

GPGPU 技術による高性能クラウドコンピューティング

High-performance cloud computing with GPGPU technology



加藤 真平 (Shinpei KATO, Ph. D.)

東京大学 大学院情報理工学系研究科 准教授

(Associate Professor, Graduate School of Information Science and Engineering, The University of Tokyo)

情報処理学会 IEEE ACM

研究専門分野：サイバーフィジカルシステム オペレーティングシステム
クラウドコンピューティング 高性能計算

あらまし 画像処理専用プロセッサ (GPU) を画像処理以外の汎用目的 (General-Purpose) の計算処理に応用する GPGPU (General-Purpose computing on GPU) 技術の発展が目覚ましい。スーパーコンピュータでの人工知能 (AI) や車載コンピュータでの自動運転なども、GPGPU 技術を用いるようになった。本研究では、この GPGPU 技術をクラウドコンピューティングにおいても利用可能にするための研究開発を行った。高性能な GPU を搭載できない非力なスマートフォンやその他の IoT (Internet of Things) 機器でも、クラウド上にある GPU にアクセスでき、あたかも自分のコンピュータ上で GPGPU 実行をしているかのように見せる仕組みを構築し、性能評価を行った。その結果、インターネットを通じたクラウドアクセスの遅延を加味しても、GPU を持たないスマートフォンでの処理に対して、最大 10 倍程度の処理時間短縮を実現した。

1. 研究の目的

スマートフォンや IoT 機器は普及の一途を辿っている。今後のトレンドは省電力化であり、少量のバッテリー駆動で長時間稼働できる製品が市場に出回ると予想される。省電力化は CO₂ 削減にもつながり、社会的価値の高い取り組みである一方、機能性という点では性能不足による制約を受ける。特に、昨今の AI や自動運転という新しいアプリケーションでは、多くの機

能が高性能計算を必要とする。

近年の高性能計算では、GPU を汎用計算に用いる GPGPU 技術が導入され、飛躍的な性能向上が達成されているが、GPGPU に代表される並列計算デバイスは、消費電力がボトルネックとなり、スマートフォンや IoT 機器での利用は困難な状況になった。そこで本研究は、GPU をクラウド側のサーバシステムに搭載し、それをスマートフォンや IoT 機器のようなクライアント端末からアクセスすることで、GPGPU 計算を利用できる仕組みを構築した。

2. 研究の背景と動向

GPU をクラウド環境のサーバシステム側でサポートする研究は存在する。例えば、GPU を仮想化し、複数クライアント端末から GPU にアクセスできるようにするような研究がある。この場合、クライアント端末側のプログラムは、サーバシステムに問い合わせを行い、サーバシステム側で用意された GPGPU プログラムを実行し、その結果をクライアント端末に返すという仕組みになる。よって、クライアント端末のプログラムを構築する際に、逐一サーバシステム側のプログラムも用意しないといけない。このプログラミングの手間は、開発効率を著しく低下させ、広く汎用的に使える仕組みとはならない。

本研究の特徴は、サーバシステム側は GPU にアクセスするための共通フレームワークがあるだけで、プログラムそのものはクライアント端末のみが用意すればよいという仕組みを実現した点にある。具体的には、リモートプロシージャコール (RPC) を利用して、クライアント端末で呼び出す関数をそのままサーバシステムに転送し、実行するというものである。RPC の概念は古典的だが、それを GPGPU 技術に適用した事例は本研究が初めてある。

3. 研究の方法と結果

本研究では、RPC の実装にオープンソースの GPGPU ドライバソフトウェアである Gdev を用いた。Gdev のアーキテクチャを図 1 に示す。Gdev はデバイスドライバを提供するだけでなく、GPU プログラミングに不可欠となるランタイムエンジンも提供する。

GPGPU 技術による高性能クラウドコンピューティング

High-performance cloud computing with GPGPU technology

ユーザ空間とオペレーティングシステム (OS) 空間の両方にこのランタイムエンジンを備えることで、多様なアプリケーションからの GPGPU 技術の利用を可能にしている。Gdev は筆者の成果のひとつでもあり、GPGPU ドライバソフトウェアとしては世界で唯一のオープンソースソフトウェアである。

通常の GPGPU ドライバソフトウェアは、ソースコードが公開されていないため、RPC を構築しようとすると、GPGPU ドライバソフトウェアとは別のソフトウェアライブラリ層が必要になり、クライアント端末側で既存の GPGPU プログラムを使う場合は、プログラムに修正が必要となる。一方、本研究では、Gdev

を使うことでドライバソフトウェアの内部に RPC 機構を設けることが可能になり、追加のソフトウェアライブラリが必要ないことから、GPGPU プログラムが利用するプログラミングインタフェースは変わらないで済む。つまり、既存の GPGPU プログラムがそのまま動くことになる。

これは非常に画期的な仕組みである。実際のクライアント端末に GPU が搭載していなくても、クラウド環境に GPU を搭載した RPC サーバシステムを用意すれば、GPGPU プログラムがクライアント端末で通常どおりに動作するということになる。

図 2 に、本研究で改良した Gdev の構造を示す。

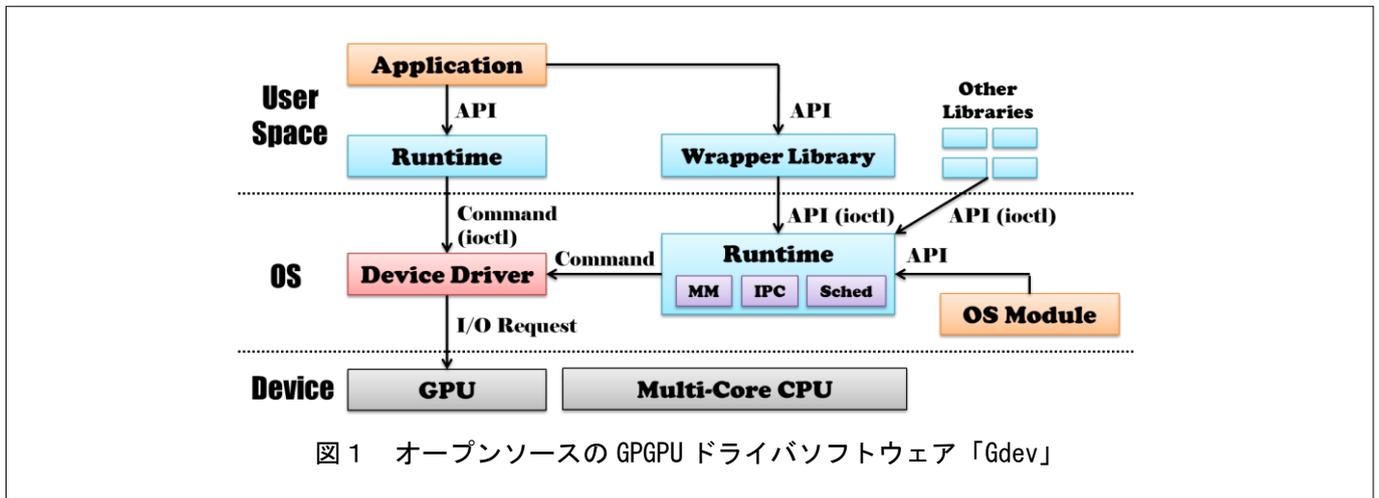


図 1 オープンソースの GPGPU ドライバソフトウェア「Gdev」

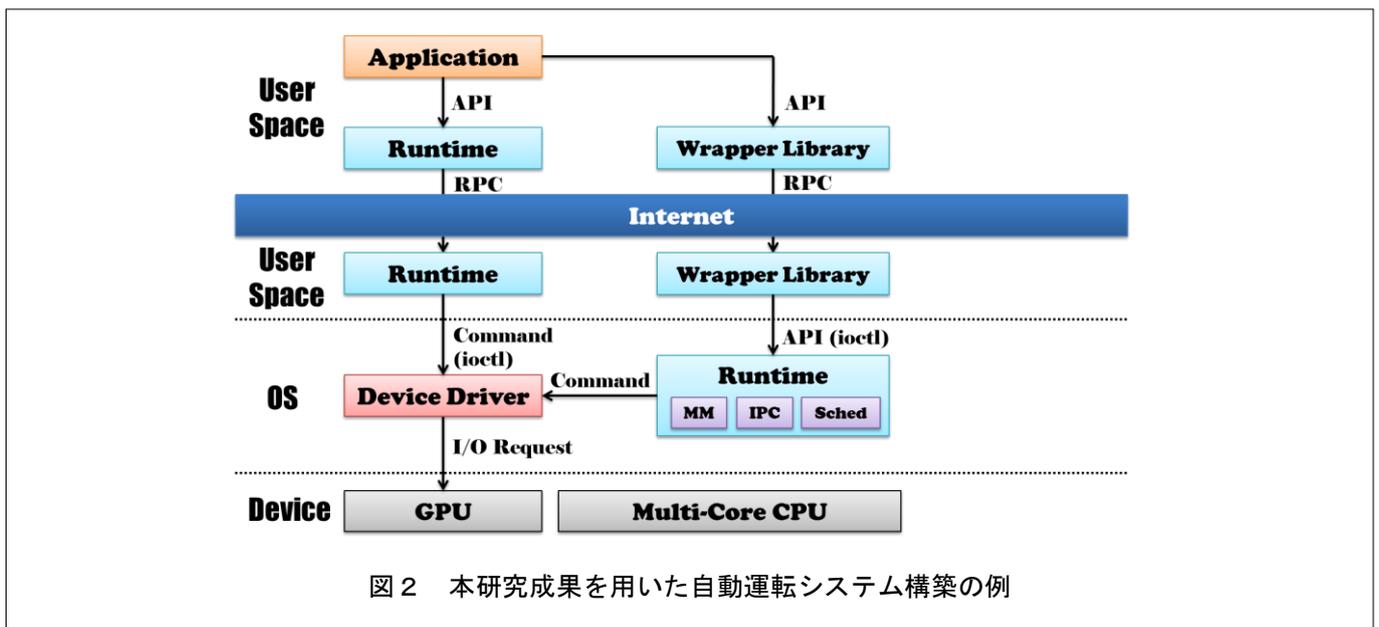


図 2 本研究成果を用いた自動運転システム構築の例

GPGPU 技術による高性能クラウドコンピューティング

High-performance cloud computing with GPGPU technology

従来の Gdev では、ユーザ空間のランタイムエンジンやライブラリから OS 空間内のランタイムにコマンドを送信していた。一方、本研究で改良した Gdev では、ユーザ空間のランタイムエンジンやライブラリが一度インターネットを介した RPC を行い、リモートにあるサーバシステムのランタイムエンジンやライブラリにコマンドを通知し、そのサーバシステム上の GPU にそのコマンドを送るようにした。

この構造により、呼び出し元のアプリケーションからは、あたかも GPU がクライアント端末にあるかのように見える。すなわち、クライアント端末に GPU が搭載されていなくても、インターネットを介してリモートのサーバシステム上にある GPU を利用できるようになる。

実際に本研究成果を既製品であるスマートフォンに導入し、いくつかのベンチマーク評価を取ってみたところ、GPU を使わないでスマートフォン単体で動作させる場合に比べて、本研究成果を導入し、クラウド環境にある GPU に RPC で遠隔利用した時の方が 2～10 倍ほどの高速化を達成できた。LTE や WiFi による通信遅延を加味したうえでの 2～10 倍という高速化であるため、本研究成果の有効性は高いと考えられる。

4. 今後の展望

技術的な面では、本研究では、複数のクライアント端末が同時に RPC 接続するようなシナリオを想定していないため、サーバシステム側のタイムシェアリング、マルチタスキング対応が急務となる。応用的な面でみれば、自動運転などは非常に興味深いアプリケーションである。

現在、研究開発段階にある多くの自動運転システムでは、GPU を搭載した車載コンピュータを想定しているが、これは消費電力の観点から、必ずしもデプロイできるとは限らない。それに対して、本研究成果を利用すれば、車載コンピュータ側には GPU がなくなるので、自動運転の普及形態によっては、本研究成果を有効活用できると考えられる。例えば、図 3 に示す通り、自動車側にはカメラとレーザーといった基本的なセンサーのみを搭載させ、計算環境はクラウド側に配置することによって、低消費電力かつ高性能な

自動運転システムを構築できるようになるかもしれない。今後は、こういった本研究成果の応用を探っていく。



図 3 本研究成果を用いた自動運転システム構築の例

この研究は、平成 25 年度 S C A T 研究助成の対象として採用され、平成 26～27 年度に実施されたものです。