



EXPONENTIAL ACCELERATION

エヌビディア CEO

ジェンスン・フアン

「計算インフレ」に対応できる次世代GPU

今年、一時の株価が年初来140%を記録したことで話題を呼んだ半導体メーカー世界最大手、エヌビディア。生成AIの活用が拡大し、コンピューターによる情報処理の高速化が求められているが、それを実現するカギとなるのは、エヌビディアの主力製品であるGPU（画像・動画処理装置）だ。より高性能なGPUへの需要が高まっている今、世界トップレベルの性能を誇る同社製品にこれまで以上の注目が集まっている。エヌビディアのCEOのジェンスン・フアン氏は、2024年6月に行われた「COMPUTEX TAIPEI（台北国際コンピューター見本市）」のスピーチで、同社が開発した次世代GPUのAI時代における有用性を語った。



スピーカー

ジェンソン・フアン

1993年にエヌビディアを共同設立し、現在に至るまで、同社の社長兼CEOを務める。1984年にオレゴン州立大学で電気工学の学士号を、1992年にスタンフォード大学で電気工学の修士号を取得した。台湾生まれの台湾系米国人。2人の子どもがいる。アメリカの半導体企業であるAMD社長兼CEOのリサ・スー氏は、同氏の親戚にあたる。

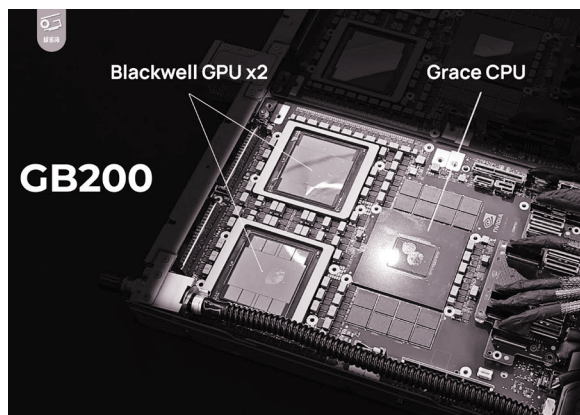


Listening Quiz

58 GPUの活用で100倍の高速化を実現

There are many algorithms that you can operate completely in parallel: computer graphics, image processing, physics simulations, combinatorial optimizations, graph processing, database processing, and, of course, the very famous linear algebra* of deep learning. There are many types of algorithms that are very conducive to acceleration through parallel processing.

So we invented an architecture to do that. By adding the GPU* to the CPU,* the specialized processor can take something that takes a great deal of time and accelerate it down to something that is incredibly fast. And because the two processors can work side by side—they're both autonomous and independent, that is—we can accelerate what used to take 100 units of time down to one unit of time. The speed-up is incredible. It almost sounds unbelievable.



エヌビディアのAI向け半導体「GB200」(写真)には、半導体大手アーム社のアーキテクチャを採用した高性能CPU「Grace」と、エヌビディアが生成AI向けに開発したGPUアーキテクチャ「Blackwell」が組み込まれている。この構造によって、情報処理をこれまで以上に高速で行うことができる

exponential:

《タイトル》〈増加などが〉急激な

algorithm:

アルゴリズム、(特にコンピュータによる)計算・問題解決の手順・方式

in parallel:

並列で、同時に

physics simulation:

物理シミュレーション

combinatorial optimization:

組み合わせ最適化

deep learning:

深層学習、ディープラーニング

be conducive to:

～を促進する、～につなげる

acceleration:

加速、高速化

parallel processing:

並列処理

architecture:

《コンピューター》アーキテクチャ、構成(法)

processor:

中央演算処理装置、プロセッサ

take...and do:

…を使って～する、…を対象に～する

a great deal of:

かなりの量の

accelerate:

～を加速させる、高速化する

down to:

(減少して)～に至るまで

side by side:

並列して

autonomous:

自立した、自立的な

that is:

～ということですが、つまり

unit of time:

単位時間、時間の単位

完全に並列処理することができるアルゴリズムはたくさんあります。例えばコンピューターグラフィックス、画像処理、物理シミュレーション、組み合わせ最適化、グラフ処理、データベース処理、そしてもちろん、深層学習でおなじみの線形代数（に使われるアルゴリズム）。それらを並列処理することによる（コンピューターによる情報処理の）高速化を大いに促進するアルゴリズムが、さまざまあります。

そこで私たちは、それを行うためのアーキテクチャを開発しました。CPUにGPUを追加することで、専用プロセッサは（処理に）膨大な時間を要することを信じられないほどの速さにまで高速化できます。しかも、2つのプロセッサは並列して作業できる、つまり、どちらも自立して独立があるおかげで、100単位時間かかっていた処理を1単位時間にまで高速化できるわけです。驚異的なスピードアップです。とても信じられないような話です。



エヌビディア CEO が本スピーチを行ったコンピューター関連の見本市「COMPUTEX TAIPEI」は台北ワールドトレードセンターで毎年6月に行われる



「COMPUTEX TAIPEI」には、エヌビディアをはじめ、インテル、AMDなどの主要なメーカーが参加する

■ Computer graphics（コンピューターグラフィックス、CG）

コンピューターを使って描かれた図形や画像。

■ linear algebra（線形代数）

代数をさらに簡略化させたもの。機械にものを学習させるには大量の計算式が必要となるが、線形代数を用いて計算を簡略化すれば、より効率的な機械学習を実現できる。

■ GPUとCPU

どちらも半導体。GPU（graphics-processing unit）は、主にコンピューターで処理した画像などを出力するために使われる。3DゲームやVRを楽しむ際には高性能のGPUが必要。一方CPU（central processing unit）は、コンピューター全体の計算処理を行う。「CPUほど複雑で多様な処理はできないが並列処理には強い」というGPUの性能がディープラーニングに適しており、AI開発においては比較的GPUが重視される。

エヌビディアは「アクセラレーテッド・コンピューティング（CPUと並列して動作するGPUを活用し、計算処理を大幅に高速化するアプローチ）」を推進している。このアプローチがAIの深層学習トレーニングで活用されれば、従来のCPUだけでは数週間から数か月かかるところを、数日以内に完了することが可能になるという。