Vector Accelerator Array in constrained memory bandwidth.

Arun Jeevaraj
jeevarajarun@gmail.com

Department of Electrical and Information Technology
Lund University

Supervisors: Flavius Gruian

Course Code: EDAN85
September 12, 2017

Printed in Sweden
Tryckeriet i E-huset, Lund

Abstract

The aim of the project, is to explore ways to accelerate the computation with vector accelerators given there are free resources in a FPGA and, within the memory bandwidth constrain. The Off-chip communication to and fro DDR Memory is slow, and on the other hand allows to store a large amount of data. This is one of the main bottlenecks for computations involving large amount of data.

For exploring these situations, a case study of the particle beam behaviour simulation is considered, where the particles is assumed to drift in vacuum. Their phase vector information will be stored at regular distances based on Hamiltonian equation of motion on the free space.

Compression techniques and fixed point simulation is considered to reduce the bandwidth utilization on memory. There by allowing for a higher throughput in computation.

Table of Contents

Chapter 1

Introduction

In system on chip(Soc) solutions, where there is a large amount of data involved for processing, using off-chip memory such as DDR memory is inevitable. The DDR memory have limited bandwidth and the access to the memory can be a bottleneck for many of the solutions as shown in the figure 1.1. Given there are free resources to spare in FPGA, a way to overcome the memory bandwidth constraint is the main purpose of this project.

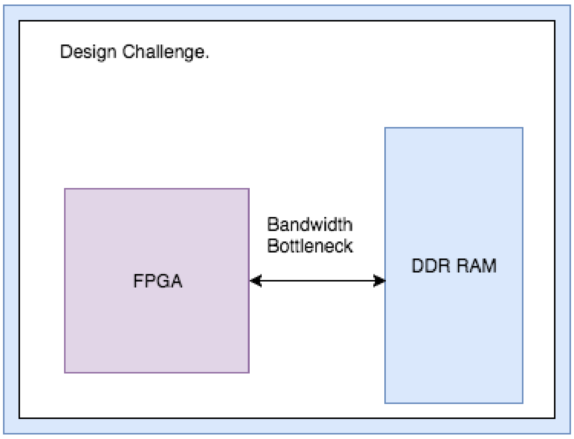

Figure 1.1: Design Challenge

A design target algorithm is considered, where high density particle drifting in space is simulated based on Hamiltonian equations defined for free space. The particle's phase vector is calculated as a function of distance and is computed for each particle at the given location. A detailed description of the theory is provided in the next section. For a high particle count, this can amount to a high memory traffic.

The project aims to increase the throughput of the model, with the help of using parallelism offered by FPGA and reducing the latency for accessing the DDR memory. A simulation model is first built using python, and then it will imported to microblaze based Soc solution with FPGA. After the functionality check, the design is analysed. And optimizations are applied to see, if there is an improvement of throughput and compare the results against the cost of hardware resources.

1.1 Physics Simulation model

For the target algorithm, a case study from physics particle simulation is considered. The basic core features of the physics engine is described below. To describe the dynamics of the particles in a beam is fundamentally based on two steps; to represent all the equations of motions affecting the particles and to solve them. Newtons second law of motion 1.1 represents an equation of motion once, the forces are known.

$$\frac{dp}{dt} = F, \tag{1.1}$$

Though this method can be used, most of the times the electric field and Magnetic fields in an accelerated beam are specified as a function of location rather than time. Hence a Hamiltonian Mechanics based methodology is followed, as it allows for representing the particles as a function of distance along the beam line. This makes it simpler to represent the moving particles through a potentially complex sequence of electric and magnetic fields. [1] provides with derivations of the equations and explains how the Hamilton's equation generalized as in equation 1.2 where x_i are the co-ordinates of the particle(i = 1...N in an N dimensional co-ordinate space), p are the components of momentum and H is the Hamiltonian.

$$\frac{dx_i}{dt} = \frac{\partial H}{\partial p_i}, \frac{dp_i}{dt} = \frac{\partial H}{\partial x_i}, \tag{1.2}$$

can be used to derive for a relativistic particle moving in an electromagnetic filed, with the Hamiltonian defined as in

$$H = c\sqrt{(p - qA)^2 + (mc)^2 + q\phi} \tag{1.3}$$

Substituting the electromagnetic potential in a particular region of space into Hamiltonian allows to derive the equation of motion of the charged particles moving through the region. So it is possible to find solutions to these equations for commonly used structures/components in the linear accelerators. The region of space ,where there are no electric or magnetic fields present is considered as the drift space, where particles are drifting. The derivations on how the equations are solved can be referred from the text, [1] and is not reproduced here as such.

The linear transfer map in matrix form can be expressed for a particle as in equation 1.4,

$$\bar{x_1} = R_{drift} \times \bar{x_0} \tag{1.4}$$

,where x_i is the phase space vector with components constructed from the dynamic variables, such as the position and momentum in the 3 dimensional space.

$$\bar{x} = \begin{pmatrix} x \\ px \\ y \\ py \\ z \\ \partial \end{pmatrix} \tag{1.5}$$

and the R_drift can be represented as in equation 1.6

$$R_{drift} = \begin{pmatrix} 1 & L & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & L & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & \frac{L}{\beta_0^2\gamma_0^2} \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \tag{1.6}$$

here, β_0 is the velocity of the reference particle scaled by the speed of light, L the unit distance under consideration and γ_0 is given by the equation 1.7.

$$\gamma_0 = \frac{1}{\sqrt{1-\beta_0^2}} \tag{1.7}$$

with the aid of these principles, the case study under consideration, is to test a million particles moving across a drift spaces and their information sampled at an interval for 1000 frames.

1.2 Simulation case study

The initial state of the particles is randomly generated as a uniform distribution, and then it is transformed into an elliptical region of space. The particles are assumed to move through a region of space without any electric or magnetic field. A million particles constitute a frame and if represented as double precision floating

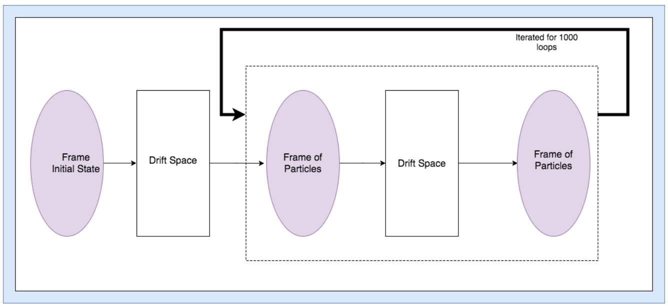

Figure 1.2: Simulation case under consideration.

point data, it will need around 50 MBytes of data per frame.

For the case study under considerations, the equation 1.4 is used to find the next frame information of the particles. The same process is iterated over for 1000 loops as shown in the figure1.2. Because of the iterative nature and the huge amount of data/frame. The information needs to be stored and accessed from a DDR RAM storage outside the chip.

Chapter 2

Architecture

The hardware architecture model is a system on chip solution with Mircroblaze centric control. The microblaze soft-core processor is mainly used for performing unity test over each hardware blocks and also for managing the DMA block. An UART over USB module will be used for communication between a computer and the FPGA board.

The DMA block controls the transfer of data from the DDR memory to the accelerated array block and back.

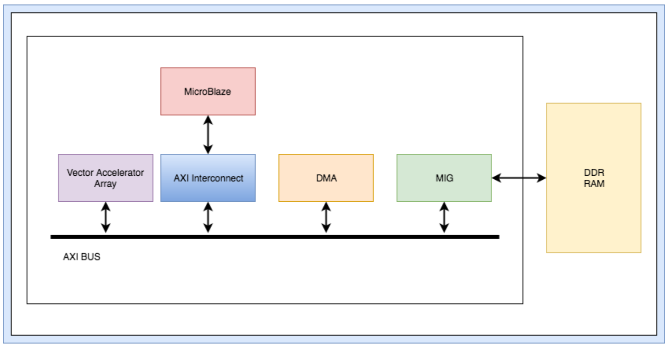

Figure 2.1: Target Hardware Architecture.

If there is time left, an ethernet interface will be developed so that, the frame data can be send from the PC over LAN.

2.1 Design Methodology

The planned design methodology is to build a basic functional model first and then improve the model with planned optimizations. A good back end framework is first developed so that, the system can be analysed for bottlenecks and time can be better utilized for critical parts of the design. For the vector accelerator array, a HLS based model using C/C++ will be developed. It allows to switch between data types and the provides a good design exploration environment.

Chapter 3

Time Schedule

fig 3.1 is the proposed Time plan and schedule of work. The report writing will be done in parallel with the work, as I believe it provides with clarity over the results and often give a solid base to take the work in the right direction.

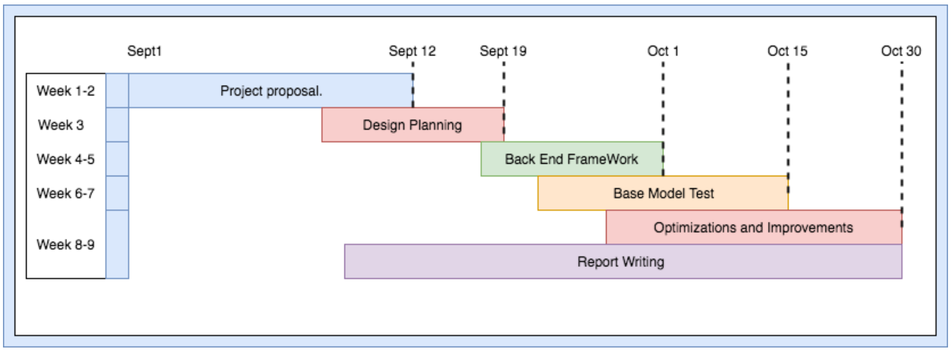

Figure 3.1: Time Plan and Work Schedule.

References

[1] Andrzej Wolski *Beam Dynamics In High Energy Particle Accelerators*, https://www.amazon.co.uk/Beam-Dynamics-Energy-Particle-Accelerators/dp/178326277X/ref=sr_1_1?s=books&ie=UTF8&qid=1501960352&sr=1-1&keywords=andrzej+wolski+beam+dynamics