

Krzysztof Rymski

**The Legend of Bluespec: A Link to the
components**

Computer Science Tripos – Part II

Churchill College

May 9, 2022

Declaration of originality

I, Krzysztof Rymki of Churchill College, being a candidate for Part II of the Computer Science Tripos, hereby declare that this dissertation and the work described in it are my own work, unaided except as may be specified below, and that the dissertation does not contain material that has already been used to any substantial extent for a comparable purpose. I am content for my dissertation to be made available to the students and staff of the University.

Signed Krzysztof Rymki

Date May 9, 2022

Proforma

Name: **Krzysztof Rymski**
College: **Churchill College**
Project Title: **The Legend of Bluespec: A Link to the components**
Examination: **Computer Science Tripos – Part II, July 2022**
Word Count: **10793 (TODO)¹**
Line Count: **10793 (TODO)¹**
Project Originator: **Jonathan Woodruff**
Supervisor: **Jonathan Woodruff**

Original Aims of the Project

Create a tool to build top-level hardware modules. Comprised of connected high-level Bluespec components via direct connections and buses. The main interface to this tool was to be a text format that is more concise than `.qsys`, produced by Intel's Quartus Prime. There was also an optional goal of creating a GUI. Evaluation would qualitatively compare the usability of my tool against Intel's Platform designer, as well as quantitatively in terms of the number of tokens and lines against Intel's text format.

Work Completed

I created a Python tool that extracts data from packages using `bluetcl` a Tcl interface to the compiler of Bluespec. It can perform operations on types and modules found in packages, allowing for inference and correctness checking of user input. I developed an interface that can synthesize top-level modules from concise JSONs. Finally, I created a web-based GUI using the React-Flow library that allows users to define a graph of connections between modules and features many types of autocompletion based on type relationships. This tool can produce Bluespec code representing a designed top-level module.

Special Difficulties

None.

¹This word count was computed. by `detex diss.tex | tr -cd '0-9A-Za-z \n' | wc -w`

Contents

1	Introduction	6
1.1	Motivation	7
1.2	Already existing tools	7
1.3	Proposed solution	8
2	Preparation	9
2.1	Understanding Bluespec	9
2.1.1	Rules	9
2.1.2	Modules and interfaces	10
2.1.3	Typeclasses	10
2.1.4	Provisos	10
2.2	Choosing tools	11
2.2.1	Haskell	11
2.2.2	Tcl (pronounced "tickle")	12
2.2.3	Python	12
2.2.4	Postmortem	12
2.3	Creating a grammar	13
2.3.1	What grammar is needed ?	13
2.3.2	Where to find this grammar?	13
2.3.3	Technical aspects of the grammar	13
3	Implementation	15
3.1	Crawling the packages	15
3.2	Loading Bluespec packages' information	16
3.2.1	Lexical and syntactic analysis	16
3.2.2	Understanding abstract syntax tree (AST)	16
3.2.3	Organizing data and functionality	17
3.3	Building the top level module	17
3.3.1	Instantiating new modules	18
3.3.2	Connections	18
3.3.3	Buses	18
3.3.4	Specifying exported interfaces and methods	19
3.4	Resolving types	19
3.5	Synthesizing the module	22
3.5.1	JSON interface	23
3.6	GUI	26
3.6.1	Backend	27

3.6.2	Frontend (React)	27
3.6.3	Frontend (React Flow)	27
3.6.4	Frontend (MUI)	27
3.6.5	Functionality	28
4	Evaluation	33
4.1	Overview	33
4.1.1	Example 1: Deadly simple FIFOs	33
4.1.2	Example 2: Weight of the AXI4	33
4.1.3	Example 3: Simulation	33
4.2	Example 1: Deadly simple FIFOs	34
4.2.1	Platform Designer	34
4.2.2	The Legend of Bluespec: A Link to the components	35
4.2.3	Conclusion	36
4.3	Example 2: Weight of the AXI4	36
4.3.1	Platform Designer	36
4.3.2	The Legend of Bluespec: A Link to the components	37
4.3.3	Conclusion	38
4.3.4	Other differences	38
4.4	Example 3 - Working AXI4 masters and slaves	39
4.5	Numbers	39
4.6	Other things to consider	40
5	Conclusion	41
5.1	Results	41
5.2	Future work	41

Chapter 1

Introduction

Making hardware is hard, and similarly to other problems in computer science, we make it easier by adding layers of abstraction. There are many layers, and the diagram below shows a simplified overview of them. The layer this project focuses on is going from modules to integrated systems. In this layer, we already have code for modules like memory, CPU cores, interconnects etc. What we need is an ability to compose them in such a way that they create a whole device. My project will focus on doing that.

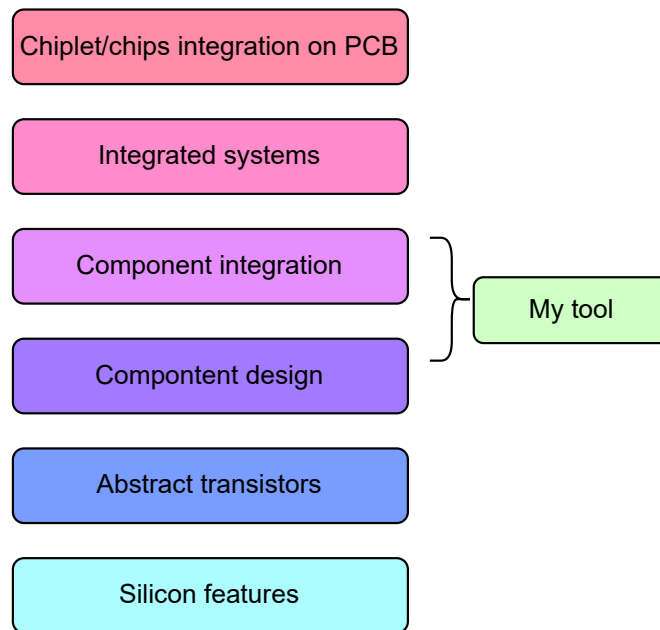


Figure 1.1: Stage one of the nodes used to instantiate new modules

1.1 Motivation

Vocabulary

- Verilog - Hardware description language
- Bluespec - Hardware description language
- Intel Quartus Prime (IQP) - programmable logic device design software produced by Intel. (source [Wikipedia](#))
- Platform Designer - Tool inside IQP used to connect high level components.

Verilog is an old language with a minimal type system. It forces people to encode more complex interfaces using finicky naming conventions. To simplify the process of connecting components with many wires, people tend to use high-level tools like Platform Designer, which has built-in logic allowing for easy connection of components. This approach is not perfect and as we will see during the evaluation, a better option might be to move away from Verilog and use a more modern language like Bluespec. Bluespec's type system allows us to detect connectable things in a systematic manner and allows us to define how to connect them. In fact, large bodies of Bluespec modules are already defined in this manner. Currently, there are no tools for automated system composition for native Bluespec modules. If we want to do this high level integration of Bluespec components then we need to transpile Bluespec code into Verilog to make them importable to tools like IQP. Unfortunately, because of the typing richness of Bluespec (comparable to Haskell), arbitrary code cannot be transpiled, and before code can be transpiled, it needs to be stripped of any arguments or other kinds of polymorphism. This process of transpilation also destroys high-level interfaces, and they must be recreated manually after importing Verilog code into IQP.

1.2 Already existing tools

During evaluation of this I will compare my tool with Platform designer from IQP. Platform Designer is a GUI tool for high-level integration, designs generated by it are kept in verbose text file format. Unfortunately, those files tend to be megabytes long for larger projects, making them hard to read and edit, for humans. There are other problems with using Platform Designer, like the cumbersomeness of importing new components.

Market share and justification for focusing entirely on comparisons with Intel Quartus Prime

It is a bit difficult to accurately judge the market share of the IQP, but it is produced by one of the biggest designers and manufacturers in the world, and it is also used by Qualcomm, according to [HG Insights](#). Using Google trends we can see that in recent years (since 2016) competitor Xilinx Vivado has overtaken IQP in interest at the global scale, in most developed countries like the US, Europe, and parts of Asia. There is roughly a 50/50 split between IQP and Vivado. My personal observations suggest that Xilinx Vivado became highly popular in India which is a large country with a history of chip design, and this might be strongly skewing data in favor of Xilinx's product. I chose to compare against IQP because we learned to use it in our course.

1.3 Proposed solution

During my project, I created a system for finding connectable things and supplying some other data for high level GUI applications. There are two main ways of interacting with this system. The first one uses the JSON files describing the systems. Those JSONs are simpler and more human-readable than comparable files used by IQP.

The reading of such JSONs is rather unimpressive on its own, and one might argue that it would be better to use Bluespec code instead. The interesting part of this project lies in the backend that can verify the correctness of this file and produce a set of typing information about produced Bluespec code.

But to truly appreciate the work I have done during this project, I also implemented GUI, which uses the same backend as the one used to interpret JSON files and is able to supply the user with typing information and possible choices when connecting components.

One can think of it as a Bluespec native high-level integration tool that uses a simple language server in the backend. This way Bluespec users can do high-level integration in the domain of untranspiled Bluespec code. Then after a satisfactory top-level module is created, it can be transpiled to Verilog and then imported into IQP.

Chapter 2

Preparation

This section will give you a basic understanding of the concepts found in Bluespec (most importantly *provisos* and *typeclasses*). It will explain my reasoning behind the choice of language I used to develop this project in. Finally, it will talk about grammar needed to parse information about types.

2.1 Understanding Bluespec

Before I started this project, my only knowledge about Bluespec came from [Cambridge Bluespec Tutor](#) which covers basics, and it was completed by me a few weeks before the start of this project. To understand this project, I have learned about most of the type system of Bluespec, but for you, the reader, most of it is not relevant. Therefore, I prepared a short tour of things that will be important to understand in this project.

2.1.1 Rules

In Bluespec, all computations are done in the form of rules. Each cycle we will take a subset of all the rules that we are going to execute this cycle. The rule is fired (executed) in the cycle only if it is ready (or will be ready) and it is not conflicting with other rules (If this happens, the compiler must issue a warning and pick an arbitrary rule to fire from a subset of conflicting rules). Each rule can fire at most one-time per cycle. For a rule to be ready to fire, it needs its implicit and explicit conditions to be true. A rule can be fired in a cycle even if it is not ready at the start of a cycle, for example, if you add an item to an empty queue and then pop can happen in the same cycle.

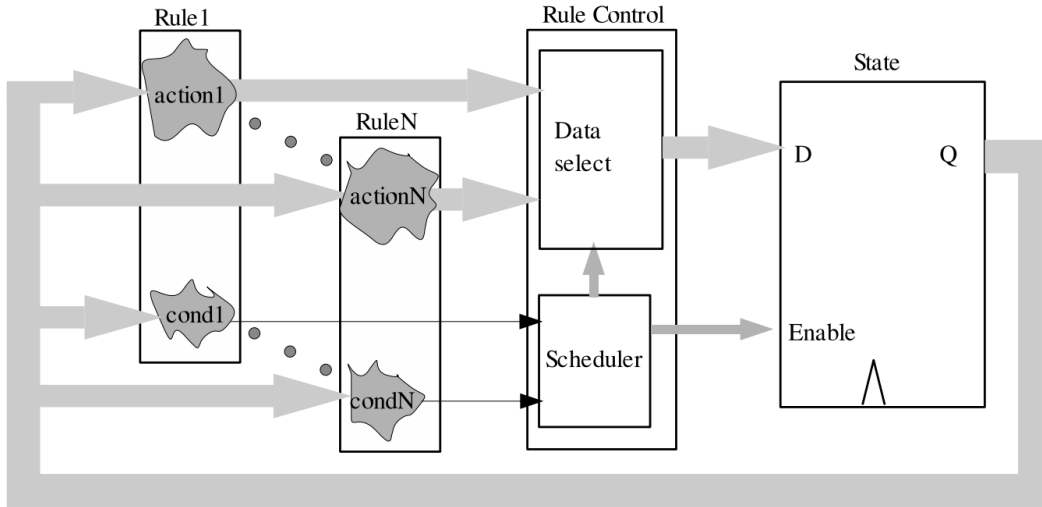


Figure 1: A general scheme for mapping an N-rule system into clocked synchronous hardware.

Image taken from Bluespec reference guide [1]

2.1.2 Modules and interfaces

Module is a unit of organization of logic. Modules don't have types, instead they implement some interfaces. This means that you can have multiple modules with different behaviors that implement the same interface. It might be helpful to think about it, as similar to different physical connectors (like USB-C, PCI-e). Interfaces are made out of two types of things:

- Methods - that allow for interaction with the module, they come with few variants, but this is not important for now.
- Subinterfaces - That allows for more generalization, for example if you want to implement a module that contains two copies of the same interface, you can just create a new interface that contains two instances of that subinterface and assign them accordingly.

2.1.3 Typeclasses

Typeclasses in Bluespec are used to group types for which specific functions or modules are implemented. The typeclass instance is (TODO describes what it is) and it can have provisos attached to it.

2.1.4 Provisos

One big feature of Bluespec typing is its support for provisos. They create constraints on types that can be used with a given function and give information about the expected type of the output of the function during compilation. For example, we can have a polymorphic function **extend** of a type $\text{Bit\#}(a) \rightarrow \text{Bit\#}(b)$ where $\text{Bit\#}(x)$ is a vector of x bits, such a function accepts any vector of any length a and always returns a vector of length $b = a + 1$. To make sure these relations stay like this, we would attach a proviso $\text{Add\#}(a, 1, b)$ which means that $a + 1$ must be equal to b .

Mentioned above **Add** is one of the Size Relationship provisos, other such provisos are **Mul Div Max Min Log**. Those provisos cannot be nested, but we still can express complex arithmetic constraints, thanks to size relationship type functions. They essentially represent the result of the operation on values that were supplied. There are only 8 of them, and they are **TAdd TSub TDiv TMul TMax TMin TLog TExp**. To better explain this, let's take a look at this simple example: **Add#(TExp#(a),TDiv#(a,8),c)** is a proviso that says that the following equation $2^a + \lceil \frac{a}{8} \rceil = c$ on parameters **a,b,c** must be satisfied. Otherwise, it will mean that the supplied parameters are not valid and the compilation won't succeed. These equations can also be solved to deduce unknown parameters.

There is also another type of provisos that I will call *typeclass based* ones. Typeclass based provisos allow us to either ensure certain functions are defined for a given type or to deduce something. For example, **Bits#(a,b)** means that for type **a**, there must be functions **pack** and **unpack** allowing for conversion to and from a vector of **b** bits. Moreover, this can be used to get the length in bits of given objects, on which we can later put constraints using Size Relationship provisos. In my project, I will often check if there exists an instance of the **Connectable#(a,b)** typeclass for a given pair of types **a** and **b** to check if function **mkConnection** (used to connect two interfaces together) is defined. All of this can be a bit confusing, but if one looks closely, the system of provisos and typeclasses is remarkably similar to Prolog, the programming language. We have variables that are effectively the same as in Prolog, we have values that are integers, and we have parametric types and functions that are like Prolog's compound terms. Typeclasses are like groups of rules with the same name, and provisos attached to Typeclass instances are like predicates attached to rules.

One significant difference is that provisos can be solved in arbitrary order, which means that unlike in the Prolog, we need the ability to solve sets of equations on integers to resolve them.

2.2 Choosing tools

I effectively had 3 choices for a language to write this project in. Here are some justifications for why I have chosen Python.

2.2.1 Haskell

The compiler of Bluespec is written in Haskell. Therefore, there are several reasons why Haskell would be the right choice. Here are some of them:

- I can reuse the already written logic.
- I will not diverge in terms of logic from the Bluespec compiler.
- I will not need to write grammar to load packages.

But there are also a few particularly good reasons why I did not use Haskell.

- I never used Haskell before, and my experience with functional languages is equal to what was shown in the foundations of the computer science course.
- I never used Bluespec before, and my experience is limited to completing the Bluespec tutor provided by the computer science department.

- Even though I knew Haskell and Bluespec, industrial compilers are rather large and complicated. I also received advice to avoid modifying/extending them as a part of the project.

While each of those reasons on its own is something that can be managed. All of them combined would create an unacceptable amount of risk of not delivering the project. There was also a chance that the use of Haskell would not create a good story for the project. (This is quite subtle).

2.2.2 Tcl (pronounced "tickle")

This language is used as a scripting language in both Intel and Xilinx tools, While I will be reading packages using Tcl scripts provided by the creators of the Bluespec compiler (BSC), my understanding is that those scripts are just handy wrappers for some Haskell functions. This is also a foreign language to me, with a minimal presence online and negligible learning resources, making it difficult to learn. It is also quite obsolete, and there is not much tooling for it making it undesirable as a time investment (As an example of the desirable to learn technology. I learned JavaScript for this project JavaScript just to create a GUI).

2.2.3 Python

Firstly, this is a language I have experience working with. Secondly, it is widely supported, and there is extensive tooling for it. Its flexible typing system allows for rapid experimenting, which was especially handy when I was not sure what the structure of the data would look like. It is also worth mentioning that Python has support for typing, and at some point, I decided to start adding typing hints around. My rule of thumb that I used was that if Pylance (Python language server) cannot deduce the type of something, I need more hints. Other reasons for using Python are:

- I will have many interesting problems to solve. Like creating and interpreting grammar.
- Code written in it can be used by a large community of python developers.
- I can use many tools like Lark (for grammar), Django (as a backend server), and SymPy (to solve size relationship provisos) that will speed up the process.

The two arguments against using Python are:

- Haskell would be better suited for this task. I explained above why it does not make sense for me to use Haskell.
- Python is usually slow. Fortunately, this is not a big issue thanks to caching and the use of PyPy (implementation of Python using JIT compiler).

2.2.4 Postmortem

Looking at those choices now (after finishing the project and knowing all encountered difficulties), I could have also considered C#, due to its speed and similar typing flexibility to Python. I also think Haskell would be a better choice if this were more like my pet project on which I would work for a few years to make it a commercially viable product.

2.3 Creating a grammar

Bluetcl

Bluetcl is a tool written in Tcl language that allows for inspection of Bluespec packages. It is a first-party tool that is included with the Bluespec compiler. There is not much documentation about it, but from my understanding, it is a wrapper around Haskell's code. From the user perspective, it is just a bash script that starts the Tcl terminal compiled with some libraries. From my perspective, I use it as an arbitrary command-line program, as this method of use is intended by the authors and contains some documentation.

2.3.1 What grammar is needed ?

Because I am using Bluetcl as a command-line tool, communication with it is done via text. While synthesizing commands is easy, parsing outputs is a bit more difficult, and I decided to do it systematically. To do this, I need the grammar of all possible outputs that I am interested in. Bluetcl produces a range of outputs, but two of them that are particularly useful are descriptions of functions and descriptions of types.

2.3.2 Where to find this grammar?

Unfortunately, this grammar is not documented anywhere, so I had to reverse engineer it. This approach might not be perfect and might not cover every input, but it is good enough to parse every standard package and all packages used to compile Flute (RISC-V CPU with 5 stage pipeline). Here are other reasons to justify this approach:

- Heaps' law suggests that the number of unique words in a given body of the text is proportional to approximately the square root of the number of words in the text. Therefore, it is fair to assume that something similar will be true if we consider the number of unique grammar rules.
- This grammar, while different from the grammar of Bluespec language, maps a subset of Bluespec grammar, so we can supplement our deductions with a reference guide for Bluespec language. Therefore, accounting for cases we have not seen, but are possible.
- It's also unlikely that exotic things will be needed as a part of top-level modules.

2.3.3 Technical aspects of the grammar

This is EBNF grammar, I parse it using Lark library for Python, and I am using Earley parser, as it is capable of arbitrary length lookahead. The grammar I created contains roughly 90 rules, and I will not include all of them here, but I will show an example to give a feel of what is happening. (TODO: Think about dumping them into an appendix)

A useful feature supported by Lark is the ability to have regular expressions in the grammar. I am mentioning this as it is effectively like having a parser inside a parser. It is also worth mentioning that there exists a handy tool for debugging and creating grammar. It is located at this website: <https://www.lark-parser.org/ide/> and it can run the parser online and show the created AST tree.

Example of a parsing rule and the resulting AST

Here is an example of parsing a small section of output that describes the location of a piece of code that is being described. I decided to show it as it's simple and shows different features that I needed to use.

----- Rules used -----

identifier_u: /[A-Z][\w\$'_]*/

tcl_position: "{" "position" "{" tcl_path NUMBER NUMBER
["{" "Library" identifier_u "]" "]" }

tcl_path: ["%/" /((\.{1,2})|(\w|-|\s)+)/
["/" /((\.\.|\.)|((\w|-|\s)+))/]* "." /\w+/"

----- Text to parse -----

{position {%/Libraries/Connectable.bs 25 1 {Library Connectable}}}

▼ tcl_position

▼ tcl_path

Libraries

Connectable

bs

25

1

▼ identifier_u

Connectable

Chapter 3

Implementation

Before I explain how top modules are built, I will first explain how Bluespec packages are crawled to extract typing information as it is essential for building top level modules.

3.1 Crawling the packages

Figure 3.1: Overview of process of crawling packages



To interact with Bluetcl, I have written a script using the Pexpect library. This script works by creating a subprocess of Bluetcl and exposes functions that allow for the performing of queries? The core of this script is a function called `call` that takes as input a string that is a command and returns the output stripped of warnings or raises an exception if an error occurred (for example: in the case where a package was not found). To remove warnings, I make some assumptions.

- I only use a fixed set of commands (like list packages, describe a type).
- For those commands the important output is always on the last line. (This was checked empirically)
- Output of command is always followed by `%` a character that never occurs in the rest of the output and marks the end of the output.

Thanks to them, I do not need a sophisticated system for parsing errors and or warnings. I can just do a quick search for the error and if needed forward this to the user.

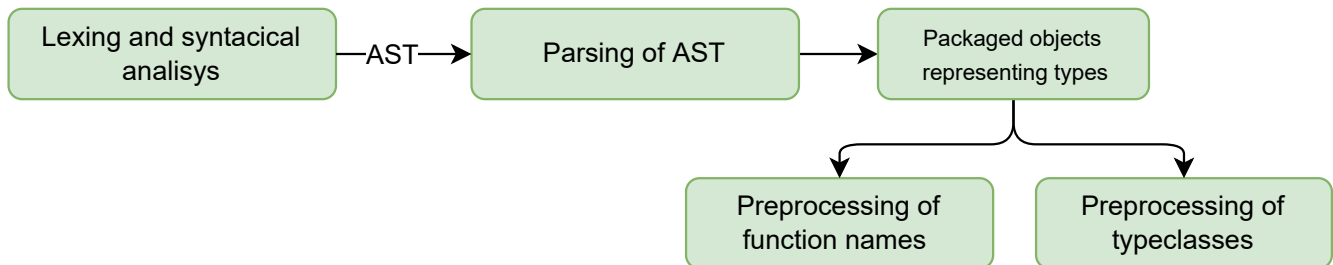
In total, this script allows me to:

- Initialize subprocess
- Add folder to search path of Bluetcl
- Load package (Bluetcl takes care of finding and loading dependencies)
- Get the list of loaded packages

- List functions in a package
- List types in a package
- Get information about types and functions in the package

3.2 Loading Bluespec packages' information

Figure 3.2: Overview of parsing bluetcl data



As mentioned earlier, the output of bluetcl is text, and I need it in some better-organized form. In this section, I describe the process of doing this.

3.2.1 Lexical and syntactic analysis

I decided to use the Lark library for lexical and syntactic analysis. This library takes grammar and strings of text as input and returns an abstract syntax tree. The only difficult part here is creating grammar, but we have taken care of it during preparation.

3.2.2 Understanding abstract syntax tree (AST)

Given AST I needed to turn it into a more useful form. To do this I used **Transformer** the class provided by Lark. The way it worked is that I created a subclass of **Transformer** that implement a function for each non-terminal / rule (Each non-terminal can have multiple rules and Lark allows for renaming of individual rules, for example, this allows parsing of different non-terminals using the same function or parsing single non-terminal using different functions depending on the rule that was used). **Transformer** will then perform a bottom-up transformation of AST, and for each rule, it will call the respective member function, with parsed children (non-terminals used by rule) as arguments. At the end of the transformation process, I have a list of objects that represent data from Bluetcl.

One of the bigger inconveniences I encounter is the fact that Lark passes parsed children, as a simple list. This does not provide information about what exactly each child is, this is not a problem if the rule contains a fixed number of non-terminals, but otherwise, I employ a mix of length-based inference, and wrapping of data from children into tuples, to remove ambiguity.

What was quite useful at this stage is Python's type system flexibility, I was able to gradually build up **Transformer** class and added typing information only after I had a firm understanding of what type something needs to be. For some things, it was not immediately obvious what is

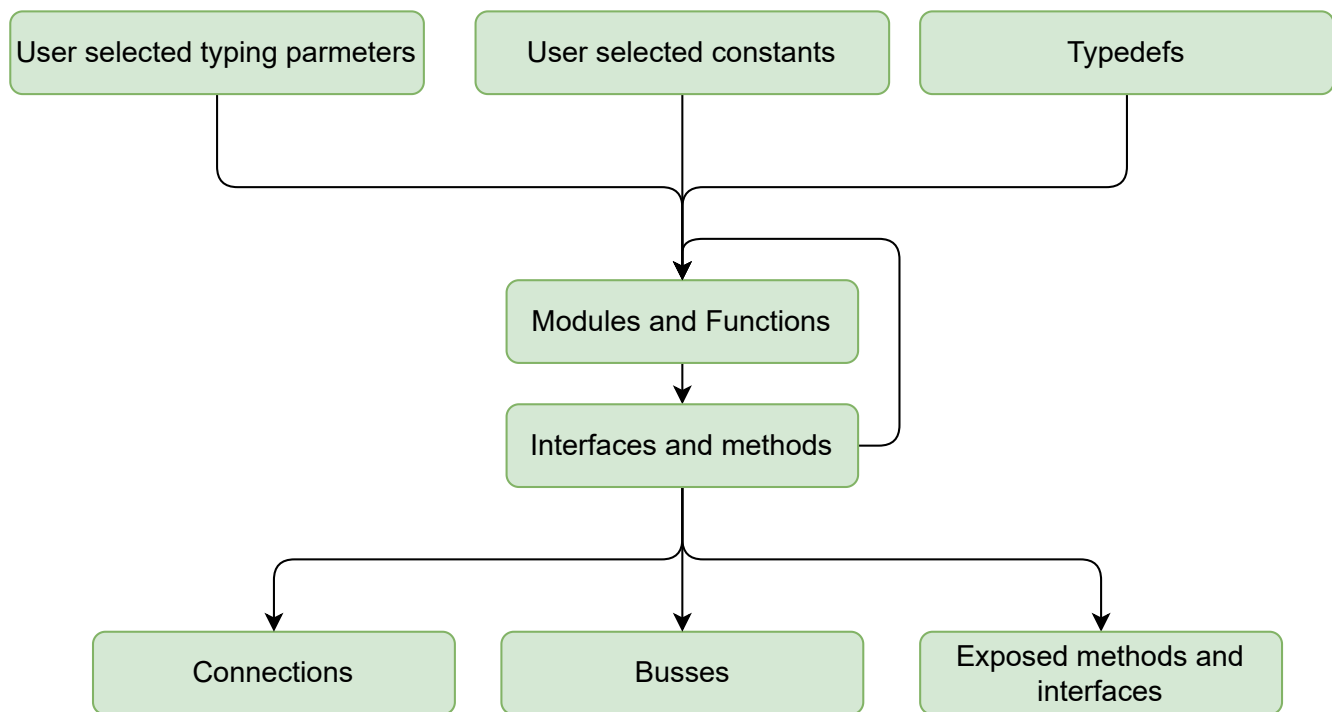
the best way to represent them, and what often happens is I would first leave them as lists of unparsed things. Then after I got to the point where I needed to use them, I would decide what is an optimal way to represent them.

3.2.3 Organizing data and functionality

I created a class called `TypeDatabase` that provides a set of functions that manage loading packages and parsing packages' information. It is also capable of loading and storing previously stashed state from a “pickle” file. This saves a lot of time as parsing and communication with `bluetcl` isn't particularly fast. This class is also home to a set of functionalities used to resolve types. Based on my testing, it took about 370s to load roughly 250 packages that I found included with the Bluespec compiler and created to compile Flute CPU. Loading those packages from cached files takes 5 to 20 seconds.

3.3 Building the top level module

Figure 3.3: Overview of the flow of data in the top level module



The above diagram shows the flow of data during the creation of the top-level module. Users can specify all nodes of this diagram with the exception of *Interfaces and methods* produced by *Modules and functions* as this is specified by typing information. Building a top-level module can be broken down into a few operations.

- Instantiating new modules.
- Connecting modules using `mkConnection` function.
- Connecting modules using buses, using functions of certain characteristics.

- Specifying exported interfaces and methods.

Operations of Instantiating new modules, Connecting and Creation of buses are quite similar, so I will start by describing instantiating new modules and then how it fits with other operations.

3.3.1 Instantiating new modules

To instantiate the new module, we need to do the following things:

1. Convert inputted by user strings to type identifiers. This includes converting the names of interfaces and their members to their type identifiers.
2. Checking if the arguments provided are of valid types.
3. Checking if typing information satisfies provisos.
4. Add newly created interfaces and methods to the namespace of known names.

From this list items 1 and 4 are implemented using a few dictionaries that keep track of known names and their types. I also have a system that keeps track of which instantiated things depend on which other instantiated things. Thanks to this system, I am able to create a graph of dependencies, which is used to order things in the right order. As for the things mentioned in points 2,3, they are completed using the later described functions for type resolution.

3.3.2 Connections

Adding a new connection is easy, as this can be done using the same logic as *instantiating new modules*. What is more interesting is that I keep track of all possible connections that can be made. This is later used to supply autocompletion in GUI (If the user uses JSON based interface, I can just print this data). Tracking of those connections is done in a naive way by simply checking all possible $O(n^2)$ connections. A question that arises is: "Can I do better?". While I do not have concrete proof of this, execution of programs written in Prolog can be reduced quite nicely to checking whether two types are connectable. Therefore the whole problem of finding all pairs of connectable things can be reduced to the following problem.

Given n *input values* (number of interfaces and methods in instantiated modules) and k *input programs* (number of instances of typeclass `Connectable`) find all pairs of two *input values* for which at least one *program* halts with the resulting value being true. Unfortunately, Prolog is a turing complete language, and it's hard to do much better than naive brute force. This sounds like a lot of work, but thanks to look-up dictionaries, usually I need to check only a few *programs*. I also spread out the work over subsequent additions of new modules and did not recompute already known results. This way, the user will perceive lag roughly proportional to the number of modules already instantiated. I also noticed that with a larger number of modules, PyPy's JIT compiler will step in and it will produce more optimized code, therefore keeping computation time low.

3.3.3 Buses

When it comes to buses, most logic goes to synthesizing code for them, and not figuring out whether given data is valid. Because of this I can fall back on logic designed to instantiate modules. To synthesize a bus, I need to synthesize a routing function. A routing function is a function that

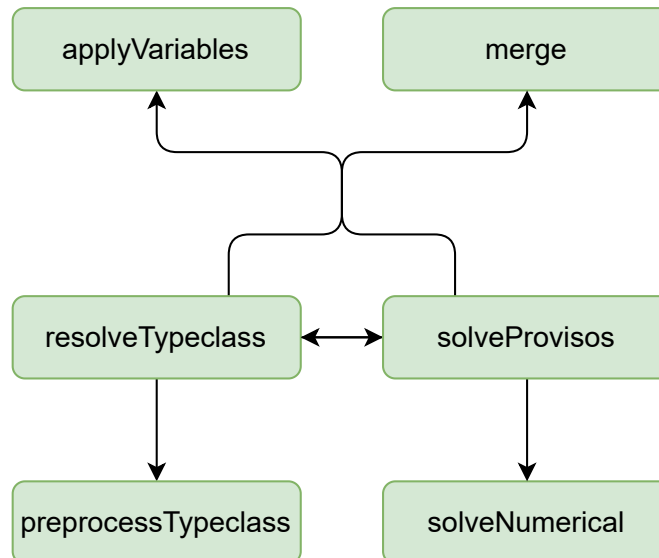
takes a route and produces a one-hot vector that specifies which slave is used for a given address. To create a routing function, two things must be created. First, I need to create a new function and add it to the database of known functions. To do this, I just reused my parsing tooling. As for creating a code for it, I generate a user template and fill it with `if` statements for each address region (I don't need to do fancier logic as in hardware, it will be optimized anyway). I also generate comments giving information about which slaves occupy which address region.

3.3.4 Specifying exported interfaces and methods

This functionality utilizes data generated previously, and most of the logic behind it focuses on supporting a few special cases that are meant to keep produced code clean.

3.4 Resolving types

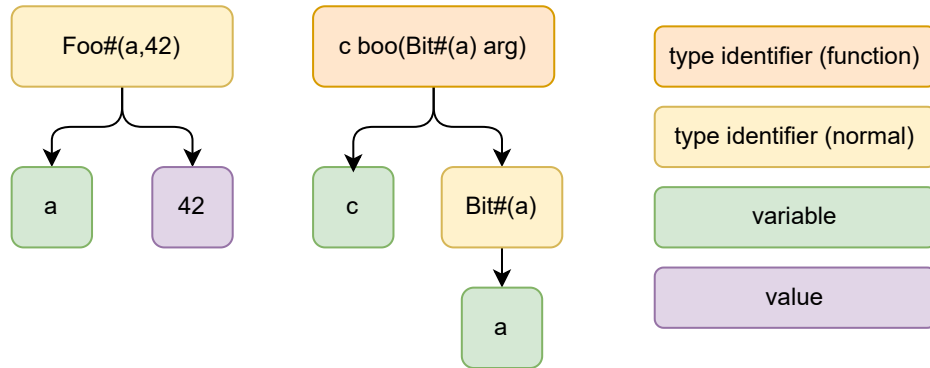
Figure 3.4: Dependencies between functions



This section assumes that one has read about typeclasses and provisos from [previous chapter](#). Functionality described here is what enables this whole project to do more than just string comparisons and formation.

Types of most things can be represented by a type identifier. Type identifiers can be thought of as a tree data structure where every non leaf node is another type identifier, and leaf nodes are values (integers) or variables. Example: type identifier `Foo#(a,42)` has a name `Foo` and two arguments: variable `a` and value `42`. Type identifiers can also describe functions. For example: `Bit#(b) boo(Bit#(a) arg)` where, `boo` is a function with one argument named `arg` (it is a type identified by a type identifier `Bit#(a)`), and it returns an object of some type `c`. In the tree representation it looks like this:

Figure 3.5: Break down of a type identifiers as a tree



In my code there are 4 main functions that, combined, allow for type resolution:

- **merge** function takes two type identifiers and returns a dictionary (map) from names of variables to their values. It's similar to Prolog's unification operator.
- **applyVariables** function takes a dictionary and a type identifier and returns a type identifier with all variables replaced by their values. This is more of a utility function.
- **resolveTypeclass** function takes a type identifier that is a proposed instance of a typeclass and returns a dictionary of learned variable values if the proposed instance is a valid one otherwise it raises an exception. In Prolog terms, this would be like looking at which rules can be applied and returning resolved variables from applying that rule.
- **solveProvisos** function takes a dictionary of variables and a list of provisos and returns a dictionary with variables learned by solving provisos or raises an exception if it's impossible to satisfy all the provisos. This is like checking predicates in Prolog.
- **preprocessTypeclass** function that takes a typeclass and produces a look-up dictionary of typeclass instances for a given group of names present in the instance of the typeclass.
- **solveNumerical** function that takes a list of size relationship provisos and returns a dictionary of learned variables.

How those functions can be used

Using all of those functions, I can answer many questions like:

- What known thing can be used as *nth* argument of a given function?
- Is it possible to call a given function with a given tuple of arguments ?
- What is the resulting type of function given certain arguments?

There are also more subtle uses of functions like **merge**, for example, when I need to deduce the exact types of members of certain interfaces.

Merge function

This function works like two synchronized instances of DFS algorithm that go through the trees (type identifiers) and apply logic based on nodes that are beginning and currently visited. This function needs to account for quite a lot of special cases as each node can be one of the following: variable, value (integer), type identifier (normal), and type identifier (function). Fortunately, half of the cases are symmetrical to the other half. In broad strokes, here is the logic behind those cases:

- If one of the nodes is a variable, and the other is a value or another type identifier, then we set the value of the variable to a value of the other.
- If both nodes are values, we check if they are equal.
- If both nodes have the same kind of type identifier, we recursively compare their children. (If they are not functions, we also compare their names)
- If both nodes are variables, then we need to check if it's possible to unify the values of those variables. If so, we assign the value of those variables to the result of unification.
- In other cases, like when we need to unify type identifiers with value, we raise an exception.

One minor problem that I encountered was that I had to keep track of which variables are equal, but I was able to avoid this problem altogether using shared pointers.

Apply variables function

This function is quite simple. It simply recursively traverses a type identifier and replaces all variables with their values.

Preprocess typeclass function

Earlier, I compared typeclasses to groups of Prolog rules with the same name, one thing I did not mention is that typeclasses can also contain something called dependencies, they are something like marking inputs and outputs in Prolog. Those dependencies are used in typeclasses like `Bits#(a,b)` which are used to determine something, and they make sure that the resolution of a typeclass is always unambiguous.

For example, typeclass `Bits#(a,b)` has a dependency `a determines b` which means that you can use it to determine `b` from `a`, but if you know only `b` then you cannot use it to determine `a` because there may be multiple types that can be packed to the bit vector of the same length `b`. An example of a determinable pair of `a` and `b` would be `a = Bit#(32)` from which we can determine that `b = 32`. It is also worth mentioning that we can have multiple tuples of dependencies. For example, we can have typeclass `Boo#(a,b,c,d)` with dependencies `(a,b,d) determines c` and `(a,d) determines (b,c)`. They mean that we need to know either variables `(a,b,c)` or `(a,d)`. From an implementation perspective, this is just extra complexity.

Now, I observed that most typeclass instances use non-function type identifiers, so I can create a look-up dictionary for every tuple of names of type identifiers in the instance. This way I find a set of typeclass instances that can be used to resolve a typeclass. I also need to account for instances that have functioned as one of their parameters as this introduces nasty polymorphism and I decided simply to put those instances in a special group called universal instances that are always checked in the resolve typeclass function.

Resolve typeclass function

This function in Prolog terms is like taking a predicate of a rule and then trying to check if it can be satisfied. This can be quite slow if there are many rules, but thanks to the lookup dictionary created during preprocessing. We can narrow down the set of rules to check. Then we can iterate over all typeclass instances (rules) and apply the following algorithm:

1. Merge candidate instance with one
2. If merging succeeded, attempt to solve provisos.
3. If provisos were successfully solved, return the dictionary of resolved variables, otherwise continue iteration.
4. If all iterations have failed, fail.

This procedure is a bit slow, so I speed it up by creating lookup dictionaries that take as a key a dictionary and as a value a typeclass instance. This method can cut down search time significantly, but there are still rules that use polymorphic functions, and those cannot be preprocessed away so easily. Therefore, I call them universal instances and search through them every time. I could invent many heuristics to speed up this process. In general, this is still as difficult as the halting problem.

Solve numerical function

This function takes a set of size relationship provisos and a dictionary of variables. Then it converts those provisos into a set of equations on natural numbers. If the set is unsolvable, this function raises an exception. Solving is done using the SymPy library. From an implementation standpoint, I just need to traverse the tree of type identifiers describing proviso and for each [name of proviso](#), I convert it into a mathematical expression. Exact expressions were taken from the Bluespec reference manual[1].

Solve provisos function

As established earlier, provisos can be of two types, one is based on typeclasses and those can be resolved using `resolveTypeclass`, and the other ones are size-based and require solving sets of equations on integers and this can be done using `solveNumerical` function. One big problem is the lack of order in which provisos must be solved. Therefore, I handle this by creating a set of provisos to be solved. Then I repeatedly iterate over this set and remove the solved provisos. If the set does not change after iteration, and it is not empty, this means no further progress can be made, and the function raises an exception.

3.5 Synthesizing the module

This is probably the simplest part of the project. I just iterate over various kinds of things defined by the user and run some string formatting code. Even for buses, thanks to polymorphism in the Bluespec, I can quite easily write a routing function.

The main goal of the project is to be able to take a JSON with some data and synthesize a top module in the Bluespec language from it. This goal is more like a justification to create a language

server like API. With this approach, I have some simple core goals, but in the later sections, we will see how thanks to all the work I have done, this API can be used to support a GUI that aids the user in creating a top module. Therefore, I will focus in this section on how the API works, and only later, on how it is used when reading JSON or as a backend to the GUI based app.

3.5.1 JSON interface

I have decided that my “human-readable format” will be JSON. This format is simple, and support built-in in many languages. The only problem with JSON is that it is more verbose than in some aspects Bluespec, and in theory, it would make more sense to use Bluespec. However, this could introduce user confusion as I am not able to support arbitrary Bluespec features in my backend. My understanding is that supporting something only partially could lead users to think that my tool is broken because it does not match their assumptions about what it can do. While it can be argued that if I had written this project as an extension of the Bluespec compiler, I could have supported all features of Bluespec (thanks to already written logic in the compiler). I explained, at the beginning of this document, why doing this was infeasible for me as a part II project. At the same time, using JSON allows me to support simplified semantics for things like bus creation and features not found in Bluespec like the specification of folders with additional libraries.

Structure of JSON file

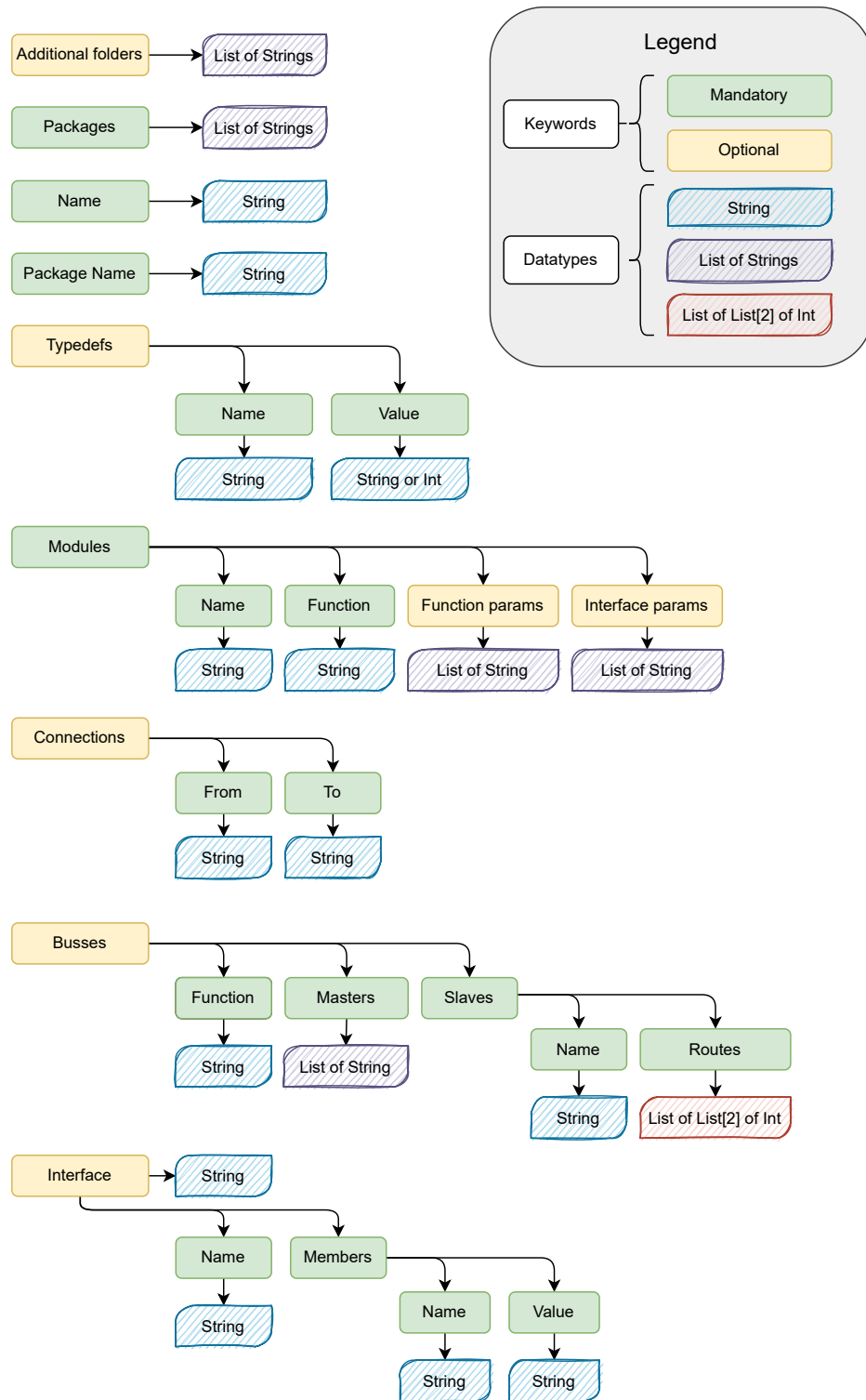


Figure 3.6: Disclaimer: In this diagram, I am using uppercase names but the JSON file uses lowercase ones also all spaces are replaced with underscores. I am also using s in the name to signify a list of dictionaries with those keywords

More detailed descriptions of keywords:

- `additional_folders` - list of folders that will be added to the Bluetcl search path.
- `packages` - list of packages that will be used
- `name` - name of the module (defaults to “top”)
- `package_name` - name of the created package
- `typedefs` - list of typedefs defined by a user
- `typedefs[i].name` - string starting with uppercase
- `typedefs[i].value` - integer or string
- `modules` - list of modules and functions to be instantiated
- `modules[i].name` - string being the name of the module (used later to access such module)
- `modules[i].function` - lowercase name of the function used to create the module
- `modules[i].func_params` - list of strings that will be parsed, this can include numbers, names of interfaces of previously declared modules, **tagged** `<Value>` where `<Value>` can be replaced with struct’s type identifier. This keyword is optional if the function does not take any arguments.
- `modules[i].interface_params` - parameters passed to the interface of the result of the function, this keyword is optional if the full type of the interface can be inferred from the type of function used to create a module, arguments, and provisos.
- `connections` - list of dictionaries containing definitions of connections
- `connections[i].from` - lowercase string with access path of a left side of the connection
- `connections[i].to` - lowercase string with access path of a right side of the connection
- `buses` - list of dictionaries describing buses
- `buses[i].function` - lowercase name of a function used to create a bus
- `buses[i].masters` - list of access paths of interfaces to use as masters
- `buses[i].slaves` - list of dictionaries describing slaves
- `buses[i].slaves[j].name` - access path of an interface used as slave
- `buses[i].slaves[j].routes` - list of lists length two of integers describing starts and ends of address ranges on which slaves will be accessible. Starts of ranges are inclusive, and are exclusive. (This data is used to generate routing function)
- `interface` - either a string describing an access path of an interface to be exposed by a module or a dictionary describing a more complex interface made out of multiple member interfaces and methods

- `interface.name` - name of a type of the interface, if one is not present in loaded packages a new interface will be synthesized
- `interface.members` - list of members methods and subinterfaces of this interface
- `interface.members[i].name` - name at member will be accessible
- `interface.members[i].value` - access path of a method or an interface to be exposed at given name

Using JSON interface

If all I did was simply synthesize a Bluespec file describing a package from JSON, it would be mostly (one would need additional data to synthesize a new interface) possible to do that just by doing simple string manipulation.

I know this because my initial iterations were doing this. However, my tool can do much more than that. To show those abilities, I print additional data and verify the correctness of the given JSON. Below is a rundown of this additional data, but to truly appreciate it, I would recommend looking at how GUI is using this data.

- Dictionary of all possible connections, better interface and subinterface and methods of instantiated modules.
- Dictionary of possible valid arguments for each function used to instantiate a module or a function.
- Dictionary of possible other masters and slaves for each instantiated bus.
- Inferred types of every instantiated module and types it is subinterfaces and member methods.

All this data can be useful when using complex modules from foreign packages, like a CPU core that does not have any direct connections, but has many subinterfaces that can be connected to things like memory, external devices, etc.

3.6 GUI

[Click here to jump to the overview of the GUI. \(Page 38\)](#)

Adding GUI was an optional goal mentioned in the project proposal. I originally intended to use a game engine, but the Bluespec compiler is a Linux tool and game engines like Unreal Engine 4 or Unity3d are optimized to be used on Windows, so I decided against that idea. Instead, I found a library called React Flow, that provided a simple API for creating graph-based user interfaces. Therefore, I decided to create a GUI as a website. This was quite an adventure as it was effectively my first contact with web development. (I have done backend in .NET and C# for a group project last year, but it was mostly focused on writing queries to the database and I had no contact with frontend development)

3.6.1 Backend

For the backend, I had a choice between Django and Flask (as those are the main python libraries for backend development). After reading about them in theory, Flask was supposed to be better for small projects, like this one. However, I decided to use Django because I wanted to learn techniques that will be more useful in the future. I followed an official tutorial on Django and after initial pain with security, adding new functionality was easy.

3.6.2 Frontend (React)

The frontend was written in React.js as you might have guessed from the name of the library I wanted to use. According to [Stack Overflow 2021 survey](#), React.js is the most popular frontend technology in 2021, so again I considered time spent on learning it to be a worthwhile investment. One thing that took me quite a bit to get used to is everything being a function (React.js supports classes, but it is an old paradigm and I wanted to learn to do things the “correct” way). This makes working with variables a bit tricky, as changes to state variables cause re-rendering of the whole component, and if a variable is not a state variable, then its value is going to be lost after re-rendering. If not careful, it is easy to cause a feedback loop causing infinite re-rendering and subsequent crash of the application.

3.6.3 Frontend (React Flow)

Vocabulary

Handle - Is a component of a node that is used to start or end a connection.

React Flow from a developer perspective requires defining nodes (they can be arbitrary React components), and some metadata about displaying a graph that is things like how edges look, how a user can add new nodes, etc. The exposed API is quite simplistic. For example, the position of a handle is constrained to the middle of the side of a node, and this library will generate CSS to put it there. Therefore, I needed to write some finicky CSS to align the handle with text displaying its name.

Another thing that is problematic about React Flow is that it will cause the node component to update every time it is moved. What I found is that if you wrap larger subcomponents of a node in `memo` from React, then performance stays quite good even if multiple nodes are moving, as React will cache rendered subcomponents.

On the bright side, after I figured out workarounds for those issues, I was left with many other things being handled by the library. This included automatic graph formatting and pretty Bézier curves for edges.

3.6.4 Frontend (MUI)

Default HTML buttons and text boxes are ugly. So, I decided to use the React UI component library called MUI (previously Material UI). I do not have much to say about it except that use of it was frictionless, at least in my humble opinion it made my application look modern. I am mentioning this as one of my complaints about Intel’s platform designer was that its UI is decades old.

3.6.5 Functionality

After teasing how I did things, let us talk about the features of my GUI. Using this GUI, a user can do the following things:

- Instantiate new module.
- Connect two interfaces.
- Create a bus.
- Select methods and interfaces to be exported as part of a created module.
- Look at the created Bluespec file.
- Run compilation followed the simulation and looked at the output of both.

To do all those things, the user will need to interact with the graph building tool. Below is an overview of the components users can choose from that will describe the top-level module.

Instance node

This node is most important as by using it the user can instantiate new modules, convert between interfaces, and use functions in general. It can be in 3 main stages. During the first one, it is an empty vessel, using which one can specify its name and function/module that will be used. In the second stage, a user specifies inputs and typing information. Finally, after clicking *confirm update* node transitions into stage three and the interfaces and methods of a created object become usable.

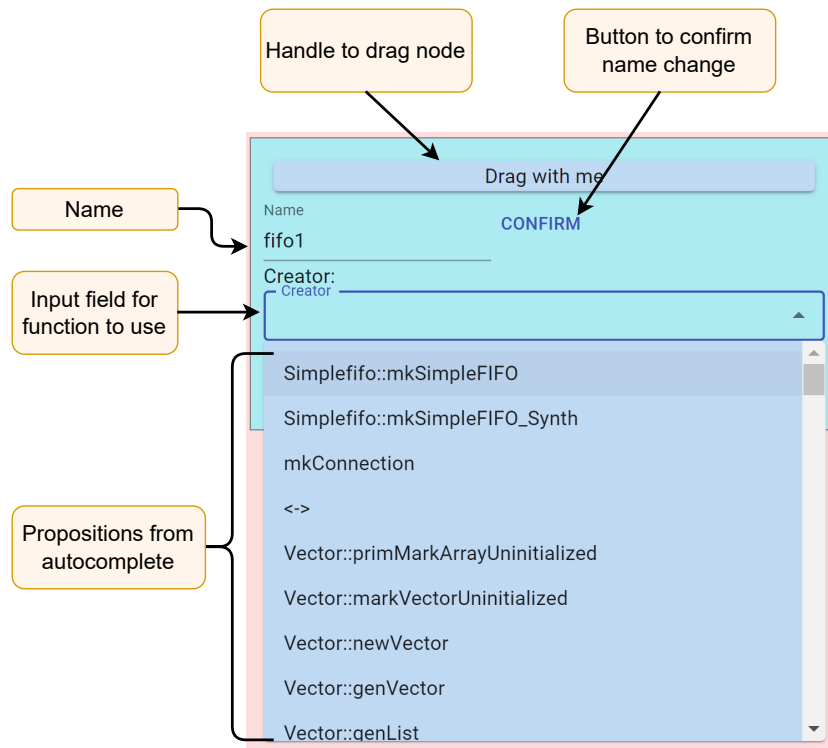


Figure 3.7: Stage one of the nodes used to instantiate new modules

At stage one we will see the first field with autocomplete, it is the simplest one and simply lists all the functions/modules users can use. It contains stand-alone functions as well as functions found as part of typeclasses. The node will transition to stage two when one of those is picked. At stage two, the user can specify typing information. To do this, they are presented with a tree-like input structure. Values inputted there can be of either uppercase string describing type identifier, integers or lowercase strings that will be used as a variable that is meant to be resolved by the backend. There is also a list of inputs that go directly to the function. These inputs contain fields with autocomplete that will filter all typedefs found in loaded packages, interfaces and functions found in previously initialized nodes. Some handles allow for the drag and drop style of creating connections. If the user chooses a value via a text field that is a member of a previously initialized node, then the connection will be generated automatically. After a user is satisfied with their choices, they press the *confirm update* button. At stage three, the user is presented *information about the error or lack of such* this bar can be in one of three states. Its functionality is shared between nodes.

- Waiting for confirmation - When the backend is processing data.
- Instantiation successful - if everything went smoothly.
- Error with text that tries to describe an error that occurred. This usually occurs after the user-supplied data is not able to satisfy the provisos of a function used.

In case of lack of errors, a user is given an overview of the interface of the object created with its subinterfaces and methods and their types. Attached are the handles to start the connection using DnD.

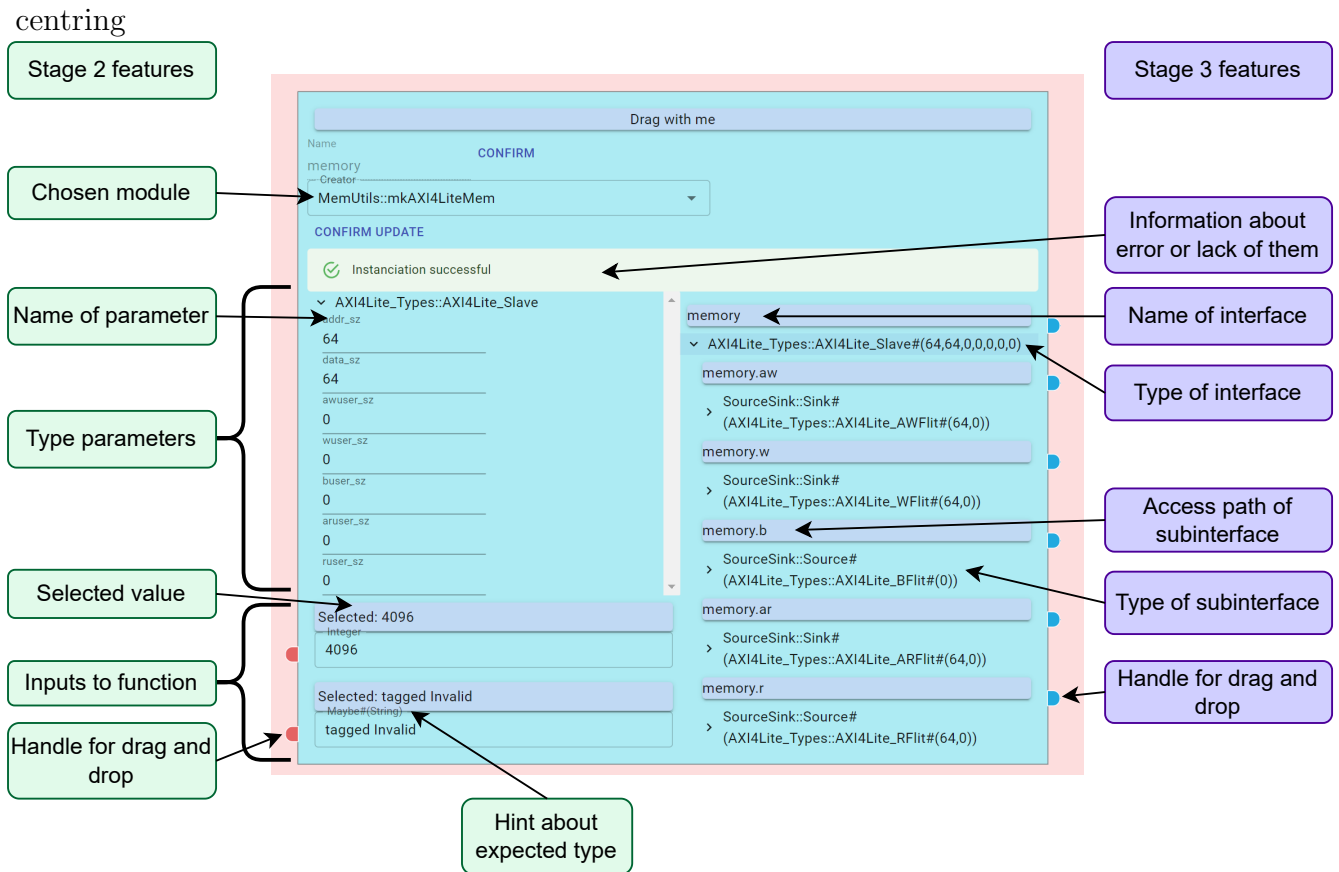


Figure 3.8: Fully instantiated node with descriptions of features, divided into stages two and three

Connection node

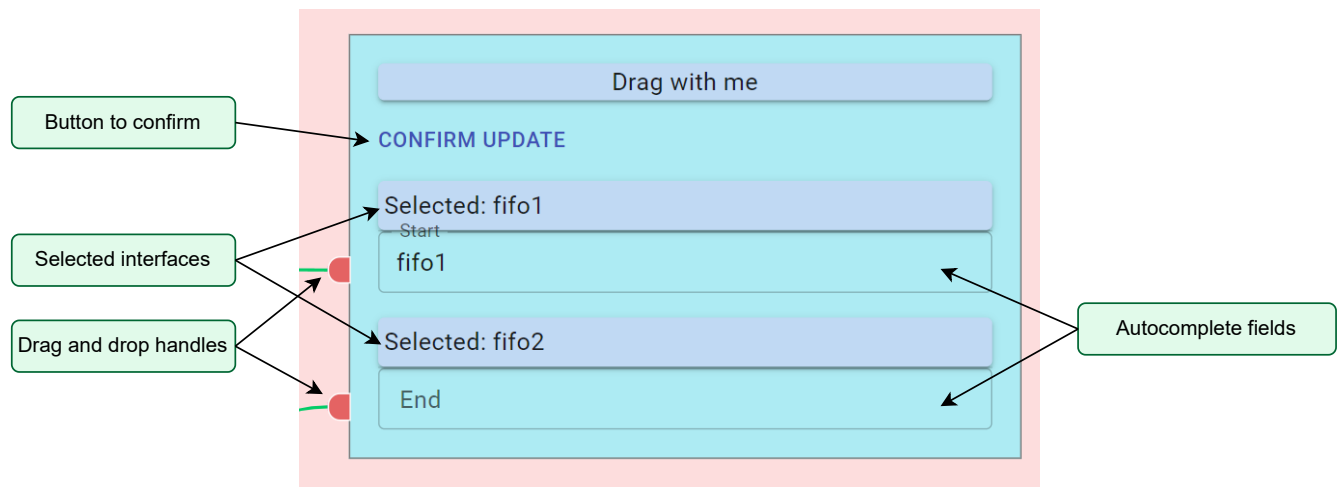


Figure 3.9: Connection node, connecting two FIFO modules

This is the simplest node, it contains two inputs, like the ones found in the instance node. The key change hides in autocompletion, *start* field will show all modules that can be used to start a connection, whereas *end* will only autocomplete the things that can be connected to the thing specified in the start field. The power of this node lies in the fact that it uses typing information

in the backend, and it will work with new types if there is an applicable instance of typeclass `Connectable` which should be common as it is part of the standard libraries.

Bus node

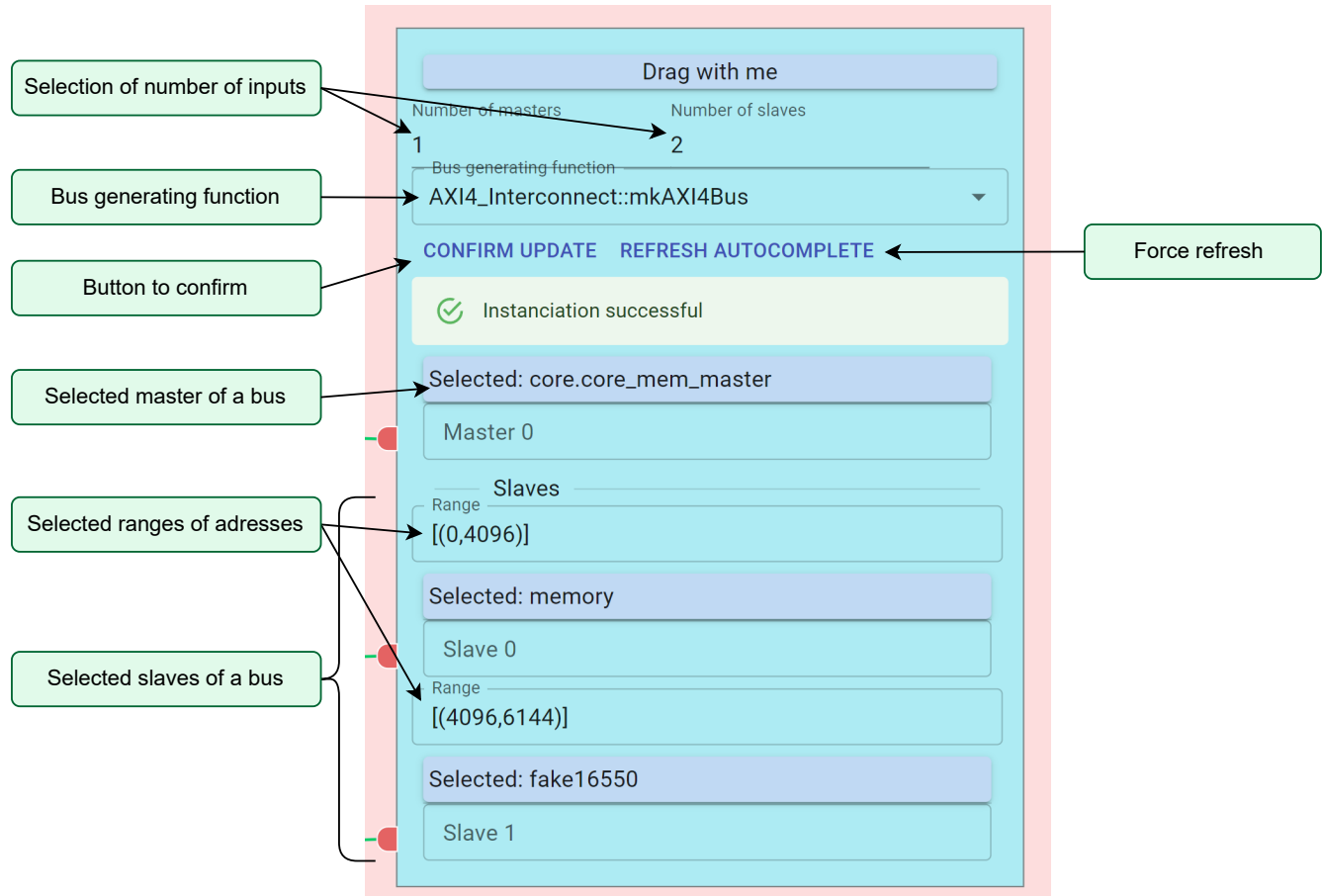


Figure 3.10: Bus node connecting one master and two slaves using AXI4 interface

Similarly to the connection node, most of the components were reused, and interesting are the details. This node only works with a subset of functions that have predefined properties (take tuple (function returning vector, vector, vector) as arguments). Fortunately, users do not need to look for those functions, the autocomplete field will only show you valid ones. (It might be possible that some non-bus function will slip through as I do not want to be overly aggressive and require things like having “bus” in the name of the function, but from my testing, my simple check narrows down options to less than ten from few thousand found in all packages, I worked with.).

Autocompletion for masters and slaves is a bit more nuanced, as it has two levels of precision. At the first level, it will show all masters and slaves that can be used with chosen function, but not necessarily together. However, after the first successful initialization of a node (using *confirm update*), options for masters and slaves will be narrowed down to only ones that can work with masters and slaves selected during the first initialization. This node also contains fields to select a range of addresses on which slaves will be visible.

Exported interface node

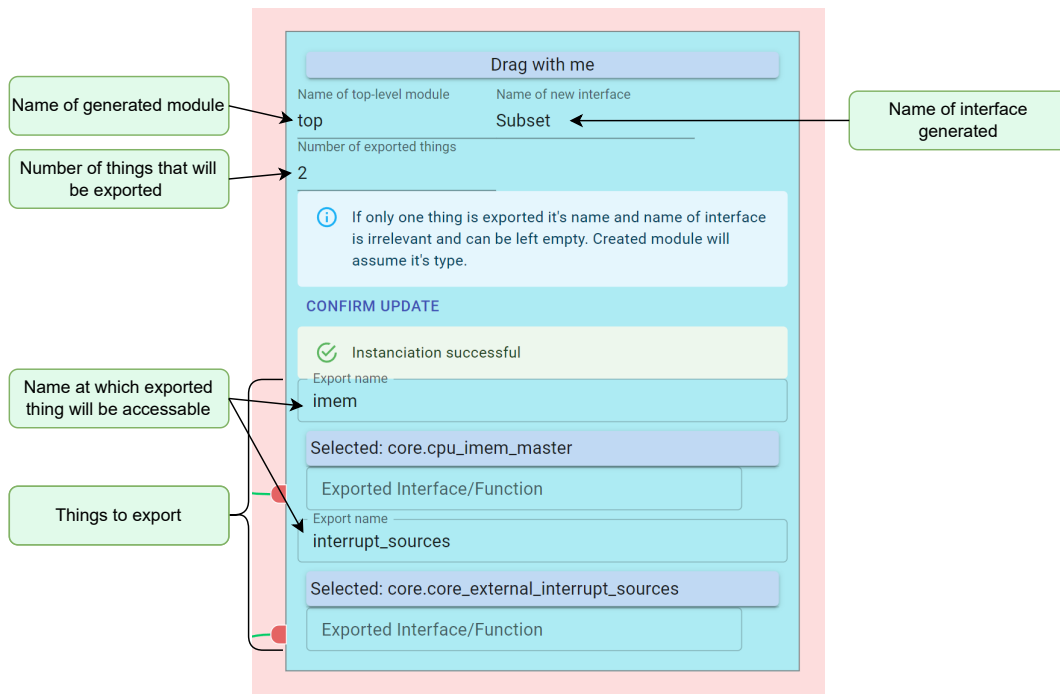


Figure 3.11: Overview of features of a node, allowing to export things from a created module

This node allows users to specify the name of a function by creating a module they designed. Name of the interface which will be used by this module if it will need to export more than one thing, and most importantly, all the exported interfaces. If the name of the interface is already known, the system will try to populate the members of that interface with the members provided, if the name is not known it will simply create a new interface using all the typing data from the members. It also contains a simple feature which will generate a name based on the access path if the user does not provide one.

The intention behind this node is that one might want to create their system partially in Bluespec and partially in Verilog, as there are vast libraries of IP written in Verilog. Using this node, one can create a new interface on the fly that does not have much use in the Bluespec world, but if transpiled to Verilog it will only contain things user-specified, making the process of importing it into tools like Platform Designer much easier.

Chapter 4

Evaluation

During the evaluation, I am going to compare *The Legend of Bluespec: A Link to the components* and *Intel Quartus Prime*, with its Platform designer. IQP obviously has more bells and whistles, but the testing scenarios are specifically concerned with the process of connecting components, written in Bluespec. The goal is to show problems with integration using transpiled Bluespec, and how they can be avoided by staying in the native Bluespec. Evaluation will also highlight the usefulness of integrating the type system into high-level component integration.

4.1 Overview

My evaluation will be divided into three examples:

4.1.1 Example 1: Deadly simple FIFOs

FIFOs are one of the building blocks of components. They are usually the first thing implemented when learning hardware design. This example is meant to show inflexibility, and the problems that arise due to lost information when working with transpiled Bluespec.

4.1.2 Example 2: Weight of the AXI4

In this example, we will design a quite simple System-on-chip (SOC). The SOC will be composed of a Flute core (5 Stage RISC-V CPU), a memory module, and a fake 16550 interface that is meant to represent an I/O. This example will show the best scenario for the Platform designer as connections will use the AXI4 interface which is fully supported by IQP.

4.1.3 Example 3: Simulation

This example is more of a formality. For this example, I created a simple AXI4 master and AXI4 slave, that communicate via the bus. It is meant to show that indeed my generated routing function works. I am using already made libraries, so I do not need to worry about the correctness of the bus-generating function.

4.2 Example 1: Deadly simple FIFOs

4.2.1 Platform Designer

To create a system with two connected FIFOs, we first need to import the FIFO module into the platform designer. Here we encounter the first problem. As mentioned before, Bluespec cannot generate Verilog code for polymorphic functions. Therefore, we must decide on the width of the data stored in such FIFO before transpiling to Verilog. After exporting the Verilog file, it needs to be imported into the Platform designer. This tool performs an analysis of a file, which results in the decision that all wires found are parts of the Avalon slave interface. This is an obvious error, and we need to correct it by hand.

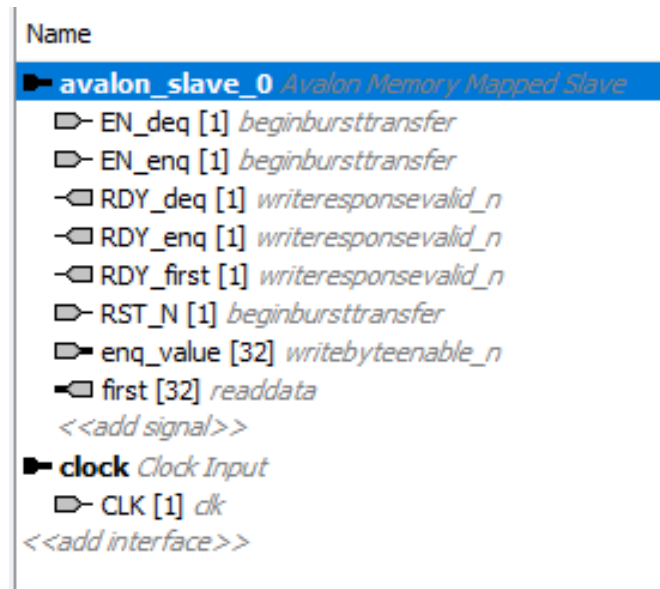
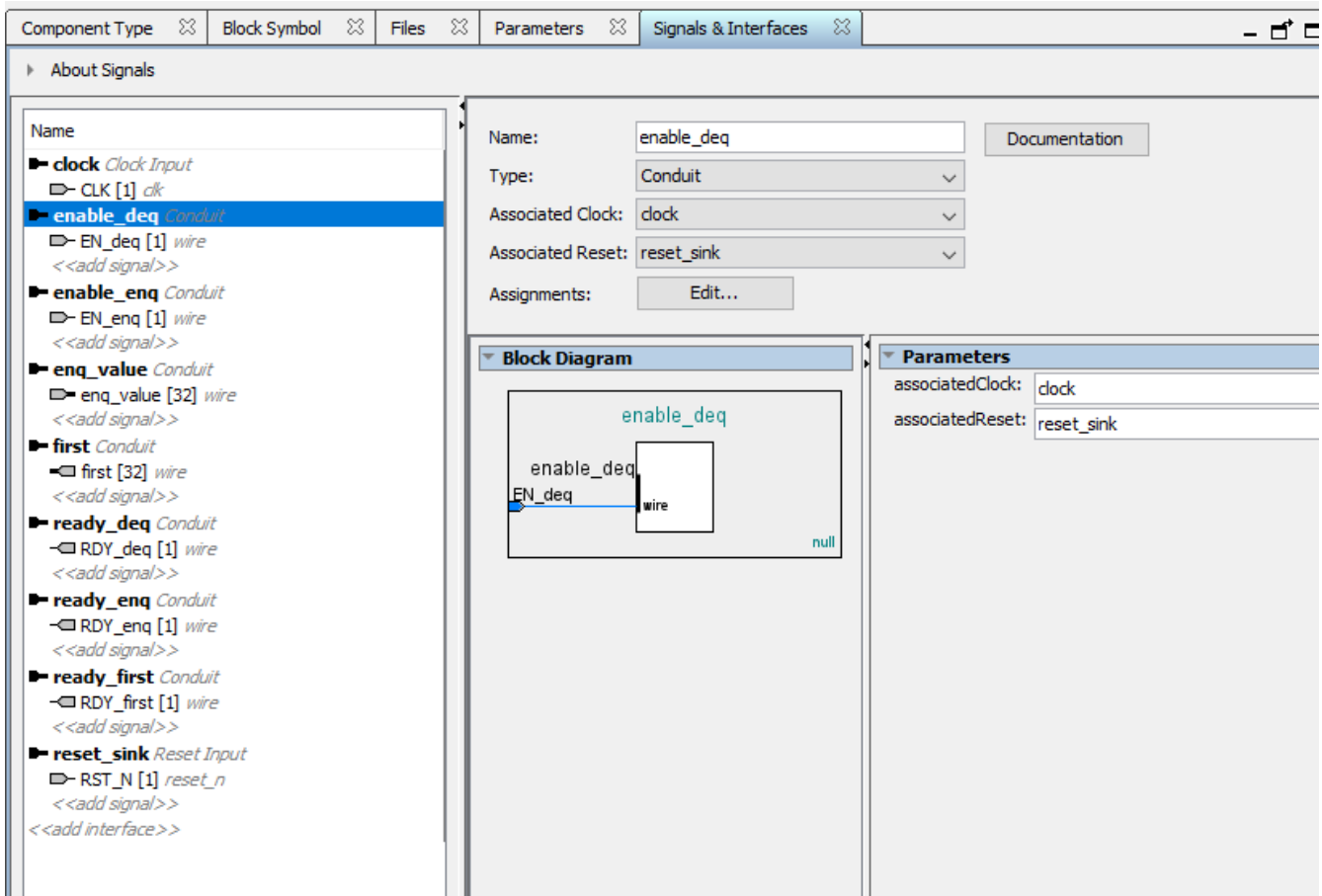


Figure 4.1: Zoom in on the arrangement of wires after import

This is unfortunately easier said than done. There are a few options for assigning those wires, but neither of them is perfect. First, we will need to choose the type of interface that we will represent. In Platform designer, we have over 30 options to choose from, but most of them are variants of complex protocols like AXI4. The only option that will make sense for us is *Conduit*, which is effectively a meaningless blank canvas. We could create a conduit for each method, but after consultation with my supervisor, he recommended creating a conduit for each wire. I did as he recommended and this produces the following set of interfaces: This requires quite a bit of manual work, as listed below.

- Create 8 new interfaces (QSys figured out what to do with the clock wire on its own, so we have one for free) this includes giving them a name and assigning a reset signal.
- Move each wire to its respective interface and change the signal type of each wire to something generic like “wire” from some rubbish given to it as a part of the assignment as Avalon slave interface.

Figure 4.2: After assigning wires and Interfaces

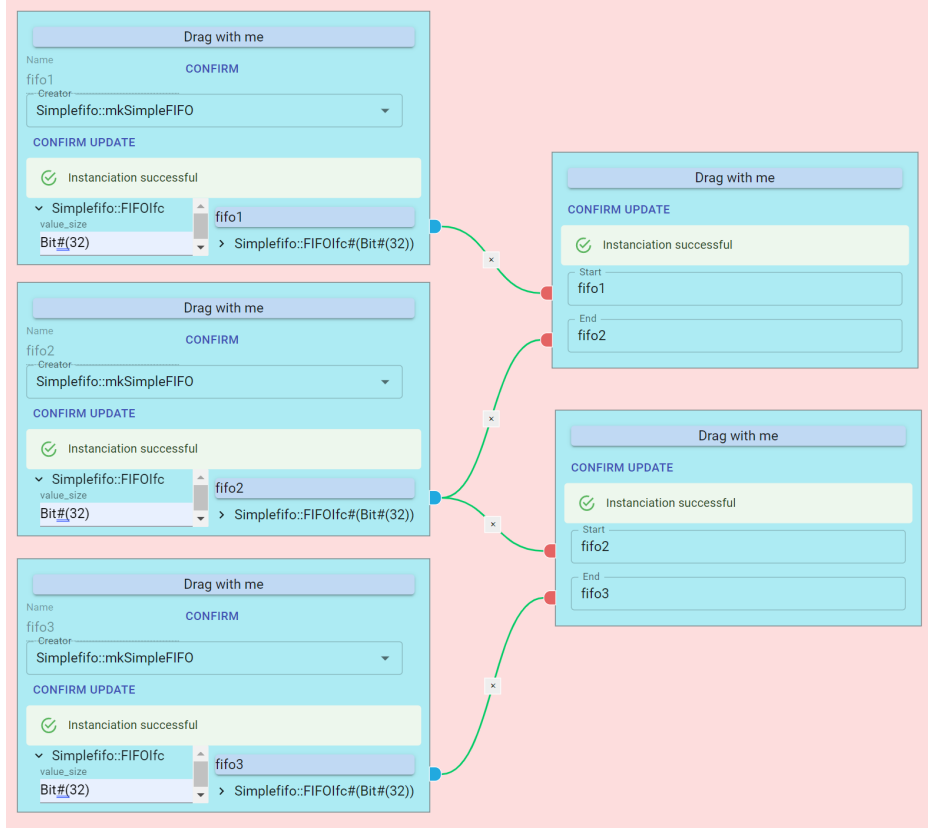


With all of those wires, we can connect obvious things like `first` to `enq_value` and clock signals, but then we run into a problem with all the wires that represent the Ready-Enable micro protocol because we would like to enable the enqueue in the second queue only if both wires `ready_deq` and `ready_first` are high in the first queue, otherwise because we don't know how our FIFO is implemented we might run into a problem where we might get bugs (for example we lose data because we `ready_deq` became a high cycle faster than `ready_first` or in an opposite scenario, we would read stale/corrupted data). Unfortunately, we cannot create such logic using a platform designer and one way of solving this would be to create a special connector module that would do this logic (I'm going to come back later to this to explain why this approach might be infeasible in the general case, but in short, we would need such connector for every width of the object on the queue, and it shows a pattern of lost information when transpiling).

4.2.2 The Legend of Bluespec: A Link to the components

Using my tool, the whole phase of importing and setting up wires can be skipped, as all data information needed is grabbed directly from packages. Connecting FIFOs is as simple as creating a new connection node and selecting the chosen FIFO's interface. Extending two connected FIFOs to three is as simple as adding one more FIFO and one more connection node. It is also worth mentioning that users can select the type of things stored on the queue, therefore avoiding problems with accidentally "casting" data to the wrong type.

Figure 4.3: Three connected FIFOs using my tool



4.2.3 Conclusion

In this example, we learned that FIFOs are so simple that IQP is not able to connect them. This is explained by its lack of understanding of the FIFO interface. Whereas, my tool can understand that FIFOs are connectable (according to the type system) and connect them. It can also work with polymorphic FIFOs, ensuring greater flexibility and type safety.

4.3 Example 2: Weight of the AXI4

As explained before, this example is about creating a simple SOC.

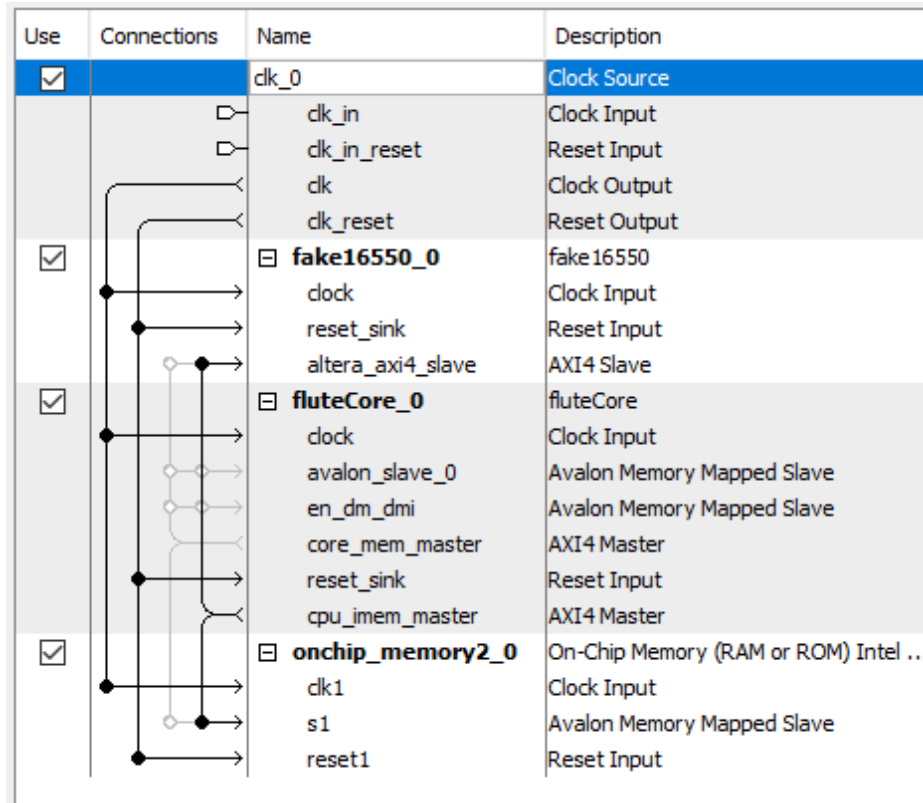
4.3.1 Platform Designer

Again, we will start by importing our modules. This time we need to import two of them, one for fake 16550 interfaces and one for flute core. As for the memory module, we will use one from the IQP library to save time and effort.

Even though AXI4 is an interface supported by IQP, the Platform designer is not able to recognize groups of wires that make up AXI4 masters and slaves on the imported modules. Therefore, we need to do it manually. This would not be so painful if not for the fact that we need to set the type of each wire according to its use. Fortunately, the Bluespec compiler named them nicely, for example, `cpu_imem_master_awready` has type `awready`. Each of those AXI4 master interfaces

has 37 wires, and if we account for all other interfaces of the Flute core, we get 150 wires that individuals need to be assigned to interfaces. Similarly, our fake 16550 interfaces also use 40 wires in total. When assigning this through GUI, it took me around *16 minutes* to roughly 80 wires that make up one AXI4 master and one AXI4 slave interface on the Flute-core and the Fake-16550 (The other 100 wires were unnecessary, so I ignored them). (For context, tools like Xilinx Vivado have heuristics that can do a lot of this automatically, but my understanding is that it still only works for selected known interfaces)

Figure 4.4: Connected System-on-chip using Platform designer



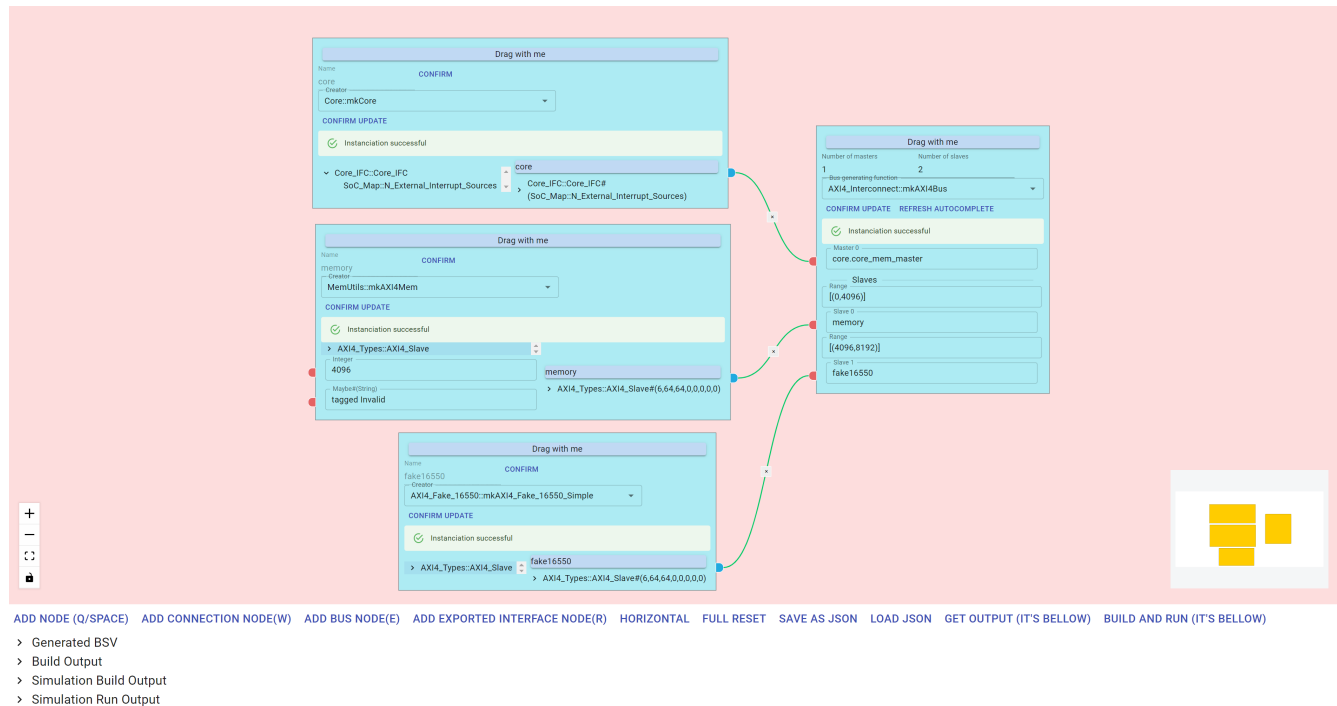
After the initial pain of importing modules, we arrive at the point where the Platform designer shines. It can show possible connections between AXI4 masters and slaves, on a 2D-grid-like layout. Connecting two interfaces is as simple as clicking the cross-section of two wires on the screen.

4.3.2 The Legend of Bluespec: A Link to the components

Again, using my tool we do not need to do anything to import modules, as all the typing information is provided by the package. To connect those, we create a special bus node, select the function for creating a bus(`mkAXI4Bus`), select the number of masters and slaves, and then assign each accordingly. This in total takes less than *2 minutes* (this includes specification of AXI4 interfaces and if one were to use modules prepared for synthesis locked to specific dimensions (like the ones imported to IQP)) this time can be comfortably brought under *1 minute*. While I am unable to

show a nice 2D-grid of possible connections, I have [two-stage autocomplete](#). That can simplify the process of choosing things to be connected to the bus.

Figure 4.5: Connected System-on-chip using my tool. This image is also meant to be an overview of the whole application



4.3.3 Conclusion

When it comes to connecting things, both systems are easy to use. The main difference between those systems is in the difficulty of importing modules. Thanks to the use of typing information, my system does not need additional metadata from the user, therefore saving them time and effort. My system can also extract parameters and expose them to the user, therefore keeping modules flexible. The importance of this will be explained later.

4.3.4 Other differences

Some subtle differences are worth talking about but were not pronounced by this example. Platform designers can automatically convert between AXI4 interfaces and Avalon ones, and within those families, it allows for some degree of flexibility when parameters do not match exactly. When I create buses, I am limited by the typing system, and the way functions for creating buses are defined. Therefore, in my system, all masters and all slaves must be of the same types. If the user wants to do similar conversions, they would need to use functions (like `fromAXI4toAXI4Lite` that convert between interfaces) and perform those conversions by themselves. My system has obvious drawbacks, but at the sample time, it allows users to define new conversions that implement arbitrary logic.

Another slight difference is the fact that in Platform designer, each master has its own bus, and while slaves can be connected to any number of masters they need to be connected individually. Forcing the user to specify $O(n^2)$ connections.

4.4 Example 3 - Working AXI4 masters and slaves

This example follows similarly to example 2, where the process of importing modules is a hassle using Intel's tool, especially as I made it so each master and slave takes a single integer as a parameter, this is meant to represent some configuration data. Importing modules using my tool is again as simple as providing the location and name of the library with those modules. After this, both systems allow for easy connection of modules. To check whether what I generated is correct, I have run the simulation using `bluesim` tool included with the Bluespec compiler. I also noticed a few differences between the two systems when it came to assigning addresses. For example, in Bluespec it is assumed that masters and slaves share the same with of the wire for address, whereas in Platform designer it is assumed that the address space of the master is greater or equal to the sum of the address space of all slaves. (TODO will check with supervisors the implications of this)

4.5 Numbers

This section presents a quantitative comparison of text formats used by both systems. Sometimes one might not want to bother with using GUI and wants to quickly edit some variables or connections. To facilitate this, the text file format should be short and simple, otherwise, the user will be forced to sift through a lot of fluff. Therefore, I will compare those file formats using line and token counts. To count the number of tokens, I am using the following regexp `(\w|\.)+`. This regex will match sequences of alphanumeric characters with dots and underscores. I decided to ignore things like braces as they can account for a lot of tokens, at the same time people might have different preferences for them.

Figure 4.6: Tokens used to instantiate and connect 2 FIFOs (blue), extra tokens to create and "connect" one more FIFO (red)

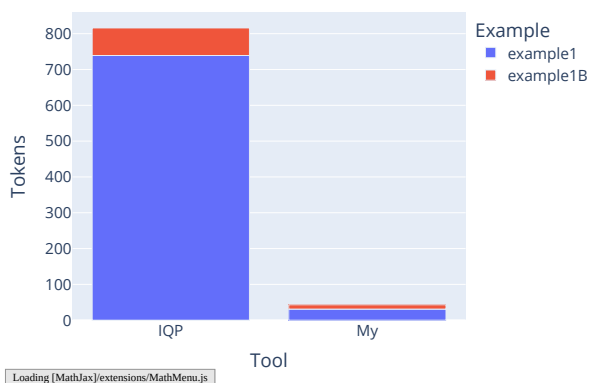
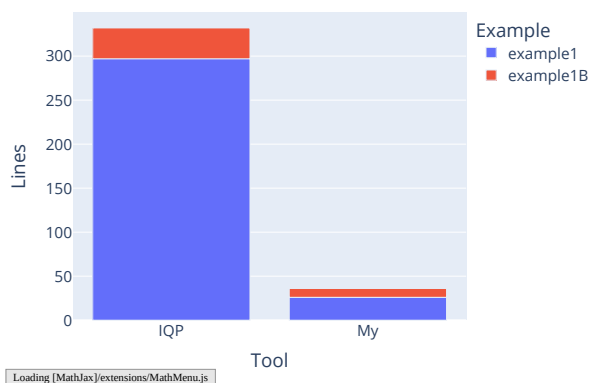


Figure 4.7: Lines used to instantiate and connect 2 FIFOs (blue), extra Lines to create and "connect" one more FIFO (red)



As we can see, the Platform designer uses roughly an order of magnitude in a more verbose text format. What is important from this chart is that all that verbosity is not a one-time cost, but more like a constant multiplier.

I also included charts with lines and token counts for other examples to show that this verbosity is a general trend.

Figure 4.8: Tokens used by each system for each example

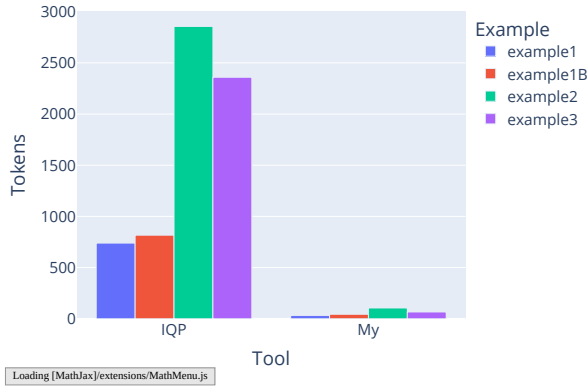
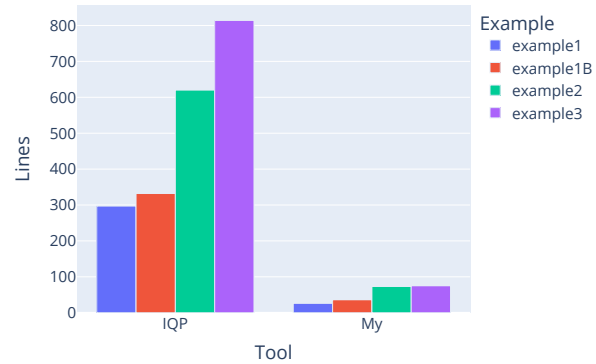


Figure 4.9: Lines used by each system for each example



4.6 Other things to consider

One important thing to consider is the workflow. Bluespec compiler cannot generate parametric designs. This means that if we wanted to have a parametric module with 3 variables each taking one of 5 values, we would already need to import 125 modules, and I showed how importing even a few of them can be a hassle. In contrast, with my tool, users can not only use polymorphic functions and modules but can also use the type system to resolve unknown parameters.

IQP is a gigantic tool, and it would be dishonest to say that my tool is strictly better. However, thanks to the ability of my tool to quickly export interfaces, one can use both tools in tandem. The user can start by defining a blob of connected Bluespec modules using all the good stuff in my tool, then they can export only the interfaces they will need. This way, the user must always import only one module to the IQP. Such blobs can then be connected to an IP defined in Verilog or to pins on the FPGA.

Chapter 5

Conclusion

5.1 Results

I hope that I demonstrated how *The Legend of Bluespec: A Link to the components* can be used to speed up the prototyping of high-level modules in Bluespec. It can provide the same basic functionality of allowing for connection and easy creation of buses as a comparable tool made by Intel. At the same time, thanks to being Bluespec native it can drastically cut downtime needed to import modules, especially in cases where one wants to use polymorphic modules. While my tool is not commercial grade, thanks to the easy ability to synthesize interfaces for export, it allows for a workflow where everything Bluespec native can be connected using it, and then this part of the system can be imported to a tool like a Platform designer for connection to modules not available as a Bluespec library and connection of wires I/O on the FPGA etc.

5.2 Future work

This tool can be extended in many ways, but the most interesting would include:

- Upgrade of the backend to be a fully functional language server.
- Ability to add comments that partition the graph (like in UE4).
- Ability to map arbitrary Bluespec code into graphs like this.
- Integration with a debugger to visualize the flow of data, and values of data.

Bibliography

- [1] Bluespec SystemVerilog Reference Guide http://csg.csail.mit.edu/6.S078/6_S078_2012_www/resources/refguide.pdf
- [2] Bluespec SystemVerilog User Guide <https://web.ece.ucsb.edu/its/bluespec/doc/BSV/user-guide.pdf>
- [3] Template project using React Flow (I don't think there is more than few lines from the original left in my code, but it was useful as a starting point) <https://github.com/recoding-io/react-mind-node>

Computer Science Tripos

Part II Project Proposal Coversheet

Please fill in Part 1 of this form and attach it to the front of your Project Proposal.

Part 1

Name: CRSID:

College: Overseers: (Initials)

Title of Project:

Date of submission: Will Human Participants be used?

Project Originator: Signature: -----

Project Supervisor: Signature: -----

Directors of Studies: Signature: -----

Special Resource Sponsor: Signature: -----

Special Resource Sponsor: Signature: -----

Above signatures to be obtained by the Student

Part 2

Overseer Signature 1: -----

Overseer Signature 2: -----

Overseers signatures to be obtained by Student Administration.

Overseers Notes:

Part 3

SA Date Received: SA Signature Approved:

Project Proposal

kr469

September 2021

1 Introduction (Project description)

In my project, I will develop a tool for easy device building from modules in Bluespec. To do this I will need to create a library that can use Bluetcl tool to extract data from .bo package files. My project will use existing .bo (and .ba) package files which represent modules in their pre elaboration form. (.ba are after elaboration but they behave in similar manner they contain more information but because they are translated to hardware description language they can't capture concepts like polymorphism), Bluetcl is tool made in Tcl language which is a bit obscure and by creating this library I will also create essentially a python wrapper for Bluetcl which could be useful for further automation of the generation of Bluetcl files.

After extraction, I will have an internal representation of different modules and their interfaces, this will be later used to synthesize top-level files, thanks to typing information I will be able to make sure to synthesize only connections between modules that are possible (at least according to typing information).

One of the larger difficulties of this project will be the integration of shared busses, they allow for the connection of multiple slaves and masters in some orderly fashion, they come with extra complexities coming from the fact that some devices have different widths of connectors, but you might still want to connect them together via bus and module converting widths.

After all of this, we can tackle the main goal of this project which is creating an interface for synthesizing those top-level files. There are currently alternatives like .qsys files but they can easily run in tens of megabytes, and while they store a lot of other information that my format will not, it doesn't change the fact that for humans such bodies of text are effectively not readable nor writable. I plan to create a simpler format that is more human-oriented. Ofc this file format won't be read by any other software but because those synthesized top-level files will be written in Bluespec they can comply with Verilog(thanks to Bluespec compiler).

If all of this goes smoothly, I should have quite a bit of time left over for an optional goal, If I will be taking on QSys text file format I might as well go after QSys itself this means creating GUI for adding and connecting components. Everyone who did practicals or otherwise had a chance to work in QSys knows that the development of UI of this program stopped somewhere in the mid-2000s. It's ugly and difficult to navigate, with small targets for mice (some are few pixels wide). Due to the fact that this is an optional goal I plan here to focus on developing something quickly so I plan to use established game engines like Unity3d or Unreal Engine to do this UI as I have already worked with them and they allow for compilation for multiple platforms.

Evaluation will be done in form of a qualitative comparison between my solutions, QSys and QSys Pro. The criteria I will evaluate it on are ease of modification and readability. For more quantitative data I will use the number of lines and tokens used. I will prepare one or more tasks like Given two connected FIFOs, how much change is needed to connect 3rd FIFO and then solve them using all 3 solutions then explain differences. My

project will be successful if created tools will allow for solving those task more easily while still producing valid Bluespec that behaves in a reasonably similar manner (function/device comparison is very difficult and beyond scope of this project, as long as logic produces the same behavior it's fine, we don't care about exact timings ect.)

I plan to use an iterative software development strategy where every 2 weeks or more often I will push out incremental changes or new features in accordance with the timeline.

2 Timeline

Whenever I mention here word package I mean work package lasting 2 weeks. Start 15.10.2021

- 29.10.2021 - Understanding Bluespec and Bluetcl, during those 2 weeks I will read up on Bluespec and how it fits into the System Verilog ecosystem. At the end of those two weeks, my goal is to produce a document explaining the core of Bluespec typing, and the toolchain I will use to get typing information, and where this information will be fed back. This will be an exercise for me and a good introduction to what's going on for people unfamiliar with Bluespec reading this.
- 12.11.2021 - Busses, during this fortnight, make some reading on busses what they require and work on how to integrate them into my project. Goal produces a short document detailing how busses behave, how to represent them, what information is needed to synthesize a bus, and possibly some explanations of unforeseen difficulties. Thanks to doing this reading-up work early I will be able to use an iterative/spiral development strategy.
- 26.11.2021 - Extracting type information using Bluetcl, the goal here is to produce code that is capable of calling Bluetcl with parameters that allow for extracting data from .bo packages. Here time will be also spent creating an internal representation that allows for a large amount of flexibility. (Interfaces in Bluespec are a bit complicated, they have multiple functions and those functions can have multiple ports for ready / data signals.) The goal of this package is a bit vague, but let's say that I will extract interface types, and interface functions, with parsed types.
- 10.12.2021 - Continuation of above. The goal here is that parsing is fully functional, and it creates an internal representation of modules, interfaces, functions, ports, types it can find in the package.
- 24.12.2021 Synthesizing Bluespec, Given my internal representation I will synthesize a Bluespec that represents a top-level file, with modules used and connections between them. Goal has functionality in code that allows synthesizing: Creation of module, a connection between two modules, possibly some nonstandard connection that requires more fiddling(adding automatic widening/shortening in connections between busses and slave/masters). Plus progress report.
- 07.01.2022 Creating text file interface, Goal use some existing file format like Yaml / XML / JSON to store and read how synthesized top-level file needs to look like. Extra time here will be spent to make sure this file format is both human-readable and human writable, as this will be the main way of interfacing with my system.
- 21.01.2022 - Evaluation Part 1, creation of files in other editors and write up on differences between them and my file format.
- OP1 (Optional goal) GUI / Interactive text, while editing text is fine, and can be done even from terminal it would be nice to have some way of creating a top-level file that uses typing information to guide the process. Goal: Create a GUI/Interactive text support.

- 04.02.2022 - (Optional goal evaluation) Here I will evaluate my GUI vs QSys, in terms of ease of use and flexibility.
- 18.02.2022 - Writing dissertation
- 04.03.2022 - Writing dissertation, final work package.

3 Determining special resources and checking their availability.

No special resources are needed.

4 Securing the services of a suitable Supervisor.

Jonathan Woodruff agreed to supervise my project.