

Maliちゃん
@tanakmura

今日の内容

- Abstract
- AndroidでOpenCL
- Mali-T604ベンチマーク
- 結論

自己紹介

- @tanakmura
- Id: w_o
- 某F社でプログラム速くしたりコードを読みにくくしたりバグを入れたりしてる

高速化の仕事の紹介

- 高速化の仕事って一体何やってるの？

仕事内容1: add命令の計測

- add命令のレイテンシを計測する
- 特に浮動小数addのレイテンシは
 - CPUごとに違う
 - アプリケーションによっては影響大きい
 - →すごく重要な仕事
- add命令をひたすら並べて時間を計る

仕事内容2: add命令の計測

- add命令のスループットを計測する
- 命令のスループットはアプリケーションの性能に大きな影響を与える
 - →すごく重要な仕事
- add命令をレジスタがかぶらないようにひたすら並べて時間を計る

仕事内容3: add命令の計測

- VLIWにいくつadd命令を埋められるか調べる
- コンパイラでどうやってそれを書くか調べる
- アプリケーションの性能に(ry
 - すごく重要なしg(ry
 -
- add命令をスロットがかぶらないようにひたすら並べて時間をはかる

その他の仕事内容

- load命令のry
- mul(ry
- 起動レイテンシ
- メモリ転送レイテンシ

というのは7割くらい冗談ですが

clminibench

- OpenCLでもそれらを計りたい場合用ベンチ
- まあこんな感じで動く

こっから本題

ARM Mali

- ARM社が出してるGPU
- https://twitter.com/arm_link/status/245703054912000003
- 公式がMaliちゃんと呼んでいたのものでMaliちゃんと呼ぶべき

-  **アーム株式会社**
@arm_link フォロー
- Snowball - a catalyst for mobile innovation
blogs.arm.com/software-enabl... A9500には、Maliちゃんも載ってます。
stericsson.com/products/a9500...
- [返信](#) [リツイート](#) [お気に入りに登録](#) [その他](#)
- 2012年9月12日 - 10:59

ARM Mali T604

- OpenCLが動くらしい
 - しかし公式SDKはヘッダが入ってるだけで使えない…
-
- しかしNexus 10のlibGLES_mali.soのシンボルを見ると、clCreateContextなどの文字が
 - <http://forum.beyond3d.com/showthread.php?t=63071>
- →libGLES_mali.so をリンクすればOpenCLが利用可能

clminibench for Android

- Android端末のGPUのadd命令のレイテンシを測定可能な画期的なベンチマークソフト
- ポータブルなAPIであるOpenCLを使うことで、世界中のNexus 10でコンピュータの性能を測定可能

以下結果をお楽しみください

CIGetDeviceInfo

- Compute unitの数 = 4
- PREFERRED_VECTOR_WIDTH_FLOAT = 4
 - float4つかったほうがいい
 - 公式ドキュメントにもそう書かれている
- OpenCL 1.1
- キャッシュあり(read/write)
- local_mem_type = GLOBAL
 - local memoryは意味ない
- GPUだけ。CPUで動くOpenCLは付いてない

Fma throughput

- float : 8.8GFLOPS
- Float4 : 19.0 GFLOPS
-
- Compute Unit = 4, float4, FMA, 533MHz
- $4*4*2*0.533 = 17$
 - are...?
-

FMA throughput

- CPUはCoretex A15 1.7GHzx2
- = $4*2*1.7*2$
- = 27.2GFLOPS
- あっ…
-
- GPUの可能性とは一体…

Fma1 dep throughput

- ワークアイテムたくさん
- 各ワークアイテムでは全依存
- 4.8GFLOPS
-
- あとで書いてるけどレイテンシが30と長いのでワークアイテム並べるだけでは埋めきれないのでは
- 手で命令スケジューリングが必要

Enqueue kernel latency

- 780usec
- てもとのこのAPUだと130usec

Enqueue memory XX

- API呼び出しの時間
- 公式ドキュメントにはCL_ALLOC_HOST_PTRが使えるからあんまこのへんのは使わないと書いてある
 - Host unified memory = 0だが…？
- Latency : 100usec
- Throughput : read=3.2GB/s write=5.8GB/s
 - writeのほうが速い理由は知らない

Global read bandwidth

- CL カーネルでメモリアクセス
-
- float でread 0.77GB/s
- float4 でread 2.81GB/s
-
- float4 でアクセスしろとドキュメントに書いてる
- EnqueueMemoryXXのほうが速いが...

float add latency

- addを依存全部つなげて時間計測
-
- float1 : 30clk
- float4 : 56clk
-
- float4とfloatで別の命令があるのか？
- ちなみにfloat2は56clk
 - Ivy : 3
 - ApuのGPU: 4
 - HD7750: 5

memory load latency

- `off = mem[off] // mem は0初期化`
-
- Gmem: 132clk
- Lmem: 350clk
-
- Imemのほうが遅い…？

まとめ

- Nexus 10 はOpenCL使える
- メモリ転送はなんか遅い
- FLOPS値はCortexA15x2より遅いとか気にしてはいけない