Performance Analysis of the Simplex Method on OpenCL Hardware Accelerators

Bradley de Vlugt*, Maysam Mirahmadi†, Serguei L. Primak*, Abdallah Shami*
* Electrical and Computer Engineering
† Western University
{bdevlugt, ashami2, slprimak}@uwo.ca
maysamah@ca.ibm.com

Abstract—This work proposes an energy-efficient hardware accelerated Linear Programming (LP) solver. The system is based on the Simplex Algorithm for solving LP problems and is operable on Field Programmable Gate Arrays (FPGAs), Graphic Processing Units (GPUs), and Multi-Core Computer Processors (CPUs). The system is targeted towards the dense problems in radiotherapy applications as they represent a challenge to modern solvers.

Performance benchmarking reveals speed ups relative to a sequential implementation that approach 2 and 10 on a CPU and GPU for random, dense problems. The FPGA exhibits unity speed up but proved to be the most efficient in terms of Simplex iterations processed per unit energy with an efficiency 5 times greater than the CPU. This is a notable speed improvement and power saving in comparison with current technology for solving dense problems as the GPU code can solve problems with speeds up to 50 times faster than a sparse solver.

I. INTRODUCTION

High performance implementations of the Simplex Algorithm facilitate computational optimization in medicine, engineering and business with linear programming (LP). Emerging dense LP problems in radiotherapy applications pose a challenge to modern Simplex solvers designed with sparse data structures. The importance of these applications led to proposals for hardware based solvers in recent literature [1], [2]. The algorithm’s dense form exhibits data-level parallelism that is ideal for processing on modern hardware accelerators.

Radiation therapy is a cancer treatment method that uses several beams to deliver a radiation dose to tumour cells. The distribution of the radiation dose to the target area can be modified by assigning different weights to the beams. This distribution must be carefully adjusted to avoid delivering a harmful dose to healthy areas. Recent literature has formulated the problem of finding the optimal beam weights for a given target area as an LP problem [3], [4]. The models attempt to minimize the dose delivered to non-target areas subject to constraints based on beam interactions.

II. LINEAR PROGRAMMING AND THE SIMPLEX ALGORITHM

A. Linear Programming Problem Formulation

Linear programming finds the extreme value of a linear objective function subject to constraints. The mathematical form of a linear programming problem is (1), where $c$ is the objective function coefficients, $x$ is the decision variables, $A$ is the constraint system and $b$ is the constraint bounds.

$$\text{min } cx$$
$$\text{s.t. } Ax \leq b$$
$$x \geq 0$$

Algorithms that solve linear programming problems iteratively check feasible solutions for optimality. The Simplex Algorithm restricts this search to corner points, or basic feasible solutions (BFS), of the convex polyhedron representing the feasible region of the problem and traces a path along its edges towards the optimal solution. At each iteration, a new BFS is chosen to improve the objective by replacing a variable with poor objective contribution with a better candidate. These variable changes propel the algorithm along a path formed by the edges of the feasible region to incrementally decrease the objective.

B. Radiotherapy Problem Formulation

The radiotherapy LP problem is represented by (2) and is derived as follows [3]. The target area is subdivided into voxels and the dose delivered to voxel $i$ by each beam, $j$, at unit exposure time is measured to form the dose matrix, $D$. The set of voxels is subdivided into target, $t$, and normal, $n$, voxels.

$$\min (D^T_n c_N)^T w$$
$$D_{t} w \leq x_t^U$$
$$D_{t} w \geq x_t^L$$
$$D_{n} w \leq x_n^U$$
$$w \geq 0$$

The dose applied to $t$ is lower bounded by $x_t^L$ and upper bounded by $x_t^U$. The dose delivered to $n$ is only upper bounded by $x_n^U$. The beam weights are represented by $w$ and are components of cost vector $c_N$. Since the radiation delivered by many beams will overlap to cover the full treatment area, the dose matrix of the LP model will be dense.
III. Benchmarking Results

This section presents the benchmarking performed on the hardware implementation. The design was tested with 100% dense, random problems on an Intel Core i7 4930k CPU, a Nvidia GeForce GTX-780 GPU, and a Nallatech PCIe-385N equipped with an Altera Stratix V FPGA. A range of problem sizes from 256 by 256 to 8192 by 8192 variables and constraints was tested. Since the dose matrices in the radiotherapy problems studied in [3] are of size 1989 by 24000 and 36 by 800000 they are represented by the larger end of the test set.

A. Speed Up

Fig. 1 shows the speed up achieved by the OpenCL solver over the sequential, dense solver. The GPU was the fastest device of the three tested. This advantage became pronounced when the problem size increased beyond 756 by 756 variables and constraints due to higher utilization of the device’s memory bandwidth and streaming multiprocessors. The speed up became constant when the problem size increased beyond 4096 by 4096 due to saturation of the available memory bandwidth. Though the FPGA was the slowest solution, with close to unity speed up, its low power consumption is an asset that was not considered in this particular test.

The performance of the proposed OpenCL solver was compared to SoPlex [5], an advanced sparse LP solver, over the same range of random, dense problems. This showed that the GPU code was up to 50 times faster than the sparse solver for these dense problems.

B. Energy Consumption Estimate

The speed at which a solver can process linear programming problems per unit energy is displayed in Fig. 2. Higher values represent faster processing for each Joule consumed. This result indicates that the FPGA outperforms the CPU and GPU with its low energy consumption. Although it has the lowest speed, it was the most efficient platform.

Fig. 1: Speed Up for the OpenCL Solver over the Sequential Solver

Fig. 2: Energy Efficiency of the OpenCL and Sequential Solvers

IV. Conclusions

In this paper, an accelerated system for solving LP problems raised in radiotherapy treatments is proposed. The system is based on the Simplex Method for solving LP problems and is implemented with OpenCL to enable the comparison of its performance on multiple devices. Benchmarking indicated that the FPGA reduced the solver’s energy consumption by 80% and the GPU provided an order of magnitude speed up.

ACKNOWLEDGMENT

This work was supported in part by the Southern Ontario Smart Computing Innovation Platform Consortium (SOSCIP) and IBM Canada Ltd. The authors would like to thank Dr. Mary Fenelon, Dr. Sean Wagner, Mr. Laszlo Ladanyi and Mr. Blair Adamache from IBM for their valuable comments and suggestions.

REFERENCES