**A project report submitted in partial fulfilment for the degree of**

**BSc (Hons) Software Engineering**

**School of Psychology and Computer Science**

**University of Central Lancashire**

Programming Language for homemade CPU architecture - Gaungau

Rhys Garman

**29th July 2022**

# Abstract

This project set out to solve the problem of people not be able to easily see what is going on under the hood of modern day computing. Most constructs are hidden under layers of abstraction, and it is incredibly difficult, and sometimes seemingly impossible to see what is going on when you, for example, run a program.

The way in which the project sets about solving this is by creating a whole new CPU architecture, and a programming language associated with it. The programming language is based on high-level syntax, so that any user picking up the project will have a shallow learning curve and will be able to get good results very quickly. It is also designed this way to ensure that new people coming into computing (although not necessarily the targeted demographic) can also have some good results which might lead on to some independent learning.

The two primary objectives are to make a powerful enough language to enable the 8-Bit computer to be Turing complete machine, meaning that theoretically any solution can be made with the language; and to make the compiler as transparent as possible to enable the user to do as much “digging” into what is going on as possible.

I am very happy with the result, not only does the compiler enable the computer to function to it’s full potential as a Turing machine, but also it is very user friendly, and module to allow for deeper understanding at the pace of the user. During the use of the project by the end user, there are no particular objectives to hit, just to use it, and take the time to enjoy and understand what exactly is happening when they write a line of code.

# Attestation

I understand the nature of plagiarism, and I am aware of the University’s policy on this.

I certify that this document reports original work by me during my University project. I also confirm that I adhere to the University’s legal and ethical guidelines for undergraduate projects in Computing.

Signature: \_\_\_\_RHYS GARMAN\_\_\_\_\_\_\_

Date: 29/07/2022

# Acknowledgements

A huge thank you to my supervisor Matthew Horton for helping me through this project, and providing valuable feedback and motivation to do my best.

I would also like to thank BAE Systems for sponsoring my through the degree, and providing the time and assistance to make the degree the best I can.

Furthermore, my thanks also go to Ben Eater who inspired me to start the project and provide areas in which to improve and undertake self-learning to even further my understanding to so much of an extent that I felt confident to be able to teach to others.

# Table of Contents

[Abstract i](#_Toc110007989)

[Attestation ii](#_Toc110007990)

[Acknowledgements iii](#_Toc110007991)

[Table of Contents iv](#_Toc110007992)

[List of Figures vii](#_Toc110007993)

[List of Tables viii](#_Toc110007994)

[List of Listings ix](#_Toc110007995)

[1 Introduction 1](#_Toc110007996)

[2 Background and Related Work 2](#_Toc110007997)

[2.1 Introduction 2](#_Toc110007998)

[2.2 Teaching Low Level Programming 2](#_Toc110007999)

[2.2.1 Why Teach Programming 2](#_Toc110008000)

[2.2.2 Best Ways to Teach 2](#_Toc110008001)

[2.2.3 A practical approach 5](#_Toc110008002)

[2.3 Low-Level Programming 6](#_Toc110008003)

[2.3.1 Low-Level vs. High-Level 6](#_Toc110008004)

[2.3.2 Digital Logic 7](#_Toc110008005)

[2.3.3 Relevance of Low-Level Programming 7](#_Toc110008006)

[2.3.4 Applications of Low-Level Programming 7](#_Toc110008007)

[2.4 Microcontrollers 9](#_Toc110008008)

[2.4.1 Understanding 9](#_Toc110008009)

[2.4.2 Applications 9](#_Toc110008010)

[2.4.3 Relevance 10](#_Toc110008011)

[2.4.4 Uses in Education 11](#_Toc110008012)

[2.5 Summary 11](#_Toc110008013)

[3 Project Planning 13](#_Toc110008014)

[3.1 Introduction 13](#_Toc110008015)

[3.2 Methodology 13](#_Toc110008016)

[3.3 Requirements 13](#_Toc110008017)

[3.3.1 Must 13](#_Toc110008018)

[3.3.2 Should 13](#_Toc110008019)

[3.3.3 Could 14](#_Toc110008020)

[3.3.4 Would 14](#_Toc110008021)

[3.4 Potential Solutions 14](#_Toc110008022)

[3.5 Tools and Techniques 14](#_Toc110008023)

[3.6 Legal, Social, and Ethical Issues 16](#_Toc110008024)

[3.7 Summary 16](#_Toc110008025)

[4 Design 17](#_Toc110008026)

[4.1 Introduction 17](#_Toc110008027)

[4.2 High-Level Design 17](#_Toc110008028)

[4.2.1 Tools and Syntax 17](#_Toc110008029)

[4.3 Low-Level Design 21](#_Toc110008030)

[4.3.1 Syntax 21](#_Toc110008031)

[4.3.2 Logic 22](#_Toc110008032)

[4.3.3 Performance vs. Universal solution 22](#_Toc110008033)

[4.4 Binary – Where Software meets Hardware 23](#_Toc110008034)

[4.5 Summary 25](#_Toc110008035)

[5 Implementation 26](#_Toc110008036)

[5.1 Introduction 26](#_Toc110008037)

[5.2 Main Processes 26](#_Toc110008038)

[5.2.1 Lexical Analysis 26](#_Toc110008039)

[5.2.2 Syntax Analysis 28](#_Toc110008040)

[5.2.3 Code Generation 31](#_Toc110008041)

[5.3 Garbage Control 32](#_Toc110008042)

[5.4 Error Handling 33](#_Toc110008043)

[5.5 Operating System 34](#_Toc110008044)

[5.6 The Hardware 36](#_Toc110008045)

[5.7 Points of Interest 39](#_Toc110008046)

[5.7.1 Functionality vs. Performance 39](#_Toc110008047)

[5.7.2 Parsing Recursion 40](#_Toc110008048)

[5.7.3 Polymorphic Behaviours 40](#_Toc110008049)

[5.8 Summary 40](#_Toc110008050)

[6 Testing 41](#_Toc110008051)

[6.1 Introduction 41](#_Toc110008052)

[6.2 Functional Testing 41](#_Toc110008053)

[6.3 Non-Functional Testing 44](#_Toc110008054)

[6.4 User Testing 45](#_Toc110008055)

[6.4.1 Test subject 1 45](#_Toc110008056)

[6.4.2 Test Subject 2 45](#_Toc110008057)

[6.4.3 Test Subject 3 46](#_Toc110008058)

[6.5 Evaluation 46](#_Toc110008059)

[7 Evaluation, Conclusions and Future Work 47](#_Toc110008060)

[7.1 Project Objectives 47](#_Toc110008061)

[7.2 Self-Evaluation 47](#_Toc110008062)

[7.3 Project Evaluation 47](#_Toc110008063)

[Must 47](#_Toc110008064)

[Should 47](#_Toc110008065)

[Could 48](#_Toc110008066)

[Would 48](#_Toc110008067)

[7.4 Applicability of Findings to the Commercial World 48](#_Toc110008068)

[7.5 Conclusions 48](#_Toc110008069)

[7.6 Future Work 49](#_Toc110008070)

[References 50](#_Toc110008071)

[1 Appendix 1 – Project Proposal 55](#_Toc110008072)

[1.1 Project Context 55](#_Toc110008073)

[1.2 Specific Objectives 55](#_Toc110008074)

[1.3 Potential Ethical or Legal Issues 55](#_Toc110008075)

[1.4 Resources 56](#_Toc110008076)

[1.5 Potential Commercial Considerations - Estimated costs and benefits 56](#_Toc110008077)

[1.6 Proposed Approach 56](#_Toc110008078)

[1.7 References 57](#_Toc110008079)

[2 Appendix 2 – Technical Plan 58](#_Toc110008080)

[2.1 Title 58](#_Toc110008081)

[2.2 Summary 58](#_Toc110008082)

[2.3 Deliverables 58](#_Toc110008083)

[2.4 Constraints 58](#_Toc110008084)

[2.5 Key Problems 59](#_Toc110008085)

[2.6 System and Work Outline 59](#_Toc110008086)

[2.7 Project Activities 60](#_Toc110008087)

[**2.8** Risk Analysis 61](#_Toc110008088)

[**2.9** Options 61](#_Toc110008089)

[2.10 Potential Ethical or Legal Issues 62](#_Toc110008090)

[**2.11** Commercial Analysis 62](#_Toc110008091)

[2.12 Employability Contribution 63](#_Toc110008092)

[2.13 References 64](#_Toc110008093)

[Appendix 3 – Title of Appendix 66](#_Toc110008094)

# List of Figures

[Figure 1 9](file:///D:\DDocuments\Uni\Final%20Year\Final%20Project\Interpreter\Docs\Report\CO3808%20Report%20-%20RGarman.docx#_Toc110008095)

[Figure 2 15](file:///D:\DDocuments\Uni\Final%20Year\Final%20Project\Interpreter\Docs\Report\CO3808%20Report%20-%20RGarman.docx#_Toc110008096)

# List of Tables

[Table 1 – Functional Test Results 41](#_Toc110008097)

[Table 2 – Non-Function Testing 44](#_Toc110008098)

# List of Listings

# Introduction

This project involves the creation of a compiler for my bespoke custom-designed 8-Bit CPU. The 8-Bit computer has been worked on for several years, and has now reached a stage where the software can become a serious consideration.

Java will be the primary language used to create the compiler. This has been proven possible in an implementation by Strom and Aas (Strom & Aas, 2001). This is because it allows anyone to use the compiler if the project was to be rolled out, as java is compiled on a VM, so it can run on any existing architecture (e.g. Windows, and iOS).

The point of a compiler is to convert high-level (human-readable) code, into machine code or binary. These binaries can then be loaded onto the EEPROM (electrically erasable programmable read-only memory) on the CPU and executed. The language will be created with its own syntax, structures, garbage control, etc.

The language will be created with the high-level user in mind, by this, the base syntax will be based more around languages like Java and PHP, even though it is a direct link to the machine code like C or C++ is. By doing this, it will allow anyone to code their own program, and it will be easy to see exactly what is going on. There will be nothing happening in the background, and everything will be visible to the user, from the interpretation to the binary output, and potentially further options to do with garbage control constraints, and manual addressing.

The issue surrounding the lack of more acute compilers is addressed in (Tanenbaum et al., 1983) which was also a strong motivator towards the project

# Background and Related Work

## Introduction

This chapter explores the literature that has been written surrounding the topic of low-level programming. Three themes relevant to this project will be discussed: the ways in which programming should be taught; an overview of the concepts and techniques within the low-level programming domain; and the application of microcontrollers. In addition, this chapter considers how this existing knowledge and literature overall will have an impact on this project, and to get the most out of the time and resources available.

## Teaching Low Level Programming

### Why Teach Programming

It is well known that computer programming is a difficult thing to do and is not simply one subject. It is incredibly broad and has many transferable skills. This is because it forces people to think from a point of view where they have no previous knowledge, and with a completely literal mindset. One example of this is multiplication. This is a very simple task for a human to achieve, but a computer cannot perform multiplication natively. Instead, the computer must add numbers a given number of times. For example, consider the equation:

This is a simple operation, but it forces the programmer to analyse everything in absolute depth. Some of these transferable skills will undoubtedly be reasoning skills, problem-solving, and self-efficacy in mathematics, and this is exactly what was investigated by Psycharis and Kallia (2017) in their paper on the effects of computer programming on high school students. They tested this by having two groups of 33, the control group being taught ‘Development of Applications in Computer-Programming Environments’, and the experiment group being taught Chemistry and Electrology. In addition, both groups had the same mathematics and physics courses in common. They had the same material delivered, and by the same teacher. The results of the study showed a significantly higher level of self-efficacy in mathematics, problem-solving, and reasoning skills, but unfortunately, no discernible difference was found in the performances across the two groups. Although, it should be noted that there was a higher mean score in the group with the computing integrated into their syllabus. As such, while the study was categorised as inconclusive, it could be argued that there is sufficient evidence of notable improvement in mathematics and physics from the learning involved with computer science. This elevates the importance of this project in teaching people low-level computing in a new way, as it is shown to have a positive effect on other subjects. This is excellent motivation to add as much as possible into the learning aspect of this project and ensure that the problem-solving and reasoning skills are being addressed for the impact of the project to be its greatest. Further investigation into the methods of teaching low-level programming will therefore be valuable in the design phase of this project, to ensure that the educational materials included are linked back to broader topics. This will help users of the project expand their knowledge beyond simply how the computer operates.

### Best Ways to Teach

Low-level programming is widely discounted when it comes to teaching computer programming, although sometimes for good reason. This could perhaps be the way in which it is delivered and the response to the material. Certain approaches such as a full understanding of the tools needed and full demonstrations are mostly forgotten about. Gries (1974) illustrates this point with an example case of cabinet making. In his scenario, he describes an instructor letting students use each of the tools required for cabinet making for a few minutes in isolation. While the students are now knowledgeable in all of the component skills they need, none of the students could be expected to apply this knowledge to fully constructing a cabinet on their own. This is a great example of the parallels between software and working on physical materials. It is important to note that in programming, we are presented with tools, and if not used correctly, can waste time, and cause vast amounts of frustration. Gries continues to describe ways in which to teach programming correctly, but the main points to take his paper are to: Understand the problem, devise a plan, carry out the plan, and look back. With these steps, one can “develop a level of discipline which most programmers don’t have” and transfer these skills to other aspects of one’s life, such as reasoning skills, problem-solving, and self-efficacy in mathematics Psycharis and Kallia (2017). The list of best practices to follow set out by Gries is a great way to break down a problem, and is a list that the project will very much look to replicate and encourage when developing the associated learning materials and programming language. This will also be used as motivation to fully teach every tool that is made available to the user of the product. The beauty of low-level programming is that each command does one thing, and it is building these commands into layers and layers of functionality that allows users to accomplish amazing and complex things. This is the joy that this project is setting out to communicate to the user, and to be able to facilitate this learning to any ability.

Furthermore, while providing the correct foundation to learning about the tools used in programming, it is also hugely important to cater to the experience level of the user. An advanced programmer who has programmed their whole life will not be interested in declaring variables, or how to use a for loop; in contrast, a novice programmer will become overwhelmed and put off if they are faced with advanced object orientation or nested recursion. Dreyfus et al. (1987) state that there are 5 steps from novice to expert: novice, advanced beginner, competent, proficient, and expert. It is imperative that each of these stages are being catered for. The journal goes on to explain in much more detail about the transition from being a novice programmer, all the way to becoming an expert, and most importantly, how the thought process changes over this time. Thought process is a huge factor in the project, as all it is setting out to do is enable people to fully understand what is going on behind the screen in front of them. This again is where the use of low-level code comes into its own. As it is so close to the hardware it is being run on, it is not a case of “how to declare a variable” it is more, “how the variable is being declared”. By rearranging this initial question, we are able to create a new base level for anyone in the demographic to begin or move ahead if they already know. As such, the one material can cater to all stages. This way of thinking must be reflected onto the user of the project by fully understanding the material they are reading. The documentation however, must be written carefully to enable anyone to understand and have fun with.

Another consideration is to look at a mental model first approach. This is whereby the actual programming comes very much second to the thought process behind what is going on. Robins et al. (2003) discuss that "programs are usually written for a purpose and that a mental model of this problem domain must precede any attempt to write an appropriate program". This is also confirmed by Deek et al. (1999) who describe a first-year computer science course based on a problem-solving model, where language features are introduced only in the context of the students’ solutions to specific problems. A study was undertaken to test this theory of using a mental model before starting to write any code, by Xinogalos (2014), the results of which showed that 40% of the 50 students involved in the voluntary study answer “very much” to “How much did the use of a pseudo-language help you in your introduction to programming?”. This mental model approach, whereby the students must first gain an understanding, not only of the problem they need to solve, but also to apply the limitations present in the tools they have available to them is great. Having this plan and allowing the use of programming based solely on the mental model they have built is a great way to allow the students to think about how the computer is thinking. This also backs up Gries (1974) who states some rules to software development: Understand the problem, devise a plan, carry out the plan, and look back. This mental model approach incorporated the initial understanding of the problem and devising a plan, in one. It forces the linear approach to abide by the rules set out by Gries. As stated, these rules laid out by Gries are ones that that project should reflect and implement. it should also look to encourage the use of Deek’s mental model approach when thinking of low-level concepts, as dry running the programs in your head is an incredibly useful way to generate a simple layer of logic that is to be built upon throughout the implementation of the problem’s solution.

### A practical approach

Low-level programming is all around us, and despite this, is widely forgotten about. It is the only way in which we interact with a computer, and it is the fastest way to run any program on a computer. Low-level programming is a dying art. This is mainly due to high-level languages and structures taking the leading role in education and software development. They are more tailored to how people think and so are much simpler to pick up. Despite this, low-level programming is still incredibly relevant, now more than ever, and should not take a back seat. It is not until late education that we are taught about high-level language being translated into something which can be understood by the computer being used. In a study undertaken by Smith and Webb (2000), they looked at how meaningful learning requires pre-existing knowledge to build upon. As we are not taught how the computer actually operates when the topic of low-level programming emerges, the learner is forced to memorise information without proper understanding, and this will leave one’s mind very quickly. Their solution to this was to create a glass box interpreter based upon C, which would show a novice user exactly how the compilation process is undertaken. This is a strong parallel to this project, as I too wish to create this glass box model of an interpreter. The difference is that my glass box will be based on a custom programming language, as opposed to an existing language such as C. I think that I am like-minded to these researchers, and am out to solve the same problem. Their journal goes on to explain a great deal about the functionality of their glass box implementation and I am going to investigate drawing parallels to these functionalities, this is because the results of the study when introduced to real students was very positive, and they are features which I believe will improve my current design greatly. I spoke about some of the features in prior documentation, but this will add context and proof that this idea will work and will yield results. While the results from this study have been overwhelmingly positive, the way in which the responses were gathered was via a multiple-choice questionnaire. While these are quick and easy to set up and get results fast, there are a few issues with this form of feedback gathering; some people may simply guess the answers and still come out with a majority of correct answers.

In contrast to having a practical looking into the software side of things, one of the advantages of low-level programming is that is very close to the hardware that the code is being run on. This gives the opportunity to see the results of the code written, very easily. Microcontrollers (or microcontrollers) can have code embedded directly onto them and be used to create advanced circuitry that can be fun to make and see the results immediately. In recent years, this technology has become extremely affordable and easy to use in a DIY format. Modern embedded C compilers employ built-in features for keeping programs short and manageable and, hence, speeding up the development process Bolanakis (2017) The difficulty here, however, is that these built-in functions will not be universal, and very specific to the hardware being used, while the language I am using is extremely specific to this project, I wish to draw as many parallels to other microcontrollers, so the skills will be transferable. This approach is called Code More to Learn Even More and directs the reader toward a low-level accessibility of the microcontroller device Bolanakis (2017). This “motto” describes the way in which low-level code (certainly for educational purposes) should be written. This way, the user will have a full understanding of what is going on. The code that is actually written is incredibly simple and can act as the pre-existing knowledge that is built upon, as described by Smith and Webb (2000). It is what the user will do with this foundation knowledge that is important, as with low-level programming, there is very rarely one approach to a problem. This forces the user to really think about the problem, and once it is broken down into small enough pieces, it is completely manageable. This is the mindset that I want to be able to bring across in this project, if anyone using my language can begin to think in that way, then I will consider it a success.

## Low-Level Programming

### Low-Level vs. High-Level

Low-level and high-level programming languages are very much different kettles of fish and have very little in common. As such, it is impossible to say which is “better”, nor is this the goal of this section. With that said, there are pros and cons to both and different uses of the types of programming. These are the differences that this section is out to explore. Frampton et al. (2009) look to demystify the magic involved with programming. They state that “The power of high-level languages lies in their abstraction over hardware and software complexity, leading to greater security, better reliability, and lower development costs” This is an incredibly true statement meaning that the low-level programming languages have direct control over the hardware that is connected to a given system, and the power of the high-level languages are all based off utilising this power. The issue is that this power in abstraction removes the willingness of the compilation process to be seen by the programmer. This is shown by their later statement “opaque abstractions are often show-stoppers for systems programmers, forcing them to either break the abstraction or more often, simply give up and use a different language”. Here, they are explaining that as high-level programming is built upon low-level instructions, only commands that are deemed of use by the people making the high-level language, can be used by the developer. In some situations, the developer will need to do things that are simply impossible with a given programming language. This may force the developer to either break the abstraction (either by editing the high-level language or by writing raw assembly) in order to achieve what they need to. This is incredibly time-consuming, and expensive due to the specialism required to achieve this. Failing that, in some cases, the programming language does not and will never be able to do the operations required of it, and an entirely different language will be required to be used. The opacity is something that this project is out to address. The interpreter needs to be able to be completely transparent and explain each of the decisions being made. This is an essential skill for people who want to become specialised in this field.

In contrast, while the use of low-level programming is incredibly powerful, and can enable the developer to do almost anything with the hardware in the system, it is also the most difficult type of programming to do. This is the reason high-level programming was created: to make programming easier for humans, while also enabling computers to execute the instructions being given to it. This problem has been recognised and was addressed by Métrailler and Mudry (2015) who recognised the advent of off-the-shelf programmable embedded systems such as Arduino, which all are aimed at educating people how to use hardware, but not necessarily low-level programming. As such, some limitations exist, and this is something that Métrailler and Mudry look to rectify by building a block-based programming language being compiled directly to incredibly low-level assembly/machine code. This eliminates the need to learn a language such as C/C++ which contains extremely difficult concepts that the young demographic of these products will be spooked by and feel put off. This block-based programming was made popular by the Scratch Foundation (n.d.) who created Scratch, which is a coding language with a simple visual interface that allows young people to create digital stories, games, and animations. It is very intuitive and teaches people the basics of programming while doing something like creating a game, or animation, which (for its demographic) is a fun activity. While with this project, it is not realistic to have a visual interface, the idea of using something familiar to teach another concept is something that will very much be taken advantage of. The compile will accept the custom language which will be based on the syntax from a high-level language. In addition, the demographic of this project is someone with some programming experience, and it is this audience the language will be designed for.

### Digital Logic

Digital logic is the home of all low-level programming. It is common knowledge that there are billions of transistors inside even a typical mobile phone, but the average person is not really sure what this actually means. Digital logic is a form of thinking about how electrical signals are handled by a circuit. A common implementation of this is the TTL chip which stands for Transistor-Transistor Logic. Meaning that if you open up the chip, you will find these billions of transistors. If these are added together, you can form logic gates, such as AND, OR, XOR, etc. and by grouping these together, amazing constructs such as static memory, decision making, and mathematics can be performed. It is true that all a computer is is the abstraction of smaller things to create bigger and better machinery. Half of this project is all about creating a simplified TTL architecture to run commands on. This is a similar concept to what Rodriguez (1994) has investigated with his very oversimplified architecture. This too is very much an educational tool, and the results of which had very positive results, and due to its simple construction, was cheap and able to be used by any student that wanted it. Rodriguez did show some of the issues from his study, however. The main areas of negativity are the complexity of assembling the board, and general poor implementation. Apart from this, it overall shows a great result and encourages true promise to the project, which draws parallels to Rodriguez’s designs in that the project will also be cheap in construction, but will also learn from the mistakes made here, and focus on being simple to pick up and use.

### Relevance of Low-Level Programming

Low-level programming is still used a great deal in the modern-day, it is just a lot more specific than other higher-level languages which are easier to learn, and so more people are able to enter the industry with this experience. Some applications, however, will still require the experience of low-level programmers to create, from operating systems to general system architecture. While it is the norm to look at languages such as C++ and C# to create these types of applications, some hardware will still require intervention from languages like assembly to operate efficiently and correctly. The mere existence of recent books in assembly is a clue to this. Examples include Modern X86 Assembly Language Programming, Kusswurm (2014) and Assembly Language Step-by-Step: Programming with Linux, Duntemann (2009). These books were created to teach modern applications of assembly language, and most importantly, on modern machines, and architectures. Kusswurm’s book is actually written for an x86 architecture, which is the main architecture for 32-bit based operating systems (the name is from the 32 bit Intel 8086 microprocessor). These books are incredibly good reads, with a great deal of information about the usage of assembly in the modern-day. They are primarily based on learning assembly (not to a young audience), however, some of the techniques here will be perfect draw parallels to for this project. Furthermore, it would be pointless to make a completely new version of the assembly syntax, as such, this project will not look to reinvent the wheel, but to replicate the basic syntax, but simplified slightly for the target demographic.

### Applications of Low-Level Programming

Any computer or device uses low-level programming, after all, low-level programming is not a language in itself, it is more of a concept. Machine code is the only language that computers can use. Take Windows 10 as an example. While it is unfair to call it one piece of software, it is in fact layers of code all working to manage the PC’s resources. It can be thought of as a large piece of software called an operating system. It was created with the use of multiple languages, the kernel (the lowest part of the OS) was developed in almost pure C (a very low-level language). OS Today (2019) wrote about this and explained the primary use of C and assembly is for such a purpose.

More specifically, low-level languages such as C or Assembly can also be used for microcontrollers. They have the ability to directly access the hardware peripherals under the microcontroller’s control. This is because the languages allow for the actual manipulation of the architecture itself. This has advantages more than just being able to use all of the software, but it also improves the performance of the code to be run as the logic is not compiled or interpreted by another person’s software. It is your logic that is being run on the computer. When it gets to this low level, every clock cycle can add complexity and thus increase the time it takes for the program to run. LoopPerfect (2019) has a great explanation of this, the outline of which is that the reason higher-level programming is less performant is that generic low-level commands are required to be able to compile any generic program. Microcontrollers are a fantastic tool and are a very common use of low-level programming which this project aims to help facilitate the learning of.

## Microcontrollers

### Understanding

A picture containing graphical user interface

Description automatically generated The point of microcontrollers is not to be a fully universal tool for every application, but to be incredibly specific to the task at hand. A microcontroller used in a vending machine should not be expected to have some software changes and run in a NASA satellite. As such, it is imperative that the correct controller and hardware are chosen for the job. The perks of having such a specific system are that they will be cheap to create and manufacture in bulk, meaning that the technology can be more affordable and not have unused features. For example, if you buy the biggest and most expensive multitool out there, but you only want to open the occasional package, you won't use 90% of the other tools in its arsenal. It will be too expensive and large for the job it needs, instead, purchase a single retractable knife, it performs one job, it is cheap to purchase, and gets the job done just as well as the expensive one, if not better. A microcontroller is the knife in this example, a cheap and cheerful solution, and with a few subtle tweaks and additional hardware, can be the solution to any problem to any problem. The ability to chose which controller to use, and which hardware to use as the peripherals is the difficult part, and much research should be put into selecting the right tools for the job, as with any solution. Yılmaz et al. (2017) wrote an article to explain exactly this. They do this primarily by comparing various different types of controllers, peripherals, and architectures. The most interesting and relevant to this project however was the comparison of development boards:

Figure

These are great boards, mainly for educational purposes, but can also have industrial applications. While some of these boards are very well known and used a lot in education, they do very much hide a lot of the ins and out of what they are actually doing, but such is the nature of higher-level architectures. The raspberry pi, for example, is not a microcontroller, as it has a full operating system, “a computer” is a much better description of what it is. The wide range of technologies available in this area is made very obvious from this study. While the project is not aimed at advising the user about which microcontrollers and hardware to use, I do feel as though understanding the limitations given by low-level programming will enable them to make their own decisions and be able to carry out their own research as to what is available to them to solve the problem they are out to face.

### Applications

Microcontrollers are processors which have the capability to be used on a wide range of devices, from calculators to vending machines. While such devices like this on the outside are very much a hardware product, they would be completely useless if it were not for the software burned into the microcontroller itself. This is demonstrated by Tanshi and Bello (2014) who show the vast amounts of applications of the µController. The article is mainly focused on the implementation of a basic calculator using a µController as the core. As such, a big focus of the paper was about the hardware requirements for such a device, and this is where the power and flexibility of microcontrollers come into play. The µController is only as powerful as the peripheral devices it has available to it. Here, Tanshi and Bello are using hardware to create a user interface for a calculator, but could just as easily make one for others such as automatic transmission cars, fuels pump meters, washing machines, digital cameras, DVD’s, mobile phones, vending machines, etc. Just from this, it is clear that microcontrollers are all around us, and they will only ever become more powerful, flexible, and most importantly, still in demand. This demand drives up the need for people who are able to program such devices in the future. People who are used to only programming in a high-level environment will simply not have the know-how to make devices like the ones listed by Tanishi and Bello operate. One might argue that modern microcontrollers should have the ability to use high-level techniques, but this adds complexity, and cost. This thus size to the units themselves and begins to degrade the very point of having such a low level and specific device. The goal of this project is to explain to users how low-level programming can be achieved, thus helping to develop the next generation of low-level programmers and to spark interest in this dying art.

In addition to consumer products being powered by microcontrollers, research projects have been undertaken by countless academics and companies to better their products or research with the use of microcontrollers. This is due to the relevance, availability, and simplistic nature of the microcontroller which shows a true reflection on how good the technology is. Shao (2006) created an Improved Microcontroller-Based Sensorless Brushless DC (BLDC) Motor Drive for Automotive Applications. He developed a cheap and effective new way to monitor the back EMF produced by electric motors which can more accurately monitor the motor positioning, while also providing overcurrent protection. Zwerg et al. (2011) created a new form of microcontrollers for energy harvesting applications. It draws only 0.000082Amps (82µA) which makes this technology perfect for applications in energy harvesting where every nano amp drawn, counts. Babiuch et al. (2019) Even take a look at the applications of the internet of things with embedded technology, which is something that everyone will soon be using, and most of the time without the user's knowledge. As shown, there are countless applications for this type of technology, which are aiding in the progression of many different industries that are used by all types of people all over the world. The applications will only continue to grow with the development of new technologies and industries that we don’t even know about today. As such, it is imperative that people remain clued up in this area to keep the up-and-coming technology developing. This project is aimed to achieve this education upkeep, as low-level programmers are becoming few and far between with high-level technologies being easier to understand.

### Relevance

With the knowledge of what microcontrollers are, and how old the technology is, one may begin to wonder what the point of learning this old form of thinking is. With great advancements in technology, surely the use of these small and specific devices will phase out, and no one will need to know any low-level programming languages or logic. Mitchell (2020) made an argument for this being far from actual fact. His primary arguments consisted of the incredible complexity of larger, more powerful microcontrollers, along with the vast increase in price which is attached to this. He concludes by stating that the industry shows no sign of a reduction in 8-bit usage, and their low-cost nature combined with simplicity still makes them highly relevant. In addition, Pal (2017) covers some very interesting information about the reasons for using an assembly language (a very low-level type of programming). He shows how the use of assembly offers complete control over the system’s resources, direct access to hardware, a deep understanding into how the computer actually operates, complete transparency, and more. These points are extremely potent in this project, as it aims to address each of these positives for learning low-level programming. The whole point of the project is to be as transparent as possible while also providing a somewhat powerful outcome. All of the hardware will be available all of the time, meaning the user can fully grasp what everything does and will have the ability to play around with all of the parts. As shown, this technology is still rife in the modern world, despite it being old. This gives me great confidence in the project and only strengthens the fact that this is needed and there is a serious gap in the market here.

### Uses in Education

The point that microcontrollers are not taught very much in education due to their complexity, is a major driving factor of this project. “Microcontrollers account for the majority of processors produced today, yet their capabilities are seldom explored in modern computer science curriculum” This is a sad quote from a study by Margush (2006) which goes into the importance and relevance of assembly language. This quote is one thing that this project intends to address. The study by Margush continues to talk about different ways in which to compile assembly code. One example of this is for the Intel 8080 microcontroller “Windows XP platforms to write, assemble, debug, and execute Intel 8088 assembly language programs”. The issues that are raised here regard the way in which the OS allows access to low-level things like I/O and interrupts This is a theme that is shown across a wide range of assemblers (the compilers for assembly code). This problem with transparency is again a problem that prevents people from gaining a further understanding as to what goes on inside the computer. The usual phrase used is that “the code is turned into something the computer will understand”. The project will ensure that there will be complete transparency in the workings of the compiler, showing the user the output of stages as it goes and shows the final output, so the user has a complete view as to what is going on.

## Summary

This section discusses the current literature surrounding the area of education and applications of low-level programming and how the lessons learned here can be applied to this project.

The education of low-level programming has become rather limited and is barely taught at all due to the invention of high-level languages. While this is not necessarily a bad thing, it does introduce some potential limitations for advanced developers who have reached the extent of the functionality of high-level programming when it comes to fully utilising the resource of a PC; Integrating hardware into their projects; or simply executing an operation which is not possible given the constraints on the high-level programming due to its very nature of abstraction. It has been shown that there are great benefits of teaching programming and starting low-level principles will improve anyone’s literal and logical thinking which is incredibly transferable to other STEM-based subjects.

The application and usage of microcontrollers is also a primary talking point of this chapter. Their relevance in the modern day is still very much evident with almost every device in the world running some form of microcontroller with low-level software written on it. As such, the need for people who understand this type of programming will only increase. The performance of the low-level direct hardware addressing is a large driving factor for this. With hardware, for example, the process cannot simply be paused, the controller must be available to process any input it provides, or critical failure could occur.

There is most definitely a place in this world for high-level programming, and this chapter proves this, but it also analyses the other side of the argument whereby low-level languages should not be forgotten about and thought of as simply another layer of abstraction that is too complicated for anyone to understand. This would be incredibly sad, as the future of computing would be much less performant, and it would be near impossibly to acquire full implementation of the latest and greatest in peripheral technology.

To finalise, both forms of programming have a place in this world, but respect should be given to both, and not to forget about the very foundation of modern-day computing, as without it, even today, almost every modern commodity would not exist.

# Project Planning

## Introduction

This chapters takes a look into the approach taken for the project along with requirements that are needed to be hit in order to consider the project an overall success.

In addition, some of the first and most basic decisions will made with regard to the approach of implantation and justification as to why I have chosen them. Furthermore, I will also consider some of the legal and ethical issues surrounding the project.

## Methodology

The methodology to be used when running a project is a huge one, and one that cannot be taken lightly. While it is incredibly important to create and hit all of the requirements to make the project a success, given the complex nature of the project, and the strict timescale attached, it may become necessary to void some of the less important requirements. These requirements will be mainly those which are not required to make the project work, but instead are useful to have and makes the use of the project easier. These are indicative of the use of Agile, as the MoSCoW approach of requirements gathering provides some leniency as to missing some of the less important requirements.

This project will however be worked on only by one person, and as such, fully utilising the agile approach is not an option, as this is aimed more at team working and constant management with a customer (*Manifesto for Agile Software Development*, 2001) .With that said however, I am looking to utilise some elements of the agile way of working such as the MoSCoW approach of requirement gathering (as discussed above), solo scrums, and time estimates which allows for frequent re-evaluations of progress to avoid the quality of the overall project being the area which takes the hit in quality if time is becoming a major constraint.

## Requirements

My goal is simply to create a programming language which acts as a tool to make my custom built 8-Bit computer into a Turning Complete Machine (*Turing Completeness*, 2021). As such, my primary requirements will be to provide the user with enough tools to enable any solution to be created for any problem. This is achieved by using the 3 basic principles of programming: sequences (instructions), selections (if statements), and loops (while, for loops etc.). With these tools, in theory, any solution can be created to solve any problem (Turing, 1937).

Using the above mentioned MoSCoW technique, I will set out my requirements:

### Must

1. Be a language that use high level syntax with parallels to existing languages to enable a shallow learning curve for new users of the language and aspiring low level software engineers
2. Enable the steps of compilation to be explored at the users own pace and thought process and for a bespoke experience every time
3. Convert the high-level language into the bespoke and simplified assembly commands, and the resulting binary which is compatible with my custom 8-Bit computer
4. Be considered at least a Turing machine capable of solving most any problem within the obvious constraints of an 8-Bit system.

### Should

1. Be a single executable for use in the user’s very own compiler every time giving them control and understanding of the steps needed to be taken in program compilation
2. Be able to provide the user with useful and correct feedback as to syntax errors when compiling the program
3. Have built in garbage control to prevent data leaks and allow for easy memory management within the language

### Could

1. Have additional built-in functions for the language specifically for this machine
2. Have a P/S 2 keyboard interface converted to ASCII numbers for general use in programs

### Would

1. Have a built in FIFO stack to enable the use of function calls

## Potential Solutions

Given the demographic of this project, languages such as C, C++ or GoLang would be obvious choices, what with their already low-level nature, and compilers even nowadays are written in these languages too. The difference here is that these languages are low level in a way that is close the hardware they are already running on (the modern computers of today). This is why low-level languages are so specific to the machine they are running on. With this constraint of being close to the wrong type of hardware being removed, the only hardware that the code needs to be close to is the hardware of the 8-Bit computer. I have no issue with utilising the modern processing power of the computers of today in order to enable the learning of hardware programming. I am therefore more inclined to use languages such as Python, Java, or C# which are very commonly used, heavily documented, and easy to get going with. At this point, my own personal preference will come into play, and Java is my preferred choice.

Not only is it the language I am most familiar and comfortable with, due to its compilation method (using a Java Virtual Machine (Hirt & Lagergren, 2010)) it can run on nearly every modern computer, meaning that anyone will be able to run my code. In addition, Java has a great many powerful tools and features built in.

## Tools and Techniques

Java has a great many powerful tools and features built in, and I intent to take full advantage of this. My main focus is to make the compiler as modular as possible, and with the interest of meeting my first “should” requirement, the project will not be a complete compiler. The project will consist of all of the tools required to create a compiler, but it will up to the user to create it however they see fit. The primary tools that I will supply to the user are:

* Lexical Analysis
* Syntax Analysis
* Code Generation

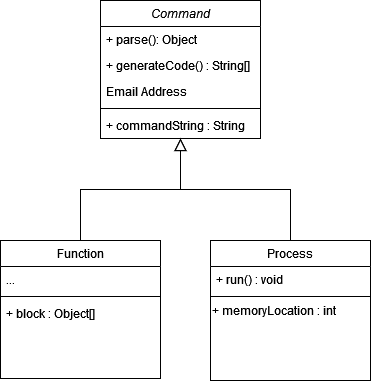
You will notice that these are also the 3 steps to most every compiler. The user will either be able to use built in debugging features within their IDE to stop the compiler running at those given stages and analyse the inputs and outputs, or will be able to call output functions within the executable and output the results directly to the console. This modularity will allow users to create their own compiler and work at their own pace to full understand what is going on in the usually shut box (compiler).

The primary tools and techniques that I wish to utilise within Java to make this achievable are:

* Object Orientation
* Recursion
* Polymorphism

I plan to keep everything in organised objects, and the hieratical ordering of the commands after the lexical analysis will be stored in arrays of these objects. Furthermore, I plan for these objects to be stored in blocks, in the same way blocks of code work in any other language, usually depicted by the open and close curly brackets “{}”. I will then use recursive techniques to parse each of the functions and commands in the blocks at the syntax analysis stage, which means it is possible to have an infinite number of blocks within blocks and so on.

One constraint of Java which is also what makes it such a good language is the use of types. One example of when these are used is in the creation of data structures like ArrayLists and Queues. If the language only allowed the user to undergo one function, this would be no problem, as there could simply be a single array of a custom object type. However, if this was the case, the language would not be good enough to enable the 8-Bit computer to be considered a Turing machine, and would not be acceptable for this project. Instead, I will have an abstract class containing a generic “Command” for example, and the master array can simply contain a list of these classes of which the function and process class can extend, and be stored in that same array to keep everything together.



Figure

The elegance of polymorphism here is evident, as both of the child classes (“function” and “process” in this plan) have a “parse” method inherited from the parent abstract class, but each one will override this parse function and do different things, although fundamentally they can both be considered a “Command”.

In order to implement these tools and techniques, a variety of existing tools must be utilised. The main tool is the IDE – the IDE of choice will be VS Code, this is a lightweight version of Visual Studio, and allows for additional features to be added if required. The idea here is that you are not forced to have anything installed – taking up storage space, and potentially slowing down the computer (*Documentation for Visual Studio Code*, 2021). As Java will be the main language, the Java addon will be installed which integrates the JDK in with the main code editing area.

In addition, as a way of loading OS and program data onto ROM chips, an Arduino will be used. As such, the official Microsoft Arduino extension will be installed onto the VS Code environment.

## Legal, Social, and Ethical Issues

With any software that relies on the input from users (the GNG source code in this case) there are many possibilities for incorrect inputs to be made. It is my job to have nothing unhandled if syntax or logical errors are made in the source code. The source of the compiler will be unavailable to anyone using the tool, so they will not be able to adjust if they find an error in my work. Therefore, intensive testing will be required to ensure nothing is left unhandled. In addition, the easy option for the developer of the compiler is to provide feedback that is not relevant, or an umbrella “an error occurred”, but messages such as this are not help to anyone, and should be avoided.

In addition, as the demographic of the end product is a student, I do not want to discourage them from learning more about the topic or using my product and deciding that it is too difficult and look to change their ideas for their future. I will do this by having accompanying learning materials which will explain the process. In addition, the compiler will be fully modular and can be hacked around to get different results and allows for the user to get a full understanding of what is happening.

## Summary

Planning the project has been very useful in making some large decisions surrounding the operation of the project, and what tools are available to assist in the project’s development.

An outline of the decisions made are as follows:

* Use a solo agile approach with personal SCRUM’s and MoSCow requirements gathering
* Use the VS Code IDE for general development purposes due to its extensive array of tools available and modularity
* Make good use of advanced programming techniques such as polymorphism, and recursion
* Keep the program linked to the requirements, so not to bloat the project with unnecessary features.

# Design

## Introduction

The purpose of a compiler is simply to translate one script to another lower-level script which is closer to the hardware of the computer (Nisan & Schocken, 2021). As such, there are really two major areas which need to be designed: The higher-level syntax that the user will interact with; and the lower-level syntax and logic that will be act as instructions for the hardware to run. Both are as important aspects as each other, and they will need to be gotten right in order for this project to be a success.

The high-level syntax (for gaungau) will need to be easily readable by the programmer, and should have some resemblance to existing language so it can be a pick up and go experience, rather than a pick up, read a manual, and then go.

The lower-level syntax however is where some performance techniques will need to be used that do not limit the power that I hope to offer from this tool. A balance must be struck, and a great deal of time, trial and error, and patience will be needed to develop the best solution for any given logical scenario.

This chapter takes a look at the design choices involved with the primary aspects of the language and assembly language. Although further choices will need to be made during implementation, it is important to go into a large project such as this with some plan to hit all of the requirements decided on earlier in the planning process.

## High-Level Design

### Tools and Syntax

In order for the project to be considered a success, I must meet all of my requirements, the basic ones which I base my entire project around is to create the functionality in my programming language to allow the user to use sequences, selections, and loops. These are very simple, and almost every algorithm can be broken down into these key attributes.

The Backus–Naur Form (BNF) is a way of describing syntax grammars with no context, meaning that while the danger of garbage in garbage out with this notation is true, if done correctly, it will describe the different areas of my syntax for the language in a standard and know format.

#### Sequences

##### Syntax Decisions

This is simply a single command which could involve arithmetic, or manipulating memory on the computer. As such, in order to hit this criteria, I will have the ability to set variables to values and do arithmetic within the same line. An example of this in Java is:

int num1 = 5 + 9;

This will obviously simply set the memory address associated with “num1” to the sum of 5 and 9 (=14). While this again, is a very simple task, it is one of the fundamental things you can do with a computer.

The above example is in Java, and this is a language widely used by a great many people with syntax that is easily readable and would be a great language to base my initial design off, however, I will be making a few tweaks as the functionality of Java greatly exceeds my plans for the gaungau language. One example is that my language will only be able to work with integers, and not other types such as chars, or floating point numbers. This might be a good way in which I can improve the project in the future, but for my purposes as of now, I will focus on the use of integers. Due to this fact, I will remove the need for defining the datatype of the variable. This decision will increase the ease of use of the language, however, it might not be obvious initially that this is required, and will need to be made paramount in the documentation.

While this is all true of Java, in other languages such as python, the datatype does not need to be set when defining a variable (*3.10.5 Documentation*, 2022). A user coming from a python background might find this an easy change to make.

Another design choice I decided to make (pretty much just because I can) was to make the equals sign of setting the variable a bit different. Instead of using just the standard “=”, I am going to use “:=” instead. This is in reference to the Dafny language (Koenig & Rustan, 2016) which I grew to like.

As a result of all of my design choices, the syntax I am settling on is to use the following:

Num1 := 5 + 9

Notice here, there is no datatype choice, there is the different assigning syntax, but apart from that, it is most like any other language, and anyone can read the intension of this line is to set the variable “Num1” to the sum of 5 and 9 (=14). This is just one example, but covers most every possibility when it comes to variable assignment such as:

* Assigning variables to each other
* Assigning variables to calculations containing other variables
* Assigning variables to even longer calculations such as 1 + 2 + 65 + 30 – 70 – 4 + 2 + 2 + 6 – 8…

I am happy with how this syntax reflects this feature, and if implemented correctly will meet this requirement

##### BNF

<symbol> ::= <expression>

<expression> ::= <integer>

| <symbol>

| <integer> “+” <expression>

| <integer> “-” <epression>

| <symbol> “+” <expression>

| <symbol> “-“ <expression>

<symbol> ::= <charater>

| <character> <symbol>

<character> ::= <integer>

| <letter>

#### Selection

##### Syntax Decisions

In programming, selection (or branching) is the ability for the computer to make decisions based on existing conditions and values (*Selection in Algorithms - Selection - KS3 Computer Science Revision*, 2022). This is by no means the computer making its own decision in some kind of AI capacity, but in a term more widely known: an if statement. If statements are incredibly widely used in most every language, and the syntax is almost identical for each language following the rule “IF<condition><block>”. I intend to use this same structure for my language and not deviate too much from this, in order for the learning curve to be as small as possible.

Furthermore, another consideration is the use of the “else” statement. One could argue that the use of the else statement is entirely possible using only if statements as seem below in the Java example:

if(num1 == 5){

    // Do something

}

if(num1 != 5){

    // Do this instead

}

As we can see, there is no else statement in use, and using num1 as 14 from the previous examples, the second if statement will be executed in the exact same way an else statement would be used. While this much is true, I can’t think of a single language that can support an if statement with that type of syntax, that cannot support an else statement. With this in mind I will be implementing an else statement in my language, and will try to keep the syntax as close to the following as possible:

If(num1 == 5){

# Do something

}else{

# Do something else

}

That is the very syntax in gaungau that I wish to implement. It is identical to other high-level languages out there, and will be quick for any new user to pick up and create solutions with.

##### BNF

<if statement> ::= “if(“ <condition> “)” <block>

<block> ::= <statement>

| “{“ <statement> “}”

| “{“ <statement> <statement> “}”

<statement> ::= <variable>

| <statement>

#### Looping

##### Syntax Decisions

Looping is the ability to loop over and repeat code but with a potentially new state than the previous run. All it means is that the block or statement of code should be run while a given condition is true.

Below is the java implementation of a while loop. Here, the block “// do something” will be run as long as the “counter” variable is less than 10.

int counter = 0;

while(counter < 10){

        // do something

}

Currently, this would be an infinite loop, as there is no logic inside of the block, and the counter will always be 0 which is less than 10. This type of infinite loop can be dangerous on modern systems, as they are executed so quickly, that all of the system resources go to it. In addition, if in each iteration, a new resource is created, it can fill up RAM, and quickly force a disk partition to be used in some cases, leading to the computer becoming incredibly slow, and eventually can become completely inoperable. While in modern computers, this is a big risk, and could potentially cause data loss, and damage to hardware; it will not be such an issue with this project. There are many reasons for this, but the primary reason is that the user has full control over the hardware, and can simply stop the clock if needed. In addition, there is only the ne program being used, and so even if the program uses all of the computers resources, it will not become slower at all. All of the memory is random access, so takes the same amount of time, and will continue to run at the same speed until the clock is stopped. The only risk here is that there are only 255 available addressable memory locations. If at each iteration of the while loop, a resource is created, after the 255th iteration, data loss will occur.

Loops are a very useful tool, but also have the potential to cause a great deal of struggle in debugging if used incorrectly.

The syntax I will use for looping is very similar to that of other languages (like the Java one above):

count := 0

while(count < 10){

# do something

}

The actual while loop and condition are identical, only the comment, and variable syntax differ here. This makes the learning curve rather shallow, and anyone with some experience with other languages will be able to pick this up very easily.

##### BNF

<while loop> ::= “while(“ <condition> “) <block>

<block> ::= <statement>

| “{“ <statement> “}”

| “{“ <statement> <statement> “}”

<statement> ::= <variable>

| <statement>

#### Custom Tool

The result of all of the main processes listed above, is going to be a custom-made instruction set of assembly instructions, based loosely on the syntax on an x86 system. This will in turn be converted directly into binary which can be run on the custom 8-bit computer. While this is a completely bespoke language and cannot be run on a standard modern computer of today such a windows machine, I want to enable the people who want to take their understanding one step further, to do so. The user will have full access to the outputted assembly code, and them are free to manipulate this as much as they see fit, however, free hand editing is not feasible really when it comes to sharing their code with friends, or to a classroom of fellow classmates. I want to enable full transparency, and for a powerful function to exist which can edit the very assembly code without a check but the compiler other than a basic syntax check. This function has the ability to throw off an entire program, and cause data loss, counters to clear accidentally, and literal control over every aspect of the computer. On the other hand, it will also allow someone who knows what they are doing to create more performant programs than even the compiler will be able to produce. The main reason however is simply to allow the user to play around with all of the hardware they have in front of them without constraints of error messages, or someone else telling them what to do which is fundamentally a large factor of modern-day high-level programming.

As the function will enable the alteration of the raw assembly code being produced, I will name this function “raw()” it will accept one parameter which will be the single instruction that will be inserted blindly into the code generation.

For example:

Num := 0

While(Num < 10){

Num := Num + 2

Raw(OUT) # Could equally be print(Num)

}

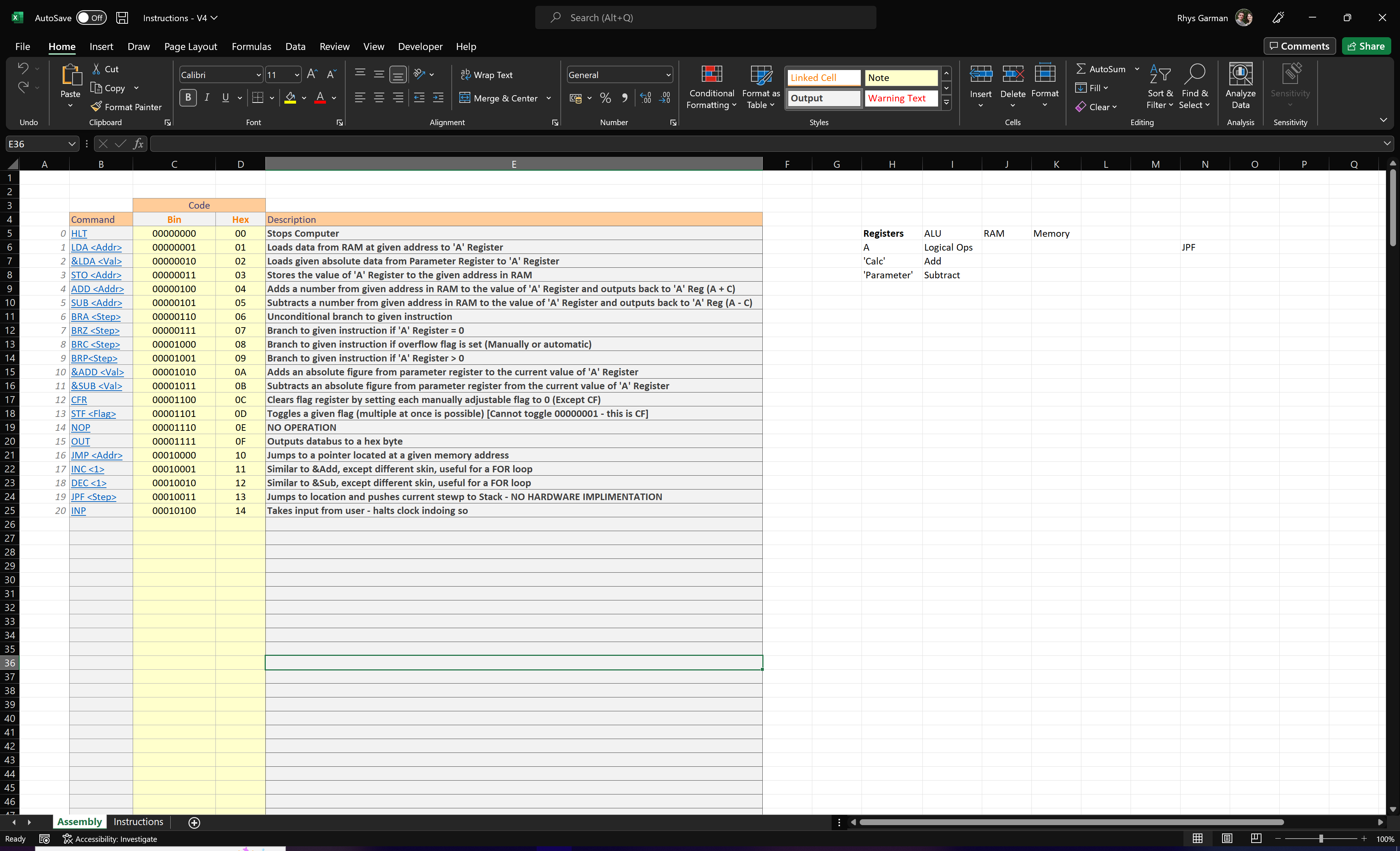
Here is a simple algorithm to figure out 10 / 2 however, the outputting of the number is done via the use of the physical assembly instruction “OUT”. I intent to include a basic syntax check, so the parameter "GOLDFISH” would return an error, this is only to ensure that the program is full runnable no matter of the outcome. This is a simple example, but shows how powerful it could be if used by the right person.

## Low-Level Design

### Syntax

The syntax of low-level language is not so well known, and as such, there is less that I can use to base my design decisions on. That said however, while not widely used on the outside nowadays, a very common assembly language is the x86 instruction set (*Nasm Instruction Set Reference*, 2020). The premise of this is a set of (usually) three letter instructions (with a few exceptions) which are character representations of the operation code associated with the operating system to tell the computer which set of binaries to execute at a given time. The compiler will produce assembly codes following this same paradigm as it has been proven to work and will aid is reducing the learning curve into assembly languages, and assemblers like NASM (*Nasm Instruction Set Reference*, 2020) and MASM (Corob, 2022) which are run on windows systems.

In addition, there is parallel between my planned low-level assembly syntax, and that of the commonly used assembly simulator called LMC (Little Man Computer) (*Little Man Computer - CPU Simulator*, 2014). This means that any code written for LMC, (with a little bit of modification) can be run on the computer with the same expected outputs.

The plan below for the assembly instructions planned to be implemented in this project. With just combinations of these instructions, theoretically, any program to solve any program can be created (with the associated lmitations of an 8-Bit system e.g. available memory, and clock speeds)

### Logic

The low-level code was written to cater for only one function, and control over one module at a time. This means that it is easier for the user to understand exactly what each instruction is doing. The binaries attached to each instruction will be made available, so people who wish to dig deeper, can dig all the way to “bedrock”.

This is in direct contrast to the goals of high-level languages which wrap lower level instructions into function calls, and statement (Cneude, 2021). This is not a bad thing, but it does mean you have less control over exactly what goes on. Take for example, the if statement. I plan to use only 3 checks, and combinations of these can make any condition that is used. Take for example “3 == 4”. To check if these two values are equal, the program will:

* Calculate the difference between 4 and 3
* Branch if the difference is 0

Those steps are abstracted into the two characters: “==”. The assembly instructions here are:

&LDA 4

&SUB 3

BRZ [Start of Block]

BRA [Skip Block]

Is it clear now that the instructions do only one thing each, and it is the layers and combinations of these individual functions that comes together to create bigger and better things.

### Performance vs. Universal solution

One of the main disadvantages of high-level languages is that they are very inefficient as they are so far away from the hardware. This is a consideration which must be thought over carefully. The point of this project is to create a low-level platform upon which leaning can be done, do to create the associated language with many obvious built-in inefficiencies would be counterintuitive.

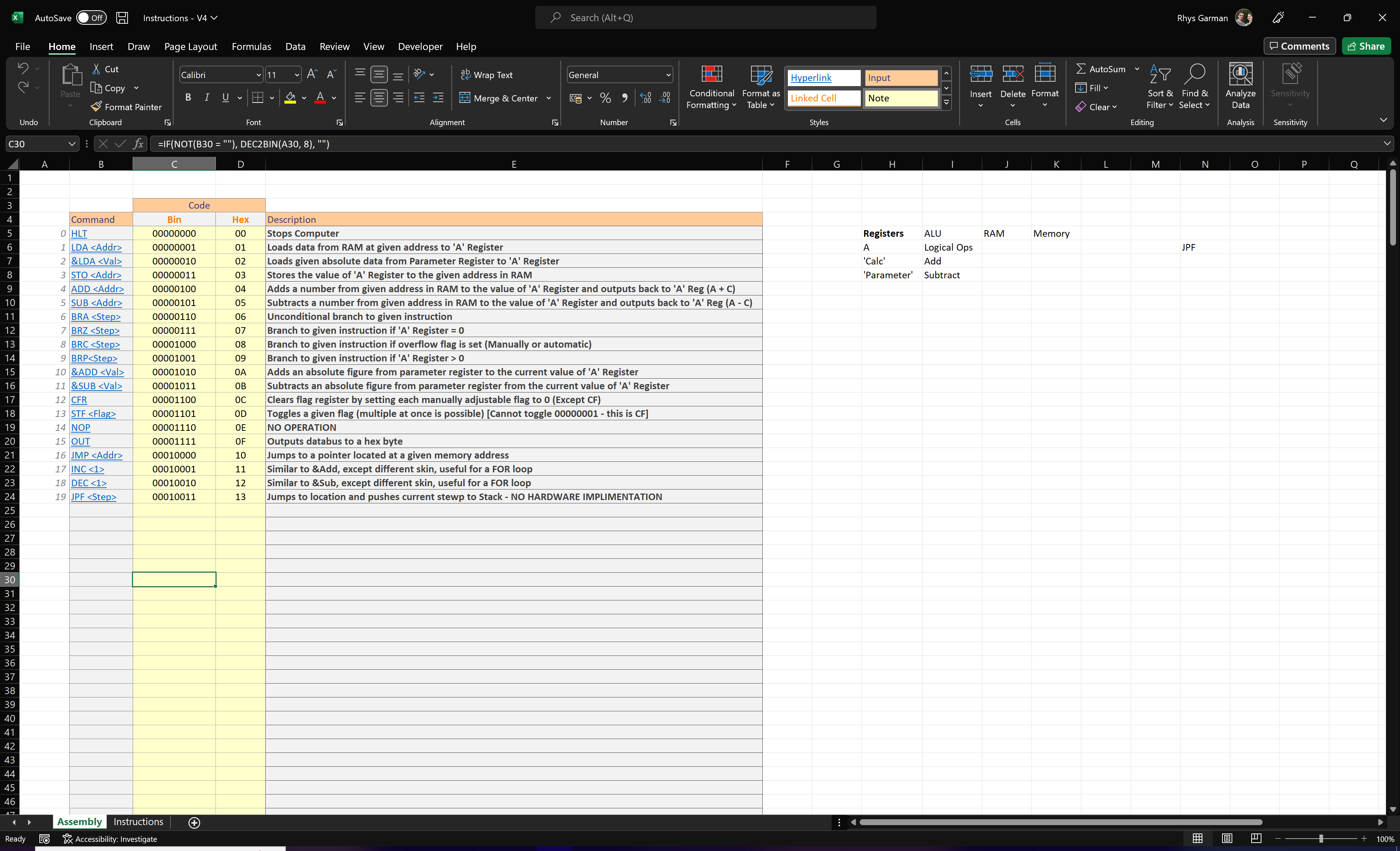
This type of problem is one which can only really be discovered when undergoing the implementation phase of the project. As such, it would be almost impossible to identify them at this stage. Saying that however, the mitigation of these issues should be considered from the off, and measures put in place not to prevent them, but to limit the use of inefficiencies.

With the language being so very close to the physical hardware of the computer, it will fundamentally be more efficient than other languages, it will however have most of the same problems faced by other languages. A huge advantage that this computer has over others, is that the hardware is designed specifically to execute this language, meaning that hardware changes could be made to make the language run more efficiently, and potentially removing the planning inefficient methods all together.

## Binary – Where Software meets Hardware

Every computer will require an operating system to run, even my little 8-Bit computer. A standard OS will be able to manage the goings on of the computer, and control programs to run as and when needed (*Computer Basics: Understanding Operating Systems*, 2020), however my OS will only be able to manage the running of one program at a time. While this is very simplistic, it is all that is needed at the moment. I plan in future to create a graphics card to be able to use any VGA capable monitor to display information on, and eventually to create a WYSIWYG interface on which to create programs and trigger the execution of programs. At such a point, a more advanced operating system will be required, however, for now, the ability to manage the one program is all that will be needed.

The operating system will need to be written in a way that the computer will understand, the user will not need any knowledge of this, it will be the final interface between the user and the computer. Everything past this point will be pure binary in a format that is almost impossible to write, and most importantly debug. The binaries which make up the OS are stored on ROM chips mounted to the computer. This may seem like a complex task for a few 1’s and 0’s to accomplish, but it is the very foundation upon which everything about this project relies. Every high-level command, like an if statement, is translated into a bunch of assembly instruction, which are translated into another list of binaries which are simply symbols of electrical charge. Pins on the logic chips in the computer can detect these voltage differences, and active/deactivate upon these triggers. The charges are stored inside the ROM chips, or more specifically the EEPROM (electrically erasable programmable read only memory) chips. These are a very basic form of non-volatile memory; in that they do not lose the data when turned off. This type of memory storage was the main bottleneck of old computer systems, it was very expensive and not hugely reliable (*Overcoming the Memory Bottleneck*, 2019). Nowadays however, one chip which can store 8192 bytes of data costs only around £2, and is the technology used here to store the OS.



The above is the complete list of assembly operations which can be run on the computer. The command on the left is the instruction in “word” for, and is one step easier to read than the raw binary. The number value on the far-left column is the opcode to the instruction, and the Hex conversion is the actual opcode referenced when converting the assembly code to instructions e.g. LDA -> 0x01 and OUT -> 0x0F.

The opcode refers to the address in the ROM which contains the list of binary instructions to be executed.

Calendar

Description automatically generated

Above is the list of binaries associated with the LDA function which takes a value from a given address in RAM and loads it into the A Register. While on face value this seems a bit overwhelming, and complicated, all of the binary is hidden behind several layers of the language, and a good experience can still be had with this learning tool even without looking at the binary. For those who want to go deeper however, the project has the ability to do so with the complete transparency that users of the artefact can become accustomed to.

This operating system was designed ahead of time, and again is open for modification by anyone who knows what they are doing. That is the power that this project is providing and will give users the full experience of how computing actually works.

## Summary

In this chapter, design choices made have been outlined and justified for areas such as creating the syntax for the custom language and the operating system. In general, the reason these solutions have been decided upon, are to make the learning of this new low-level programming as simple as possible. As this is an educational tool, the project on a whole is not there to prove anyone’s ability or skill, but instead, to teach others how to learn this dying art of understanding exactly what is happening inside the box, and offer tools which no other language allows you to do, such as actually editing the assembly code that is being generated.

# Implementation

## Introduction

The creation of the compiler and computer was a huge task, and it is not possible to talk about each component in detail here. As such, the primary aim of this section is simply to outline the main processes involved in the compiler, some of the key aspects and interesting techniques used, along with large issues encountered with solutions.

In addition to the compiler, the creation of the computer is something that was no small feat, and is touched upon here. It is important that this explained, so that the software side of the process can be matched up with the physical hardware that is being programmed.

## Main Processes

### Lexical Analysis

The lexical analysis is the first and most basic step of compilation. In simple terms, this stage is all about turning the code writing in, near enough plaint text, into a format which is standard, and can be understood by the syntax analyser (the next step). My lexical analysis took a few small and basic steps to complete its analysis:

1. Have each instruction on different lines including multiple instructions per line with the use of semi colons or line breaks.
2. Remove comments
3. Moves each brace to a new line
4. Removes blank lines

An example of the code running through the lexical analysis is:

Num1 := 0; Num2 := 8

If(1 == 1){Num1 := 5}

Else{

Num2 := 5

}

Which will be converted into:

Num1 := 0

Num2 := 8

If(1 == 1)

{

Num1 := 5

}

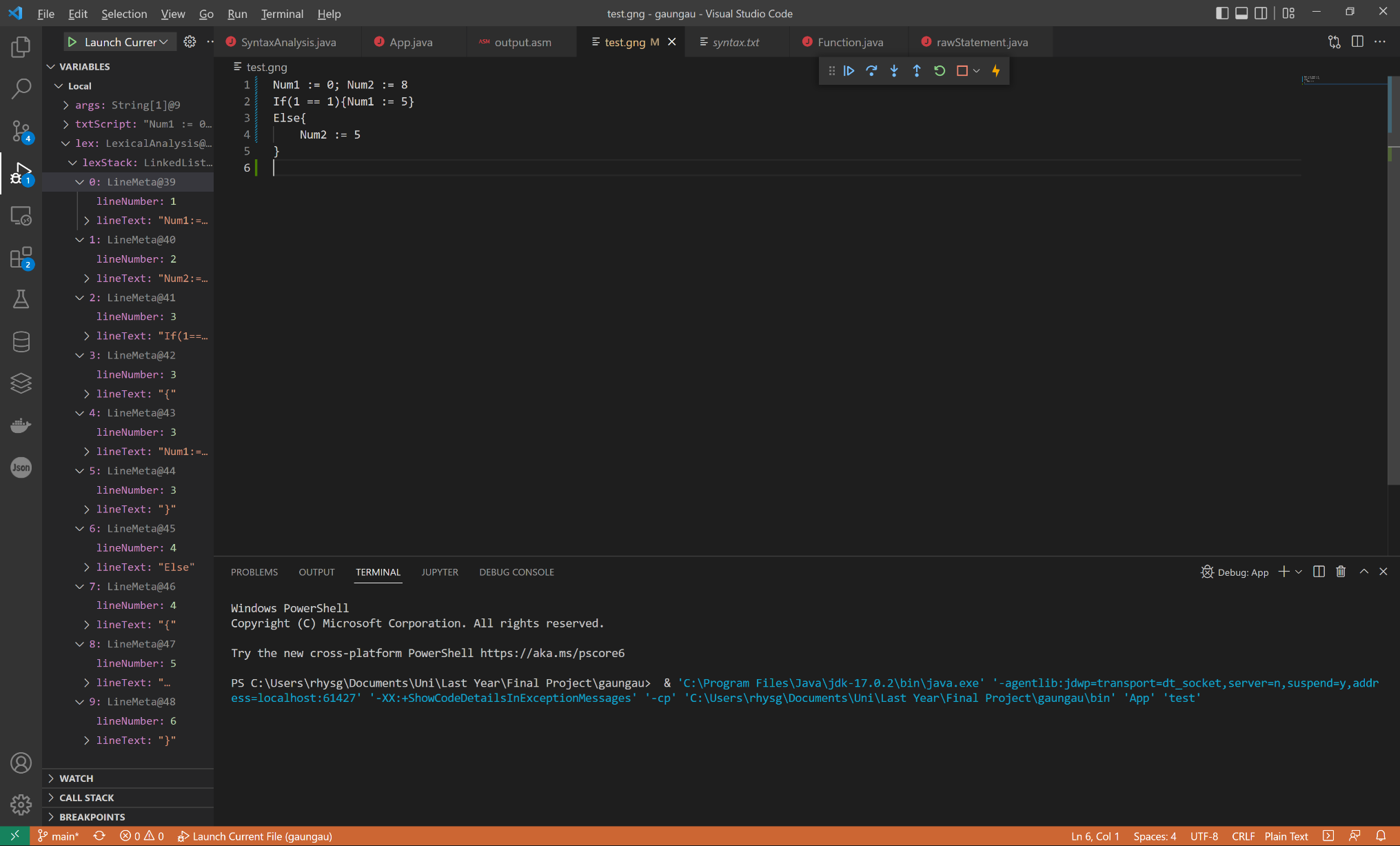
Else

{

Num2 := 5

}

I will use this example in the actual compiler, showing how the user will be able to see exactly what is going on.

We can see the code on the right entered exactly as seen above. And the lexStack in the debugger window on the left is the output of the lexical analysis.

NB: Note here, the actual ability to do this. Someone can stop the program at any point, and see exactly what is happening.

We can see that each of the elements in the lexStack contains an object called a “LineMeta” This contains the actual raw code along with the line number in the file associated with that command. This is useful for error messages, so it can figured out exactly which line on the real that is causing the issue.

We can see that each of the elements have been identified and put in its own object to be executed later.

Also note, that this language is case sensitive, so the keyword “Else” is different to the keyword “else”. At this point however, the compiler does not care. It is simply putting the code into a format which the later steps expect and can accept.

If this code was to be fully executed, there would be many syntax errors due to the upper-case keywords used throughout.

While this step is as important as any step taken to compile the code into binary, the actual code itself is fairly uninteresting, containing mainly a vast amount of string manipulation

### Syntax Analysis

The syntax analysis is where the program created is checked for any errors in syntax and logic, and the resulting valid and correct code is then converted into a hierarchical tree to show scope and blocks.

#### Phase 1 Syntax Check

Each line is checked for the presence of:

* Else statements
* Variable setting
* An open block parenthesis (“{“)
* A Close block parenthesis (“}”)
* A function
* Any keyword
* Built in functions
* Creating a new function

In that order. At this point, the compiler flags any syntax errors, or logic errors such as floating closing parenthesis. Take the following example:

x := 0

y := 1

else{

test := 9;

}

while(1 == 1){

print(y)

t := y

y := x + y

x := t

}

}

Here, there are 2 errors which should be flagged at this point, highlighted in green is an else statement with no if statement attached. This is a logic error, although the syntax is technically correct. In addition, the closing bracket highlighted in red has no matching open parenthesis, and as such will also be flagged. Helpful error messages will be displayed in the console to indicate to the user where the issue(s) are:

Syntax Error: Invalid placment of else statement at line: 5

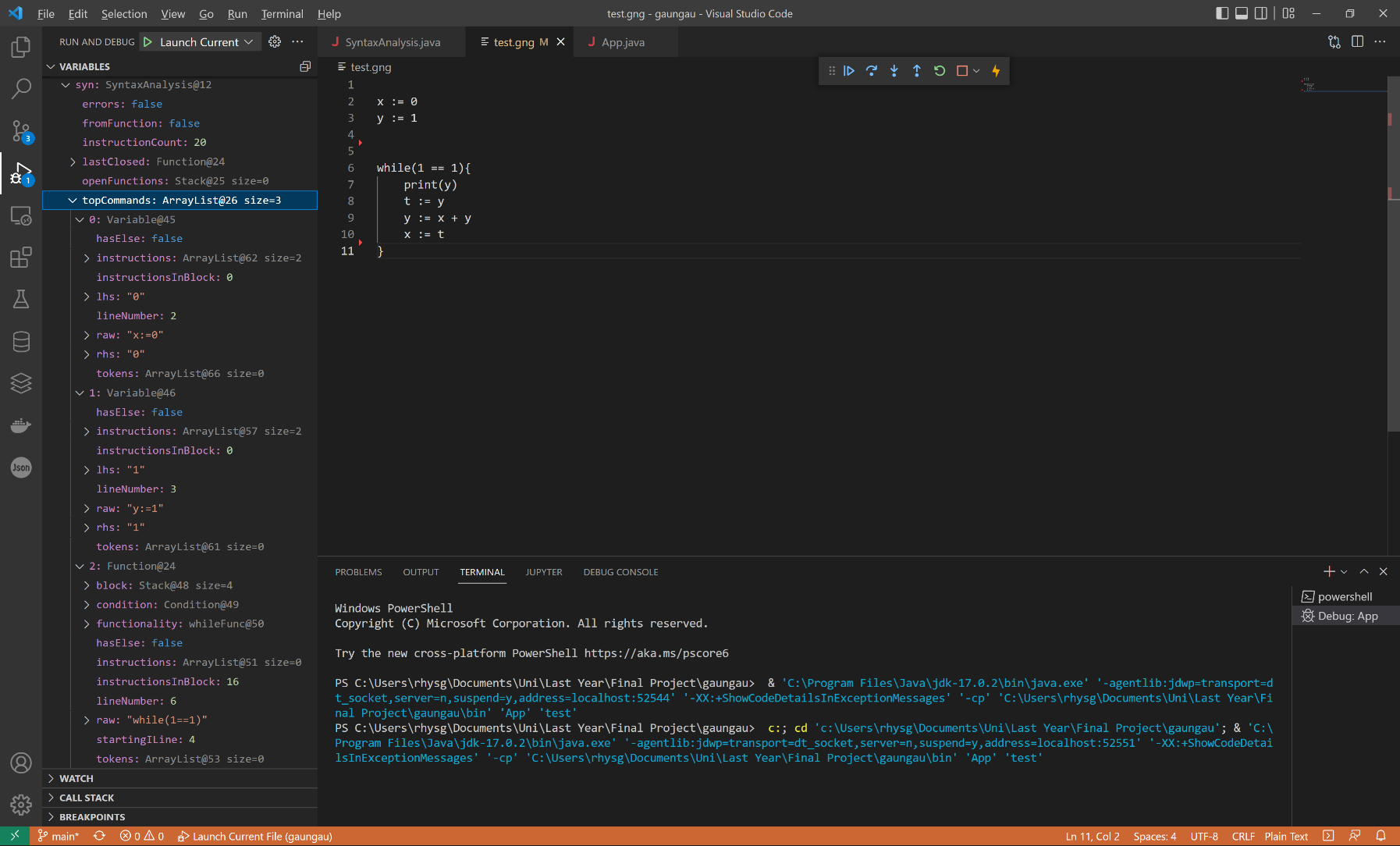
Syntax Error: Stray '}' at line: 18

#### Sorting into Blocks

After the code has been transformed into a state that is uniform and expected by the compiler, it can now take a look at sorting the code into blocks and creating the objects linked the functionality of the string representation of the function. For example, a “Variable” object will need to be created when the statement “x = 5” is written. The compiler approaches the block problem by using a hierarchical tree to represent the code (this makes the next step easier). The general rule in order to make this happen is open a block where there is an open curly brace, and every subsequent statement will be stored inside of this block until a corresponding close brace is found, at which point subsequent statements will now either start a new block, or be placed in the parent block until there are no lines left.

With the compiler being able to be stopped at any point makes it easy to show how this has been achieved. Using the following example of the Fibonacci sequence, we can see how the compiler splits the GNG code into the tree of blocks.

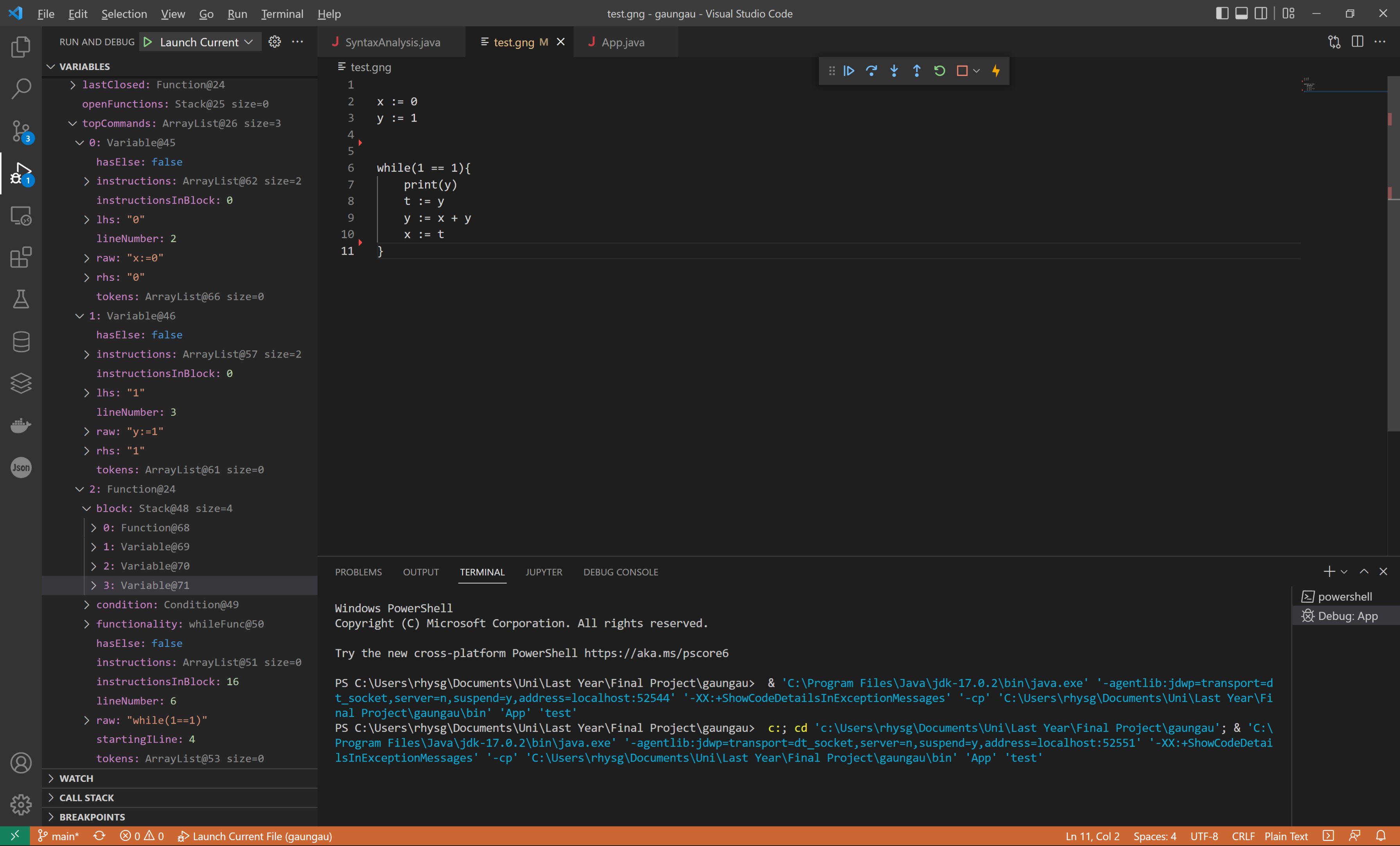
Notice that out of the 8 lines of code written in the editor to the right, only 3 commands have been identified. This is not a mistake. Note the type of the first two lines, they are both called “Variable” this is consistent with the first two lines of the program. The next command is listed as a “Function” This is because the compiler refers to While and If statements as “Functions” in the interest of polymorphism and ease of expansion later on. The function here is of course the while loop present at line 6, and this can be confirmed by looking at the “raw” property stating the content of this line. In addition, there is a property called “block”. This “block” property is the ArrayList which contains all of the subsequent commands found within this while loop block.



When the parsing stage comes around, the only initial commands that will be parsed are the ones at the top, and as such it is only these three which are parsed. The rest of the commands (inside the while block) will be parsed when the recursion occurs later.

After taking a look inside of the “block” property inside the while loop function, we see the following:

a function followed by 3 variables. This is in reference to the print statement, and the subsequent variable changes found after it.



We can now see that all of the commands have been located and stored in the correct block to reflect the structure of the code.

As previously mentioned, the syntax analysis will also convert the string representation of the intended function to the object associated with the actual functionality. Take the following example: “x = 5” will need to create a “variable” object so when the parsing and code generation comes around, the compiler will know which assembly instructions to be executed. There is rather a large if statement to catch all the differences in syntax to correctly identify the functionality which objects are needed to be created. Using the above variable example, the analysis will seek out variable syntax, in this case is “:=”.

// ======= VARIABLE ======= \\

if (statementText.indexOf(SyntaxCfg.variableOperand) > -1) { // Identifies a variable based on the presence of ':='

if (openFunctions.size() > 0) {

    openFuncAddStatement(new Variable(lineMeta));

    } else

    addTopCommand(new Variable(lineMeta));

The “variableOperand” reference on the second line is just a reference to a constant containing the actual identifier (“:=”). The current line being checked (statementText) is checked for the presence of this variable identifier with the use of the “indexOf” function. That identifier is the unique phrase used when manipulating data in RAM (altering or creating a variable) and as such, can safely assume that this line is related to variables, so the corresponding object is created and added to the tree.

#### Parsing and Phase 2 Syntax Check

The parsing process converts the existing tree (which now contains functionality objects) and parses them. The result of the parsing will be the actual assembly instructions of valid syntax. Any issues at this point will be flagged with useful messages as to how the user may go about fixing the problem. This may sound like a complex and long process, the project demonstrates a good use of recursion to parse each element in the tree in order and without special changes for each statement.

// ===== Parsing ===== \\

if (!errors) {

// All the code at this point makes sense, but not necesserily correct \\

  for (int i = 0; i < topCommands.size(); i++) {

    if (!setError(!topCommands.get(i).parse(instructionCount)))

            instructionCount += topCommands.get(i).instructionCount();

            if (errors) break;

        }

    }

}

The above code snippet is all of the code for triggering the parsing of each object. While this is technically true on the outside, a huge amount of processing takes place when the parsing is triggered. The parse function also contains polymorphic behaviours, as the “parse” function can be called on variable object, and function objects, as they both extend the “Struc” class, and override the parse function to have the parsing functionality expected of the given structure. For example, parsing a function will need to check for keywords, and built-in functions before creating the attached functionality, whereas the variable object will need to perform several name validation, and RHS validation to ensure that correct values have been passed in.

The recursion occurs when blocks are present in the code. In the Fibonacci sequence example above, there are only 3 objects on top to be parsing by the main parsing loop, but there are more commands which need parsing inside of that while loop.

for (Struc structure: block) {

if (!structure.parse(crntILine))

        return false;

            crntILine += structure.instructionCount();

}

This is the loop which will trigger the parsing of the instructions inside of a block, and if there is a block inside of a block like:

If(valid == true){

While(count < 10)

Print(count)

}

Then both the If, and While objects will contain the above loop for triggering the parse. Basically, parsing a block will trigger the parsing of its contents, and blocks within it. This means that with just these two for loops, every command in any size of GNG script can be parsed.

### Code Generation

Once the syntax analysis has established that the code is correct, it can go on to create the actual assembly instructions to be executed. It will systematically go over each object, and generate the required assembly instructions, and simply output them to a file in order.

for (Struc structure: code) {

    for (String instruction: structure.buildAndGetInstructions()) {

        if (!append(instruction)) {

            Error.customError("Could not write to file!");

            return false;

        }

    }

}

The first loop here will iterate over each object, and the second one will build and get the assembly instructions created which will return the assembly instructions in the form of an ArrayList, and so outputs each one to the file as it comes.

## Garbage Control

As this is an 8-Bit system, it only has 256 addressable memory locations, and no physical stack hardware to manage, and as such, there is only basic garbage control required. In addition, only integer values are accepted, and so datatype managing is also not needed. The main processing that occurs here is assigning memory addresses to created variables, and providing a memory location if an existing variable name is provided.

While very basic, is still essential to remove the chance of data collisions or memory addresses being overwritten unexpectedly which will cause huge issues with program execution. The code for the three main features of the garbage control are as follows:

public static int has(String varName){

if(variables.containsKey(varName))

        return variables.get(varName);

    return -1;

}

public static boolean newVariable(String name){

    if(variables.containsKey(name) || full)

        return false;

    variables.put(name, memoryHolder);

    memoryHolder++;

    full = memoryHolder >= 255;

    return true;

}

public static Integer getAddress(String varName){

    return variables.get(varName);

}

The 3 main functions here are “has”, “newVariable”, and “getAddress”.

The “has” function simply returns the address of a given variable name, or -1 if the given variable does not exist. The “newVariable” function is used when defining a new variable, and will add one if there is space (less than 255) available address. This is because address 255 (11111111) is reserved for multiple term calculations. So there are 255 (including 0) addressable memory locations available for general use.

“getAddress” is a simplified version of the “has” function which is shorter as it is only used after validation of the name has occurred. It is a last line of defence against syntax errors causing an error injection into the actual assembly output. If the compiler gets to this point, and there is still somehow an invalid variable name, a -1 would be displayed into the output assembly file. Instead of allowing this to happen, an error will be triggered.

## Error Handling

A screenshot of a computer

Description automatically generated with medium confidenceMy requirement state that the compiler will “be able to provide the user with useful and correct feedback as to syntax errors when compiling the program”. The compiler contains a custom class to handle all of the error messages which have been individually tailored to each possible issue:

These four functions are called throughout the program and pass in the message to reflect what issue has been faced. In addition, each function requires a parameter of the “lineNumber” associated with the error. This was surprisingly, a tricky thing to accomplish. This is because the first step of the compilation process is to put keys and syntax on different lines. If there was a simple one-line program consisting of:

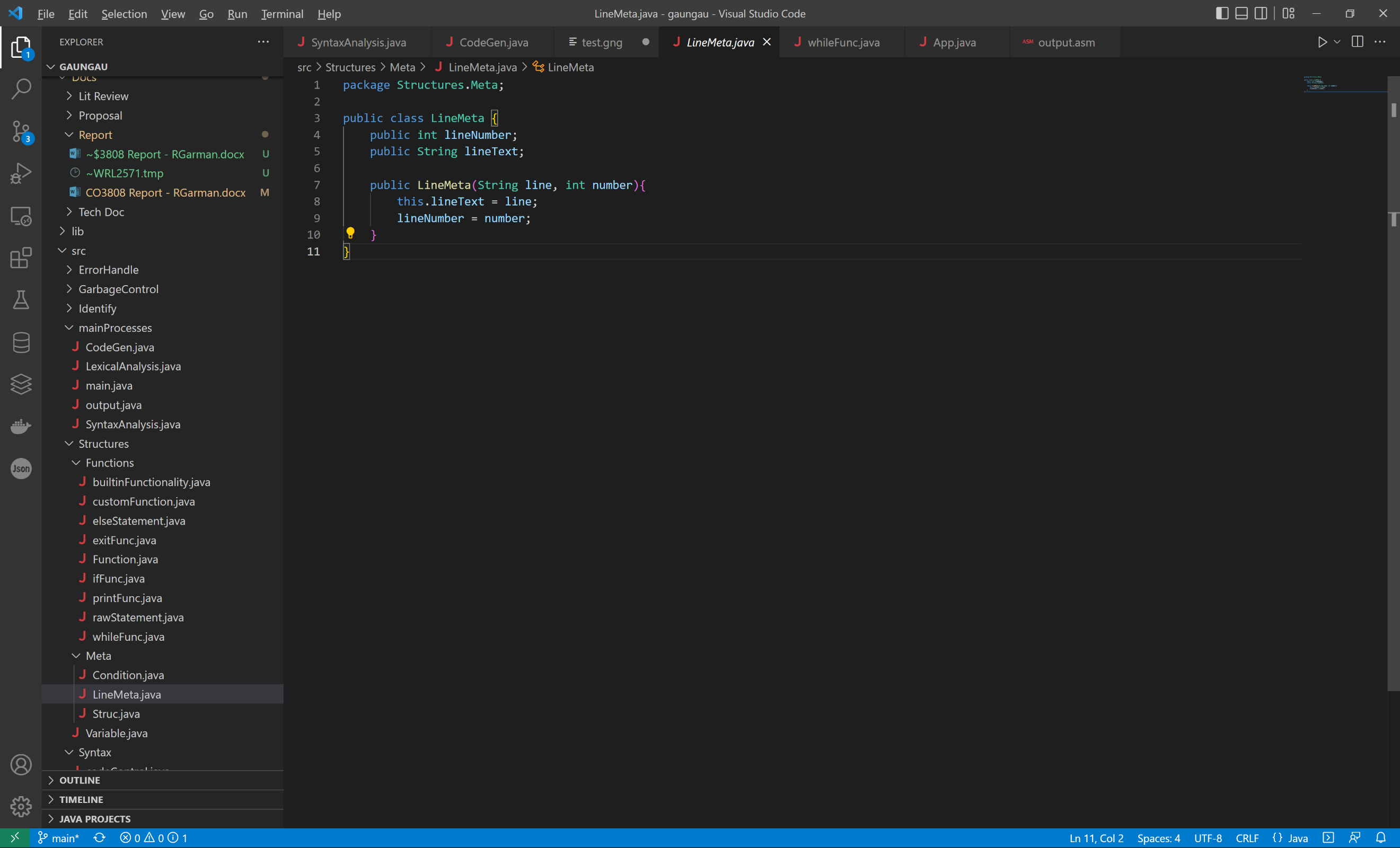
1 num1 := 6; num2:= 6; If ( num1 == doesntExist ) { num2 := 1; }

The lexical analysis would split this into:

1. num1 := 6
2. num2 := 6
3. If ( num1 == doesntExist )
4. {
5. num2 := 1
6. }

At this point, it is impossible to determine which line the error exists on (the use of the “doesntExist” variable in the condition) as the compiler thinks it is on the third line, but to the user, they can only see one line.

To solve this, the compiler uses a class called “LineMeta”. This is a class to contain the attributes about the lines:

The class simply contains the line on the real document that the code can be found, and the line itself. Now the code can re-structured by any stage of the compilation, and still have the real line number noted at each point. Due to this, any error that is brought up can have the associated line number.

## Operating System

The operating system consists of a huge list of raw binaries across 3 EEPROM (Electrically Erasable Programmable Read Only Memory) chips. The reason for needing 3, is due to the vast amount of modules present on the computer. There are 24 primary functions of the computer, and each EEPROM chip has 8 outputs, so each can control 8 functions of modules. These controls are incredibly basic, and cannot do much of anything on their own; the combination of instructions and logic is what makes computers so powerful.

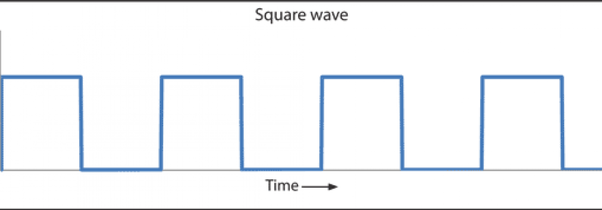
An example of a module is the ‘A’ Register, this is the primary register used for arithmetic, comparisons, and is the default location for loading data. The computer does no know this, and it is the binary in the OS which makes this the case. The operations that the ‘A’ Register can perform are simple loading a value from the data bus, and outputting a value onto the data bus. As such, the ‘A’ Register has 2 controls, and there are 3 registers on in the computer, taking up the first 6 controls. It is the granularity of each possible operation undertaken by every module that creates the need for so many EEPROM chips to be used.

The computer itself does not recognise the number 1 or the number 0, they are simply symbols, much in the same way symbols are used in everyday life to depict something else, like π meaning 3.14. Electronics use voltage differences as signals, and these difference are defined as 1 and 0, the computer uses chips which all use +5v and 0v as the 1 and 0 signals. These charges are the exact same as someone opening and closing a switch to turn on and off lightbulb, the only two differences are that the computer’s control logic is the one flipping these switches, and the signals can be used as controls as well as to power something.

Calendar

Description automatically generatedThe OS has absolute control over all of the modules, to so much of an extent that the OS (if programmed incorrectly) could cause damage to the data integrity, or even physical damage to the silicon on the chips. This could be done by overloading the data bus, which could lead to short-circuits burning out chips, and data being saved/loaded incorrectly.

The above instruction is used to load a literal value into the ‘A’ Register. The Instruction takes advantage of the fact that the EEPROM chips used have 13 address lines, 8 of these are used for the instruction itself, meaning that the computer can have a maximum of 256 commands. The remaining 5 of the 13 available bits are used for “Micro Steps” which are the commands to run within each instruction. Each of these micro steps take exactly one clock cycle to run. They are run on the falling edge of the cycle, and the setup time is from the rising edge up until the falling edge starts.



Set up

Execute

## The Hardware

As this project is such a low level, there would be an injustice in not including some of the hardware in this report. While a lot of the hardware is complex and not relevant, there is also a great deal of digital logic which is linked to the software goings on.

A screenshot of a computer

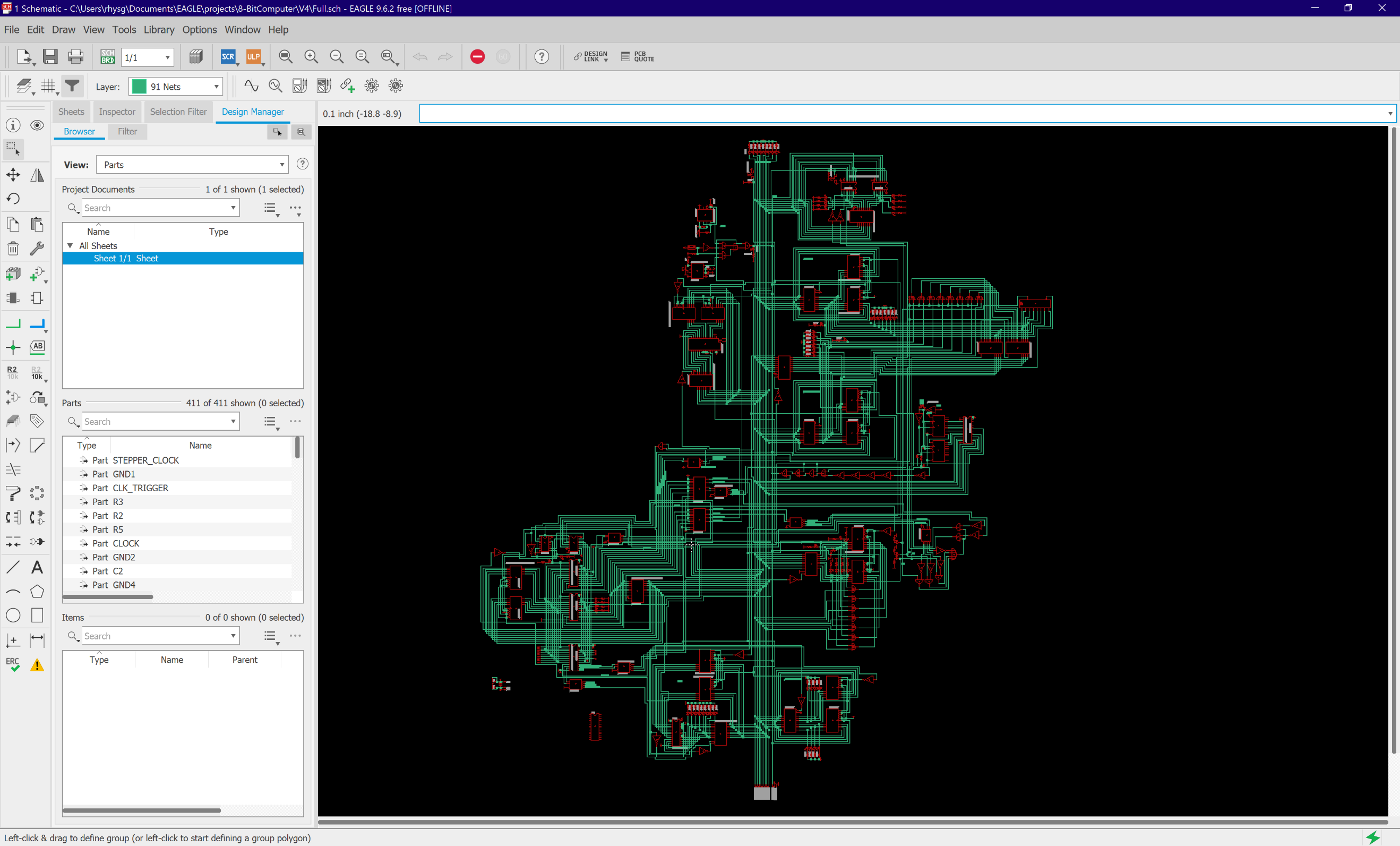
Description automatically generatedOne such example of good digital logic is the logic which controls the branching in the computer.

Diagram

Description automatically generatedThe central square is a “74ls139” chip which turns 2 input into 4 outputs. This is achievable as 11 in binary is 3, and including 0 makes 4 outputs:

|  |  |
| --- | --- |
| A | B |
| 0 | 0 |
| 0 | 1 |
| 1 | 0 |
| 1 | 1 |

From the datasheet of this chip, we can see that the 4 outputs labelled o0 – o3 go low (L) when active. This can be used as signals throughout the computer. In the branching logic, the two inputs are “Loop Ctrl A” and “Loop Ctrl B”, and the four possible branches are an unconditional branch, branch when the ‘A’ Register = 0, branch on a carry (result of calculation is greater than 255), and branch when the ‘A’ Register is greater than 0. The signal selector then passes through an AND gate along with the associated flag. Take BRZ (Branch when ‘A’ Register is 0) for example. BRZ is linked to output 2, so Loop Ctrl A will be High (H or 1), and Loop Ctrl B will he Low (L or 0). After the selector, this will tie O1 low, as this is the line being selected. This then goes through the not gate to convert the 0 to a 1, and into an AND gate, the other line on the AND is attached to the Zero Flag. This flag is high when the value of the ‘A’ Register is 0. If at this point the ‘A’ Register is 0, then the AND gate would return high, as both the zero flag and BRZ selector are high. It does not matter about the state of the other relevant flags, because all of the other lines coming out of the selector are high, and as such, they go through the not gate, resulting in one of the lines in each of the AND gates being low. Due to the nature of AND gates, with this 0 going in, they will all result in 0 apart from the one which is selected. This then enters a 4 input OR gate, and then off to the program counter to loop to a given step.



1

2

4

5

6

7

9

8

10

11

14

13

12

3

This is the circuit diagram for the whole computer. It looks like a full city of road and buildings, and thinking like this is helpful to get your bearings of the diagram.

|  |  |  |
| --- | --- | --- |
| Number | Part | Explanation |
| 1 | Program Counter (PC) | Stores the current step to be used as the memory address, and incremented when instructions are executed |
| 2 | Clock (CLK) | To keep timings and trigger certain events |
| 3 | ‘A’ Register | A general purpose register which all of the branching is based upon |
| 4 | Arithmetic and Logic Unit (ALU) | Can add and subtract integers based on the ‘A’ and ‘Calc’ Registers |
| 5 | Memory | Stores the program memory |
| 6 | ‘Calc’ Register | A secondary register to work on calculations, and can be used for temporary storage |
| 7 | PS/2 Keyboard Interface | Takes inputs from a PS/2 keyboard and can output it to the database for use throughout. |
| 8 | Instruction logic Data bus Interface | Enables the Instruction logic to interface with the data bus |
| 9 | Micro Step Counter | A secondary program counter to keep track of the steps within each instruction |
| 10 | Branching logic | To control the branches and to trigger them with the correct signals and conditions |
| 11 | Instructional Logic and OS | Contain the control signal lines, and Operating System Binaries |
| 12 | Toggleable Flag Register | Stores the built in and general-purpose flags which are toggleable |
| 13 | RAM | 256 addressable bytes with 8 bit address, and 8 bit data direct to the data buse |
| 14 | Parameter Register | A general-purpose register to store the parameter provided by each instruction. |

## Points of Interest

### Functionality vs. Performance

It is a general fact that high-level languages are less performant than low-level languages (*High-Level Languages vs. Low-Level Languages*, 2021), but it is rare that a good reason with examples is provided, other than to state that there are more layers involved, and high level is further away from the computer hardware. This project demonstrates with good examples how these inefficiencies occur and why they still exist even in today’s computer regardless of the fact that we are aware of the problems.

One such example of how these inefficiencies are using in high level languages is in my implementation of calculations during conditions and when setting values to variables. Take the following line:

If ( 5 + 6 == 3 ) // do something

The computer will need to complete the sum before being able to establish if the condition is true. One would think that the following assembly instructions would be correct for the line:

&LDA 5

&ADD 6

&SUB 3

BRZ 5

BRA 6

; do something

However, this is not the case, as the language must cater for any eventuality of program, and must keep the program running correctly, and have memory managed properly. The eventuality that must be accounted for is if there are calculations on the right hand side of the “==” for example:

If ( 5 + 6 == 2 + 8 ) // do something

After the first calculation is complete, the ‘A’ Register now contains the 11 which is need to be tested against. By now undertaking the second calculation (2 + 8), this 11 will be lost if it is not saved. Furthermore, if it is done this way around, the program will be even less performant. Lets go through what would need to happen:

* Load 5, and add 6
* Store this 11 in RAM
* Load 2 and add 8
  + As loading a value puts it in the A register, a straight load cannot be done here, as then the 10 from the previous calculation would be lost.
  + Also, a subtract cannot be done now either, as then it would be 11 – 10, it doesn’t matter much in this situation, as all the program is looking for is if the difference is 0, but in scenarios including “<” and “>”, the order is very important.
* Store this 10 in RAM
* Load the 11 from before
* Subtract the 10 from before
* Make the checks

We can see here this is a long process, and can be made slightly better, as if done in reverse, the second storing process is not required:

* Load 2, and add 8
* Store this 10 in RAM
* Load 5 and add 6
* Subtract the 10 from earlier
* Make the checks

While longer than the original statement, it is a lot shorter. The actual assembly instructions that are generated from the compiler are as follows:

&LDA 3

STO 255

&LDA 5

&ADD 6

SUB 255

BRZ 7

BRA 7

The condition length is the first 5 lines here, and compared to the first one with the bespoke logic of length 3, we can see that there has been a reduction in performance on purpose. If the code was written for the bespoke problem, then this performance issue could be mitigated, but with the universal and generalised nature of high-level languages, this type of issue will be rife.

### Parsing Recursion

This is a technique which is explained earlier in this report, however, the project shows good use and understanding of this complicated technique, and it should be noted that particularly interest was taken about the implementation of the technique during the development of the compiler.

### Polymorphic Behaviours

Following the above example of the &LDA command running, in the OS EEPROM chips, at address: “0000000000010” is the value “000000000000000000010001”. It is obvious at this point why programming languages are required, because that binary is completely impossible to debug if ever there was an issue with the code. However, within the table, we can see exactly what is happening. Where the 1’s are in the instruction, that is where the controls are being run, the fields the 1’s are referring to on the first instruction of the &LDA command are: PCO, and MAI. These are abbreviations created for the commands they control. PCO is Program Counter Out, and MAI is Memory Address In. This means the that Program Counter will output the current step the computer is on and the Memory module will take that value in. This is all that is done for that step, and at the next clock tick, the micro address counter will increment by one, and the new address being looked at inside the EEPROM chips is now “0000100000010”. This leads on to another set of binary, and new commands for the modules to listen to.

The OS is built using the method of “Fetch, Decode, Execute” (*Fetch-Decode-Execute Cycle - Instructions - GCSE Computer Science Revision*, 2020) which means that the command is fetched from memory, decoded using the EEPROM chips, and then executed by the signal logic to tell the relevant modules what to do. This is done in the first 4 steps of each command, which is followed by the unique code to the instruction, hence doing something different per instruction.

## Summary

The implantation of this project was long, and challenging. Many hurdles became very apparent, and solutions were created to combat them. Some solutions caused inefficiencies in the compiler, and some were very elegant, creating neat code which is maintainable, and accurate.

# Testing

## Introduction

Testing to ensure that the artefact is functions and accurate is as important to the development process as implementation, as there may be a working system under optimal circumstances, but it could break at the first sniff of anything that is not expected.

Both function and non-functional testing will be performed to ensure that every aspect of the project is working with no possibility of failure on the compilation side of things. Due to the nature of the project, the compiler will not run every time, but in these situations, the correct error messages should be displayed to show the user where they went wrong.

## Functional Testing

Functional tests ensure that the actual functionality of the system is working properly, this is primarily based on the validity of the outputs with given inputs.

While the outcome of each test can be simplified to a pass or fail, the validation process of each run is more lengthy than other programs. This is because the whole will need to be created by hand.

For each test, the standard “expected result” cannot be applied for the simple reason that it is incredibly difficult to know exactly what the compiler will output, as it is a very complex process, and mistakes would most likely be rife when creating the expected result. This would lead to pass or fail with test deficiency which is not ideal. Instead, GNG code will be created to be compiled, and then dry run, followed by an actual run on the computer and see if the two match.

Table – Functional Test Results

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Test Reasoning | GNG Code | Assembly Code | Dry Run Pass? | Actual Run Pass? |
| Ensures that variables are created with a unique memory address and are stored in RAM without overwriting | x := 1  y := 2  print(x)  print(y) | &LDA 1  STO 0  &LDA 2  STO 1  LDA 0  OUT  LDA 1  OUT | Pass | Pass |
| Ensures that an if statement can correctly identify a true condition and ignore the else statement | x := 1  y := 2  if(x == 1){  print(x)  }else{  print(y)  } | &LDA 1  STO 0  &LDA 2  STO 1  &LDA 1  STO 255  LDA 0  SUB 255  BRZ 10  BRA 13  LDA 0  OUT  BRA 15  LDA 1  OUT | PASS | PASS |
| Ensures that the while loop works correctly, and iterates while the condition is true | x := 1  y := 5  while(x < y){  print(x)  x := x + 1  } | &LDA 1  STO 0  &LDA 5  STO 1  LDA 1  STO 255  LDA 0  SUB 255  BRC 10  BRA 16  LDA 0  OUT  LDA 0  &ADD 1  STO 0  BRA 4 | PASS | PASS |
| Ensures that complex setting of variables show no issues | x := 1  y := 5  x := 7 + y - x - 9  y := x + y - y + x | &LDA 1  STO 0  &LDA 5  STO 1  &LDA 7  ADD 1  SUB 0  &SUB 9  STO 0  LDA 0  ADD 1  SUB 1  ADD 0  STO 1 | PASS | PASS |
| Ensures that complex conditions within the IF statement. This is to include AND, and OR gates.  NOTE: Here is where you can start to see the power and reason these compilers are needed. Such a simple few lines of code, can turn into something huge. In this case, it is 9 lines which turn into 30. | x := 1  y := 5  if(y + 5 + x == y + 2 && 5 == 4 || x < y){  raw(&LDA 1)  raw(OUT)  }else{  raw(&LDA 0)  raw(OUT)  } | &LDA 1  STO 0  &LDA 5  STO 1  LDA 1  &ADD 2  STO 255  LDA 1  &ADD 5  ADD 0  SUB 255  BRZ 13  BRA 24  &LDA 4  STO 255  &LDA 5  SUB 255  BRZ 24  LDA 1  STO 255  LDA 0  SUB 255  BRC 24  BRA 27  &LDA 1  OUT  BRA 29  &LDA 0  OUT | PASS | PASS |
| Complex test of nested while loops, and if statements with and without braces.  NOTE: again, a large difference, 17 lines of code being compiled into 51 instructions to be run, resulting in an infinite loop. The compiler will allow this happen, as it is simply converting the code to instructions, and not doing any of the calculation or checks. | x := 1  y := 5  if(x < y){  while(x < y)  x := x + 2  }else  while(y > x)  x := x - 2  if(x == y){  while(1 == 1){  print(x)  print(y)  }  }else{  raw(&LDA 0)  raw(OUT)  } | &LDA 1  STO 0  &LDA 5  STO 1  LDA 1  STO 255  LDA 0  SUB 255  BRC 10  BRA 21  LDA 1  STO 255  LDA 0  SUB 255  BRC 16  BRA 20  LDA 0  &ADD 2  STO 0  BRA 10  BRA 31  LDA 0  STO 255  LDA 1  SUB 255  BRP 27  BRA 31  LDA 0  &SUB 2  STO 0  BRA 21  LDA 1  STO 255  LDA 0  SUB 255  BRZ 37  BRA 49  &LDA 1  STO 255  &LDA 1  SUB 255  BRZ 43  BRA 48  LDA 0  OUT  LDA 1  OUT  BRA 37  BRA 51  &LDA 0  OUT | PASS | PASS |

## Non-Functional Testing

Non-functional tests are tests to ensure that everything else works. The frills, the addons, the user interaction, the interface etc.

There are not many non-functional tests to be made, however, the way in which the compiler communicates its errors to the user is a very important thing, and as such, tests will be carried out to ensure proper and useful error messages are displayed. In addition, the interface between user and compiler is by the use of a single executable file which is included into a java project. Tests will be carried out to ensure this process is functional.

Table – Non-Function Testing

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Test Reasoning | Code | Compiler response | Helpful? | Identified correctly? |
| Provides useful syntax error to the user about an incorrect variable declaration | x = 6 | Syntax Error: 'x=6' at line: 1 | Yes | Yes |
| Provides errors surrounding misuse of variables | x := 6  x := y + 2 | Syntax Error: 'y' does not exist at line: 2 | Yes | Yes |
| Notified user of improper use of while loop | x := 6  while(z == 8){  } | Syntax Error: Variable: z does not exist at line: 4  Syntax Error: Invalid condition syntax at line: 4 | Yes | Yes |
| Using unknown keywords | x := 6  call() | Syntax Error: Cannot create or call custom methods yet sorry at line: 3 | Yes | Yes |
| Simple code to ensure GNG compiler code can be used usefully | LexicalAnalysis lex = new LexicalAnalysis(txtScript);  SyntaxAnalysis syn = new SyntaxAnalysis(lex);    CodeGen Gen = new CodeGen("output.asm"); | <NO ERRORS> | Yes | N/A |

## User Testing

While the functional and non-functional testing is very important to ensure the project is suitably operational, it is considered white box testing because the person carrying out the testing is the developer of the project, and has a full view of what is going on. User testing is the fist run of black box testing, as the test subject does not know the goings on under the bonnet of the program. They simply use the project as they see fit, and the testing will determine if the design has been a success.

The user testing consisted of asking the test subject to undertake the following tasks:

* Create a variable with any name, and assign an integer value to it
* Create another variable, and create an if statement to compare the two values
* If they are the same, to output the first variable, and if different, to output the value of the second
* Finally, create a loop to output 1…10 sequentially

### Test subject 1

The result of this test was very positive. The resulting code from the instructions above, was:

one := 5;

two := 4;

if (one == two) {

print(one);

} else {

print(two);

}

counter := 0

while (counter < 10) {

print(counter);

}

The instructions were followed well, and I only explained the syntax of the variables, the rest was intuitive to them, as they have experience with programming.

The only issue that became apparent was the use of the “<=” operator. This as yet is unsupported mainly due to the time constraint present on the project. A simple fix was used in the test which was to set the counter to 0, and use the “<” operator instead.

### Test Subject 2

This test subject also has a background in programming, and did very well in completing the tasks. The only issue is that for the iteration (loop) task, they opted to use a for loop initially. Upon compiling the code, there was a syntax error, sure to the fact that the “for” loop is not implemented in the compiler at the moment, and is another example of future work.

### Test Subject 3

In the interest of fairness, although not technically the targeted demographic for the project, the third subject is not at all familiar with programming. They seemed to pick up the first two tasks involving creating variables really quite quickly, although struggled a bit when it came to the if statement, and while loop. Support was provided to them in order to complete it, and delivered a similar solution to the first two after a bit more time. At the end, they were a little bit disappointed that they did not see the outputs that they made, and this is because they did not understand the point of the project. The resulting code was run on the 8-bit later on, and they were satisfied, and became very excited, and wanted to learn more about why they did not see the output at first. This is the response that the project is set out to create, an excited and intrigued user who wants to go off and undertake independent learning, or dig deeper with the tool.

## Evaluation

As a result of testing, it has been established that the compiler does what it set out to do. The functional testing has been incredibly positive, with both the dry runs, and actual results returning correct and workable assembly solutions to the GNG code that was inputted. The non-functional testing showed that the interface in which the compiler is used by the user is also solid and can be implemented into a new or existing system with relative ease. This coupled with a possible user guide on how to do this will make a very smooth transition, and have the user writing real code in no time.

The user testing was very positive. Even with a shaky start to the third test, and a great deal of support being required, it lead them to want to learn more about why the output was not made available, and about what a compiler actually is. While the resulting code was not created as quickly, or written as well as the first two subjects (who have knowledge and a background in programming) this is the exact goal of the project, to inspire people to want to learn more about things that are covered over, and hidden to them. As a result, I would call all of the tests a resounding success and provides great evidence to suggest that this is a product that can and will be used in the real world.

# Evaluation, Conclusions and Future Work

## Self-Evaluation

Overall, I think that I performed very well during the course of the project. I began with only very basic knowledge of how a compiler operates. Along the way of creating one myself, I learned ways in which other compilers work, from basic C compilers to more advanced ones for high level languages like Python. I have learned things that I would do differently next time for example taken a more procedural approach, and leaned more on the lexical analysis to give the syntax analysis less of a job.

One major point of self-improvement was learning exactly why high-level languages are less efficient that languages closer to the hardware. It used to just be a fact, and these are exactly the things that I wanted to eliminate in this project. The main reason for the less efficiency is that the layers of instructions and order in which processes are executed are set not by the software, but by the hardware, and cannot be moved around slightly, but instead must be adhered to exactly. Deviation here will not cause a reduction in efficiency, but instead will cease the program is fail all together. As such, these inefficiencies cannot be avoided, and instead should be mitigated, and planned into the compiler instead of attempts made to create workarounds for the issues. While I was not very happy with knowingly introducing inefficiency, I quickly adapted to being able to simply mitigate their instances instead of fretting about their existence at all. This was only possible after I understood why they were required.

## Project Evaluation

There is still much that is needed to be done in order to make it a finished product which is potentially ready for the consumer, however the foundation for a project showing some serious potential is very evident.

Referring to the requirements made in section 3.2:

At the beginning of the project, the requirements to meet in order to consider the project a success were:

Must

1. Be a language that use high level syntax with parallels to existing languages to enable a shallow learning curve for new users of the language and aspiring low level software engineers
2. Enable the steps of compilation to be explored at the users own pace and thought process and for a bespoke experience every time
3. Convert the high-level language into the bespoke and simplified assembly commands, and the resulting binary which is compatible with my custom 8-Bit computer
4. Be considered at least a Turing machine capable of solving most any problem within the obvious constraints of an 8-Bit system.

Should

1. Be a single executable for use in the user’s very own compiler every time giving them control and understanding of the steps needed to be taken in program compilation
2. Be able to provide the user with useful and correct feedback as to syntax errors when compiling the program
3. Have built in garbage control to prevent data leaks and allow for easy memory management within the language

Could

1. Have additional built-in functions for the language specifically for this machine
2. Have a P/S 2 keyboard interface converted to ASCII numbers for general use in programs

Would

1. Have a built in FIFO stack to enable the use of function calls

The “Must” requirements have most definitely been met, as a language has been created with high-level syntax which is easy to understand and quick to pick up. This is evident from the user testing which was carried out, in which the subjects with previous experience with programmed seemed to have no issue with creating a basic script. In addition, the computer can be considered a turning machine, and has the ability to complete almost any solution, with the only limitation being the amount of data it can access with only an 8-bit bus for both the address and data lines.

The ”should” points have mostly been hit, as the compiler can be expressed as a single executable which can be run on its own, or have the functions used for a bespoke version of the compiler. In addition, the feedback to the user it very useful and points the user exactly to where the problem is if there is one. The garbage control however, is somewhat lacking. It works as is, up to having 255 (and 1 reserved location), but beyond that, will simply loop back around, as the compiler can reference address 257 for example (100000001), and due to the 8-bit hardware constraint, will treat this simply as 00000001 (ignoring the most significant bit) this will not be flagged by the compiler, but instead will cause data loss, and other issues down the line.

The ”could” requirements have been half met. Additional functions have been included, including the “raw” function which is incredibly power and gives the user complete control over what is executed on the computer. The keyboard interface however, is not yet functional, work is required to get this working, however, once it is, the ability to interact with the program at run time is a really great tool and gives a much closer experience with interacting with the silicone on the board in front of the user.

## Applicability of Findings to the Commercial World

From the research carried out at the beginning of the project, there are no products like this in existence that full encompass the process of programming language compilation that is not just simply a large amount of documentation.

## Conclusions

I would describe this project as a challenging success. It is something that I have not delved into in much detail before, and I think that overall, not only have I gained a deeper understanding into the inner workings of compilers, but also an appreciation for the early compiler which did not have the powerful computing techniques and tools available to them, and still created masterpieces of languages which were (and still are) used to mould modern computing.

The learning potential of the project is another main point which I am pleased about. I feel as though anyone will be able to pick up the project, and be able to write code in a way that is familiar to them, and have the tools to pick at the commands being generated and get a full understanding as to what is happening with complete transparency. This was the goal of the project, and having met it, I would safely say that the project is a success when measured against any metric, and put against all of the requirements of the project.

## Future Work

The limitations of this project are very well known, and it is base upon extremely primitive technology. The largest limitation for example is that it can only store 255 numbers (excluding the reserved address for complex calculations), and the largest number it can store is 255. This is something which could be improved upon by using software to combine memory locations to store larger numbers.

The iteration section has a lot of improvements that could be made as well for example:

* Support for the “<=” and “>=” operators
* For loops

These are common constructs which could be seamlessly incorporated into the compiler, and would be used with a very shallow learning curve from people who have a background in programming. This was made evident by my second test subject in the user testing, who instinctively used a for loop instead of a while loop for their program.

Further hardware improvements that could be made are:

* a P/S2 keyboard which is a universally used peripheral device interface, although mostly outdated and replaced by USB nowadays, is still out there, and would be a great improvement to the hardware.
* LIFO stack for function calls, in order to be able to return to the step that the call occurred, as the computer is not multithreaded, the single thread will need to remember where to return back to.

Further to these improvements would involve a restructure of the computer entirely, and possibly a more sophisticated power delivery system, as the current drawn may be significant.

References

*3.10.5 Documentation*. (2022). Python Documentation. https://docs.python.org/3/

Al-Dhaher, A. (2001). Integrating hardware and software for the development of microcontroller-based systems. *Microprocessors and Microsystems*, *25*(7), 317–328. https://doi.org/10.1016/s0141-9331(01)00124-7

Babiuch, M., Foltýnek, P., & Smutný, P. (2019). Using the ESP32 Microcontroller for Data Processing. *2019 20th International Carpathian Control Conference (ICCC)*, 1–6. https://doi.org/10.1109/CarpathianCC.2019.8765944

Bolanakis, D. E. (2017). Microcontroller Education: Do it Yourself, Reinvent the Wheel, Code to Learn. *Synthesis Lectures on Mechanical Engineering*, *1*(4), 1–193. https://doi.org/10.2200/s00802ed1v01y201709mec009

Bolanakis, D. E., Evangelakis, G. A., Glavas, E., & Kotsis, K. T. (2009). A teaching approach for bridging the gap between low-level and high-level programming using assembly language learning for small microcontrollers. *Computer Applications in Engineering Education*, *19*(3), 525–537. https://doi.org/10.1002/cae.20333

Cneude, M. (2021, July 5). *What Are Abstractions in Software Engineering with Examples*. The Valuable Dev. Retrieved 2019, from https://thevaluable.dev/abstraction-type-software-example/

*Computer Basics: Understanding Operating Systems*. (2020). GCFGlobal.Org. https://edu.gcfglobal.org/en/computerbasics/understanding-operating-systems/1/

Corob, M. (2022, January 13). *MASM instruction format*. Microsoft Docs. https://docs.microsoft.com/en-us/cpp/assembler/masm/instruction-format?view=msvc-170

Deek, F. P., Hiltz, S. R., Kimmel, H., & Rotter, N. (1999). Cognitive Assessment of Students’ Problem Solving and Program Development Skills. *Journal of Engineering Education*, *88*(3), 317–326. https://doi.org/10.1002/j.2168-9830.1999.tb00453.x

*Documentation for Visual Studio Code*. (2021, November 3). Documentation for Visual Studio Code. https://code.visualstudio.com/docs

Dreyfus, H. L., Drey-fus, S. E., & Zadeh, L. A. (1987). Mind over Machine: The Power of Human Intuition and Expertise in the Era of the Computer. *IEEE Expert*, *2*(2), 110–111. https://doi.org/10.1109/mex.1987.4307079

Duntemann, J. (2009). *Assembly Language Step-by-Step: Programming with Linux* (3rd ed.). Wiley.

*Fetch-decode-execute cycle - Instructions - GCSE Computer Science Revision*. (2020). BBC Bitesize. https://www.bbc.co.uk/bitesize/guides/z2342hv/revision/5

Frampton, D., Blackburn, S. M., Cheng, P., Garner, R. J., Grove, D., Moss, J. E. B., & Salishev, S. I. (2009). Demystifying magic. *Proceedings of the 2009 ACM SIGPLAN/SIGOPS International Conference on Virtual Execution Environments - VEE ’09*. https://doi.org/10.1145/1508293.1508305

Gries, D. (1974). What should we teach in an introductory programming course? *Proceedings of the Fourth SIGCSE Technical Symposium on Computer Science Education - SIGCSE ’74*. https://doi.org/10.1145/800183.810447

*High-Level Languages vs. Low-Level Languages*. (2021). High-Level Languages vs. Low-Level Languages. https://www.baeldung.com/cs/high-level-vs-low-level-languages

Hirt, M., & Lagergren, M. (2010). *Oracle JRockit : the definitive guide : develop and manage robust Java applications with Oracle’s high-performance Java Virtual Machine*. U.K: Packt Enterprise.

Koenig, J., & Rustan, K. (2016). *Getting Started with Dafny: A Guide*. Microsoft.

Kusswurm, D. (2014). *Modern Parallel Programming with C++ and Assembly Language: X86 SIMD Development Using AVX, AVX2, and AVX-512* (1st ed.). Apress.

*Little Man Computer - CPU simulator*. (2014). © 2014 Peter L Higginson. https://peterhigginson.co.uk/lmc/

Littler, C., & Littler, C. (2017, August). *Designing a Programming Language: Part II*. Kidscodecs.Com. https://kidscodecs.com/designing-programming-language-part-ii/

LoopPerfect, L. (2018, May 21). *Why are low-level languages fast and high-level languages slow?* Medium. https://medium.com/@LoopPerfect/why-are-low-level-languages-fast-and-high-level-languages-slow-9034c7e74da8

*Manifesto for Agile Software Development*. (2001). Agile Manifesto. https://agilemanifesto.org/

Margush, T. (2006). Using an 8-BIT risc microcontroller in an assembly language programming course. *J. Comput. Sci. Coll.*, 15–22. https://doi.org/10.5555/1181811.1181813

Métrailler, C., & Mudry, P. A. (2015). ESPeciaL: an embedded systems programming language. *Proceedings of the 6th ACM SIGPLAN Symposium on Scala*. https://doi.org/10.1145/2774975.2774982

Mitchell, R. (2020, September 2). *Are 8-Bit Microcontrollers Still Relevant?* Electropages. https://www.electropages.com/blog/2020/09/are-8-bit-microcontrollers-still-relevant

*Nasm Instruction Set Reference*. (2020). NASM X86 Instruction Set. http://home.myfairpoint.net/fbkotler/nasmdocc.html

Nisan, N., & Schocken, S. (2021). *The Elements of Computing Systems, second edition: Building a Modern Computer from First Principles* (2nd ed.). The MIT Press.

OS Today. (2019, June 22). *Question: What Language Is Windows 10 Written In?* OS Today. https://frameboxxindore.com/windows/what-language-is-windows-10-written-in.html

*Overcoming the memory bottleneck*. (2019, May 14). Scientific Computing World. https://www.scientific-computing.com/feature/overcoming-memory-bottleneck

Pal, K. (2017, February 6). *Why is learning assembly language still important?* Techopedia.Com. https://www.techopedia.com/why-is-learning-assembly-language-still-important/7/32268

Psycharis, S., & Kallia, M. (2017). The effects of computer programming on high school students’ reasoning skills and mathematical self-efficacy and problem solving. *Instructional Science*, *45*(5), 583–602. https://doi.org/10.1007/s11251-017-9421-5

Robins, A., Rountree, J., & Rountree, N. (2003). Learning and Teaching Programming: A Review and Discussion. *Computer Science Education*, *13*(2), 137–172. https://doi.org/10.1076/csed.13.2.137.14200

Rodriguez, B. J. (1994). A minimal TTL processor for architecture exploration. *Proceedings of the 1994 ACM Symposium on Applied Computing*, 338–340.

Scratch Foundation. (n.d.). *Scratch - About*. Scratch.Mit.Edu. https://scratch.mit.edu/about

*Selection in algorithms - Selection - KS3 Computer Science Revision*. (2022). BBC Bitesize. https://www.bbc.co.uk/bitesize/guides/zy3q7ty/revision/1

Shao, J. (2006). An Improved Microcontroller-Based Sensorless Brushless DC (BLDC) Motor Drive for Automotive Applications. *IEEE Transactions on Industry Applications*, *42*(5), 1216–1221. https://doi.org/10.1109/tia.2006.880888

Smith, P. A., & Webb, G. I. (2000). The Efficacy of a Low-Level Program Visualization Tool for Teaching Programming Concepts to Novice C Programmers. *Journal of Educational Computing Research*, *22*(2), 187–215. https://doi.org/10.2190/n0vv-0p48-xj9g-f8wv

Tanshi, F., & Bello, N. (2014). A Basic Approach to Designing Embedded Systems Using a Simple Calculator and C Programming Language. *Nigerian Journal of Technology*, *33*(3), 359. https://doi.org/10.4314/njt.v33i3.14

Turing, A. M. (1937). On Computable Numbers, with an Application to the Entscheidungsproblem. *Proceedings of the London Mathematical Society*, *s2-42*(1), 230–265. https://doi.org/10.1112/plms/s2-42.1.230

*Turing Completeness*. (2021, December). Turing Completeness. https://www.cs.odu.edu/%7Ezeil/cs390/latest/Public/turing-complete/index.html

Xinogalos, S. (2014). Designing and deploying programming courses: Strategies, tools, difficulties and pedagogy. *Education and Information Technologies*, *21*(3), 559–588. https://doi.org/10.1007/s10639-014-9341-9

Yılmaz, G., Ercan, C., Sıtkı, K., Harun, G., & Eray, Y. (2017). Understanding the concept of microcontroller based systems to choose the best hardware for applications. *International Journal of Engineering and Science*, 38–44.

Zwerg, M., Baumann, A., Kuhn, R., Arnold, M., Nerlich, R., Herzog, M., Ledwa, R., Sichert, C., Rzehak, V., Thanigai, P., & Eversmann, B. O. (2011). An 82μA/MHz microcontroller with embedded FeRAM for energy-harvesting applications. *2011 IEEE International Solid-State Circuits Conference*. https://doi.org/10.1109/isscc.2011.5746342

# Appendix 1 – Project Proposal

**Student Name:** Rhys Garman

**Course:** BSc (Hons) Software Engineering

**Project Title:** GAUNGAU (.gng) Script Interpreter for Custom 8-Bit CPU Arch

## Project Context

I intend to create a programming interpreter for my own custom designs 8-Bit CPU. I have been working on creating a CPU for several years, and I have now reached a stage where I can begin thinking about the software.

I am going to use Java to create the interpreter, this has been proven possible in an implementation by Strom and Aas (Strom & Aas, 2001). This is because it allows anyone to use the interpreter if the project was to be rolled out, as java is compiled on a VM, so it can run on any existing architecture (e.g. Windows, and iOS).

The point of an interpreter is to convert high level (human readable) code, into machine code or binary. These binaries can then be loaded onto the EEPROM (electrically erasable programmable read only memory) on the CPU and be executed. I will be creating my own language with its own syntax, structures, garbage control, etc.

Due to my prior experience with higher and lower-level programming, I will attempt to create this language with the high-level user in mind, by this, I mean I will base my syntax more off languages like Java and PHP, even though it is a direct link to the machine code like C or C++ is. By doing this, it will allow anyone to code their own program, and I will make it easy to see exactly what is going on. There will be nothing happening in the background, and everything will be visible to the user, from the interpretation to the binary output, and potentially further options to do with garbage control constraints, and manual addressing.

The issue surrounding the lack of more acute compilers are addressed in (Tanenbaum et al., 1983) which is also a motivator toward this project

## Specific Objectives

1. The interpreter must convert a custom programming language to the binaries required to be run on my 8-Bit computer
2. The interpreter should provide a deeper understanding as to the way in which programs are compiled and run computers and show the user each step of the way to execution
3. The interpreter should be able to offer corrections to the syntax if a syntax error is present. This should be displayed to the user when they attempt to compile their code.
4. The interpreter would automatically upload the binaries to the 8-Bit EEPROM, so the user does not need to handle the binaries if they didn’t feel up for the job

## Potential Ethical or Legal Issues

With any software that relies on the input from users (the source code in this case) there are many possibilities for incorrect inputs to be made. It is my job to have nothing unhandled if an incorrect input is provided. The source of the interpreter will be unavailable to anyone using the tool, so they will not be able to adjust if they find an error in my work. Therefore, intensive testing will be required to ensure nothing is left unhandled.

## Resources

As this project is purely software, the only resources that I will require are man hours, and software to write the interpreter. My preferred IDE is VS Code, this caters for my every need and also has the ability to run and debug Java as a free addon.

On an estimated hourly rate of £20 per hour. With a time estimate of 350 hours, I predict this will be £7000 worth of man hours.

While this section of the overall package is software, I have previous developed the computer for the code to be run on, and this will be a required resource as well.

## Potential Commercial Considerations - Estimated costs and benefits

The main cost of this project will be the actual 8-Bit computer. This is the only physical thing that is needed. This is to be purchased, as a kit, and it would be cruel to make the user purchase the interpreter as well. As such, the interpreter will be made available for free. I will be building the interpreter in Java, this means I can distribute the executable without sharing the source.

While the pricing of the 8-Bit computer itself has little baring on this project, it will be close to the final retail price of the package that this project contributes to.

* PCB costs £70 for 5 (£14 apiece)
* Componnts: £50
* Total price for one package is £64.

Most likely, this would be retailed for around £85 per unit. Meaning a profit per unit of £21.

## Proposed Approach

The way in which any script is interpreted is the same across all programming languages (Isaac Computer Science, n.d.):

1. Lexical analysis
2. Syntax analysis
3. Code generation

I will use this same approach when interpreting my programming language. As such, my first step will be writing an implementation for the lexical analysis. As this is an educational tool, I will write these in stages, so that the user will be able to see the outcome of each step and how to affects the code they are writing. Lexical analysis is all about removing code that will not be executed, such as comments and blank lines, after this, the commands will be tokenised,

Next is the Syntax analysis implementation, again, I will add this as a separate stage so the user can see the code as the lexical level, and then at the syntax level. This is the stage where the tokens can be checked that they follow the rules of the language and what the keywords actually represent.

Finally, the code generation is run, this is where the commands are converted to their respective binary codes and outputted to the user to be burned to the CPU.

As you can see, no logic has been done on the code, it has simply been converted from one form to another. While it would be incredibly easy to do all of the computation on the machine being used to interpret the code, that would void the purpose of the project. (Hadžiavdić, 2021) (‘Modern Compiler Design’, 2000)

## References

Hadžiavdić, S. (2021, June 23). *How to Approach Writing an Interpreter From Scratch*. Toptal Engineering Blog. Retrieved 1 December 2021, from https://www.toptal.com/scala/writing-an-interpreter

Isaac Computer Science. (n.d.). *Stages of compilation*. Isaaccomputerscience.Org. Retrieved 2 December 2021, from https://isaaccomputerscience.org/concepts/sys\_trans\_stages?examBoard=all&stage=all

Grune. (2000). Modern compiler design . Wiley.

Strom, O., & Aas, E. (2001). An implementation of an embedded microprocessor core with support for executing byte compiled Java code. *Proceedings Euromicro Symposium on Digital Systems Design*. Published. https://doi.org/10.1109/dsd.2001.952346

Tanenbaum, A. S., van Staveren, H., Keizer, E. G., & Stevenson, J. W. (1983). A practical tool kit for making portable compilers. *Communications of the ACM*, *26*(9), 654–660. https://doi.org/10.1145/358172.358182

# Appendix 2 – Technical Plan

**Name:** Rhys Garman

**Course:** BSc (Hons) Software Engineering

**Supervisor:** Matthew Horton

## Title

GAUNGAU (.gng) Script Compiler and Interpreter for Custom 8-Bit CPU Architecture

## Summary

My proposed project is a full compiler and interpreter for my custom made 8-Bit CPU. It will also act as an educational tool for people to see exactly how the code their write on the daily, is actually turned into machine language that their computer can understand and execute.

There will be a great deal of incredibly low-level programming involved, with very little human readability, as we cannot understand binary without some form of interpretation. I will however be managing binary directly at certain stages which is just about as difficult as computing gets. Not only this, but I will also write a program which translates a custom, higher level, human readable programming language into the raw binaries to be executed on a computer.

All of this is in an attempt to give people a deeper understanding as to why programs need to be written as they are. I feel this is a dying art, and a tool such as this does not exist, and I feel would be incredibly useful for students and hobbyists who want to learn more about the bridge between hardware and software, without reading pages and pages of old looking and difficultly worded documentation online.

Due to the nature of the project, a methodology such as test driven, or agile would not be suitable. This is because it is so low level, no plugs can be used. I am also working on this project on my own, and agile is a methodology for team working. I will however take certain aspects from agile (Agile Alliance, 2021) and use them in my methodology, such as putting the customer first, and taking time to reflect on the work done and how the process can be improved during the project development.

## Deliverables

As the actual 8-Bit computer is a physical hardware deliverable, I do not think it fit for this project. As such, the main deliverable I will produce is the compiler itself. This will be in the form of a java executable (Strom & Aas, 2001) (Hadžiavdić, 2021) (along with the source code). In tandem however, to prove that it works, I will also produce a demonstration of the code being written, compiled, and finally run on the physical machine. Although this will purely be for demonstration purposes.

## Constraints

The primary constraint is the processing ability and general ability of the 8-Bit computer. The length of the program is only example. As the computer has only an 8-Bit address bus, only 256 instructions can be loaded at one time. As such, the number of lines used in the program will be limited, and users must think hard about the number of instructions they are using which adds complexity to the programs they are writing.

In addition, there are only 13 base instructions, and any other functionality is simply a combination of these base instructions. In high level programming, we are used to being able to search google and finding thousands of built-in functions, features, and commands. But there will only be a set few that can be used.

These are to be understood by the user, as 8-Bit computing is VERY different to modern computing working on 32, and 64-bit architectures, with vastly superior processing power. I am not looking to challenge IBM or NASA; I am simply showing the real-world process of code compilation and execution on a contained platform that is in a way that is easy to understand.

## Key Problems

The key objective of this project is to allow people to understand the process of having programs compiled and executed on a computer. As such, this is also the biggest problem that the solution must negotiate: this is a very difficult concept to grasp. It is an incredibly difficult task to understand what is going on behind the scenes. Even the best programmers out there, have no idea what happens when press run on their code. This is NOT a bad thing. Nowadays, people do not need to know this information, but most programmers will not be satisfied with the answer of “it just works” (I certainly wasn’t). It is the intuition of a good engineer to ask why, and this project is looking to provide the answer and inspiration to learn even more about how compilation works on different higher end computer and architectures.

## System and Work Outline

I will be creating a compiler to turn a high-level programming language that I will create, into low level binary code which will run directly on a custom 8-Bit computer which has been designed and built by me.

I will use Java to create the compiler, as this is widely available, generic for any machine architecture (Mercer, 2021), and I have previous experience using this language.

I will create a demonstration on video showing that the program is run on the 8-Bit computer to prove the success of the project.

The project MUST be able to convert a piece of code written in a custom programming language, into binaries to be run on my custom 8-Bit computer.

The project SHOULD be able to show the user each step of analysing and converting the code from high level, to binary. In addition, the compiler should be able to offer potential corrections to syntax errors if they are made, this is because these types of errors are incredibly common, and sometimes are difficult to spot. (‘Modern Compiler Design’, 2000)

The project COULD have the ability to upload the generated binaries to the 8-Bit computers’ memory automatically, as this saves the user needing to this. This should be done via USB.

The project WON’T make the process harder than it needs to be, as this will make an environment which could lead the user to not want to pursue this field. I want to be able to spark an interest and understanding as to what is going on behind the scenes of any computer.

## Project Activities

|  |  |  |
| --- | --- | --- |
| **Task** | **Estimated Time** | **Result** |
| **Set up Git Repo structure** | 1 Day | Have a working git hub repo and a well organised structure for version control |
| **Create Base class to allow reading of the script to the compiler** | 2 Days | The result will be a base foundation layer of code that will pass the content of a script (most likely from the args parameter of the java execution) to the analysis section of the compiler |
| **Begin Lexical analysis on the script** | 5 Days | An implementation of the lexical analysis will have been created to begin the process of compiling the script to binary (Isaac Computer Science, n.d.) |
| **Create implementation for Syntax analysis** | 5 Days | An implementation of the syntax analysis will have been created to continue the process of compiling the script to binary |
| **Confirm the binary commands and parameters on a bit wise level for each instruction** | 1 Day | The OS for the computer has been created (in binary) the built-in commands have binary equivalents which are required to be noted and made sense of in order for the code generation to be accomplished accurately |
| **Correlate the bit wise instructions with the command in the language** | 0.5 Days | I will add an additional layer of abstraction to enable the user to see in more depth the sub instructions that are being used in breaking down the high-level code into lower level and then eventually binary |
| **Create implementation for the code generation** | 3 Days | An implementation of the code generation will have been created to finalise the process of compiling the script to binary |
| **Implement the ability for the user to see the code at each stage of compilation** | 5 Days | Enables the user to see a visual output as to the breakdown of their code and how it is slowly being transformed into binary |
| **Test code is compiling and executing correctly** | 10 Days | Vigorous testing is essential, and the result of this stage will be comprehensive testing to ensure there are no unhandled errors. The beauty here is that I am not voiding any errors, as errors will be rife in the inputs, but I must handle and inform the user of any errors that are present in their code |
| **Create demonstration video of the code being compiled and executed** | 5 Days | A final full run through of the compiler’s functionality, along with the binaries being uploaded to the 8-Bit computer itself and being proved to run the correct program with expected outputs being demonstrated |

## Risk Analysis

|  |  |  |  |
| --- | --- | --- | --- |
| Risk | Severity | Likelihood | Action |
| Complexity of project requires more time than I can allow | HIGH | Medium | I will ensure that I stick to the objectives I have set to ensure that I can count my project as a success. In addition, I will attempt to not deviate too far from these objectives as well, in the hope of not doing work that I do not need to do |
| Unpleasant UX | LOW | Low | Due to the nature of the project, there is no nice and friendly interface to communicate with. The command line will be essential to use. For someone who is not familiar with a CLI (Command line interface), this could be a blocker. However, the demographic for this product is someone with an interest in computer science and will most likely be used to using the command line, and thus will not be an issue. |

## Options

Text

Description automatically generatedWhen selecting the programming language to write the compiler in, there were several options, my first thought was to use a web interface and compile the code in a language such as PHP. This is because of my vast prior experience using this language and I feel strong in this way of programming. This however presented more issues than answers. One such problem was it would be impossible to work on the project offline which defeats the purpose of old school computing, it would also present a lot of issues with transferring the produced binaries to the computer. I decided eventually that a Command Line Interface would be the way to go, as it would be simple to have a file generated and the binaries to be loaded onto the computer’s memory. It is possible to run PHP on the command line, however, this is not common and is not simple to set up. Java is a well-designed package which allows support in the environment variables too (Mercer, 2021), so the compiler could be run from any directory on the machine. This is the option I have decided to take forward.

Another decision I had to make was the format of the programming language itself. I could make things easy for myself and create the language in a similar layout to assembly language, but not many people know of this layout, and are unfamiliar with how it logically works (see right)

Instead, I have decided to make the language feel more like a higher level language that most programmers nowadays are familiar with, such as java or php. I think this will enable people to pick up the language much quicker and have a much more enjoyable experience.

## Potential Ethical or Legal Issues

With any software that relies on the input from users (the source code in this case) there are many possibilities for incorrect inputs to be made. It is my job to have nothing unhandled if an incorrect input is provided. The source of the interpreter will be unavailable to anyone using the tool, so they will not be able to adjust if they find an error in my work. Therefore, intensive testing will be required to ensure nothing is left unhandled.

In addition, as the demographic of the end product is students, I do not want to discourage them from learning more about the topic or using my product and deciding that it is too difficult and look to change their ideas for their future.

## Commercial Analysis

If I do decide to fully commercialise this, I would most likely start down the route of a Kickstarter project. This is a crowd-funding organisation to help small start-ups with initial funding. It is a more risk-free route to set up the selling on of this project and potentially take it to full production depending on how well it does initially.

|  |  |  |  |
| --- | --- | --- | --- |
| Factor name | Description | Is this a cost or a benefit | Estimated Amount |
| PCB | The printed circuit board for the 8-bit computer | Cost | £14 per unit |
| Electronic Components | The cost of the components to make the computer | Cost | £50 per unit |
| Man hours (Compiler development) | The time is takes to develop the software for the computer and compiler | Cost | 300 Hours at |
| Selling the unit | Selling on the unit (including profits) | Benefit | Selling at £90 - £100 would bring the project to £26 – £36 per unit. |

## Employability Contribution

This is an area of work that I thoroughly enjoy and wish to take forward in my career after some time at BAE Systems. This is something which not many people understand, and I am excited as to what is possible with this in-depth knowledge that I possess.

In addition, if this project goes how I am thinking, it is something which I might attempt to commercialise sometime in the future.

## References

Agile Alliance. (2021, August 10). *Agile Manifesto for Software Development | Agile Alliance*. Retrieved 16 December 2021, from https://www.agilealliance.org/agile101/the-agile-manifesto/

Hadžiavdić, S. (2021, June 23). *How to Approach Writing an Interpreter From Scratch*. Toptal Engineering Blog. Retrieved 1 December 2021, from <https://www.toptal.com/scala/writing-an-interpreter>

Isaac Computer Science. (n.d.). *Stages of compilation*. Isaaccomputerscience.Org. Retrieved 2 December 2021, from <https://isaaccomputerscience.org/concepts/sys_trans_stages?examBoard=all&stage=all>

Mercer, J. (2021, May 17). *Why is Java so popular for developers and programmers?* FRG Consulting. https://www.frgconsulting.com/blog/why-is-java-so-popular-developers/

Grune. (2000). Modern compiler design . Wiley.

Strom, O., & Aas, E. (2001). An implementation of an embedded microprocessor core with support for executing byte compiled Java code. *Proceedings Euromicro Symposium on Digital Systems Design*. Published. <https://doi.org/10.1109/dsd.2001.952346>

Tanenbaum, A. S., van Staveren, H., Keizer, E. G., & Stevenson, J. W. (1983). A practical tool kit for making portable compilers. *Communications of the ACM*, *26*(9), 654–660. <https://doi.org/10.1145/358172.358182>