

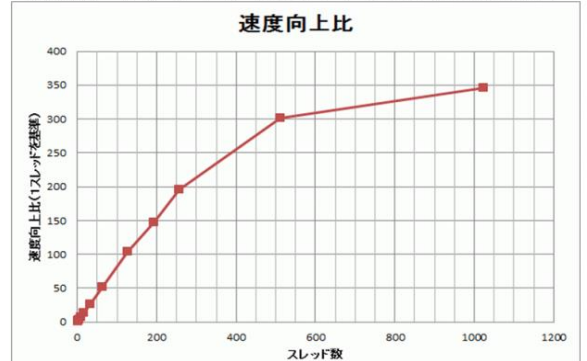
今回は、汎用並列ハードウェアの限界について、いくつかの問題を考えてみたい。

- (1) GPGPUのコアをどうやったら使いきれか、考えてみて欲しい。基本的にコア数程度の並列性が出せるかどうか、という点が、疑問である。

Device 0: "GeForce GTX TITAN Z"	
CUDA Driver Version / Runtime Version	6.5 / 6.5
CUDA Capability Major/Minor version number:	3.5
Total amount of global memory:	6143 MBytes
(15) Multiprocessors, (192) CUDA Cores/MP:	2880 CUDA Cores
GPU Clock rate:	876 MHz (0.88 GHz)

並列処理・高速化

汎用グラフィックプロセッサでの並列処理



- (2) GPGPU では、倍精度浮動小数計算資源が少ない。これを回避する方法を考えてみてほしい。

一般に、数値計算プログラムはほぼデフォルトで、倍精度浮動小数を使うだろう。単精度では有効桁数が不足することが多い(7桁?)し、いちいち考えるのが面倒ということもある。今通常はすべて倍精度で計算している演算を、単精度で済むものをきちんと選んで単精度で計算するとしたら、どのようなやり方が考えられるか？

(注: たとえばグラフィックスの処理では、一部を全面的に単精度で済むと考えられ、だからこそ GPU が単精度浮動小数でしか計算しないようになっている。)

更には、現在は倍精度でなければ結果の精度の保証ができない計算を、何らかの工夫で単精度で済ませる手立はあるだろうか？

- (3) NVIDIA のアーキテクチャでは、GPU 用のメモリと「ホスト」(計算機本体)のメインメモリは別々に持っており、その GPU 側で計算するためにはデータを転送しなければならない。これに対してどういう対応が考えられるだろうか？ 現在は、基本的(原始的)なプログラミング環境(CUDA)ではプログラマが陽に指示して転送することになっており、プログラミング上の手間は面倒であるが、プログラマが性能上のチューニングをすることが可能である。他方、一部のプログラミング言語・環境では転送を自動化している。どう考えたらよいだろうか？

次の中から1つの課題を選んで、自分で調査・研究もしくは実験をし、レポートを書いて提出してください。

<課題候補 1>

コンピュータ科学で使われるアルゴリズムを1つ選んで、並列化による高速化の可能性を議論してください。そのアルゴリズムに対して、①並列化の手法提案、②並列化の効果の予想(並列性の予想)、③実装、④実験による評価、のステップを追い、物理的な並列度をいくつか変えて実測し、アムダールの法則の直列・並列部分の割合を推定して、「どれだけ並列化できたか」を示してください。また、全く並列化を考えない(当初の)プログラムの実行時間と比較して、並列化の仕組みを入れたために遅くなっていないかも確認してください。どうしても手が回らなければ①②でも及第とします。アルゴリズムの選択は、身近なもの(卒論・修士研究等で使うもの)でもよいし、それが無ければ教科書に載っているような、たとえばグラフの最短経路の検索やその他の最適化問題、凝ったソート(少なくともバブルソートではないレベル)などを対象にしてもよいでしょう。

<課題候補 2>

実習の回で試した区分求積法の例題を、上記<課題候補 1>のアルゴリズムの代わりに使って、同様のことを議論してください。この場合は①と③は新たに考えることはないので丁寧に説明をすればよいですが、②と④については自分独自の議論を展開することを期待します。また<課題候補 1>と同様に、並列度を変えてアムダールの法則の直列・並列部分の割合を推定してください。更に、その割合がどうして生まれるか、割合をより100%並列に近づけるためにプログラム上で何ができるか、いろいろと提案し実験し結果を評価し、その内容を報告してください。この課題では、アムダールの直列・並列の割合推定まで行って及第とし、並列度向上の提案・実験・評価の報告がまとまれば完璧です。

<期限・分量など>

期限は、1月12日(金)としますが、卒業研究等のワークロードがあるでしょうから、早め(12月中旬まで)に済ませることを勧めます。早めに提出してかまいません。

分量は、10ページ程度以上、としておきます。長くても構いませんが、読み手のことも考えて、要領よく、しかし自分の考えが十分に伝わるように、書いてください。

提出先は、山内のオフィス(4541室)の扉の書類受けへ。