

Numerical mathematics on FPGAs using C λ aSH

Martijn Bakker

Committee:

Dr.ir. J. Kuper

Dr. R.M.J. van Damme

Dr.ir. J. Broenink

Computer Architecture for Embedded Systems (CAES),
Electrical Engineering, Mathematics and Computer Science (EEMCS),

University of Twente

This thesis is submitted for the degree of

Bachelor of Science

Advanced Technology

July 2nd, 2015

Abstract

Performing computations directly in hardware can be a very challenging task for a scientist or engineer only familiar with software, but there is much that can be gained in terms of power reduction and performance improvements using FPGAs. This thesis describes the process of implementing an accelerator in which the computational part is specified using the functional hardware description language C λ aSH and discusses the feasibility of performing numerical mathematics on this accelerator by computing approximations to ordinary differential equations. The accelerator is capable of using the methods of Euler and Runge-Kutta (second order) to perform the approximations, but due to the use of a fixed-point number representation the accuracy suffers. The performance of the accelerator, implemented on a low-power, low-cost development FPGA: the Altera Cyclone V is 40% worse than an i7-950, but the power usage of the accelerator is 2 orders of magnitude lower.

Acknowledgements

- Jan Kuper - Introducing me to functional programming, C λ aSH and giving feedback
- Christiaan Baaij - Creating C λ aSH and answering related questions
- Ruud van Damme & Jan Broenink - Feedback
- Rinse Wester - Input on configuring and using the Avalon bridges

Acronyms

ASIC	Application-specific integrated circuit
FPGA	Field-Programmable Gate Array
CPU	Central Processing Unit
DSP	Digital Signal Processing
LED	Light Emitting Diode
VHDL	VHSIC HDL
VHSIC	Very High Speed Integrated Circuit
HDL	Hardware Description Language
SoC	System-on-Chip
ODE	Ordinary Differential Equation
HPS	Hard Processor System
ARM	Advanced RISC Machine (CPU development company)
RISC	Reduced Instruction Set Computer
IP	Intellectual Property
GUI	Graphical User Interface

Table of contents

1	Introduction	1
1.1	Project goals	1
1.2	FPGAs	2
1.2.1	What is an FPGA?	2
1.2.2	How does it work?	2
1.2.3	System-on-a-chip	3
1.3	Numerical solvers for ODEs	3
1.4	Functional programming	5
1.4.1	What is functional programming?	5
1.4.2	Using FP for numerical mathematics	7
1.4.3	Example: Numerical solutions of ODEs in Haskell	7
1.5	C λ aSH	10
1.5.1	Mealy machines	11
1.5.2	Advantages of C λ aSH	11
2	Methods	13
2.1	Overall structure	13
2.2	External types	15
2.3	Internal types	16
2.3.1	Internal number representation	16
2.3.2	SystemConstants and SystemState	17
2.4	Implementation of equations and integration schemes	18
2.5	Simulation	19
2.6	Synthesis and deployment	20
2.7	Loading data into the FPGA	21
2.7.1	Constants	21
2.7.2	Initial values	22
2.8	Solving the system and extracting values	23
3	Results	25
3.1	Euler	25
3.1.1	First oscillation - initial position	25
3.1.2	Second oscillation - initial velocity	28
3.2	Runge-Kutta (second order)	29

3.3	Runge-Kutta (fourth order)	32
3.4	Euler revisited	32
3.5	Performance	32
4	Discussion	35
4.1	Accuracy	35
4.2	Performance	35
4.3	Suggestions for further work	36
4.4	Suggestions for additions to C λ aSH	36
5	Conclusion	39
References		41
Appendix A	Haskell source code for numerical solutions of ODEs	43
Appendix B	Project structure	49
Appendix C	Handling data IO	51
Appendix D	Toolchain integration	53
Appendix E	Performance benchmark	57

Chapter 1

Introduction

This introduction contains the project goals and some very short introductions on the FPGAs and how they are used, numerical methods for approximating ODEs, functional programming and C λ aSH. The goal of this report is to be understandable for any technical BSc student which is not necessarily familiar with functional programming, or numerical mathematics and therefore the short introductions are included for the sake of completeness and are assumed knowledge for the rest of the thesis.

1.1 Project goals

From the start, this project has one main goal: obtaining information on the feasibility and the advantages and disadvantages of performing numerical mathematics (specifically, solving ODEs¹) directly on (programmable) hardware, the FPGA. As per usual, having a main goal spawns off several minor goals which support the main part. Both supporting goals are about simplifying the process of configuring FPGAs: an easy way of setting up projects with complicated IO requirements and furthermore, developing a toolchain integration which turns the long process of compiling and deploying your FPGA project into the execution of a single command.

Alongside these concrete goals the underlying theme is to do as much work as possible in C λ aSH, a library and compiler based on the functional programming language Haskell, developed by Christiaan Baaij at the CAES group at the University of Twente. Further elaboration on C λ aSH can be found in section 1.5.

¹In this thesis, the term 'solving' will be used for both the process of solving an ODE analytically and for computing a numerical approximation to an ODE.

1.2 FPGAs

1.2.1 What is an FPGA?

An FPGA (Field Programmable Gate Array) is a chip in which you can specify the hardware yourself. In contrast to regular programming in which you generate a long list of instructions which are executed sequentially on a fixed chip configuration, the FPGA allows you to specify exactly which wire (signal) leads where and what operation should be applied to that signal. This approach to programming can have several advantages. The first one arises from the large opportunities for parallelism. Every part of the FPGA can be executing a meaningful computation simultaneously, whereas processors are bound by the amount of physical cores they have in the amount of truly concurrent instruction executions possible. Secondly, a conventional processor only has a fixed instruction set. Using an FPGA you can define your own instructions (subcircuits), again providing a possible improvement in computational speed. According to [14], FPGAs were already capable of outperforming CPUs on very parallelizable numerical tasks on single and double precision floating point numbers in as early as 2004. Furthermore, as you are implementing your signal processing directly in hardware, there will be a fixed bound on the possible latency. This makes FPGAs ideal for purposes in high-throughput, low-latency signal processing, eg. real-time audio, video or data stream processing. Lastly, the reconfigurability of FPGAs whilst remaining close to the actual hardware allows for cost reductions in the verification of ASIC (Application-Specific Integrated Circuits) designs. It's cheaper to reprogram your FPGA than to have a new version of an ASIC manufactured.

However, the FPGA is a trade-off between implementing designs directly in hardware and being able to run multiple designs (after a reconfiguration). As a consequence of this, it still loses to ASICs with several orders of magnitude on performance [7] and CPUs still dominate in terms of versatility and on-the-fly reconfigurability.

1.2.2 How does it work?

An FPGA is built up from several distinct element types, depicted in figure 1.1

1. *Logic elements* Responsible for the actual signal processing. An FPGA may contain different types of logic elements, eg. memory, DSP and logic blocks. These blocks implement some signal processing capability which can be configured up to certain limits.
2. *Programmable interconnects* In order to be able to represent complex designs, the logic elements need to be connected in a certain way. This is what the programmable interconnects are for. Essentially, those are wires which can be turned on and off by the user as part of a design specification.
3. *IO blocks* Finally, the functionality implemented using the logic elements and programmable interconnects should be exposed to external signals in order to be useful. IO blocks can be used to control hardware pins, controlling a LED or reading a switch but also for more intricate IO facilities, eg. external memory controllers.

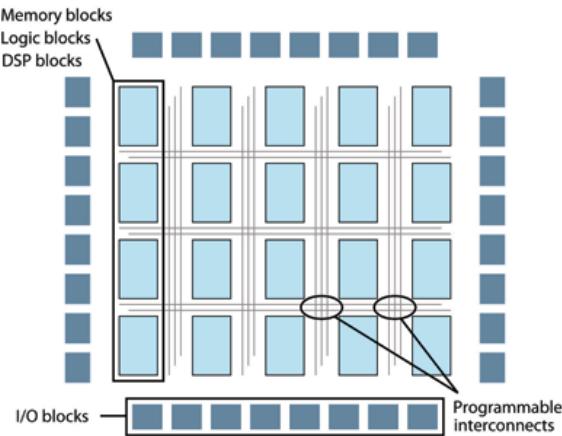


Fig. 1.1 FPGA fabric.

It might seem that programming an FPGA involves manually specifying the interconnects and exact configurations of the logic elements. However, specialized languages have been developed just for the purpose of describing the FPGA functionality at a higher level of abstraction and leave the specific routing and assignment of logic elements to the compiler. The two main advantages of these languages are that you do not have to worry about low-level problems like how the interconnects will be routed and secondly, your written specification will be portable across multiple FPGA vendors as long as the vendor supplies you with the proper compiler from your specification to a file which can be used to program the FPGA.

More information on FPGA functionality can be found in [15].

1.2.3 System-on-a-chip

FPGAs in itself can be useful, but especially for design with IO requirements that are more complex than just reading out hardware switches and controlling LEDs, more control is needed. This requirement has led to the rise of SoCs (System-on-Chip). These devices integrate an FPGA with additional hardware on a single chip. This extra hardware usually contains a CPU, which can be used to simplify the process of loading and extracting data from the FPGA. The SoC used for this thesis is the Terasic SoCKit, a development kit containing an Altera Cyclone V FPGA and a dual-core ARM A9 CPU on a single chip (Altera 5CSXFC6D6F31C6N), alongside a wide variety of IO possibilities. Further information on the SoC used is available at [9].

1.3 Numerical solvers for ODEs

The field of numerical solvers is a vast and active area of research. Furthermore, there is a lot of theory on which solver to pick for specific problems, related to stability, computational efficiency and other factors. However, the goal of this thesis is to get an impression of the

feasibility of using numerical solvers for ODEs on FPGAs. Therefore, the selection of solvers will be restricted to some very basic schemes.

The solvers used have some common properties:

- *Fixed-step* - The solvers compute a value for the ODE at a fixed step size. This means that in contrary to the continuous 'mathematical' solution of an ODE, the output of the solver will be an approximate to the actual value of the solution at the discrete set of values of the independent variable, in which the difference between the values of the independent variables is defined by the fixed step size.
- *Single-step* - The approximate value of the ODE x_k at t_k only depends on t_{k-1} , x_{k-1} and the ODE that is being approximated.

Both solvers work equally for both vectors and scalars. In case of a vector, the operations listed in 1.1 and 1.3 should simply be applied to all elements of \vec{x} .

Euler

The easiest numerical solver, without doubt, is Euler's method (equation 1.1). For every discrete point in time, the value of the next point is approximately equal to the derivative at that point multiplied by the time step added to the current value. The simplicity of the scheme has a cost: it is not very accurate and the errors accumulate quickly. From [13], the maximum error of Euler's method can be shown to be linear in the time step and exponential in the interval length (eq 1.2), in which M and L are constants depending on the equation to be solved, $b - a$ is the solution interval length and h is the time step.

$$\begin{aligned} s_{k,1} &= f(t_k, x_k) \\ x_{k+1} &= x_k + h s_{k,1} \end{aligned} \tag{1.1}$$

$$\text{maximum error}_{\text{euler}} \leq \frac{Mh}{L} (e^{L(b-a)} - 1) \tag{1.2}$$

Runge-Kutta

The Runge-Kutta methods are a family of solvers, of which the 4th order version is the most well-known (RK4). The solver used here will be a second-order Runge-Kutta method, also known as the *improved Euler's method* [13]. This second order method requires the computation of two slopes (equation 1.3), in contrast to the single one required for Euler's method.

$$\begin{aligned} s_{k,1} &= f(t_k, x_k) \\ s_{k,2} &= f(t_k + h, x_k + h s_{k,1}) \\ x_{k+1} &= x_k + h \frac{s_{k,1} + s_{k,2}}{2} \end{aligned} \tag{1.3}$$

$$\text{maximum error}_{\text{RK2}} \leq \frac{Mh^2}{L} (e^{L(b-a)} - 1) \quad (1.4)$$

Note that the maximum error of RK2 is proportional to the square of the time step (equation 1.4). As the time step decreases by a factor of 2, the maximum error will decrease by a factor of 4, given that the equation and the range stay the same. However, the maximum error still depends exponentially on the interval length.

1.4 Functional programming

1.4.1 What is functional programming?

As the name suggests, the functional programming paradigm uses functions. These functions are used to build up the program and create structure. In contrast to the imperative programming paradigm, there is no assignment - there are only expressions. It is by evaluation of these expressions that you execute your program. These expressions consist of variables, constants and operations. However, the name variable may be ill-chosen, as assignment does not exist and therefore it is impossible for a variable to vary. Once you have bound a value to a certain variable, this value may not change and the value should be the same at every point where this variable is referenced (a concept called referential transparency).

The exclusion of assignment from functional languages has several consequences. It becomes impossible to program using loops. The alternative is the use of recursive (listing 1.1) and higher-order functions (functions that have other functions as parameters or output). Especially higher-order functions have the side effect that they add clarity about the way the program functions (listing 1.2). For instance, a **map** applies the same function to every element in a list, resulting in a new list. A fold would start at an initial value and element-by-element combine the list into a single value using a specified function, eg. addition. Both these operations would be implemented using a for-loop in an imperative language, but their goal is completely different. The availability of higher-order functions allows for more clarity in code by being able to specify exactly what kind of operation you want to perform. Another effect of the lack of assignment is the lack of side effects in functional programming. As there is no way to modify a variable, there is also no way of accidentally modifying a variable such that you enter an invalid state and other parts of the program stop functioning correctly.

Listing 1.1 Recursive functions

```

1 fac :: Num a => a -> a
2 fac 0 = 1
3 fac n = n * fac(n-1)

```

Listing 1.2 Higher order functions

```

1 timesTwo :: Num a => [a] -> [a]
2 timesTwo xs = map (*2) xs
3
4 sum :: Num a => [a] -> a

```

```

5 sum xs = foldl (+) 0 xs
6
7 powers :: Num a => a -> a -> [a]
8 powers init power = iterate (\x -> x * power) init
9
10 powers100 = take 100 $ powers 1 2

```

The concept of higher-order functions also serves as an introduction to another very important concept in Haskell: the type system. Everything has a fixed type and often times, when only looking at the type definition you can already guess what the function is going to do. Consider the type signature of `map` in listing 1.3. It requires a $(a \rightarrow b)$, a function to turn something of type a into something of type b . Furthermore, it needs a list of a , $[a]$ (indicated by the square brackets) and it returns something of type 'list of b ' ($[b]$). The type signature of the `foldl` is slightly harder to understand, but it requires a function which needs an a and a b to produce a new b . Furthermore, an initial value (a) and a list of b to operate on ($[b]$) in order to return the final result, which has again type a . Function type signatures can be daunting to understand at first, but not all of them are as complicated as the `foldl`. For instance, a function which can be used to represent a differential equation, which needs a state of the system (the `ODEState`) in order to compute the derivative at that point (the `D_ODEState`). As a last example, the type signature of `Solver`: it requires some integration scheme, settings for the time (initial time and a time step), it requires an equation and an initial state of the system. All of this combined results in a list of states: the numerical approximation to the solution of the ODE. Note that in this example, the integration scheme `Scheme` itself is a function, for which understanding the type signature should pose no problem by now.

The last concept in functional programming which is important to understand is so-called lazy evaluation. In contrast to imperative programming, a value is only evaluated whenever it is needed. For instance, this allows for the concept of an infinite list, which is definitely not attainable in imperative programming as it would run out of memory. An example of this is the higher-order function `iterate`, shown in listing 1.2, as part of the `powers` function. This function generates an infinite list by repeatedly applying a function (in this case a multiplication by a constant) to its own output, starting at the initial value `init`. However, it would be impossible to actually display the entirety of said infinite list and therefore you apply another function, `take n`, which only consumes the first n elements of a list. As the total evaluation only requires the first n elements, these are the only ones that will actually be computed. [12]

Listing 1.3 Type signatures

```

1 map :: (a->b) -> [a] -> [b]
2 foldl :: (a->b->a) -> a -> [b] -> a
3
4 type Equation = ODEState -> D_ODEState
5 type Scheme = TimeSettings -> Equation -> ODEState -> ODEState
6 type Solver = Scheme -> TimeSettings -> Equation -> ODEState -> [ODEState]

```

1.4.2 Using FP for numerical mathematics

Functional languages have several properties which make them suitable for the purpose of solving problems in numerical mathematics. First and foremost, Haskell, being based on λ -calculus is very close to mathematics. The useful mathematical properties here are *referential transparency*, easy *partial function application* and being a *declarative language*. Referential transparency implies that a variable only has a constant value which is the same everywhere in the program. This prevents that changing a variable might have influence on another computation as a side effect and it corresponds to mathematical notation. For instance, in an imperative programming like C you could write $i = i + 1$, which is a mathematical impossibility and therefore not allowed in Haskell. Partial function application is another very useful concept. Often in numerical mathematics, you want to create or process a function. You need a function that has another function as return value. For instance, take a function which requires two arguments. After only applying a single argument, the object returned still needs the second argument in order to compute the final value. This is exactly according to the definition of a function: An object that still needs arguments before being able to return its final value. Being a *declarative language* means that you write code that specifies what you want to accomplish, not how to get there. This concept is again borrowed from mathematics. You put in a set of function definitions and Haskell will figure out how to actually compute the value you request according to those definitions. This property of declarativity also has the result that Haskell is a terse language whilst remaining easy to understand. Secondly, Haskell has a very strong type system. The type system has three main advantages. It becomes very easy to swap out and replace functions as long as you make sure that the types are the same. The Haskell compiler will start to assert errors immediately whenever you feed it something which does not make sense or could be ambiguous which is very useful when writing programs. By having a look at the types of a Haskell program it becomes very straightforward to see what the program does and how it works, which is very useful when attempting to understand your own or someone else's code. Lastly, a property which is often very important for numerical mathematics: Haskell is fast. According to the *Computer Language Benchmarks Game* [5], Haskell is almost on par with Java and Fortran but significantly faster than Python and Matlab (not shown), two languages which are often used for numerical mathematics nowadays. There is still a performance gap of around a factor 3 between Haskell and C (the reference), hence if speed is of the absolute highest concern C is still a valid option.

1.4.3 Example: Numerical solutions of ODEs in Haskell

As mentioned before, the types in Haskell reveal lots of information about the structure and functionality of the program. The three main types constituting the numerical solver for ordinary differential equations are listed in listing 1.4.

Listing 1.4 Main types for the ODE solver

```

1  type Equation = ODEState -> D_ODEState
2  type Scheme  = TimeSettings -> Equation -> ODEState -> ODEState
3  type Solver   = Scheme -> TimeSettings -> Equation -> ODEState -> [ODEState]
```

Equation

In essence, a differential equation is a mapping (function) from a certain state of the system to the change of this system. This is also what the type signature of Equation signifies, a mapping from an ODEState to a D_ODEState, the change in state or the derivative. This generic set up allows the specification of any ODE for solving. The implementation in pure Haskell of a simple ODE is given in listing 1.5, which corresponds to the equation $x' = -x$. However, this representation is not very elegant and a lot of the code is performing unboxing of the types. Using property that this equation is linear, it is possible to use an utility method which takes as input a matrix and returns the Haskell differential equation function belonging to that matrix. The same can be done for heterogeneous linear systems using a different utility function, which does not only takes a matrix as input but also a list of functions representing the heterogeneous part of the equation.

Listing 1.5 Example equation for exponential decay

```

1  eq_exponential :: Equation
2  eq_exponential state    = [-x !! 0]
3  where
4  x = xs state

```

SolveMethod

The SolveMethod performs the actual computations on what the next value of the solution should be: the integration scheme. In order to obtain this next state, the scheme needs three input values: It needs information on the timing constraints of the solution, in this case it needs the time step. Furthermore, it needs the equation itself and it requires the state of the system at t_n in order to be able to determine the state of the system at $t_{n+1} = t_n + \Delta t$.

The most straightforward integration scheme is called forward Euler, given in equation 1.1. Listing 1.6 depicts the translation of the mathematical expression 1.1 to Haskell. Even though some list operations have been inserted (**zipWith** and **map**), the structure is still recognizable. It computes the change in state, multiplies this with the time step obtained in line 6 and adds the initial state in line 4. Lastly, the integration scheme returns the new state of the equation, consisting of a list of x-values and a corresponding time value. Implementations of different solvers (i.e. 4th order Runge-Kutta) can be found in appendix A.

Listing 1.6 Example code for the Euler integration scheme

```

1  euler :: Scheme
2  euler time equation initState    = ODEState newX newT
3  where
4    newX    = zipWith (+) (xs initState) dX
5    dX     = map (timestep *) (equation initState)
6    newT   = (t initState) + timestep
7    timestep = dt time

```

Solver

The Solver function in listing 1.7 acts as the main interface to the program. You specify a SolveMethod, the TimeSettings (containing the time step and the time at which to stop solving), the equation itself and an initial condition. The Solver will then return a list of states of the system. This problem could be solved recursively, but a method featuring more clarity is a higher-order function, in this case `iterate`. This function keeps applying itself to its own output, starting with some specified initial value, generating an infinite list. However, we are not interested in the infinite list of solutions and therefore we only take the first N elements, in which N depends on the initial time, the final time and the time step used for the solution.

The solutions of a wide range of equations, both linear and non-linear, both homogeneous and heterogeneous and using the input matrix utility functions have been plot with suitable initial conditions to show their behavior in figure 1.2.

Listing 1.7 The main controlling function

```

1  solve :: Solver
2  solve solvemethod time equation initState = states
3  where
4      states = take steps $ iterate (solvemethod time equation) initState
5      steps = ceiling $ (tMax time - t initState) / dt time

```

Results

This example of approximating solutions to ODEs using Haskell shows the versatility the language and especially its type system. Without any trouble, it is possible to exchange solver schemes, equations, settings and initial conditions by merely changing a single identifier in the function call. A variety of equations with suitable initial conditions have been approximated: both linear and non-linear, of multiple orders, both homogeneous and heterogeneous. As an example, the equations of which the solutions (with suitable initial conditions) are depicted in figure 1.2 are listed in equations 1.5. The exact Haskell source generating the plot is shown in A, but depends on an external library for generating the plot [2].

Exponential	$x(t)' = -x(t)$
Simple harmonic	$x(t)'' = -x(t)$
Cosine hyperbolic	$x(t)' = \frac{\sqrt{x(t)^2 - a^2}}{a}$
Simple harmonic	$\vec{x}(t)' = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \vec{x}(t)$
Simple forced harmonic	$\vec{x}(t)' = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \vec{x}(t) + \begin{bmatrix} \sin(t) \\ e^{-t} \end{bmatrix}$

(1.5)

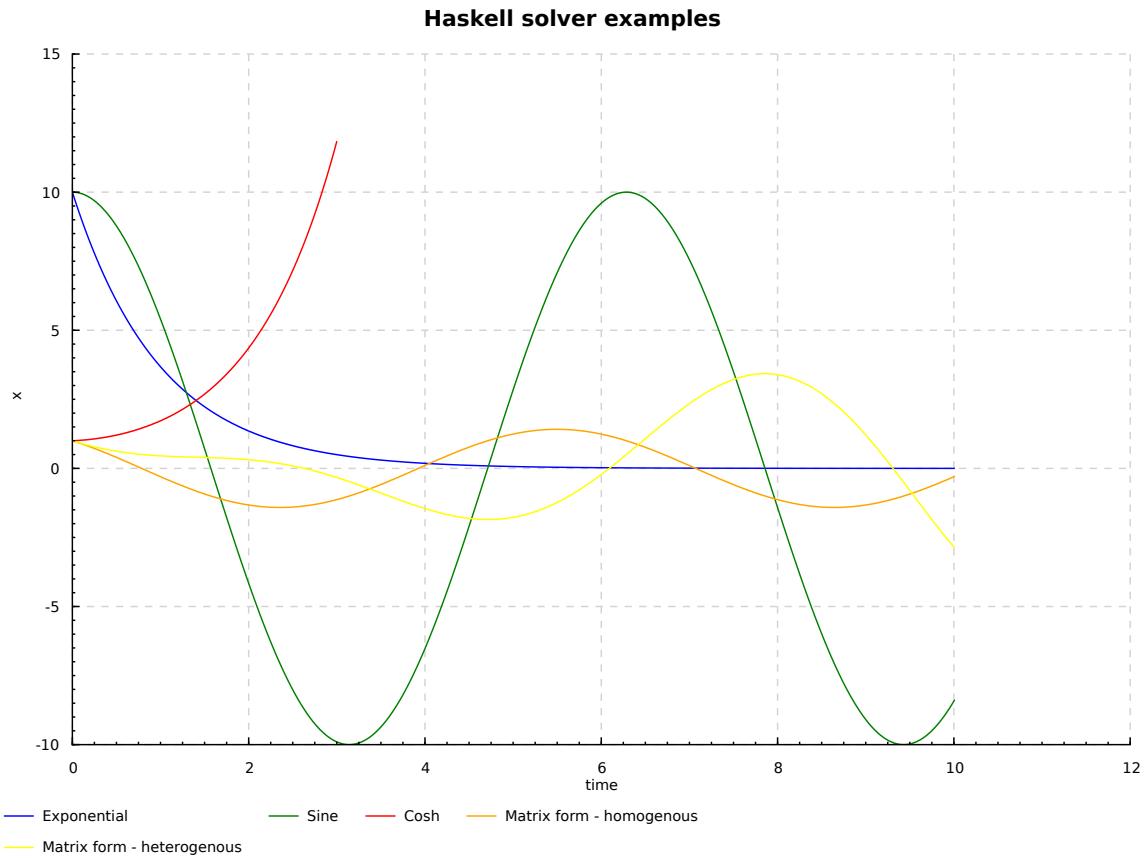


Fig. 1.2 Plots of the ODE solutions, simulated in Haskell.

1.5 C λ aSH

Functional programming has several aspects in common with hardware design, which was the original reason for the development of C λ aSH, a library and special compiler which is capable of compiling a subset of Haskell into HDL. In C λ aSH, every (sufficiently complex) hardware design consists of two parts: a combinatorial part and a synchronous part. It is the combinatorial part which can be modeled with few problems in a functional way: there is no time dependency - the output is merely an evaluation of a certain function of the inputs and for every input the output will be the same. However, complex hardware designs do not merely consist of stateless combinatorial circuits, in order to perform work some statefulness must be included. This is the task of the synchronous part of the design, keeping track of the state. The state is often implemented using latches or memory cells which can get updated with new values, computed by the combinatorial part, every clock cycle (hence the name synchronous part). Together, the combinatorial part and the synchronous part of hardware are called sequential logic: the output depends on both the current input and the past inputs.

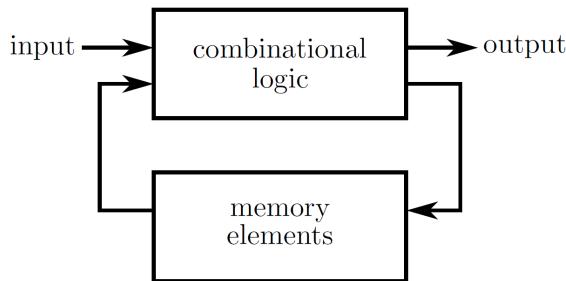


Fig. 1.3 A flowchart of the Mealy machine

1.5.1 Mealy machines

It is the synchronous (statefulness) part that cannot be modeled directly in Haskell, but C λ aSH gets around this by use of the Mealy machine [10]. On every clock cycle, based on the input and the current state the combinational logic computes an output and a new state. The updated state gets stored in the memory elements, the output forwarded to the external ports of the hardware. When looking at some C λ aSH-Haskell, this structure of generating an output and a new state based on an input and the current state is still clearly visible, for example in listing 1.8. After defining a function with the proper type signature for a Mealy machine you can pass this function to the C λ aSH built-in function `mealy`, which handles the conversion to a valid `topEntity`. In this case the specified function is `multiply-accumulate`, which, for every clock cycle, multiplies its inputs together and adds it to an internal sum which gets forwarded to the output. [6]

Listing 1.8 A basic example a multiply-accumulate specification in C λ aSH

```

1 topEntity :: Signal (Signed 9, Signed 9) -> Signal (Signed 9)
2 topEntity = mealy mac initial
3
4 mac state input = ( state ', output)
5   where
6     (x,y) = input          -- unpack the two inputs
7     state' = state + x*y  -- the new state
8     output = state'        -- output the new state

```

1.5.2 Advantages of C λ aSH

The power of C λ aSH is manifold. Firstly, you are writing valid Haskell code, which allows you to simulate and verify your designs by relying on existing debugging techniques for Haskell. Secondly, Haskell enables more clarity in code, partially by use of higher-order functions, partially by use of its record syntax, which allows you to easily group signals together in a meaningful way. This additional clarity is especially useful in complex designs in combination with the modularity of functional programs (as long as you adhere to the type signature), which allows you to easily swap out parts of design. Thirdly, notice that the only location at which the types of the signals has been specified is in the first line (Signed 9, a

signed integer of length 9). This means that, in order to have our multiply-accumulate circuit work on, for instance, unsigned integers of length 32, the only place that needs modification is the `topEntity` function type signature. The other types are inferred from here.

After specifying your design in C λ aSH, you can invoke the C λ aSH-compiler to generate HDL. Both mainstream languages (VHDL and Verilog) are supported. It is then the HDL which can be compiled by a vendor-specific compiler into a binary file which can be flashed to the FPGA.

Chapter 2

Methods

This chapter contains a description of the methods used in describing the hardware on the FPGA side using C λ aSH and VHDL. The description of the host-side (HPS) programming is included in appendix C, as even though this part of the project is crucial for obtaining results, it is merely tangential to the project goals. The methods are ordered chronologically in order to properly describe the process that lead to the FPGA side programming used in chapter 3. This approach has as side effect that the explanation of the source code is evaluated lazily: only whenever necessary. Therefore the IO system is only covered after testing and synthesis, as it is not yet needed before.

2.1 Overall structure

In any reasonably complex project it pays off to keep a clear structure: it improves understandability and allows for easier debugging. For simple projects in C λ aSH, the structure would be uniquely determined by the HDL generated by C λ aSH, but this project also relies on other sources of HDL. Most noteworthy, the interface between the HPS and the FPGA (Altera Avalon); these signals are very specific to the type of the FPGA and the interconnects it has to the HPS. Luckily, the vendor provides a way to generate HDL (the QSys system) in order to create a bridge from the host code running on the HPS to the programmable hardware, the FPGA. This process is described in appendix ???. After configuring the bridge, VHDL can be generated and we are left with an instantiable VHDL component. However, even though C λ aSH is capable of instantiating external IP components written in VHDL [11], the most sensible way of implementing such a structure (due to the easy extensibility) is by writing a specific connecting component in VHDL, which can instantiate the C λ aSH-generated VHDL. The only responsibility of the connecting component is to distribute and forward signals: it should not perform any computation. It would be a straightforward process of generating such a connecting component based on the signal names in the top-level entities of the components it instantiates, however, due to the fact that it does not change for different designs and it is not very large, the connecting component has been written by hand in VHDL. Figure 2.1 depicts an overview schematic of the complete system.

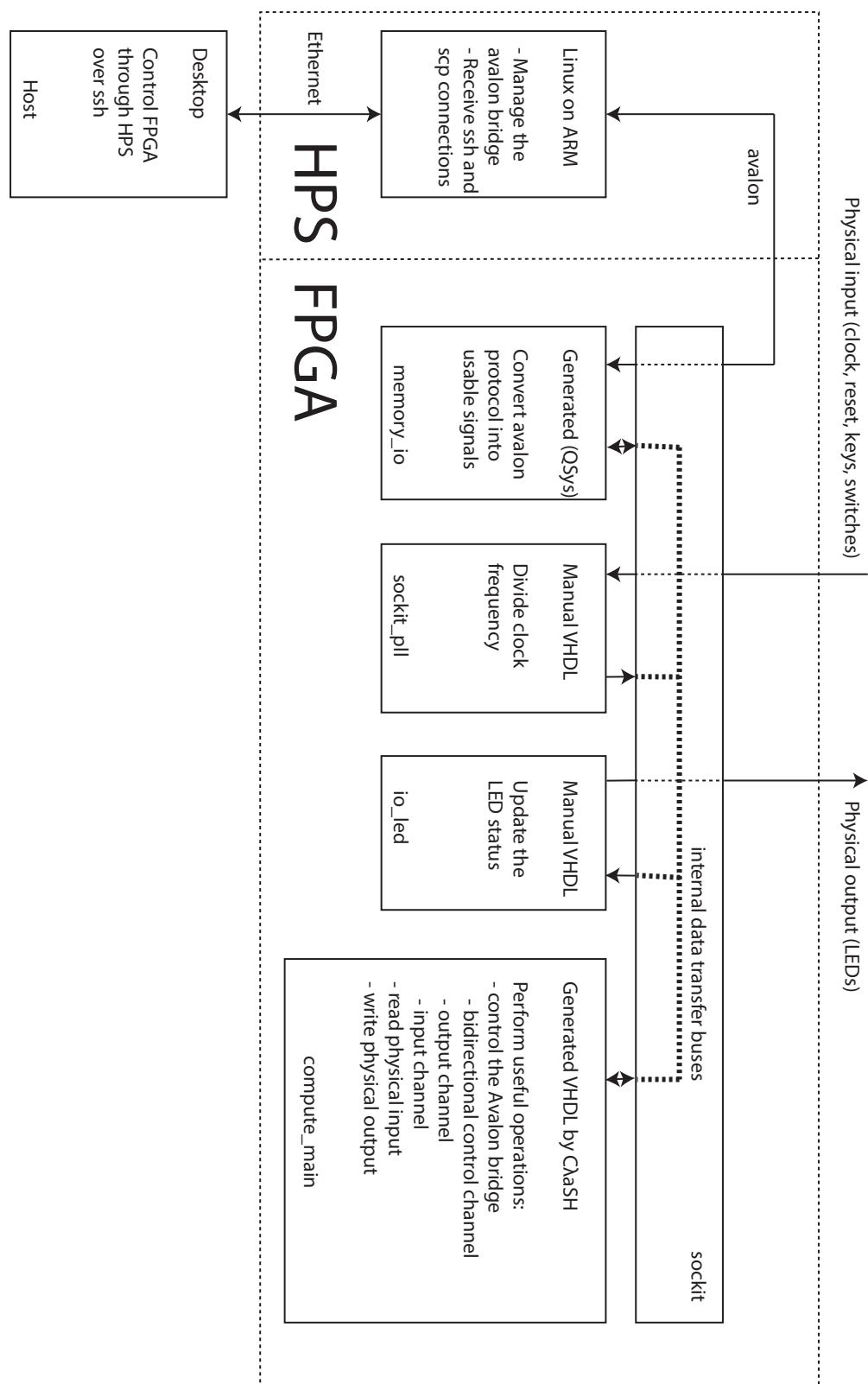


Fig. 2.1 A block diagram depicting the overall system encapsulating the C_λaSH-generated module

2.2 External types

Usually, one of the first things to do when setting up a Haskell project is defining types, and using CλaSH forms no difference. It is especially useful to start by defining an input and an output signal. In a project with simple IO requirements the input and output signals would map straight to switches, keys and LEDs for testing purposes, but as the goal is to write data from the host system into the FPGA registers, the input and output signals will be a lot more complicated. The input and output signals from the CλaSH-generated VHDL top entity are shown in listing 2.1. The input consists of three separate channels: a bidirectional control channel, the actual input channel and some input needed for the output channel. It might appear strange that the output channel requires input, but this is necessary as the communications adhere to a strict master-slave pattern, in which the HPS is the master and the FPGA the slave. The HPS has to indicate whenever it wants to receive the FPGA output (by specifying an address (out_address) and setting the out_read-bit). As a consequence of the strict master-slave protocol, there are a lot less output signals than input signals: only the control and output channel can output data. Lastly, there are the keys, switches and LEDs as additional input and output signals. This concludes the type signatures of the input and output signals in CλaSH-Haskell. However, as the goal is to generate actual VHDL, CλaSH requires some additional information for the naming of the ports of the topEntity of the VHDL module. This is specified using the topEntity -annotation. In this case this annotation is not very interesting and it's merely a restatement of the input and output signal names, but when the CλaSH-generated VHDL should instantiate other VHDL-modules or use multiple clocks, the annotation can become more complex. It is shown in listing 2.2.

Listing 2.1 Input and output signals for CλaSH

```

1  data InputSignals = InputSignals { keys_input :: BitVector 4  -- keys_input
2      , switches_input :: BitVector 4  -- switches_input
3      , control_write :: Bit      -- control_write
4      , control_writedata :: BitVector 32 -- control_writedata
5      , control_address :: BitVector 8  -- control_address
6      , control_read :: Bit      -- control_read
7      , in_write :: Bit      -- in_write
8      , in_writedata :: BitVector 32 -- in_writedata
9      , in_address :: BitVector 8  -- in_address
10     , out_address :: BitVector 8  -- out_address
11     , out_read :: Bit
12   } deriving(Show)
13
14 data OutputSignals = OutputSignals { leds_status :: BitVector 4  -- leds_status
15      , control_readdata :: BitVector 32 -- control_readdata
16      , out_readdata :: BitVector 32 -- out_readdata
17   } deriving(Show)

```

Listing 2.2 Signal names used in the CλaSH-generated topEntity VHDL module

```

1  {-# ANN topEntity
2  (defTop
3  { t_name      = "compute_main"

```

```

4     , t_inputs  = [ "keys_input"
5         , "switches_input"
6         , "control_write"
7         , "control_writedata"
8         , "control_address"
9         , "control_read"
10        , "in_write"
11        , "in_writedata"
12        , "in_address"
13        , "out_address"
14        , "out_read"
15        ]
16     , t_outputs = [ "leds_status"
17         , "control_readdata"
18         , "out_readdata"
19         ]
20     }
21 )
22 #-}

```

2.3 Internal types

So far the external types of the C λ aSH-code have been covered, but the real work is done by the internal types: the types that keep track of the state of the system and allow it to do useful work. In order to allow for easy modifications to the types upon which the FPGA operates, they have been defined once and have been referenced everywhere else. This in combination with the property of the higher-order functions in C λ aSH that they can operate on all vectors, regardless of length ensures that changing some internal types will not break the program.

2.3.1 Internal number representation

The solvers main data type is `type Data = SFixed 8 24`. The constructor `SFixed 8 24` stands for a signed fixed-point number, with 8 integer bits and 24 fractional bits. It uses the 2's complement signed number representation, meaning that an 8-bit integer part is capable of representing the integers in the range $[-128..127]$. The 24 fractional bits give this number representation a smallest representable unit of 2^{-24} . This results in accuracy up to the 7th decimal place, which is similar to the IEEE 754 single precision floating point standard when the magnitude of the number represented is larger than 1. However, in contrast to the floating point standard, fixed point numbers guarantee precision up to a certain amount of decimal places, whereas floating point guarantees accuracy for a certain amount of significant figures (within the bounds of the representation).

The main reason for the use of fixed point integers is that C λ aSH does not support floating point numbers yet, but additionally, fixed point representations are less demanding on FPGA area and result in a shorter critical path. Furthermore, the reason for choosing the total width of the number representation to be 32-bits is purely convenience: the input and

output bridges are also 32 bits wide which allows for transferring a single number per write or read.

2.3.2 SystemConstants and SystemState

The SystemConstants keep track of the variables that do not change during the process of solving the ODE. It consists of a variety of constants that need values for the integration scheme (maxtime, timestep and maxstep). Furthermore, it may contain custom constants, which are passed to the equation to be approximated. This setup allows for (limited) changes to the equation by changing the constants at run-time without the time-consuming requirement of recompiling the entire FPGA side of the project.

Moving on to the data type which responsible for keeping track of the state of the system: the SystemState. This type has two fields: the ODEState, which is has the exact same function as the similarly named type in section 1.4.3: it keeps track of the values and the time in the numerical solver. It does this using a valueVector (of type Vec 4 Data) and a time variable (of type Data). The main difference between the type of the ValueVector in this implementation of ODEState and the one in section 1.4.3 is that this one uses a vector instead of a list. In Haskell, lists can have any length (including infinite) whereas the C λ aSH-vectors have a fixed length. This property is very important when generating VHDL, as all vector lengths have to be immutable and known at compile-time in order to compile the higher-order Haskell functions (eg. map) to VHDL and afterwards to hardware.

The second field of the SystemState is a counter called step. Together with the maxstep field in SystemConstants, these govern the amount of output generated from the FPGA. A more elaborate explanation of the generation of output can be found in section 2.8.

Listing 2.3 Internal state variables for C λ aSH

```

1  type Data = SFixed 8 24
2  type UInt = Unsigned 32
3  type ValueVector = Vec 5 Data
4  type ConstantVector = Vec 20 Data
5
6  data ODEState = ODEState { valueVector :: ValueVector
7                                , time :: Data
8                                } deriving(Show)
9
10 data SystemState = SystemState { odestate :: ODEState
11                                , step :: UInt
12                                } deriving(Show)
13
14 data SystemConstants = SystemConstants { maxtime :: Data
15                                , timestep :: Data
16                                , maxstep :: UInt
17                                , userconstants :: ConstantVector
18                                } deriving (Show)
19
20 uIntMax = 4294967295 :: UInt
21

```

```
22  type Equation = (ODEState, ConstantVector) -> ValueVector
23  type Scheme = SystemConstants -> Equation -> ODEState -> ODEState
```

2.4 Implementation of equations and integration schemes

The C λ aSH-implementations of the equations and integration schemes can be very similar to the implementations in plain Haskell, from section 1.4.3. However, C λ aSH does not support any special operations yet (exponentiation, trigonometry or fractional powers) and therefore this imposes limitations on the type of equations which are representable, especially non-linear and heterogeneous equations which use special operations. In order to allow for easy specification without recompilation of a large range of equations, the default equation for testing will be a 4 by 4 matrix vector equation (2.1). This set up is capable of representing any linear and homogeneous fourth order equation with constant coefficients as well as lower-order equations with constant heterogeneous parts.

$$\begin{bmatrix} x_0 \\ x_1 \\ x_2 \\ x_3 \end{bmatrix}' = \begin{bmatrix} c_0 & c_1 & c_2 & c_3 \\ c_4 & c_5 & c_6 & c_7 \\ c_8 & c_9 & c_{10} & c_{11} \\ c_{12} & c_{13} & c_{14} & c_{15} \end{bmatrix} \begin{bmatrix} x_0 \\ x_1 \\ x_2 \\ x_3 \end{bmatrix} \quad (2.1)$$

In order to implement this equation in Haskell (listing 2.4 shows the 2 by 2 version), it is certainly possible to use higher order functions, using `foldl` `(+)` `0` `$ zipWith` `(*)` `matrixRow` `Vector`. However, this would require the construction of additional vectors. Furthermore, as C λ aSH is not yet completely optimizing, in order to obtain the highest possible performance for this hot-zone of the hardware, the decision was made to manually unroll the higher order functions into expressions containing only the unpacking of variables from the `SystemConstants` and simple arithmetic operators, `+` and `*`.

Listing 2.4 Implementation of a second order equation with constant coefficients

```
1  matrix2d :: Equation
2  matrix2d ( odestate , constants ) = dxs
3  where
4      xs = valueVector odestate
5
6      c1 = constants !! 4
7      c2 = constants !! 5
8      c3 = constants !! 6
9      c4 = constants !! 7
10
11     x0 = c1 * (xs !! 0) + c2 * (xs !! 1)
12     x1 = c3 * (xs !! 0) + c4 * (xs !! 1)
13
14     dxs = fst $ shiftInAt0 xs (x0 :> x1 :> Nil)
```

As for the integration schemes, these are incredibly similar to the implementations from section 1.4.3. The first part only contains unpacking of necessary variables from the records. For Euler's method, the actual work integration scheme is only a single line, the remaining

lines check whether the time is not yet exceeding the maximum simulation time and generate an ODEState type, which gets used stored as part of the state in the main controlling logic, from listing 2.8. A reference to the controlling logic (the topEntity), the integration schemes and the equations can be found in appendix B

Listing 2.5 Euler's method in CλaSH

```

1  euler :: Scheme
2  euler constants equation state = state '
3  where
4      --Unpack the needed values
5      c_user = userconstants constants
6      c_maxtime = maxtime constants
7
8      h = timestep constants
9      t = time state
10     xs = valueVector state
11
12     --Apply Euler's integration scheme
13     eulerxs = zipWith (+) xs $ map (*h) (equation ( state , c_user ))
14
15     --Check the time constraints
16     (xs ',t ') = if t < c_maxtime then (eulerxs , t + h)
17                     else (xs,t) -- already at maximum time
18
19     state ' = ODEState {valueVector = xs ', time = t '}
```

2.5 Simulation

After writing the Haskell code which is going to be compiled by CλaSH, it is very easy to simulate your design: CλaSH includes a simulate function which is capable of generating user-specified signals. Especially in contrast with the process of generating test benches for VHDL, using Haskell to perform simulations saves a lot of time and typing. However, choosing to perform the tests in Haskell over a VHDL testbench does have the underlying assumption that CλaSH properly translates the Haskell specification into VHDL. CλaSH is also capable of generating VHDL testbenches, but this merely shifts the assumption of correctness from CλaSH' ability to generate VHDL to its ability to generate VHDL testbenches, therefore, this option of directly generating VHDL testbenches has not been explored further.

Listing 2.6 A simulation in CλaSH

```

1  dis = defaultInputSignals
2
3  --write input
4  cis x y = defaultInputSignals { control_write = 1, in_address = x, in_writedata = y }
5  wis x y = defaultInputSignals { in_write = 1, in_address = x, in_writedata = y }
6
7  -- start computation and get output
```

```

8   scs = defaultInputSignals { control_write = 1, control_writedata = 1 }
9   fvs x = defaultInputSignals { out_address = x, out_read = 1 }
10
11  is = [ cis 4 (-1)
12    , cis 5 7
13    , wis 1 10
14    , wis 2 20
15    , wis 3 30
16    , wis 4 40
17    , scs
18    , dis
19    , dis
20    , fvs 1
21    , fvs 2
22    , fvs 3
23    , fvs 4
24  ]
25
26 test x = Data.List.take (Data.List.length x) $ Data.List.map out_readdata $ simulate topEntity x

```

The simulations in C λ aSH consist of three parts:

1. *Signal definitions* - In order to keep the rest of the code concise, it is important to first define often-used signals. For versatility, these signals can even be functions: some values still have to be defined in order to return a signal.
2. *Signal listings* - This is where signals defined in step 1 are put together in order to create a list of signals: the input of the simulate function.
3. *Simulation* - The simulate function takes a topEntity and a list inputs and returns a list of output. Due to the infinite nature of the Signal in C λ aSH, which is fed into the topEntity as input it is necessary to only take a certain amount of elements. However, as Haskell is evaluated lazily, the infinite lists are not a problem.

The output of the simulation is printed to screen, which was checked by eye for grave mistakes. If everything appeared to be correct, the design was ready for synthesis. The reason for such inexactness in verifying by simulation is that (at least for simple) Haskell programs, that they tend to have the property that they either work correctly or they do not work at all. Furthermore, the real trouble in debugging lies in the other parts of the FPGA side of the design: proper clock frequencies and the IO system.

2.6 Synthesis and deployment

After the simulation appears to be correct, C λ aSH should generate HDL which can be compiled by the FPGA vendors tools into a binary file which can be used to program the FPGA. The process of synthesis and deployment consists of several steps, of which a short overview is shown below.

1. *Generating HDL* - The C λ aSH compiler contains the optional flags $--vhdl$ and $--verilog$. These can be used to generate HDL.

2. *Project creation* - Set up the external IO systems. These will not be written in CλaSH-generated HDL, but will be generated by another tool. In case of Altera this is the QSys system. Furthermore, add the manually written HDL files, for instance the connecting component and a frequency divider.
3. *Adding CλaSH-generated files to the project* - Add all CλaSH-generated files: if both the connecting component and the CλaSH-Haskell code were written properly then the connecting component should be able to instantiate the CλaSH-HDL as their ports match.
4. *Compiling* - After everything has been added and configured properly, start the compilation. For Altera FPGAs, the result will be a .sof (SDRAM Object File).
5. *Deploying* - Use the 'Programmer' feature in order to flash the .sof to the FPGA.

The process depicted above is rather high in manual workload as it uses the Quartus GUI. A shell script automatizing the process is described in appendix D.

2.7 Loading data into the FPGA

2.7.1 Constants

In order to understand the CλaSH source code, it's important to know what the variables mean. In order to keep the lines relatively short, the variable names are rather short, but they do follow a fixed pattern. All variables starting with `i_` indicate input, `s_` a state and `c_` a constant value. As for the input variables, those consist further out of 1 or 2 characters. The first character indicates the source channel of the input (whether it is a real input : `i`, a control signal : `c` or a request for output `o`). The second optional character is either `a`, for address, or `s`, for 'set': the boolean indicating that the input is ready to be read. When this second character is missing, the data itself is meant.

Listing 2.7 Handling the input of the constants

```

1  systemConstants'  --Enter the constants into the ConstantVector
2  | i_cs == 1 && i_ca == 1      = systemConstants{ maxtime = i_c_d }
3  | i_cs == 1 && i_ca == 2      = systemConstants{ timestep = i_c_d }
4  | i_cs == 1 && i_ca == 3      = systemConstants{ maxstep = i_c_u }
5  | i_cs == 1 && i_ca >= 4      = systemConstants{ userconstants = c_user' }
6  | otherwise                      = systemConstants
7  where
8      c_user' = replace i_ca (unpack i_c :: Data) c_user
9      i_c_d   = unpack i_c :: Data
10     i_c_u   = unpack i_c :: UInt

```

The first step of getting the system to work is loading constants into the FPGA. These constants govern the time step, the maximum time for simulation, how much output to generate and it's possible to specify custom constants which can be used in the equations you are solving. In order to keep the system simple, these constants are sent over the control channel as the input channel is reserved for initial values. However, before covering the specifics of handling the constants, it is important to understand the behaviour of the

signals originating from the bridge between the HPS and the FPGA first. Whenever the HPS program writes the 32-bits value V to 8-bit address A , three things happen simultaneously, `control_writedata` takes on the value V , `control_address` gets set to address A and `control_write` gets set to true. This set up means that it is possible to differentiate the target of the control input signal based on the value of the address. Only whenever `control_write` is true the control input value can be considered valid. Lastly, as Haskell is a strongly typed language, you cannot simply insert a `BitVector 32`, originating from `control_writedata` into a `Vec 4 (SFixed 8 24)`. You first have to cast or unpack the `BitVector 32` into a `SFixed 8 24`, which luckily does not pose any problems as they both consist of 32 bits. The protocol for entering constants is depicted in table 2.1.

Table 2.1 The protocol for entering constant values into the FPGA, based on addresses

Address	Function	Specifics
0	Signaling flags	Writing 1 starts the computation Writing 2 performs a soft reset
1	Maximal computation time	
2	Time step	
3	Step limit for blocking	
4+	Custom constants	

2.7.2 Initial values

The initial values are loaded into the FPGA in the same way as the constants. The address designates the location at which the value should be stored. In order to understand how the data gets loaded into the FPGA registers it is important to If this is the case, the `valueVector` of `ODEState` in the `SystemState` gets updated: the value at `in_address` gets replaced with `in_writedata`, which gets unpack'ed into the main data type used by the application.

Listing 2.8 State machine responsible for controlling the solver

```

1  (systemState ', oul ', block')
2  --Handle the setup (reset the state, insert input values, start the computation)
3  | i_c == 2 && i_cs == 1 && i_ca == 0 = ( initialSystemState , 0, 0)
4  | i_is == 1 = ( systemState{ odestate = s_odestate_in' }, 0, 1)
5  | i_c == 1 && i_cs == 1 && i_ca == 0 = ( systemState{ step = 0 } , 0, 0)
6
7  --Handle the computation and output:
8  | block == 1 && i_os == 1 = ( systemState , pack (xs !! i_oa), block)
9  | block == 0 && s_step < c_maxstep = ( systemState_up' , 0, block)
10 | block == 0 && s_step >= c_maxstep = ( systemState{ step = uIntMax}, pack uIntMax, 1)
11
12 --Default, do nothing
13 | otherwise = ( systemState , oul , block)
14 where
15   s_odestate_in' = s_odestate { valueVector = replace i_ia (unpack i_i :: Data) xs}
16

```

```

17      s_odestate_up  = scheme systemConstants equation s_odestate
18      valueVector_wt = replace 4 (time s_odestate_up) (valueVector s_odestate_up)
19      s_odestate_up' = s_odestate_up { valueVector = valueVector_wt }
20      s_step'        = s_step + 1
21
22      systemState_up' = systemState{ odestate = s_odestate_up', step = s_step'}

```

2.8 Solving the system and extracting values

After sending the command to the FGPA to start solving the ODE over the control channel, on every clock cycle the FPGA will update the ODEState and the step variable. The ODEState gets updated by applying an integration scheme to the equation, which results in a new vector of values and a new value for the time. The step variable gets incremented to indicate that another step has passed. At some point, the value of step will exceed the value of maxstep, from SystemConstants. Whenever this happens, the FPGA stops updating the ODEState and writes all ones to the output port. This indicates that the FPGA is done processing and the results are ready to be collected by the HPS.

The step of SystemState in listing 2.3 field takes a bit more explanation. Every time the integration scheme gets applied, the step variable gets incremented. At some point, the value of step becomes larger or equal than the maxstep field from the data type SystemConstants. Whenever this happens, the system blocks until you order it to start again by setting the value of step to 0. During the time that the system does not progress, the values of the system can be read. This is done by requesting access to an address from the HPS. This request gets processed by the bridge and the hardware on the FPGA side into a high value for the out_read input signal, accompanied by a valid address from the out_address input signal. The FPGA is then responsible for actually writing the requested value to the output channel, in which the out_address is directly equal to the index in the ValueVector.

Chapter 3

Results

3.1 Euler

One of the first tests for an ODE solver is whether it handles harmonic oscillations properly. The simple oscillations (shown in 3.1 with their analytical solutions) can be implemented as a matrix vector equation (3.2), which represents two uncoupled oscillations of different frequencies. Besides the matrix of constants, the solver also requires initial values. For the first oscillation the initial position is non-zero, for the second oscillation the initial velocity is non-zero.

$$\begin{aligned} x_0'' &= -x_0 \\ x_2'' &= -4x_2 \end{aligned} \tag{3.1}$$

$$\begin{aligned} x_0(t) &= 50 \cos(t) \\ x_2(t) &= 25 \sin(2t) \end{aligned}$$

$$\vec{x}' = \begin{bmatrix} 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & -4 & 0 \end{bmatrix} \vec{x} \quad \text{with} \quad \vec{x}(t=0) = \begin{bmatrix} 50 \\ 0 \\ 0 \\ 50 \end{bmatrix} \tag{3.2}$$

3.1.1 First oscillation - initial position

The plots belonging to the first oscillation (3.1) depict three variations on solving the system. Firstly, it contains the solution generated by the FPGA. Secondly, it contains the result of the exact same combination of equation and solver, but implemented in MATLAB. The only difference between the FPGA and MATLAB implementation is the number representation. Therefore, if the FPGA solution starts to diverge from the MATLAB solution the reason has to

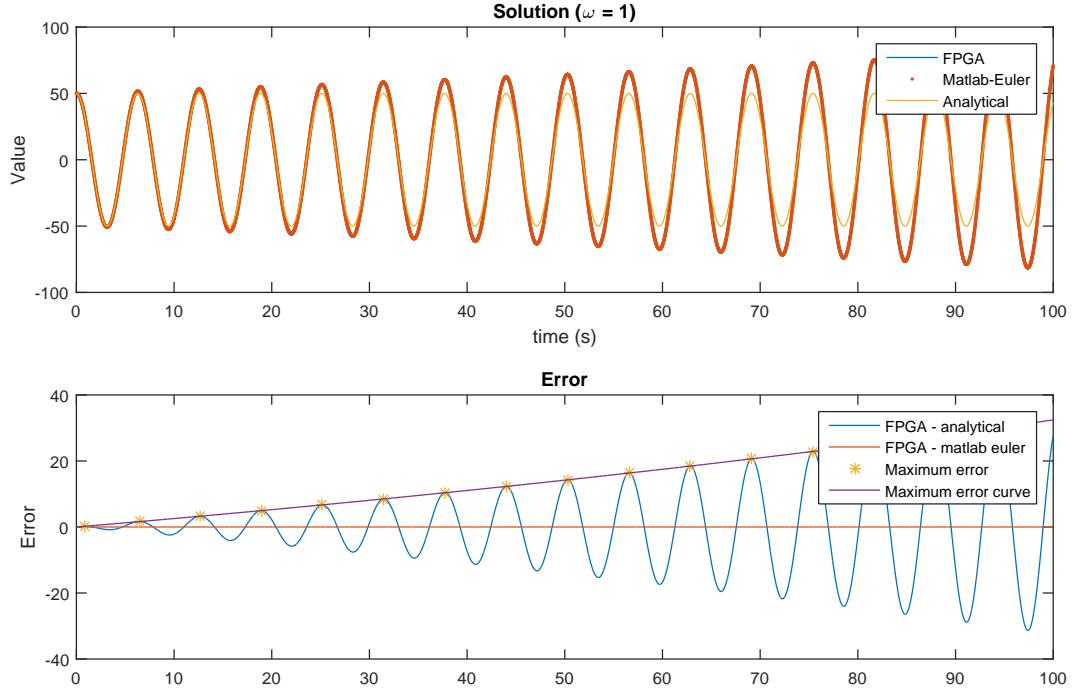


Fig. 3.1 Simple oscillation using Euler's method ($h = 0.01$)

be the reduced accuracy of the 32-bit fixed point representation used in the FPGA. Internally, MATLAB uses a (64-bit) double precision floating format [8], which is guaranteeing a precision of 15 significant figures for almost all magnitudes supported by the IEEE 754 double precision floating point standard. Lastly, the plot contains the analytical solution of the problem.

The plot shows that the FPGA exhibits the expected behaviour of solving a simple oscillation with Euler's method. Due to the relatively large step size ($h = 0.01$) the approximation quickly diverges from the analytical solution. The curvature of the solution is always opposite in sign to the solution itself ($x'' = -x$), which results in a self-amplifying effect in the magnitude of the error: an exponential error growth. This was expected, as the exponential dependency was already derived by [13] in equation 1.2. However, this does show that Euler's method is a particularly bad integration scheme for a simple oscillation. It is possible to use MATLAB's Curve Fitting Tool to fit the equation of the theoretical maximum error of Euler's method to the points of maximum error. After combining some constants, 1.2 becomes equal to 3.3. For $a = 49.75$ and $b = 0.00502$ this equation achieves a fit of $R^2 = 1$, which is a perfect fit. The error plot and the curve fit of the maximum errors are shown in 3.1.

$$\text{error}_{\text{euler}}(t) = a(e^{bt} - 1) \quad (3.3)$$

Lastly, the FPGA solution the solution of Euler's method implemented in MATLAB shows that the error due to the fixed point number representation is clearly insignificant when

compared to the intrinsic error in Euler's method: the maximum absolute difference between the two solutions is less than 6×10^{-5} .

Decreasing the time step - improving accuracy?

The expectation based on [13] and equation 1.2 is that the maximum error is indeed proportional to the time step, meaning that a hundred fold decrease of the time step also decreases the error by a factor 100. The maximum error at $t \approx 100$ for $h = 0.01$ was ≈ 30 , whereas the maximum error for $h = 0.0001$ is ≈ 0.2 , which is an improvement of $150 \times$, 50% more than expected. However, even though the error relative to the analytical solution has decreased, the divergence from the MATLAB implementation of Euler's algorithm has increased to 8×10^{-4} , a factor of 13.

As the time step gets decreased even more, eventually the improvement of having a shorter time step loses out to the reduction the accuracy in the computation. This can be seen very clearly in 3.3. Featuring a time step of $h = 1 \times 10^{-7}$, which is approaching the smallest representable value at 2^{-24} . In this case the MATLAB implementation is still following the analytical solution with little whereas the solution generated by the FPGA begins to diverge noticeably.

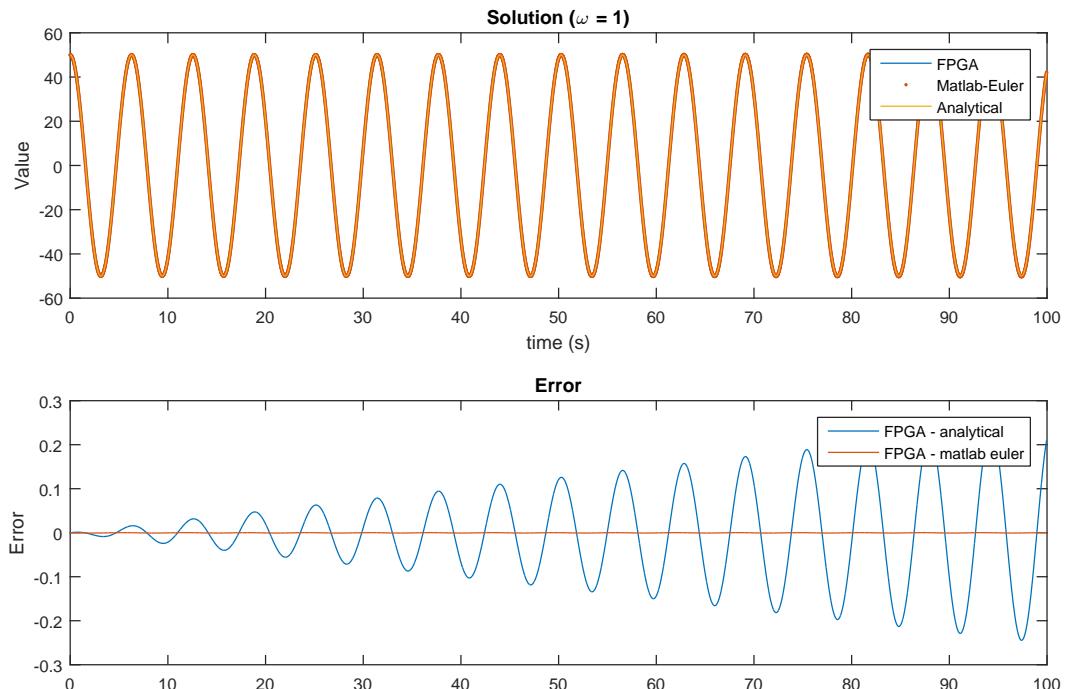


Fig. 3.2 Simple oscillation using Euler's method with a lower time step ($h = 1 \times 10^{-4}$)

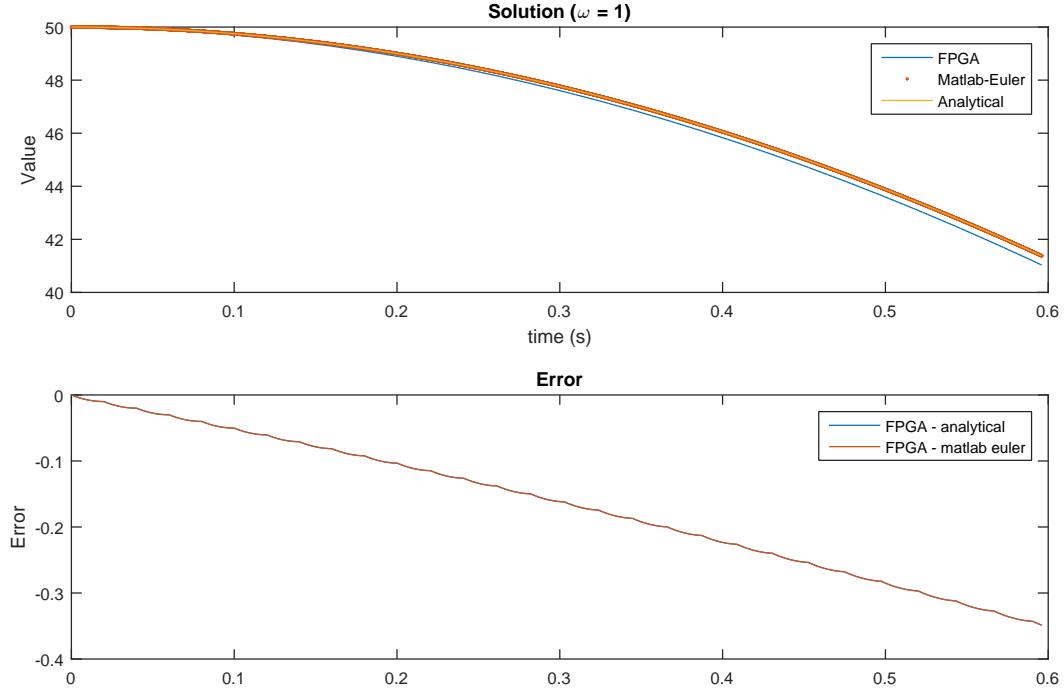


Fig. 3.3 Approaching breakdown due to the fixed point number representation ($h = 1 \times 10^{-7}$)

3.1.2 Second oscillation - initial velocity

Similarly to the previous scenario, the error increases exponentially over time and an decrease in time step is approximately proportional to the decrease in error (figure 3.5 and 3.6). However, there is an important difference between the two oscillations: the frequency of oscillation is higher. This has as result that the time step of $h = 0.01$ which worked fine for the oscillation with $\omega = 1$ does not work properly any more: it results in figure 3.4. This figure shows two interesting phenomena. Firstly, even though MATLABs Euler's method does diverge from the analytical solution, it only diverges in magnitude: it stays in phase. However, when looking at the FPGA solution you notice that it diverges from the other two, not only in magnitude but also in phase. The discrepancy in phase between MATLAB and FPGA indicates that the number representation is the culprit of the shift.

An attempt to fix this problem was made by changing the relevant part of the matrix from $\begin{bmatrix} 0 & 1 \\ -4 & 0 \end{bmatrix}$ to $\begin{bmatrix} 0 & 2 \\ -2 & 0 \end{bmatrix}$, which should have the result that the numbers remain smaller internally as the multiplication by 4 is distributed over two multiplications by 2 in separate vector dot products. Even though both matrices result in the same second order equation ($x'' = -4x$) and the eigenvalues are the same, the eigenvalues are different which results in different behaviour for the two matrices.

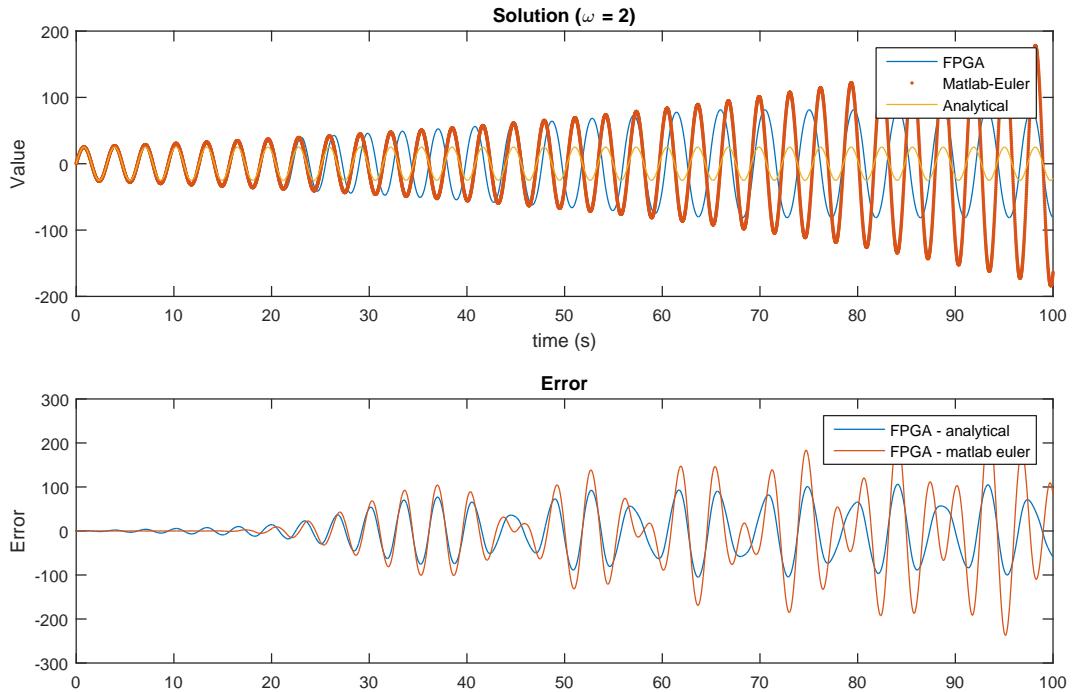


Fig. 3.4 Frequency shifting due to an insufficiently small time step ($h = 0.01$)

3.2 Runge-Kutta (second order)

The testing of the RK2 method uses a different matrix (equation 3.4) in order to verify the systems capability to correctly solve a system of 4, coupled, first order equations. The matrix has been generated randomly in MATLAB under the constraint that all eigenvalues are negative. The reason as to why this property is necessary is, again, based on the fixed point number representation. Whenever one of the values (which could even occur internally as part of a vector dot product, invisible to interface) exceeds the allowed range, the result becomes invalid. If all elements in the vector described by the ODE tend towards zero, this problem is less likely to occur (but it might still happen whenever the initial conditions that have been chosen are too large). Furthermore, the added computational steps of RK2 compared to Euler's method have the effect that the time step becomes less important - for the entire range of possible time step values the results and errors are approximately the same (shown in figure 3.7 and 3.8). Lastly, in contrast to the FPGA solution, MATLAB RK2 and ODE45 do remain close together which, once again, points in the direction of problems stemming from the fixed point numbers.

$$\vec{x}' = \begin{bmatrix} 2 & 3 & 2 & 0 \\ -5 & -5 & -3 & 1 \\ 3 & -1 & -2 & -3 \\ 4 & 2 & 2 & -3 \end{bmatrix} \vec{x} \quad \text{with} \quad \vec{x}(t=0) = \begin{bmatrix} 7 \\ 5 \\ 7 \\ 5 \end{bmatrix} \quad (3.4)$$

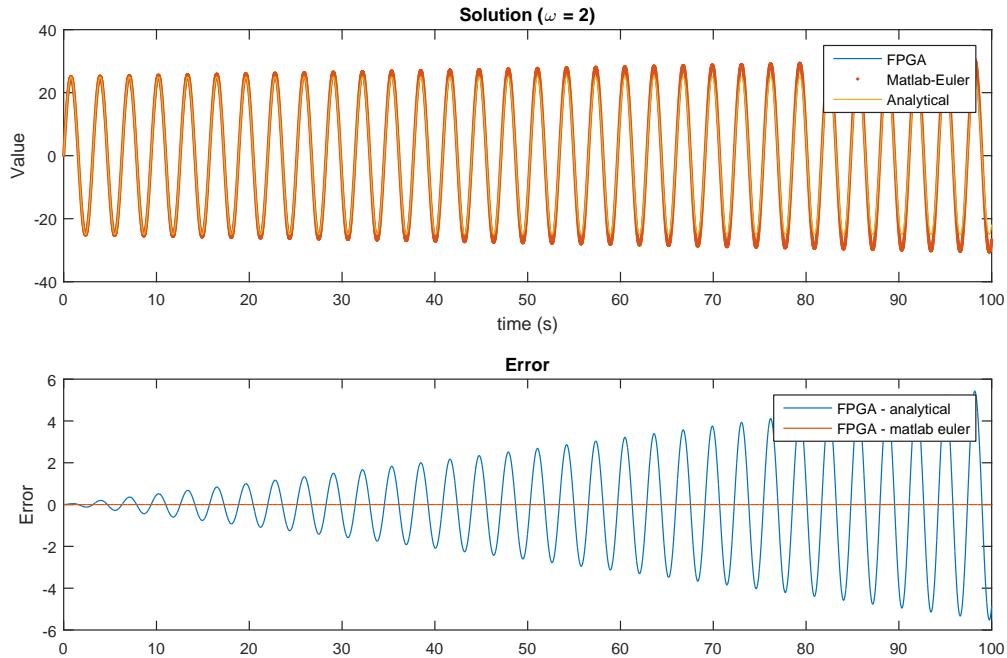


Fig. 3.5 Relatively large time steps result in large errors in the long run ($h=0.001$)

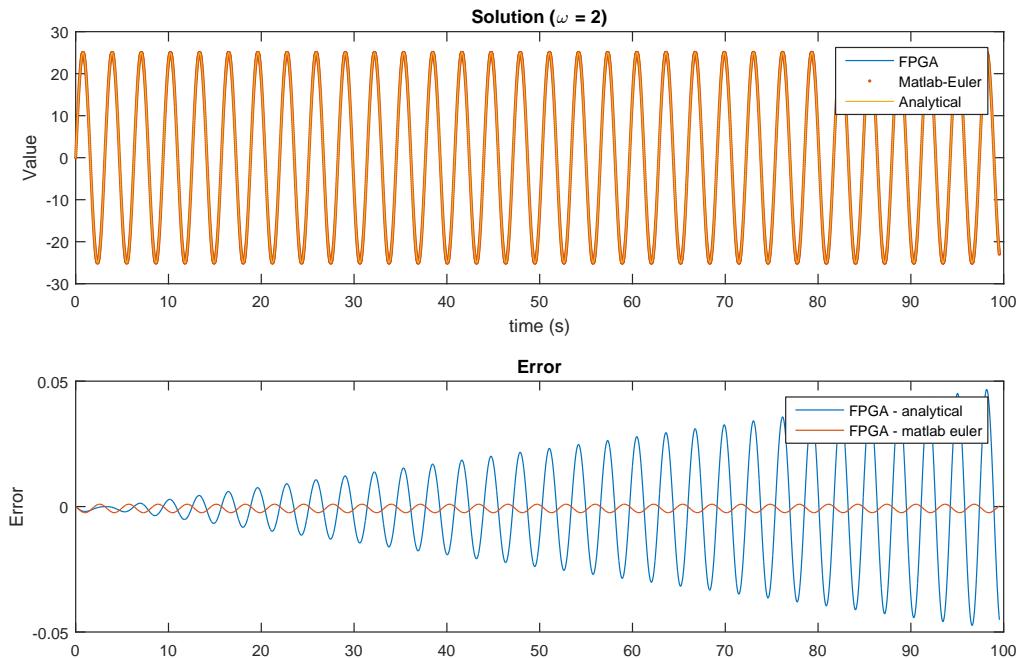


Fig. 3.6 But a decrease in time step goes a long way in reducing the error. Note that the error due to the fixed point number representation starts to become significant again, in contrast to figure 3.5 ($h = 1 \times 10^{-5}$)

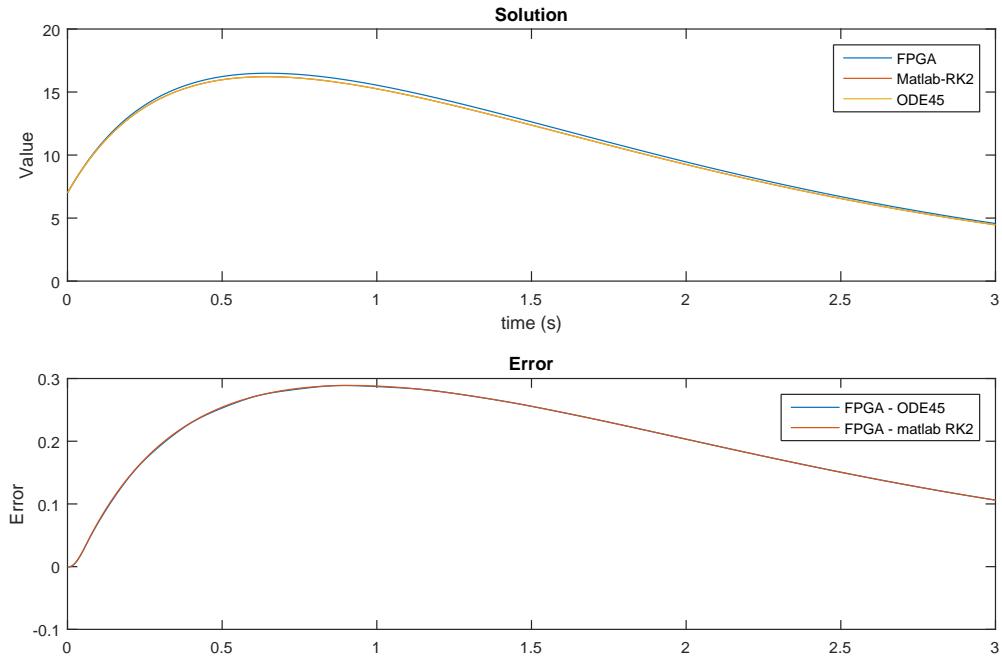


Fig. 3.7 The largest possible time step has a maximum error of approximately 0.3, for both the comparison with the solution of ODE45 and MATLABs RK2 ($h=0.01$).

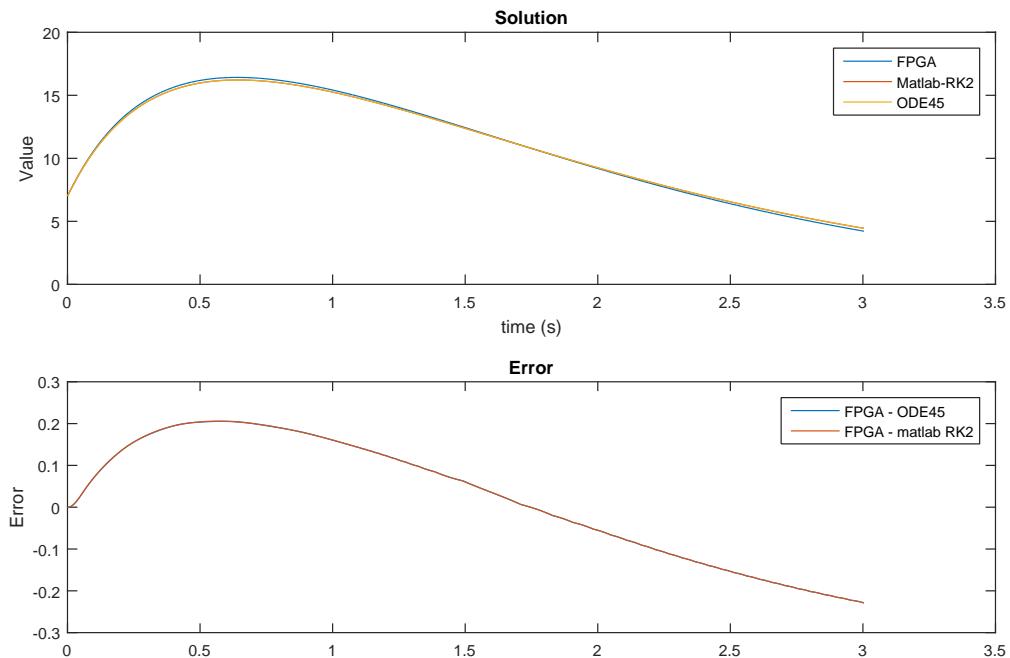


Fig. 3.8 Considering the other end of the spectrum regarding variety in time steps, the results do not differ significantly enough for the additional amount of work that has to be put in, a factor 20000 ($h=5 \times 10^{-7}$).

3.3 Runge-Kutta (fourth order)

The RK4 integration scheme ¹ introduces even more computational steps, which leads to a faster overflow of the internal numbers. This means that at least one of the derivatives overflows almost immediately, which results in plots as shown in figure 3.9.

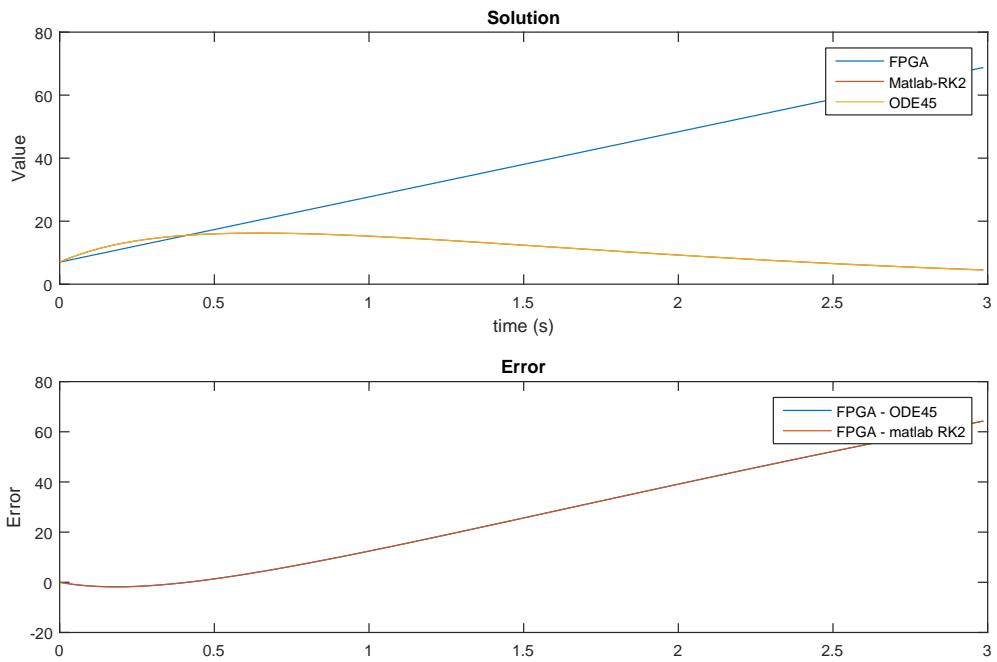


Fig. 3.9 For every value of the time step the internal number representations overflow, causing fixed values for derivatives and straight lines instead of solutions to ODEs ($h=1\times 10^{-5}$).

3.4 Euler revisited

So far it appears that accuracy is inversely proportional to the complexity of the integration scheme, which leads to the question: What accuracy would Euler's method attain for equation 3.4? As shown in figure 3.10, Euler's method only results in errors which are 2 orders of magnitude lower than the best values of the time step for RK2 and of course performs better than RK4 in this scenario.

3.5 Performance

Performance is one of the main reasons why people use FPGAs over implementations in software and therefore one would expect that a properly written FPGA solution outperforms a CPU.

¹Which took almost 22 hours or 78894 seconds to synthesize.

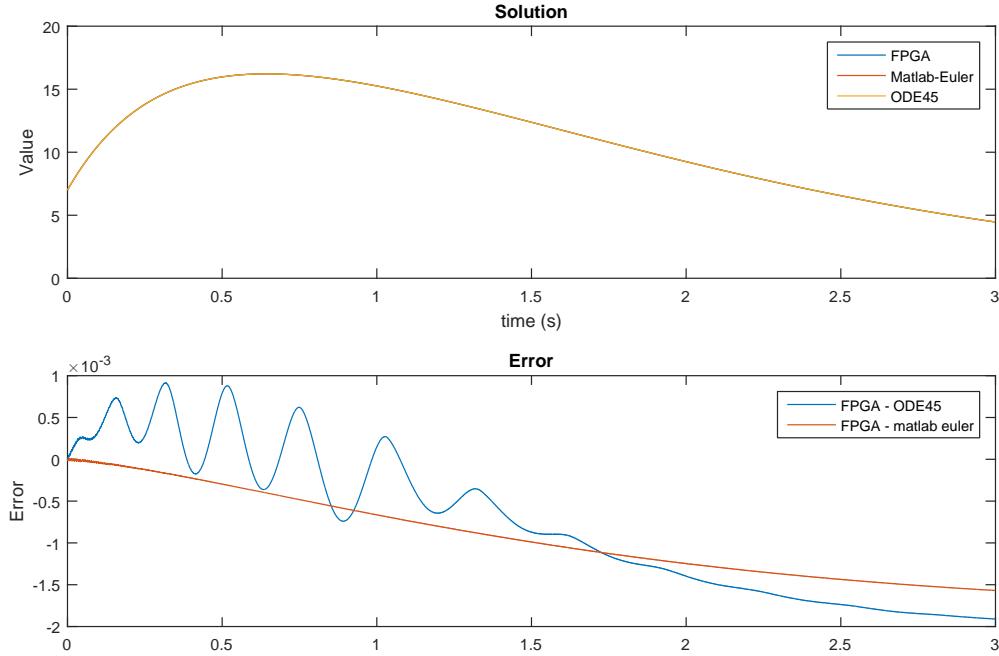


Fig. 3.10 Simplicity prevails: Euler's method attains better results than RK2 for all different time steps ($h=1\times 10^{-4}$).

The FPGA runs at a clock speed of 50 MHz and it is capable of executing a single iteration of Euler's method per clock cycle. In theory this would mean that the FPGA is capable of 50×10^6 iterations per second. However, there is still quite some overhead stemming from the system handling the data input and output, which has to wait for the HPS to supply or request the proper information before the FPGA can move on with the computations.

From table 3.1 it can be seen immediately that the difference between the theoretical and measured real-world performance is very small (the measurement indicates an 8% decrease in performance, but keep in mind that the time taken also includes writing the initial values to the FPGA, a process which is limited by the HPS). Furthermore, as expected, the performance of the FPGA starts to decrease as there are more and more events that require the HPS to read the output (output interrupts). Lastly, the most striking rows of this table are 'CPU - C++': for floating point, a single core of a stock i7-950 loses out to the FPGA by almost 2 orders of magnitude. However, when performing the computations on integers, a single CPU core outperforms the FPGA. This could be considered a more fair comparison as performing operations on fixed point numbers requires the same circuitry as operations on integers (fixed point numbers with 0 fractional bits).

More details on the testing set up and code which was used can be found in appendix E.

Table 3.1 Performance benchmarks: Euler's method, sorted by iterations per second.

Device	Iterations	time (s)	Output interrupts	Iterations per second ($\times 10^6$)
CPU - C++ - int	1×10^8	1,35	1	74,1
FPGA	1×10^8	2,18	1	45,8
FPGA	1×10^8	2,25	1×10^3	44,4
FPGA	1×10^8	2,30	1×10^4	43,4
FPGA	1×10^8	3,27	1×10^5	30,6
FPGA	1×10^7	0,39	1	25,6
FPGA	1×10^6	0,21	1	4,76
CPU - C++ - float	5×10^7	58,1	1	0,86
FPGA	1×10^5	0,19	1	0,53
CPU - Haskell - float	1×10^6	3,23	1	0,31
CPU - MATLAB - float	1×10^7	304	1	0,03

Chapter 4

Discussion

4.1 Accuracy

The results presented in this thesis are not highly accurate for all solvers and time steps used. The reason for this is two-fold. Firstly, the thesis is meant to show whether numerical mathematics is a topic which is feasible on an FPGA, which it demonstrably is. Secondly, the number representation used resulted in sub-par results, especially for higher-order integration schemes like RK4.

However, what has been shown (for certain values of the time step) is an equality between the same solver schemes, implemented in MATLAB and directly on the FPGA. Therefore, if more advanced schemes are used, together with a number representation which is suitable for a wider class of problems these could lead to results which are of high quality. A more suitable number representation would be double-precision floating point, which has as disadvantage that it is very expensive, computation-wise. Another option would be to simply extend the fixed-point representation, for instance to SFixed 24 40 or, if the FPGA allows, SFixed 32 96.

4.2 Performance

Table 3.1 shows that all FPGA implementations lose out to a modern CPU on performing the same operations. Does this mean that FPGAs are not useful in performing numerical mathematics? It certainly doesn't, as there are quite a few remarks to be made on this comparison. First and foremost - the FPGA used in this thesis is an Altera Cyclone 5. According to [4], this series is intended for "low-power, cost-sensitive design needs", whereas the Stratix series is designed for "high-performance". This suggests that use of a Stratix FPGA could clearly improve performance, either by supporting higher clock speeds, by having more hard adders and multipliers or by having more configurable logic elements. Furthermore, due to limitations in CλaSH, the FPGA design did not use a fractional PLL, only an integer frequency divider which was written manually. This meant that Quartus was unable to optimize the clock speed for the design. It may have been possible that the maximum clock speed the 4x4 matrix vector product (as used in Euler's method) supported

was higher than the 50 MHz used in testing, which would again, boost the FPGA performance. Additionally, in order to reduce compilation time for repeated testing, the optimizations in the Quartus compiler were disabled, in contrast to the C++ solution, which was produced with the highest optimization settings ($-O3$). Without optimizer, the C++ solution performed approximately 4x worse, which would have moved it to the third quartile of the tested FPGA solutions. Furthermore, according to the Quartus compiler, the design only took around 10% of the FPGA fabric. A quick, approximate calculation based on the number of operations necessary suggests that this FPGA would be capable of performing a 12×12 matrix vector multiplication using 90% of its area. This would only require a linear decrease in clock speed (a factor 3), whereas the matrix size has grown quadratically, which means that the CPU has to perform 9x as much work. Lastly, CPUs are heavily optimized for integer performance. This can also be seen from table 3.1: merely switching the data type from int to float results a decrease in performance of almost 2 orders of magnitude, which indicates that it could be very hard to compete with modern CPUs on integer arithmetic performance. The high performance of CPUs comes at a cost: the total power associated with running a single CPU core at 100% will be considerably higher [1] than the maximum power associated with this FPGA: 1.8W [3].

4.3 Suggestions for further work

This thesis has shown the very basics of solving ODEs on FPGAs. A lot of time has been spent on 'engineering challenges', which leaves some topics for future works:

- *Including block-ram as data storage* - This allows for more data, larger matrices and possibly even approximations of PDEs, but it could face problems of performance: on larger FPGAs, all data is needed in every cycle whereas the block-ram interfaces have limited bandwidth and a delay of a single cycle per read or write.
- *Implementations of more advanced integration schemes* - Adaptive (variable-step) and multi-step methods. These should probably be combined with a different number representation because as shown, for 32-bit fixed point numbers the best implemented method was still Euler's method.
- *An entirely different application* - Improvement of the IO system. The current IO system is versatile and does not only work for solving ODEs. By changing the C λ aSH code and modifying the host code slightly, the functionality can be flexible: quick computation of hashes or a hardware-implemented random-number generator are some of the possibilities.

4.4 Suggestions for additions to C λ aSH

- *Floating point numbers* - When an extremely wide fixed point number representation is not possible due to FPGA constraints, floating point arithmetic could be a solution. Furthermore, considering that numerical mathematics is often implemented using floating-point arithmetic (e.g. MATLAB) it would be important to be able to use C λ aSH to implement MATLAB algorithms with the same results.

- *More sample projects* - Short code samples can be found in the clash – compiler repository and the Hackage documentation is very complete, but getting started on complex projects can still be hard. This report has been written - both as a BSc thesis and an overview guide on using CλaSH, attempting to cover the process from the beginning: 'What is functional programming' to the end: 'Deploying your design on an FPGA', as my contribution to CλaSH.
- *More flexibility for PLL clocks* - Using a single PLL for the CλaSH design works great, but when the CλaSH design is merely a module of a greater project, things start to get hairy. Due to some non-deterministic behaviour of the Altera's own PLL (Altera PLL in the Quartus IP Library), a frequency divider was implemented manually in VHDL, which was capable of driving both the IO system as well as the CλaSH design, but as the PLL was implemented manually, Quartus could not optimize the design for an optimal frequency which could have resulted in some performance loss. As an addition to CλaSH it would be very helpful if it would be possible to feed the output clock and the locked signal, generated by the PLL module, directly into the CλaSH module.

Chapter 5

Conclusion

Despite the current limitations in accuracy it is definitely feasible to perform the process of solving ODEs directly on hardware, by programming an FPGA. The measured advantage of performing such computations on an FPGA would be the reduced power usage, however, the use of high-performance FPGAs will allow for an increase in performance which could outperform a CPU. Furthermore, an additional advantage of the FPGA is the variable accuracy, which allows for hardware-level performance on non-standard (more accurate) number representations whereas a CPU would have to emulate such operations in software.

Previously it was very hard to write programs which could leverage the performance characteristics of an FPGA without having knowledge of hardware design and/or HDL, but now C λ aSH enables a high-level specification of the hardware which is similar to mathematics: Haskell. The already set-up IO system for exchanging data (appendix C) and a toolchain which is entirely integrated (appendix D) ensure an even smoother experience in moving your design from mathematics onto hardware.

Despite the tremendous advantages offered by this work flow, there are some disadvantages. Compared to an implementation in MATLAB, which can be interpreted, the process of synthesis and deployment takes quite a bit longer due to the step of compiling the C λ aSH-generated HDL into a binary file which can be flashed to the FPGA. This process has been measured to take anywhere between 10 minutes (Euler's method) and 22 hours (RK4). Furthermore, including the additional development time of writing your algorithms in Haskell over MATLAB (mainly caused by the large advantage MATLAB has in terms of size of its standard library for numerical mathematics) results in development and synthesis times which are very much longer for FPGA-solutions.

References

- [1] Intel® core™ i7-950 processor (8m cache, 3.06 ghz, 4.80 gt/s intel® qpi). http://ark.intel.com/products/37150/Intel-Core-i7-950-Processor-8M-Cache-3_06-GHz-4_80-GTs-Intel-QPI, 2009. [Accessed: June 2015].
- [2] Chart: A library for generating 2d charts and plots. <https://hackage.haskell.org/package/Chart>, 2014. [Accessed: June 2015].
- [3] Cyclone v fpgas & socs. <https://www.altera.com/products/fpga/cyclone-series/cyclone-v/overview.html>, 2015. [Accessed: June 2015].
- [4] Altera > products > fpgas > overview. <https://www.altera.com/products/fpga/overview.html>, 2015. [Accessed: June 2015].
- [5] Computer language benchmarks game. <http://benchmarksgame.alioth.debian.org/u32q/which-programs-are-fastest.html>, 2015. [Accessed: June 2015].
- [6] clash-prelude-0.8.1 - clash.tutorial. <http://hackage.haskell.org/package/clash-prelude-0.8.1/docs/CLaSH-Tutorial.html>, 2015. [Accessed: June 2015].
- [7] Mining hardware comparison. https://en.bitcoin.it/wiki/Mining_hardware_comparison, 2015. [Accessed: June 2015].
- [8] Mathworks: Floating-point numbers. http://nl.mathworks.com/help/matlab/matlab_prog/floating-point-numbers.html, 2015. [Accessed: June 2015].
- [9] Sockit - the development kit for new soc device. <http://www.terasic.com.tw/cgi-bin/page/archive.pl?Language=English&CategoryNo=&No=816>, 2015. [Accessed: June 2015].
- [10] Christiaan Baaij. *Cλash: From Haskell To Hardware*. 2009. URL <http://essay.utwente.nl/59482/>.
- [11] Christiaan Baaij. Clash fpga starter. <http://christiaanb.github.io/posts/clash-fpga-starter/>, 2015. [Accessed: June 2015].
- [12] J Kuper. *A Short Introduction to Functional Programming*. 2015.
- [13] J Polking, A Boggess, and D Arnold. *Differential Equations with Boundary Value Problems*. Prentice Hall, Upper Saddle River, New Jersey, 2006.

- [14] Keith Underwood. Fpgas vs. cpus: Trends in peak floating-point performance. *Association for Computing Machinery*, 2004. URL <http://home.engineering.iastate.edu/~zambreno/classes/cpre583/2006/documents/Und04A.pdf>.
- [15] Wayne Wolf. *FPGA-Based System Design*. Prentice Hall, Upper Saddle River, New Jersey, 2004.

Appendix A

Haskell source code for numerical solutions of ODEs

Listing A.1 SolverTypes.hs - Defining the types that get used in the rest of the application. This is a good place to start reading if you want to understand the entire system.

```
1 module SolverTypes where
2
3   import Prelude
4
5   type NumRepr = Float
6   type D_ODEState = [NumRepr]
7
8   data ODEState = ODEState { xs :: [NumRepr]
9     , t :: NumRepr
10    } deriving (Show)
11
12  data TimeSettings = TimeSettings { dt :: NumRepr
13    , tMax :: NumRepr
14    } deriving (Show)
15
16  type SubFunction = (NumRepr -> NumRepr)
17
18  type Equation = ODEState -> D_ODEState
19  type Scheme = TimeSettings -> Equation -> ODEState -> ODEState
20  type Solver = Scheme -> TimeSettings -> Equation -> ODEState -> [ODEState]
```

Listing A.2 Solver.hs - The main calling code responsible for starting and stopping the simulation of the proper equations

```
1
2 module Solver where
3
4   import Prelude
5   import SolverTypes
6
7   import SolverEquations
8   import SolverSolvers
```

```

9  import SolverPlotter
10 import SolverPresets
11
12 ----- CALLERS
13 -- general form, stop after a certain time
14 solve :: Solver
15 solve solvemethod time equation initState = states
16 where
17     states = take steps $ iterate (solvemethod time equation) initState
18     steps = ceiling $ (tMax time - t initState) / dt time
19
20 sol_start = solve rk4 initTimeSettings
21
22 solution_expo = sol_start eq_exponential initODEState
23 solution_sine = sol_start eq_sine initODEState
24 solution_cosh = solve rk4 initTimeSettings2 eq_cosh initODEState2
25 solution_homo = sol_start (eq_linear_homo_const sinematrix) initODEState2
26 solution_hetr = sol_start (eq_linear_hetr_const sinematrix funcvec) initODEState2
27
28 testPlot = plotSolutions [s1,s2,s3,s4,s5] "Haskell solver examples"
29 where
30     s1 = (solution_expo, "Exponential")
31     s2 = (solution_sine, "Sine")
32     s3 = (solution_cosh, "Cosh")
33     s4 = (solution_homo, "Matrix form – homogenous")
34     s5 = (solution_hetr, "Matrix form – heterogenous")

```

Listing A.3 SolverEquations.hs - Definitions of the equations

```

1 module SolverEquations where
2
3 import Prelude
4 import SolverTypes
5 import SolverPresets
6
7 -- Exponential : y = A * exp(-t)
8 -- y' = -y
9
10 -- x0' = -x0
11 eq_exponential :: Equation
12 eq_exponential state = [-x !! 0]
13 where
14     x = xs state
15
16
17
18 -- Sine : y = A * sin(omega * t)
19 -- y'' = -y
20
21 -- x0' = x1
22 -- x1' = -x0
23 eq_sine :: Equation
24 eq_sine state = [x0,x1]

```

```

25  where
26    x = xs state
27    x0 = x !! 1
28    x1 = - (x !! 0)
29
30
31
32  -- Hyperbolic cosine : y = a*cosh((x - b)/a)
33  -- y' = sqrt(y^2 - a^2)/a
34
35  -- x0' = sqrt(x0^2 - a^2)/a
36  eq_cosh :: Equation
37  eq_cosh state = [x0]
38  where
39    x = xs state
40    x0 = sqrt((x !! 0)^2 - a^2)/a
41    a = 0.99
42
43  -- Arbitrary homogenous system
44  -- y' = Ay
45  eq_linear_homo_const :: [[NumRepr]] -> ODEState -> D_ODEState
46  eq_linear_homo_const matrix state = map (rowmult y) matrix
47  where
48    y = xs state
49
50  rowmult :: [NumRepr] -> [NumRepr] -> NumRepr
51  rowmult vec1 vec2 = sum $ zipWith (*) vec1 vec2
52
53  -- Arbitrary heterogenous system
54  -- y' = Ay + F
55  eq_linear_hetr_const :: [[NumRepr]] -> [SubFunction] -> ODEState -> D_ODEState
56  eq_linear_hetr_const matrix vector state = zipWith (+) (map (rowmult y) matrix) (map ($time) vector)
57  where
58    y = xs state
59    time = t state

```

Listing A.4 SolverPresets.hs - The definitions of initial values time settings auxiliary vectors and other necessities

```

1  module SolverPresets where
2
3  import Prelude
4  import SolverTypes
5
6  unity :: [[NumRepr]]
7  unity = [[1,0,0],[0,1,0],[0,0,1]]
8
9  sinematrix :: [[NumRepr]]
10 sinematrix = [[0,1],[-1,0]];
11
12 vec :: [NumRepr]
13 vec = [4,3,2]
14

```

```

15  funcvec :: [SubFunction]
16  funcvec = [(\t -> sin t),(\t -> exp (-t))]
17
18  initODEState = ODEState [10, 0.0] 0.0
19  initODEState2 = ODEState [1, -1] 0.0
20
21  initTimeSettings = TimeSettings 0.0001 10
22  initTimeSettings2 = TimeSettings 0.0001 3

```

Listing A.5 SolverHelper.hs - An auxiliary function used in the RK4 implementation

```

1 module SolverHelper where
2
3 import Prelude
4 import SolverTypes
5
6 sumLists :: [[NumRepr]] -> [NumRepr] -> [NumRepr]
7 sumLists [] factors = []
8 sumLists (xs:[]) factors = map ((head factors)*) xs
9 sumLists (xs:xss) factors = zipWith (+) (map (head factors)*) xs) (sumLists xss (tail factors))

```

Listing A.6 SolverPlotter.hs - The part responsible for actually creating the plot. This part can be omitted but the result of the program will be a very long list of ODEState's

```

1 module SolverPlotter where
2
3 import Prelude
4 import SolverTypes
5 import GHC.Float
6
7 import Graphics.Rendering.Chart.Easy
8 import Graphics.Rendering.Chart.Backend.Cairo
9
10 outProps = fo_format .~ PDF $ def
11
12 plotSolutions :: [(ODEState, String)] -> String -> String -> IO()
13 plotSolutions solutions title filename = toFile outProps filename $ do
14     layout_title .= title
15     layout_x_axis . laxis_title .= "time"
16     layout_y_axis . laxis_title .= "x"
17     plotSolutions_help solutions
18
19
20 plotSolutions_help [] = error "empty list"
21 plotSolutions_help [sol] = plotSolution sol
22 plotSolutions_help (sol:sols) = do
23     plotSolution sol
24     plotSolutions_help sols
25
26
27 plotSolution (solution, curveTitle) = plot $ line curveTitle [states]
28 where
29     states = reformData solution

```

```
30
31
32 reformData :: [ODEState] -> [(Double, Double)]
33 reformData states = map reformState states
34
35 reformState :: ODEState -> (Double, Double)
36 reformState state = (float2Double tVal, float2Double (x !! 0))
37   where
38     x = xs state
39     tVal = t state
```


Appendix B

Project structure

The entire project is hosted on GitHub - <https://github.com/Gladdy/numerical-fpga-thesis>

```
benchmark ..... Performance comparisons to C++ and Haskell
  c++
  haskell
clash
  vhdl
    Solver ..... The CλaSH-generated VHDL
control ..... C++ for controlling the FPGA from the HPS
haskell
images ..... Ready-made numerical solver images for the Cyclone V FPGA
kernel ..... Bootloader and scripts for linux kernel generation
literature
minified ..... Sample project for easy modification
  clash
  control
  sockit
  minified.zip
sockit ..... Full quartus project
thesis
verification ..... MATLAB result verification scripts
```


Appendix C

Handling data IO

Due to the library code written, the process of controlling the FPGA could not be more simple. Listing C.1 shows the user-side interface, in which the comments have made any further explanation obsolete. The background library written handles all low-level processes, for instance writing directly to a memory location and converting the doubles to fixed-point numbers. Comments have been added to explain the code in the important sections. The code is very object-oriented, improving clarity and comprehensibility. The sources can be found in the GitHub repository at <https://github.com/Gladdy/numerical-fpga-thesis>

Listing C.1 Controlling the FPGA

```
1 #include "fpgacontroller.h"
2 #include "fixedpointconverter.h"
3
4 int main () {
5     FPGAController fpga;
6     double mat [4][4] = {
7         {0, 1, 0, 0}
8         , {-1, 0, 0, 0}
9         , {0, 0, 0, 1}
10        , {0, 0, -4, 0}
11    };
12    // Set the constants
13    fpga.control.writeFP (1,125);      // maxtime
14    fpga.control.writeFP (2,0.00001);  // timestep
15    fpga.control.write (3,1000);       // outputstep
16    fpga.loadMatrix (mat);
17
18    // Set the initial values
19    fpga.input.writeFP (0,50);
20    fpga.input.writeFP (1,0);
21    fpga.input.writeFP (2,0);
22    fpga.input.writeFP (3,50);
23
24    // Fetch the solutions
25    fpga.iterate (1000,4);
26 }
```


Appendix D

Toolchain integration

A single script has integrated the entire toolchain, from the process of generating HDL in CλaSH to deploying and running the program on the FPGA. This script has several dependencies, so make sure that all these commands are available in your \$PATH.

- *GnuWin32* - The script is written for the bash shell and has been tested using the GNUWin32 implementation. It also depends on sed, ssh, scp, rm and cp, mv, which are not available by default on Windows.
- *Quartus* - For compiling the VHDL into a .sof (SDRAM Object File) and converting this into a .rbf (Raw Binary File), used to flash the FPGA.
- *A linux installation running on the SoC* Make sure that the hostname and the port for SSH access are specified properly and the board is running a correct version of Linux. The scripts which automate the set-up of a proper Linux image are also located in the repository, in the folder ./kernel.

Listing D.1 Full integration of the toolchain

```
1 #!/bin/bash -e
2 clear
3
4 HOSTNAME=84.85.97.221
5 PORT=10022
6
7 # Invocation :
8 # bash run.sh [clash synthesis upload run all]
9 # select one of the commands: either a single section or run everything ("all")
10
11
12 #
13 # CλaSH
14 #
15 # invoke the clash compiler
16 #
17 if [[ $1 == clash || $1 == all ]]; then
18   cd clash
19   rm -rf vhdl
20   clash --vhdl Solver.hs
```

```

21   cd ..
22 fi
23
24
25
26
27 #
28 # Quartus
29 #
30 # move over the generated files to the quartus directory
31 # patch the compute_main vhdl file
32 # invoke the quartus compiler to produce a sof
33 #
34 if [[ $1 == synthesis || $1 == all ]]; then
35   cd sockit
36   rm -rf hdl/*.vhdl # clear all clash-generated vhdl files (the framework is called .hdl)
37   cp ../*clash/vhdl/Solver/* hdl
38   sed -i {s/" signal system1000"/"--signal system1000"/g} hdl/compute_main.vhdl
39
40 #fix up the qsf with all clash-generated files
41 #remove all vhdl files from the qsf
42 cat sockit.qsf | grep -v VHDL_FILE > sockit_removedvhdl.qsf
43 mv sockit_removedvhdl.qsf sockit.qsf
44
45 #preprocess the filenames to be added to the qsf
46 ls hdl | grep .vhd > files.txt
47 sed -i 's/^\ set_global_assignment -name VHDL_FILE hdl//' files.txt
48
49 #add the clash-generated vhdl files to the qsf
50 echo "" >> sockit.qsf
51 cat files.txt >> sockit.qsf
52 rm files.txt
53
54 #start compilation
55 quartus_sh --flow compile sockit
56 quartus_cpf -c output_files / sockit.sof output_files / sockit.rbf
57 cd ..
58 fi
59
60
61
62
63 #
64 # FPGA
65 #
66 # Upload the binary program to the SoC/FPGA over SCP
67 # Attempt to rebuild the controlling program if possible
68 # If this is not possible on your Windows machine, find an Unix machine
69 # and install the arm-crosscompile version of g++.
70 # Build the controlling executable on that machine and deploy it.
71 # A deployment script can be found in "control/deploy.sh"
72 #

```

```

73 if [[ $1 == upload || $1 == all ]]; then
74   cd sockit
75   scp -P $PORT output_files/ sockit .rbf root@$HOSTNAME:~
76   cd ..
77
78   cd control
79
80   if [[ "command -v arm-linux-gnueabihf-g++" == "" || "command -v make" == "" ]]; then
81     echo "Cross compiling is currently not supported on Windows."
82     echo "Please build the executable on a system with 'arm-linux-gnueabihf-g++' and 'make' installed"
83   else
84     echo "Make'ing and uploading"
85     make
86     scp -P $PORT fpgacontroller root@$HOSTNAME:~
87     scp -P $PORT programFPGA.sh root@$HOSTNAME:~
88   fi
89
90
91   cd ..
92 fi
93
94 if [[ $1 == run || $1 == all ]]; then
95   ssh root@$HOSTNAME -p $PORT 'chmod +x programFPGA.sh;
96                           ~/programFPGA.sh sockit.rbf
97                           '
98
99   start_time ='date +%s%N'
100  ssh root@$HOSTNAME -p $PORT '~/fpgacontroller > output.txt'
101  end_time='date +%s%N'
102
103  scp -P $PORT root@$HOSTNAME:output.txt verification/output.txt
104  du -h verification /output .txt
105  tail verification /output .txt
106 fi
107
108
109
110
111 echo ""
112 timepassed=$((end_time - start_time ))
113 echo "Total execution time: $((timepassed/1000000))ms"

```


Appendix E

Performance benchmark

Both code samples have been compiled with the highest optimization settings (-O3)

Listing E.1 Matrix multiplication in C++

```
1 int main () {
2     const unsigned maxTime = 50000000;
3     const unsigned dt = 1;
4     unsigned t = 0;
5
6     int xs [4] = {1,2,3,4};
7     int mat [4][4] = {
8         { 2,    3,    2,    0}
9         , { -5,   -5,   -3,    1}
10        , { 3,   -1,   -2,   -3}
11        , { 4,    2,    2,   -3}
12    };
13
14     while(true) {
15
16         //Perform the matrix–vector multiplication
17         int xs_new [4];
18         for(unsigned i = 0; i<4; i++) {
19             xs_new[i] = xs[1]*mat[i ][1] + xs[1]*mat[i ][1] + xs[2]*mat[i ][2] + xs[3]*mat[i ][3];
20         }
21
22         //Euler's method, multiply with the timestep and add
23         for(unsigned i = 0; i<4; i++) {
24             xs[i] += dt*xs_new[i];
25         }
26
27         t += dt;
28
29         if(t >= maxTime) {
30             std :: cout << "done" << std :: endl;
31             return 0;
32         }
33     }
34 }
```

Listing E.2 Matrix multiplication in Haskell

```
1 module Main where
2
3 import Data.Matrix
4
5 mat = fromLists [
6     [2.0,    3.0,   2.0,  0]
7     , [-5.0,  -5.0, -3.0, 1]
8     , [3.0,    -1.0, -2.0, -3.0]
9     , [4.0,    2.0,   2.0,  -3.0]
10    ]
11
12 initialState = fromLists [
13     [1.0]
14     , [2.0]
15     , [3.0]
16     , [4.0]
17    ]
18
19 mvecProduct h m v = scaleMatrix h $ multStd m v
20
21 main = do print ((iterate (mvecProduct 0.01 mat) initialState ) !! 1000000)
```

Due to memory limitations (stack overflows) in the naive matrix multiplication test in Haskell, the limit has been set to 1×10^6 , which was still possible. This should not matter for the final benchmark results, as those compare iterations per second.