



第8章 命令実行のプロセスからメモリ・アクセスまで

GPUのアーキテクチャ研究④ Jetson NanoのGPU

Jetson Nanoの性能を引き出すオリジナルのCプログラムを作るためには、GPUのアーキテクチャやメモリの知識が必要です。本章では、Jetson NanoのGPUとメモリの構造やしぐみについて解説します。
〈編集部〉

8-1 アーキテクチャ名はMaxwell

● Jetson NanoのGPUはMaxwellアーキテクチャ
Jetson Nanoに搭載されたGPUはMaxwellというアーキテクチャをベースにしています。そのブロック図を図1に、仕様を表1に示します。

GPUの内部構造は、SMM(Streaming Multiprocessor Maxwell)というブロックが1つの大きな単位になります(他のアーキテクチャのGPUでは、単にSMとよぶことが多い)。Jetson Nanoの場合、SMMは1個であり、もっと上位のSoCやグラフィック用GPUには複数のSMMが含まれているものもあります。

Maxwellアーキテクチャの場合、SMMはさらに4つのクラスタに分割されており、各クラスタ内に、処

理ユニットとして、CUDAコア(CUDAはCompute Unified Device Architectureの略で、図1ではCoreと表記)が32個、メモリのロード・ストア・ユニット(図1はLD/STと表記)が8個、三角関数など特殊な演算を実行するSFU(Special Function Unit)が8個内蔵されています。都合、CUDAコアはSMM内に合計128個あり、これらが浮動小数点データや整数データを演算する中心的な演算ユニットになります。32ビット浮動小数点データ(FP32)の加減算、乗算、積和演算を6サイクルのレイテンシで実行し、また16ビット浮動小数点データ(FP16)ならSIMD実行が可能です。2組のFP16データに対して同時に同じ演算を6サイクルのレイテンシで実行することができます。

SMM内では命令フェッチや命令のディスパッチはSMM内の各クラスタが行い、CUDAコアなどの処理ユニットはその命令を受け取って必要な処理を実行していきます。GPUにはコアが128個あるとうたっていますが、そのコアはCPUのように命令フェッチ動作は行わないので、ハードウェア演算器と考えた方がいいでしょう。SMM内の1つのクラスタがCPUのよう

表1 Jetson NanoのGPUの仕様

項目	内容	備考
アーキテクチャ	Maxwell	—
SMM数	1	—
マルチプロセッサ数	1	—
CUDAコア数	128	—
GPUの最大動作周波数	921.6 MHz	—
FP32の最大性能	236GFLOPS	—
FP16の最大性能	472GFLOPS	—
メモリ・クロック(最大)	1.6 GHz	—
メモリ・バンド幅(最大)	25.6 Gバイト/秒	—
メモリ・バス幅	64ビット	—
グローバル・メモリ	3965 Mバイト	—
L2キャッシュ・メモリ	256 Kバイト	—
コンスタント・メモリ	65536バイト	—
シェアード・メモリ	64 Kバイト	スレッド・ブロックあたり 48 Kバイト
レジスタ数	65536本	スレッド・ブロックあたり 32768本
ワーブ・サイズ	32	—
最大スレッド数①	1024	スレッド・ブロックあたり
最大スレッド数②	2048	マルチプロセッサあたり
最大スレッド・ブロック	(1024, 1024, 64)	—
最大グリッド・サイズ	(2147483647, 65535, 65535)	—
最大メモリ・ピッチ	2147483647	—

【セミナー案内】[演習あり] 高性能デジタル回路設計とハイレベル・シンセシスの基礎 — HDL設計とC設計を使い分けできるようになろう

【講師】 森岡 澄夫氏, 8/23(金) 22,000円(税込み) <https://seminar.cqpub.co.jp/>