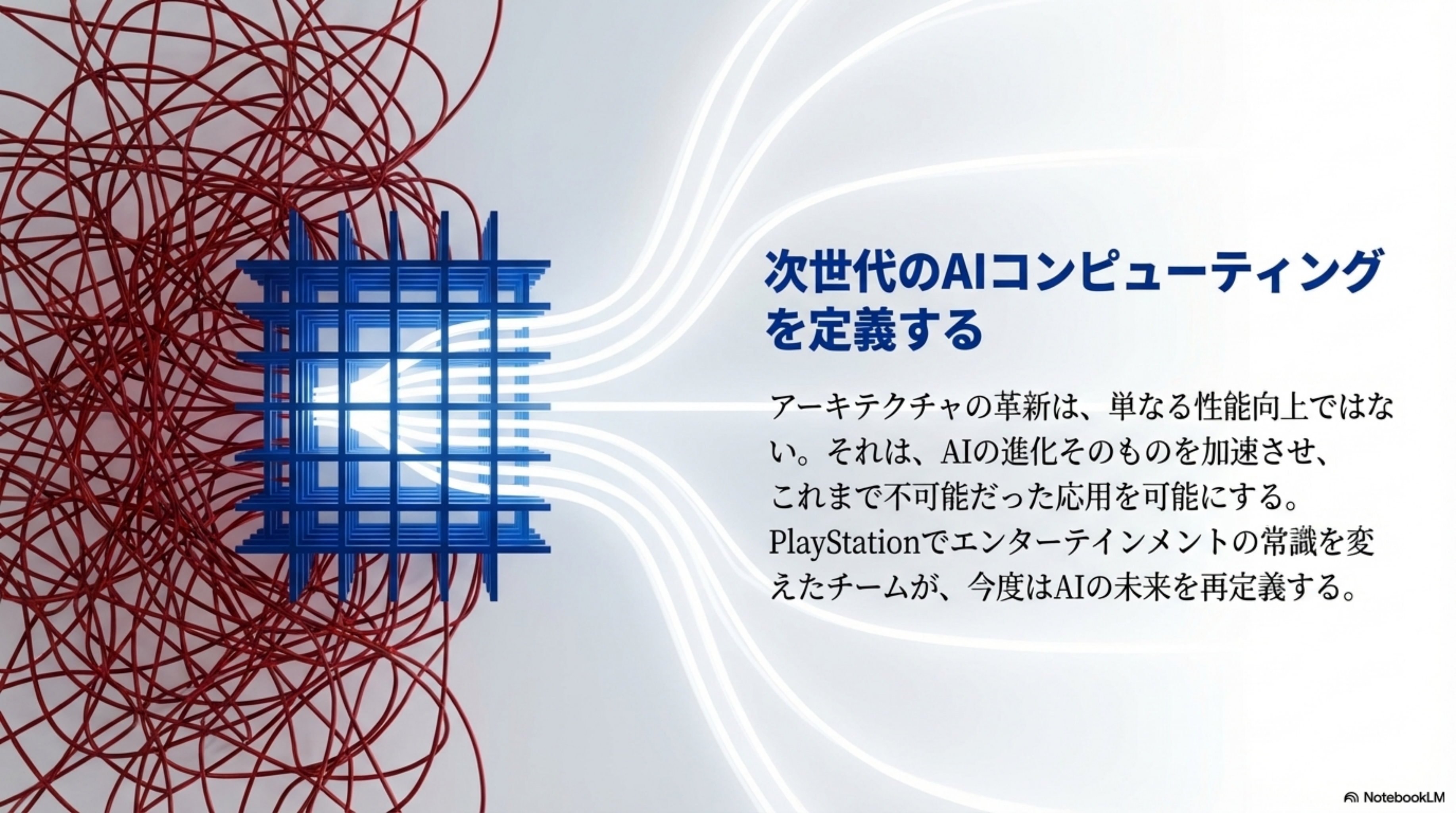
GPU	TPU	CGLA
電力効率 ✗	電力効率 ✓	電力効率 ✓
汎用性 ✓	汎用性 ✗	汎用性 ✓

市場の課題: フォン・ノイマン・ボトルネックによる電力とコストの問題は限界に達している。

既存の選択肢: GPU (汎用だが非効率) vs TPU (効率的だが不自由)。

Renzuの提案: CGLAは、電力効率と汎用性を両立する唯一のソリューション。

追い風: GoogleがTPUを外販し始めたことで、「NVIDIA以外の選択肢」を市場が求め始めたことは明白。



次世代のAIコンピューティング を定義する

アーキテクチャの革新は、単なる性能向上ではない。それは、AIの進化そのものを加速させ、これまで不可能だった応用を可能にする。PlayStationでエンターテインメントの常識を変えたチームが、今度はAIの未来を再定義する。