AURÉLIO MIRANDA SANTOS RODRIGUES GABOLEIRO

BSc in Computer Science and Engineering

RAMSES: A CONFIGURATION LANGUAGE FOR AUTOMATIC CODE GENERATION IN AN INDUSTRIAL CONTEXT.

A DOMAIN-SPECIFIC LANGUAGE APPROACH FOR SAFETY-CRITICAL EMBEDDED SYSTEMS

RAMSES: A CONFIGURATION LANGUAGE FOR AUTOMATIC CODE GENERATION IN AN INDUSTRIAL CONTEXT.

A DOMAIN-SPECIFIC LANGUAGE APPROACH FOR SAFETY-CRITICAL EMBEDDED SYSTEMS

AURÉLIO MIRANDA SANTOS RODRIGUES GABOLEIRO

BSc in Computer Science and Engineering

Adviser: Miguel Goulão
Associate Professor, NOVA-LINCS, NOVA School of Science and Techonology

Co-adviser: Dominique Blouin
Associate Professor, Télécom Paris, Institut Polytechnique de Paris

Abstract

The automatic generation of code from templates is a widely adopted approach in the industry to reduce costs and increase software reliability. However, this generation has to be highly configurable to meet specific requirements, such as project coding practices, compatibility with APIs and performance optimizations.

RAMSES is an Architecture Analysis and Design Language (AADL) code generation tool that fully automates the process of converting AADL models into code to support the design of embedded and cyber-physical systems. The most significant advantage of RAMSES is its ability to automatically generate code from high-level models, eliminating implementation details and providing better portability and reusability. However, as industrial systems become increasingly diverse, the need to adapt to specific industrial environments requires an adaptable configuration of the generated code.

This project proposes the design and implementation of a configuration language for RAMSES, enabling code generation to be customized according to specific requirements of each industry. The steps involve defining the syntax and semantics of the language, integrating it into RAMSES and testing it through industrial scenarios.

Throughout the project, various scenarios will be considered to demonstrate the effectiveness of the solution in comparison with other tools in this context. The aim is to provide an intuitive and helpful tool that can be used to make the language adaptable.

Keywords: Code Generation, AADL, RAMSES, Industrial Automation

Resumo

A geração automática de código a partir de modelos é uma abordagem amplamente adoptada na indústria para reduzir custos e aumentar a fiabilidade do software. No entanto, esta geração tem de ser altamente configurável para responder a requisitos específicos, tais como práticas de codificação de projectos, compatibilidade com APIs e optimizações de desempenho.

O RAMSES é uma ferramenta de geração de código para Architecture Analysis and Design Language (AADL) que automatiza totalmente o processo de conversão de modelos AADL em código para apoiar a conceção de sistemas integrados e ciber-físicos. A vantagem mais significativa do RAMSES é a sua capacidade de gerar automaticamente código a partir de modelos de alto nível, eliminando os pormenores de implementação e proporcionando uma melhor portabilidade e reutilização. No entanto, à medida que os sistemas industriais se tornam cada vez mais diversificados, a necessidade de adaptação a ambientes industriais específicos requer uma configuração adaptável do código gerado.

Este projeto propõe a conceção e implementação de uma linguagem de configuração para o RAMSES, permitindo que a geração de código seja personalizada de acordo com os requisitos específicos de cada indústria. As etapas envoluem a definição da sintaxe e semântica da linguagem, a sua integração no RAMSES e o seu teste através de cenários industriais.

Ao longo do projeto, serão considerados vários cenários para demonstrar a eficácia da solução em comparação com outras ferramentas neste contexto. O objetivo é fornecer uma ferramenta intuitiva e útil que possa ser utilizada para tornar a linguagem adaptável.

Palavras-chave: Geração de Código, AADL, RAMSES, Automação Industrial

Résumé

La génération automatique de code à partir de modèles est une approche largement adoptée dans l'industrie pour réduire les coûts et augmenter la fiabilité des logiciels. Cependant, cette génération doit être hautement configurable pour répondre à des exigences spécifiques, telles que les pratiques de codage du projet, la compatibilité avec les API et les optimisations de performance.

RAMSES est un outil de génération de code de Architecture Analysis and Design Language (AADL) qui automatise entièrement le processus de conversion des modèles AADL en code pour soutenir la conception de systèmes embarqués et cyber-physiques. L'avantage le plus important de RAMSES est sa capacité à générer automatiquement du code à partir de modèles de haut niveau, en éliminant les détails d'implémentation et en offrant une meilleure portabilité et réutilisation. Cependant, les systèmes industriels devenant de plus en plus diversifiés, la nécessité de s'adapter à des environnements industriels spécifiques requiert une configuration adaptable du code généré.

Ce projet propose la conception et la mise en œuvre d'un langage de configuration pour RAMSES, permettant de personnaliser la génération de code en fonction des exigences spécifiques de chaque industrie. Les étapes consistent à définir la syntaxe et la sémantique du langage, à l'intégrer dans RAMSES et à le tester à l'aide de scénarios industriels.

Tout au long du projet, divers scénarios seront envisagés pour démontrer l'efficacité de la solution par rapport à d'autres outils dans ce contexte. L'objectif est de fournir un outil intuitif et utile qui peut être utilisé pour rendre le langage adaptable.

Mots-clés : Génération de code, AADL, RAMSES, Automatisation industrielle

CONTENTS

List of Figures

GLOSSARY

API   Application programming interface (API) is a connection between computers or between computer programs. It is a type of software interface, offering a service to other pieces of software. *(pp. 2, 3, 15)*

AADL   Architecture Analysis and Design Language is a textual and graphical language that can be used to design and analyze the software and hardware architecture of performance-critical real-time systems. *(pp. 1–6, 8, 13, 14, 17, 18, 20, 21, 23, 25, 34)*

CPS   Cyber-physical systems are mechanisms controlled and monitored by computer algorithms, tightly integrated with the internet and its users. In cyber-physical systems, physical and software components are deeply intertwined, able to operate on different spatial and temporal scales, exhibit multiple and distinct behavioral modalities, and interact with each other in ways that change with context. *(pp. 1, 2)*

DSL   Domain-specific language (DSL) is a computer language specialized to a particular application domain. Its created specifically to solve problems in a particular domain and is not intended to be able to solve problems outside of it (although that may be technically possible). *(pp. 3, 16, 18, 22, 25, 27, 28)*

JSON   JavaScript Object Notation is a lightweight data-interchange format. It is easy for humans to read and write. It is easy for machines to parse and generate. *(pp. 36, 40)*

M2T   Model-to-Text (M2T) is a form of model transformation where structured models are automatically converted into textual artifacts, such as source code, configuration files, or documentation. *(pp. 5, 9, 39)*

Acronyms

AUTOSAR   AUTomotive Open System ARchitecture *(p. 9)*

CSS   Cascading Style Sheets *(p. 37)*

DI   Department of Computer Science *(p. 2)*
DSML   Domain Specific Modeling Language *(p. 18)*

EMF   Eclipse Modeling Framework *(pp. 9, 23)*

HTML   HyperText Markup Language *(pp. 36, 37, 40)*

IEC   International Electrotechnical Commission Standard *(p. 9)*
ISO   International Organization for Standardization *(p. 15)*

MBE   Model-Based Engineering *(pp. 4, 5)*
MD   Model-Driven *(pp. 1, 10)*
MISRA   Motor Industry Software Reliability Association *(pp. 9, 10, 12, 13, 15)*

RAMSES   Refinement of AADL Models for the Synthesis of Embedded Systems *(pp. 1–3, 5–13, 17, 18, 20, 23, 25)*
ROS   Robot Operating System *(pp. 30–32, 37–39)*

SUS   System Usability Scale *(p. 23)*

TSX   Task Load Index *(p. 23)*

YAML   YAML Ain't Markup Language or Yet Another Markup Language *(p. 6)*

1

Introduction

This chapter presents the work done in this dissertation, setting the context, purpose, and motivation of the study. It gives a context for the configurable code generation problems of Refinement of AADL Models for the Synthesis of Embedded Systems (RAMSES) and outlines the methodology and structure that guide the development of this thesis.

1.1   Context and Motivation

There is always a need for innovation, and consequently, technological progress is accompanied by an increase in system complexity [24], either software or hardware. This complexity is accompanied by enormous challenges in creating the solutions, particularly when software and hardware are adjacent to one another, such as is the situation when working with the robotics programming field.

For new users, robot programming can be especially daunting due to its extensive knowledge requirements and intricate integration of CPS [23]. Such systems, comprising computer-based programs, networks, sensors, and actuators, highlight the significant contribution of software development, which is responsible for the majority of the production cost of CPS [35]. Robotics, as a constituent of CPS, entails unique challenges in software and hardware integration, making problem identification late in the process very expensive [15].

Model-Driven (MD) approaches have proven to be effective solutions in this situation, offering the advantages of generation of high-quality code and results consistently [36]. The placement of the model at the center of the production process ensures that the developers are given a higher level of abstraction, while the complexity during the development of new systems is reduced.

1.1.1   AADL and RAMSES

Of all these MD techniques, the AADL is a strong modeling language well-suited to embedded systems [20]. It enables accurate description of hardware and software architecture

to support early validation and analysis of non-functional properties.

As shown by Borde et al. [14], the RAMSES project extends AADL to automatically generate source code for embedded systems. RAMSES, being a model-to-text transformation tool, enhances CPS software development quality and productivity by preventing human coding errors and accelerating the path from design to deployment.

> "With the ever increasing complexity of cyber-physical systems, RAMSES ensures trustworthy automation from design to deployment."

This project is a joint collaboration between Department of Computer Science (DI) NOVA, NOVALINCS, and Télécom Paris, unifying systems engineering know-how, formal methods, and embedded code generation.

1.1.2 Why Configurability is Necessary

While RAMSES utilization presents multiple advantages, there remains an issue concerning code generation: its decisions are hardcoded and rigid. Little is under control of the developer for elements such as code appearance, binding against specific APIs, or maintaining firm standards-based company guidelines. In industrial environments, where the need arises to reuse today's libraries and frameworks to uphold current standards, such a lack of flexibility presents a bottleneck, this inhibits not only initial integration, but also long-term system evolution and co-evolution, which are crucial for sustaining CPS over time [18].

Flexibility to customize generated code is imperative to encourage increased adoption by industry and to facilitate integration in diverse development environments [25]. With configurability, RAMSES can be set up to generate code that not only meets functional requirements but also conforms to organizational coding conventions and leverages accessible software assets.

1.2 Problem Statement

RAMSES does not have flexibility in its code generation process currently. Its generation strategies such as coding style conventions, library use, and API selection are fixedly embedded within its transformation rules. This lack of flexibility limits its application in industrial environments where projects rely on pre-existing company libraries and specific coding standards.

Currently, adapting RAMSES to different industrial contexts requires modifying its internal model-to-text transformation logic. This approach can increase maintenance effort and complicate integration with existing workflows, potentially limiting the broader adoption of the tool in diverse development environments.

1.3   Objectives and Contributions

The primary objective of this work is to enhance RAMSES configurability through the development and implementation of a configuration language. This language would externalize the parameters of code generation so that developers can tailor them to specific needs.

The key contributions of this thesis include:

- The definition of a DSL or configuration language for parameterizing the RAMSES code generation process.

- The introduction of mechanisms to enable customized C code generation, supporting various coding styles, library integrations, and API choices.

- The facilitation of reusing company libraries or APIs that already exist, enabling smoother integration of RAMSES into industrial development processes.

> **Contribution Summary:**
>
> - A configuration language for RAMSES
>
> - Flexible and customizable C code generation
>
> - Library reuse and integration support

1.4   Structure of the Thesis

This thesis is structured as follows:

- **Chapter 1:** Introduces the context, motivation, problem statement, and objectives of this research.

- **Chapter 2:** Provides a detailed overview of the state of the art, including Model-Driven Engineering (MDE), AADL, and existing code generation tools.

- **Chapter 3:** Identifies the challenges and requirements of creating a configurable code generator. Presents a prototype and evaluates feasability, implementation details and user feedback.

- **Chapter 4:** Describes the development plan for the configuration language.

3

2

Background and Related Work

In this chapter we will explore various tools related to the goal of this thesis: code generation configuration. While also explaining Acceleo, the main development tool used.

2.1 Model-Based Engineering (MBE) and AADL

Model-Based Engineering (MBE) has become a central methodology for the design of complex embedded systems . By putting high-level abstractions at the center, MBE enables engineers to manage system complexity through formal models rather than low-level code from the start [36]. This abstraction is particularly critical in embedded systems, where hardware constraints and timing requirements must be closely integrated with software behavior.

Several tool-supported methodologies, like NDT-Suite [22], show even more how MBE can be applied to real-world software engineering projects by offering methodological guidance and model-driven automation.

In the context of embedded systems, MBE facilitates early validation of design decisions, much earlier than hardware exists or code is written [21]. Engineers can model interactions, analyze performance bottlenecks, and verify compliance with safety and reliability standards, all at the model level.

One of the most important participants in this strategy is the AADL [20]. AADL is a formal hardware/software co-design modeling language. It gives precise semantics to model the architecture and behavior of embedded systems, ranging from processor bindings and memory layouts to communication buses and task scheduling.

AADL not only provides expressive architectural modeling constructs but also supports early analysis of non-functional properties such as timing, reliability, and safety constraints. Its formal semantics and tool support enable systematic verification of these properties during the design phase [20]. This is very well suited to industries such as aerospace, automotive, and defense, where such considerations are a given.

4

With AADL adoption, developers are able to early validate system architectures, preventing downstream integration risks and costly late-stage design modifications.

In this thesis, AADL is utilized as the base modeling language. Its formality and tool support, particularly within RAMSES, will facilitate automatic translation of abstract designs to the execution code and bridging of the system design and implementation gap.

2.2 RAMSES: A Code Generator for AADL

RAMSES is an M2T transformation tool with the ability to generate code from AADL models. Part of the greater Eclipse ecosystem, RAMSES automates the transformation of architectural models into deployable source code, effectively achieving the MBE dream of model-driven automation[3].

RAMSES now supports code generation in both **C** and **C++**. This makes it possible to use it in a broad variety of embedded development settings, depending on whether the target environment needs low-level procedural programming or more structured, object-oriented design paradigms.

The tool does this by systematically correlating AADL model elements to their corresponding code structures. Processors, threads, communication channels, and data components declared in AADL are mapped to their code counterparts, so much of the boilerplate and scaffolding code otherwise written by hand being done automatically.

Automation through RAMSES accelerates development and reduces human error, especially in large-scale embedded projects.

Even with these advantages, improvements can still be made. RAMSES' transformation logic is currently hardcoded, so developers have little control over customizing or fine-tuning the code structure generated without having to alter the tool itself. This rigidity becomes a performance bottleneck in projects that involve customized code structures, strict following of certain coding guidelines, or multi-variant code generation.

2.3 Code Generators in AADL and Beyond

While RAMSES plays a central role in the AADL toolset, it is by no means alone in the world of model-based code generation. There are long-established solutions both inside and outside the AADL universe with their own capabilities and niches.

2.3.1 Simulink Code Generation For Embedded Systems

Simulink is a flagship Model-Based Design solution, particularly in control systems engineering, developed by MathWorks [4]. In comparison with the tightly integrated

AADL inherent RAMSES, Simulink is backed by a graphic modeling framework of dynamic systems, and the production of code becomes straightforward with software like Simulink Coder and Embedded Coder.

Key aspects of Simulink code generation are:

- **Model-Based Design:** Control systems can be graphically designed, simulated, and validated by engineers before code generation.

- **Template-Based Generation:** Code is generated from pre-defined templates to enable integration into existing software platforms.

- **Customization and Extensions:** Developers can customize generation patterns and integrate generated code into larger legacy codebases.

Simulink is especially well-suited for rapid prototyping and tight integration with hardware-in-the-loop testing, and thus it is a favorite among automotive and aerospace industries [5].

2.3.2 OpenModelica: Modelica-Based Code Generation for System Simulation

OpenModelica is an open-source Modelica language-based modeling, simulation, and code generation software used intensively for system and physical modeling [32]. It generates simulation binaries and C code that precisely represent Modelica models and support complex system dynamics and numerical analysis [33].

Configuration options are available through Modelica annotations and compiler flags, allowing control over simulation parameters and some aspects of code generation. These are, however, mostly simulation-related settings and not related to control of the level of source code organization, naming, and coding style.

Code generation in OpenModelica prioritizes the correctness and performance of the resulting simulation code and provides little support for adherence to a given coding standard or legacy code base [34]. The major facility of the tool is to create efficient executable simulation models rather than to be highly configurable with respect to code generation output.

2.3.3 OpenAPI Generator: Configurable Code Generation Beyond Embedded Systems

OpenAPI Generator is an open-source tool that generates client SDKs, server stubs, and documentation from OpenAPI specifications [31]. Supporting over 40 languages and frameworks [30], it is widely used across software projects.

Generation is controlled via configuration files (JSON or YAML Ain't Markup Language or Yet Another Markup Language (YAML)) that specify package naming, class prefixes, data type mappings, and code style, enabling consistent architectural and coding

standards [27]. The tool's template-based system uses customizable Mustache templates to define code output, allowing adaptation to legacy code, custom logging, or specific frameworks. Plugin mechanisms and hooks enable further customization during generation [29, 28].

This flexible, configurable approach contrasts with RAMSES's more rigid, hardcoded transformations.

2.3.4 RAMSES vs. Other Code Generators

To better understand how RAMSES holds up against the competition in terms of code configuration, Table 2.1 was created based on singular tool testing and interviews with developers.

Table 2.1: Code gen configuration feature comparison based on author's analysis of documentation and tool behavior

Feature	Simulink	OpenModelica	OpenAPI	RAMSES
Identifiers[1]	Yes	No	Yes	No
Legacy Code Integration	Yes	No	Yes[2]	No
Generational Hooks	Yes	Yes[2]	Yes	No
Traceability	Yes	No	Yes	Yes[2]
Reporting	Yes	No	Yes	Yes[2]
Protected Areas	Yes	No	Yes[3]	No
Inline Functions	Yes	Yes	Yes[4]	No
Dead Code Elimination	Yes	Yes	Yes[4]	No
Comments	Yes	No	Yes[2,4]	No
Compliance Support	Yes[2,3]	No[3]	No[3]	No[3]

As can be observed in Table 2.1, features like dynamic identifier naming and legacy code integration are crucial, yet only present in some of the analyzed solutions, these features work well to achieve compliance standards imposed in the industry [16, 17], which is not natively supported by most apps like OpenModelica or RAMSES.

Generational hooks[1] are extremely useful since they allow for automated code customization during code generation, this explains why they are common in most code generators evaluated, though absent in RAMSES. Protected areas serve a similar function, however they allow for the modification of generated code, simply put, generational hooks modify the code before it is written, protected areas modify the code after it is written.

---

[1]Names of Functions, Classes, Variables, etc
[2]User-driven process (not entirely automatic)
[3]Normal code generation *might* generate compliant code, but its not very certain.
[4]Limited functionality or abstraction.
[1]Predefined extension points that allow developers to insert custom code or override generated behavior without modifying the generator itself.

Traceability is a crucial feature that enforces coherence between model and code, essentially defining the guiding thread[2] of the project. This is very sought after in a modeling context and ties well with report generation and code comment management, unfortunately, all three features are implemented only at a basic level in the analyzed tools.

Code optimization is also an important area, the ability to remove unnecessary code or shorten it proves to be useful in varied contexts, most tools provide a way to automatically optimize the generated code. RAMSES lacks such capabilities.

Overall, Simulink, a commercial high-end software, surpasses its rivals in all key aspects of code generation configurability. Its support the most sought after configurations give it is an end-to-end solution widely used in applications needing both flexibility and performance, but not on AADL.

Conversely, OpenModelica is missing a number of key points of configurability, reflecting both its complementary focus and level of maturity for code generation functionality. OpenAPI Generator, although even providing a more user-driven process in some cases, it still misses on some key features. RAMSES, in turn, presently falls short on all features, with inflexible, hardcoded transformations that curtail its usability and controllability by users.

This comparison reveals, yet again, the motivation for this thesis: researching means by which RAMSES can be enhanced through the introduction of greater configurability and extensibility, and thereby narrowing the gap with more mature tools in the domain.

The features found in table 2.1 are taken, not just from the code generators observed, but also from the wants and needs of the industry. The following feature model, present in Figure 2.1 was the outcome of that research.
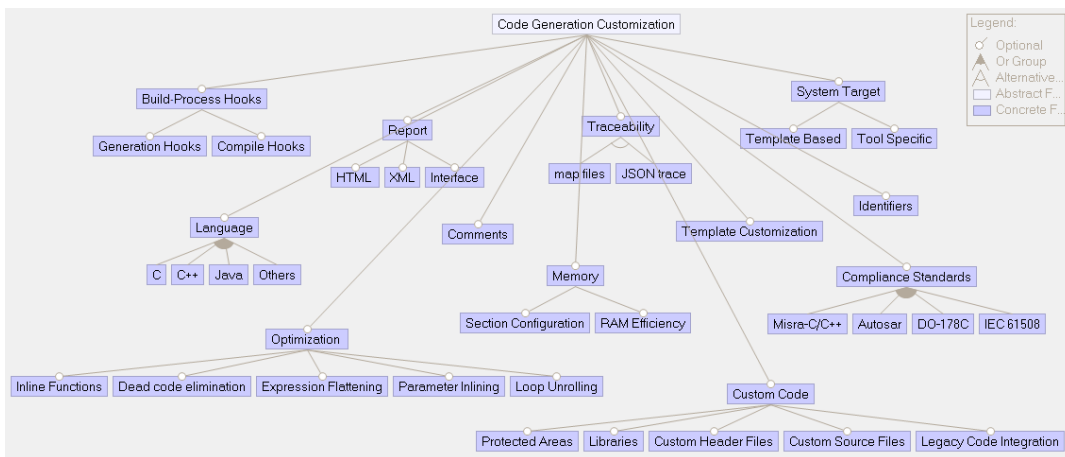

Figure 2.1: Code Gen Config Feature Model

With this, we can have a clearer look at the broader picture. The features suggested are not just the result of analyzing existing code generators like Simulink, OpenModelica,

---

[2]The coherent, central line of reasoning, logic, or narrative that connects all parts of a project or document.

and OpenAPI Generator, they are also derived from a synthesis of industry demands and recurring pain points observed in real-world development environments.

Figure 2.1 organizes these configurability aspects into a feature model, grouping them into categories such as Optimization, Traceability, Compliance Standards, Memory Configuration, and Custom Code Integration. The model also highlights optional, alternative, and concrete features that modern code generation tools must support to be competitive and practical across diverse application domains, from automotive (e.g., AUTomotive Open System ARchitecture (AUTOSAR), Motor Industry Software Reliability Association (MISRA)) to safety-critical systems (e.g., DO-178C, International Electrotechnical Commission Standard (IEC) 61508).

Additionally, this model also possesses some constraints when dealing with certain features, namely:

- MISRA-C/C++ implies C or C++

- AUTOSAR implies C

- DO-178C implies C or C++

- IEC 61508 implies C or C++ or Java

- Template Customization implies Template Based

This essentially means that most code standards are locked to one or more programming languages, if those programming languages are not selected, the standards does not apply. Similarly, Template Customazation can only be applied to Template Based solutions.

2.4 Acceleo and Model-to-Text Transformations

To counter the configurability limitations observed in tools like RAMSES, we turn to specialized M2T transformation technologies. Among these, Acceleo is highly promising.

2.4.1 Acceleo: An Overview

Acceleo is an open-source, template-based Eclipse family M2T transformation tool. Its thought model is based on the mapping of formal models (typically in Eclipse Modeling Framework (EMF)) to text artifacts like source code, documentation, or configuration files [6].

Major benefits of Acceleo are:

- **Template-Based Transformation:** Specified templates describe how the elements of a model should be translated into textual form.

- **Strong Eclipse Integration:** Acceleo offers robust integration with the Eclipse IDE, providing instant feedback, syntax coloring, and incremental generation.

- **Structured Code Generation:** Well suited for generating structured, maintainable C/C++ code from high-level models.

- **Project bias:** RAMSES currently uses Acceleo as a code generator so its logical to improve the work done as opposed to the integration of a new technology.

> Acceleo gives developers the ability to tweak code generation patterns, making the generated codebase more flexible and maintainable.

2.4.1.1 Acceleo example

Acceleo code is generated from .mtl templates, where it is defined how model elements are translated into text. Suppose the model contains a Class with the name MotorController. A basic Acceleo template might look like:

```
[template public generateClass(c : Class)]
public class [c.name/]
{
 // class body
}
[/template]
```

This would generate:

```
public class MotorController
{
 // class body
}
```

This example illustrates how Acceleo uses MD templates to produce structured and consistent source code from high level models.

2.4.2 Acceleo's Role in This Thesis

For this project, Acceleo serves as the basis for enhancing RAMSES' configurability. Through delegating transformation logic to Acceleo templates, we have the aim of:

- Isolate transformation rules from RAMSES' internal code.

- Allow easy extension and modification of code generation patterns.

- Facilitate adherence to industrial standards such as MISRA C/C++.

This plan promises to transform RAMSES into a more flexible and maintainable toolchain component from one that is rigid code generating.

2.5 Existing Work on Configurable Code Generation

The search for flexible and customizable code generation is not unique to this thesis. In most domains, tools and techniques have been created to solve the problem of generating high-quality, customizable code from models.

2.5.1 Template-Based Approaches

Template-based code generation remains the foundation in this field. Some good examples of such tools are **Acceleo** and **Simulink templates**:

- **Acceleo** allows explicit control of the structure and style of the generated code, making it highly suitable for projects in which compliance with some coding standards or architecture patterns is essential [6].

- **Simulink Templates** offers programmers the means to declare patterns of reusable code, with uniform look and feel across several projects and support for custom toolchains and legacy systems [7].

These approaches allow programmers to shape the auto-generated code towards project-specific applications without downgrading underlying models bridging the gap between automated generation and hand-coding, combining efficiency with flexibility.

2.5.2 Hook Functions in TargetLink

TargetLink, another market leader in code generation tools, comes with the concept of **hook functions**: pre-compiled points of extension within the generated code that allow developers to plug in their own logic [8]. The facility is most handy in a number of situations. For example, it eases the integration with legacy APIs or platform-dependent libraries and allows developers to add extensions without altering the primary generated code.

In addition, hook functions have the benefit of being customizable without compromising maintainability or upgradability of the generated code. When models evolve, code under it can remain unchanged while introducing custom logic using these extension points. This solution offers a clean trade-off between extending the generated code and offering its long-term maintainability with less effort for future upgrades.

2.5.3 OpenModelica and Multi-Variant Generation

OpenModelica introduces a higher degree of configurable generation with its support for **multi-variant code generation**. Through this, engineers are able to:

- Create multiple variants of code based on a common base model.

- Tailor outputs for various deployment contexts, hardware configurations, or performance constraints.

This variability is completely indispensable in automobile or aircraft production companies, for example, where a single product line might encompass several hardware targets or safety classes.

2.5.4 The Case for Configurability in RAMSES

Despite its strengths, RAMSES currently has no mechanism for fine-grained extension and configuration. Specifically:

- Transformation rules are hard-coded, which restricts flexibility.

- There is no native support for multi-variant generation or integration points like hook functions.

Including configurability in RAMSES would offer several benefits. It would facilitate the generation of custom code for different deployment environments, making it easier to adapt to specific hardware environments or performance requirements. In addition, the flexibility would simplify maintenance and development of the transformation logic, allowing the tool to better support changing development needs. Finally, by making RAMSES more configurable, it would be easier to interface with industry standards and legacy systems, rendering the tool flexible and applicable in high-speed industries.

> By adopting template-based generation, RAMSES can evolve into a dynamic, future-proof tool to meet growing embedded system development demands.

Towards MISRA C/C++ Compliance

A central element of code generation in configurable code generation, particularly in the field of safety-critical application domains, is to generate **standard-compliant code**. Strict requirements for safe, portable, and reliable embedded software are presented by the MISRA C [16] and C++ [17] standards.

Compliance to MISRA plays several principal roles: it enhances software safety by minimizing the likelihood of undefined behavior and runtime errors, guarantees that development processes meet the high standards demanded by industries such as the automobile and aerospace industries where in some instances compliance is mandatory, and is readily compatible with existing toolchains, as most static analysis tools are tailored to enforce MISRA rules.

As we integrate configurable generation facilities into RAMSES, we shall ensure that code generated is MISRA C/C++ compliant.

Flexible code generators need to not just conform to project requirements but also apply vital industry standards such as MISRA to guarantee safety and reliability.

While current solutions to code generation customization exist, they don't apply to AADL directly and the most powerful out of all the solutions studied, Simulink, is closed source, which does not solve our inherent problem: build a configuration language for RAMSES. This would make the configuration language the only open-source codegen configurability layer for AADL.

Challenges and Requirements for a Configurable Code Generator

This chapter explains why configurability is no longer optional, and how traditional code generation pipelines must evolve. In addition, a prototype of the configuration is also presented along with a system evaluation plan and research limitations.

3.1 The Inflexibility of RAMSES: A Barrier to Industrial Integration

RAMSES has served as a robust model-to-code generator for AADL based systems, yet it suffers from a fundamental architectural constraint: it assumes a uniform target environment. This assumption does not hold in real-world industrial projects, where system heterogeneity, legacy integration, and domain-specific standards define a constantly shifting context.

The core issue lies in RAMSES' transformation pipeline: it entangles policy decisions (naming, structure, integration style) with generation logic. These decisions, hardcoded in ATL transformations, reflect the assumptions of RAMSES' authors more than the needs of end users [9]. Altering them involves modifying the transformation source itself, often a hard and error-prone task [26].

Consider a simple use case: a company mandates that all task-level functions use 'snake_case' and include a 'COMPONENT_' prefix. RAMSES, which might generate 'ComputeTask' by default, offers no way to enforce those rules. A change in naming becomes a traversal through ATL templates and helpers. This is not scalable, and in safety-critical software, it is not acceptable.

Furthermore, beyond naming, decisions about memory allocation models, system initialization flows, and error handling behaviors are equally rigid. There is no declarative layer that allows users to steer generation outcomes according to organizational needs or evolving constraints.

14

3.2 Industrial Realities and Pressures

To understand why this rigidity is problematic, we must shift perspective from the generator to the organization that consumes it. In industry, generated code is not ephemeral: it is versioned, peer-reviewed, statically analyzed, tested, and in some cases certified. It coexists with handwritten code, interfaces with platform-specific services, and must evolve alongside requirements.

Compliance, Traceability, and Certification

Generated code must often comply with domain-specific standards such as:

- **MISRA C/C++:** Imposes constraints on memory usage, naming, control flow, and portability [16] [17].

- **DO-178C:** Requires traceability, tool qualification, and clear derivation from high-level requirements [2].

- **International Organization for Standardization (ISO) 26262:** Enforces safety-related development practices and documentation [19].

In these environments, code generation must do more than "just work". It must be explainable, auditable, and deterministic. Developers must be able to trace a generated function back to a model element and forward to a specific runtime behavior.

Integration with Legacy Codebases

Most industrial systems are not built from scratch. Code generators must work alongside:

- Legacy libraries with non-negotiable APIs that can potentially evolve overtime, as they are third party.

- Hardware abstraction layers that impose structural patterns.

- Existing software architecture rules (how modules communicate, how tasks are organized).

A code generator that cannot adapt to these constraints is often sidelined in favor of manual glue code or post-processing scripts. These scripts, in turn, introduce maintainability challenges and break traceability chains.

Developer Ergonomics and Maintenance

Even the most advanced generator will eventually produce code that is read (and possibly modified) by a human developer.

> If developers can't read or rely on the generated code, they will stop trusting it
> altogether.

Poor formatting, ambiguous naming, or surprising control flow all reduce the utility of generated artifacts, leading teams to "lock" generated files and prohibit modifications: an anti-pattern that defeats the promise of model-driven engineering.

3.3 Why Configuration Matters

To resolve the issues above, we must introduce a new abstraction layer: one that separates **what** is generated from **how** it is generated. This is the role of a configuration language.

A configuration language provides a structured way to express **generation policy**: the set of rules, conventions, and constraints that tailor code to its industrial context. Importantly, it allows these policies to be:

- **Externalized** from the transformation logic.

- **Composable** and layered across project variants.

- **Validated** for correctness before code generation begins.

Such language enables a fundamental shift: from a monolithic, one-size-fits-all generator, to a configurable and extensible platform that adapts to its environment.

3.4 Characteristics of an Effective Configuration Layer

Designing this kind of language is challenging. It needs to balance expressiveness, ease of use, and seamless integration. Based on industrial feedback and analysis of RAMSES issues, the following characteristics are proposed:

- **Declarative, Not Imperative**: Users should describe **what** they want ("all functions must use snake_case") rather than **how** to achieve it. This aligns with the model-driven philosophy and supports better configuration analysis.

- **Human-Readable and Tool-Accessible**: The configuration format (proposed DSL) should work well with version control, support difference review, and remain readable to engineers. It must also be machine-readable for validation and generation.

- **Functionally Equal to the Default Code**: The newly generated code should remain functionally the same as without the configurator. Meaning that it should produce the same practical result with or without the configuration, excluding performance metrics.

- **Configures the Generator, Not the Model**: Configuration keys should influence how the code is generated, without requiring changes to the input models. The DSL operates alongside the model, guiding the generators behavior (naming, structure or implementation strategy) based on domain concepts such as `thread`, `port`, or `data component`. Configuration keys should remain semantically aligned with modeling elements but their primary role is to modify generator logic, not to modify or extend the models themselves.

- **Validated and Error-Tolerant**: Invalid configurations should produce clear diagnostics before generation starts. Where possible, defaults and fallbacks should be available to prevent blocking workflows.

- **Extensible and Portable**: The configuration will be easily extensible thanks to its Ecore background, allowing for easy inclusion of new features. At the same time, since the language model is detached[1] from RAMSES, it can be easily included in other projects, allowing for a quicker development of a configuration language once this one is built.

These characteristics will make the configuration language robust, maintainable, and suitable for industrial use, allowing for flexible customization of code generation without compromising model integrity or introducing unnecessary complexity.

3.5 Beyond Code Formatting: What Configuration Should Control

While naming and formatting are important, a powerful configuration system must go further. The following dimensions should be within scope:

1. **Artifact Naming and Structuring:** Control over file names, folder layout, and identifier styles.

2. **Component Mapping Rules:** Declarative rules that assign AADL components to target platform concepts (RTOS tasks, processes, etc).

3. **Code Instrumentation:** Hooks for logging, tracing, or runtime checks (insertion of 'assert' or instrumentation macros).

4. **Conditional Feature Flags:** Ability to enable or disable parts of the generator (generate test stubs, insert wrappers).

5. **Code Documentation:** Generation of reports or toggle of code comments.

---

[1]The language model (Ecore) itself is not locked exclusively to RAMSES and can be used in other projects, however, that implies having to code the logic for each feature according to project specifics.

> **Configuration is More Than Style**
>
> While formatting is the most visible aspect of configurability, its true value lies in
> controlling semantic properties of the generated code: platform binding, integration,
> traceability, and lifecycle.

3.6 Proof of concept

In order to better understand the whole concept of the DSL to be built during this thesis,
a prototype that encompasses a specific feature needed in the RAMSES tool was built
with the intent of showcasing the usage of the configuration language, in a controlled
environment.

3.6.1 How it works

The selected feature to be implemented by the prototype was the Identifier Modification
option, which essentially means having a higher control over the generated function,
variable, and class names in the generated code. This features was chosen since it is fairly
straight forward to implement and also decently portable, as discussed in section 3.4.

Lets take the following example:

```
class MyNode : public rclcpp::Node {
 public:
 MyNode() : Node("my_node") {
  (...)
 }
```

A simple Node class has the class identifier `MyNode`. In the case of RAMSES, MyNode
would be inherited from the AADL model, meaning that, in order to change the code
nomenclature of MyNode, we would have to either manually change the generated code or
changing the name of the node in it's corresponding AADL model file and then regenerate
the code.

Both options are feasible in this case, however, with growing system complexity this
small change becomes incrementally harder and more prone to errors.

The developed prototype focuses directly on that front, providing an adjacent config-
uration model that works alongside the main one with the goal of delivering that missing
configuration option, here's how it works:

1. The placeholder code is built, in this case it was from a taskset style project[2]

2. The target naming convention is selected in the configuration model

---

[2]This project is a Domain Specific Modeling Language (DSML) that communicates task sets and automat-
ically generates C code using Acceleo, targeting the RT-POSIX real-time glsAPI.

3. The final code is generated with the now correct identifier nomenclature.

For a clearer understanding of the flow of the prototype, lets take a look at figure 3.1:
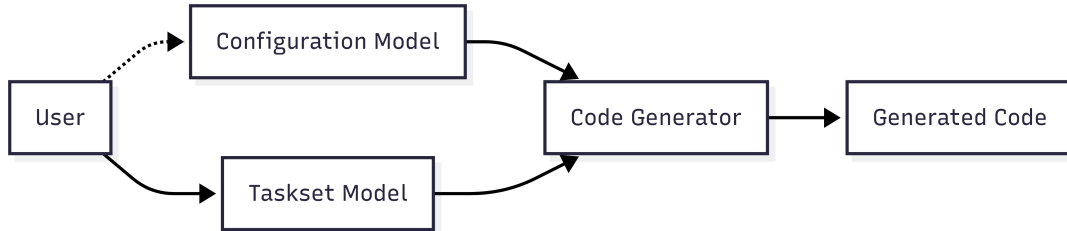

Figure 3.1: Flowchart of the prototype

The user interacts with the main model (Taskset in this case) and *can* interact with the configuration model aswell, if thats the case, the user will see changes in the code in accordance to the options selected in the configuration model.

The configuration model is not tightly coupled to the main workflow and is not mandatory, it works alongside the main model but provides extra options and choices to the user in terms of code customization, effectively elevating a bit more the code quality of the given taskset generated code.

3.6.2 Practical Example

Let's take a look at a practical example of the prototype. Since the placeholder code has already been defined and altered to allow for dinamic identifier changes, all we have to do is select the naming convention on the configuration model.

Figure 3.2: Selection of the naming convention

As we can observe from Figure 3.2, the prototype presents a rough and unrefined way of user interaction, in order to change the configuration the user must change the configuration model itself via its properties, which is not the end goal for the final product but works well in testing/controlled environments.

The current selected option in Figure 3.2 is the default option, which will not modify the file (see Figure 3.3)

Taking in the result from Figure 3.3 we can get an idea of the default, unaltered code generation. However, our goal is not simply to generate code. We can select a different naming option via the configurator model, save, and regenerate the code. The change is clear.

```
23  void T1();
24
25  // Initialization of data for task T1 (periodic, period = 1000 ms)
26  thread_config_t T1Info = { 1000, 0};
```

Figure 3.3: Result of the code generation in the default form

```
23  void t_1();
24
25  // Initialization of data for task T1 (periodic, period = 1000 ms)
26  thread_config_t t_1_info = { 1000, 0};
```

Figure 3.4: Result of the code generation with the snake_case naming option

Comparing Figure 3.3 with Figure 3.4, we can clearly see where the configuration model applies its modifications: always on the identifiers and never on anything else, especially not on the keywords[3].

The idenfitiers (in this case the name of the function[4] and the name of the variable[5]) change seamlessly according to the selected naming convention in the configurator. This is the expected behaviour for the features to be implemented, they should provide additional value while also keeping the base generated code available if the user does not wish to modify it at all.

3.6.2.1 User Feedback

In order to test the functionality and receive a third party opinion, a simple user testing was done with two users, specifically RAMSES developers that work directly with AADL. The procedure was simple and direct, users were instructed to the premise of the prototype, then they would perform tasks (in this case just one task: change the identifier name to X) and then they would be questioned in the usefulness of the feature, the applications it could have and overall thoughts.

Upon doing this testing, the users defined the feature as a flexible addition to coding generation with many application to real life and industry scenarios. Users also noted that this functionality improves information treatment between teams as different people have different ways to write models names, the prototype proves that naming normalization can also improve team cohesion.

Users commented that this functionality supports the ability to create multiple versions of functionally equivalent code just with naming differences. This also aligns with the required feature of supportign coding standards for the selected language[6], given that

---

[3]Keywords like *Void* are common and easy to identify. In this observed case however, in both Figures 3.3 and 3.4 we see that *thread_config_t* is also present but this is a *Struct*, which, while not an exclusive keyword, it serves a purpose like *Void* or *int* and is **not** an identifier.

[4]T1

[5]t_1_info

[6]Currently C++, however there are plans to include C.

some languages require identifiers to use a certain nomenclature [10, 11], which can clash with AADL node names.

Overall the prototype, although simple, was well received and proved to give value for the AADL to code pipeline.

3.6.3 Technical Details

Even though its application is specific, the code developed for the prototype is fairly generic. This means that most of the logic and code used for this prototype can be ported directly to RAMSES without much issue, effectively implementing the feature, albeit testing will still be performed.

The naming convention change logic is pretty straight forward considering identifiers, and those come from model properties. From there the properties that represent identifiers are formatted as they are being added to the code, effectively making sure the main model stays untouched while the properties extracted are modified into the desired outcome. This makes sure only the specific selected properties are modified and the rest of the code generates as normal.

The algorithm used for the naming transformation has many fallbacks that make it so even if a property name is especially complex, cases like short names are handled with care and specific styles like snake_case, while having complex patterns, can be easily defined with Regex. [38] As the configuration model is built on Ecore, the modification of naming styles is not possible directly by the user. However, if instead we wanted to **add** a new naming style, that is very much possible, however it required a quick modification of the Configurator's Ecore model and the Acceleo naming format algorithm, as in Figure 3.5

Figure 3.5: Flowchart of adding a new naming convention to the configurator

Figure 3.5 explains the steps in a clear but still abstract manner, the Ecore model needs to be modified to include the new naming style, which in turn will generate the coding logic that we can use in Acceleo to transform the identifiers by improving the current name formatting algorithm to accommodate the new change.

3.6.3.1 Multiple Models in Acceleo

Working with 2 Ecore models was not very straightforward. Although Acceleo provides the resources needed to work with multiple Ecore models, Eclipse itself does not allow for multiple input models in a single run configuration [12]. To work around this, a custom *main()* method was made to register the metamodels into the resources, effectively allowing for the usage of both metamodels. This issue, while small, presented a fairly

21

decent deal of complications for the development of the prototype, specifically because
Acceleo documentation on the matter was scarce and community forums were unhelpful.

Nevertheless, this solution allows not just for the usage of multiple models in this
prototype but also in any other projects that need an Acceleo program with 2 or model
Ecore models as inputs.

3.7 System Evaluation

To assure that the DSL meets expectations [37], the system will be evaluated in various
forms, mainly:

- The DSL exists and produces results.

- Before and after comparison (with code configurator vs without).

- Usability evaluation.

- Configuration variants evaluation.

- The newly generated code remains functionally the same as before[7].

- Various testing procedures during and after development.

Ideally the end product should meet these requirements to be considered a success.
Each one of the requirements ensures a different, core quality of the configuration system,
such as correctness, usability, or robustness.

DSL existence

This requirement will be fairly straightforward to prove since it implies a direct usage
of the DSL to produce results. This can be proven alongside the other requirements as
they depend on it. However, a system wide testing will quickly prove the existence of
the configuration language and its impacts, specifically with the testing of all features
developed and their results.

Before and After Comparison

The core idea of this configuration language is to modify the output code of the code
generator according to user needs. The best way to prove that this goal was achieved is
to compare how users would perform the same task before the configuration language
was implemented and after. This direct comparison will evaluate the direct impact of
the configuration language on development and will ensure that the developed solution
benefits the system.

---

[7]Meaning that it performs the same practical task as before the DSL implementation.

To strengthen the evaluation, a small controlled experiment involving developers familiar with model-driven engineering and code generation tools (EMF, Acceleo) will be performed. Participants will be asked to complete predefined tasks under both conditions.

Usability

The configuration language should be easily usable, even by end users with limited programming experience, to ensure this requirement is met, clear documentation, consistent syntax, and terminology aligned with the application domain should be prioritized. This requirement will be confirmed by comparing the interaction with the configuration language of users with varying levels of technical expertise within the RAMSES ecosystem, including RAMSES developers, testers, and other stakeholders familiar with AADL, allowing the comparison of usability and learning curve.

User Evaluation Methodology

Participants will be given predefined configuration tasks (changing naming schemes, visualizing traceability, etc) and asked to perform them with minimal guidance.

The evaluation will collect both quantitative and qualitative data:

- Task completion time

- Number and type of errors

- Subjective feedback on ease of use

- Observed learning curve

- Type and gravity of misunderstandings

To further support analysis, two standardized questionnaires will be used:

- The System Usability Scale (SUS) to measure perceived usability.

- The NASA Task Load Index (TSX) to evaluate the perceived workload and cognitive effort involved in using the DSL.

This setup will help assess whether the DSL meets its usability and accessibility goals across a broad user spectrum.

Configuration Variants

To assess the flexibility and robustness of the configuration language, distinct configuration files will be tested. These scenarios will vary in complexity and structure to evaluate whether the generator can handle diverse inputs while maintaining correctness and coherence in the generated code, effectively testing the limits of the configuration language.

Functionally Equal

To make sure the code generated still retains the same practical functionality, a few things need to be taken into account, firstly, the code generator needs to be modified just enough to produce the desired results of structure, naming, etc while still keeping the same functionality as before the implementation, essentially keeping the semantics intact. Monthly testings, including Regression and Integration testing, will also ensure that the code remains functionally the same by running selected examples and doing direct code comparison, however, comparisons will focus on runtime behavior rather than raw text equivalence. This will give us the confidence that the new code behaves the same as the old one.

The baseline for comparison will be the unmodified RAMSES generator. This allows for a clear assessment of what configuration capabilities are introduced in terms of productivity, flexibility, and clarity.

Testing Procedures

Tests will be conducted during and after development to ensure the configuration language behaves as expected and that the practical usage of the generated code remains the same. A variety of tests were chosen, specifically:

- Unit Testing which will be performed by the developer. These tests focus on individual transformation rules or components of the generator to make sure developed features work as intended.

- Regression Testing are also handled by the developer as they are used to ensure that new changes do not break existing functionality. They target both previously implemented configuration options and code generation rules.

- Integration Testing will also be performed by the developer and checks whether different features of the generator work together correctly. It focuses on feature interactions and overall system behavior, especially when configuration options are combined.

- Validation Testing is conducted involving actual users and project stakeholders. Its purpose is to validate that the features implemented match user expectations and project goals. Participants will evaluate whether the generated code behaves as expected and whether the configuration language is understandable and practical for its intended use.

Each of these testing strategies plays a specific role in evaluating the system, they ensure that the features work and users understand them. The outcomes of these tests is extremely important for the final evaluation of the prototype.

3.8 Research Limitations

Although the configuration language includes many features and has been designed with flexibility, usability, and maintainability in mind, this work has some limitations.

- **Limited Scope of Configurable Aspects:** The proposed DSL focuses primarily on naming conventions, traceability, and coding standards. More advanced features such as behavioral configuration, cross-component constraints, or real-time performance tuning fall outside the scope of this work [1]. Similarly, certain features might be prioritized or excluded based on client feedback.

- **Restricted Evaluation Sample:** Usability and functional testing rely on a small set of representative examples and a limited number of user interactions [13]. While including users with varying levels of technical expertise is being kept in mind, the results may not generalize across all industrial contexts or teams.

- **No Legacy Tool for Direct Comparison:** Since no previous configuration language exists for RAMSES, or AADL for that matter, before and after comparisons will rely on manual, potentially subjective assessments of workflow complexity. This limits the ability to measure productivity gains with high precision.

- **Integration Constraints:** The configuration system is designed to integrate into the existing RAMSES toolchain, which constraints how the generator behavior can be modified. Deep generator reworks or architectural transformations are intentionally excluded to maintain compatibility.

- **Third Party Library Issues:** The inclusion of third party code that is not affected by the configuration language might break compliance standards since the transformations only affect RAMSES-generated code. This limitation is hard to overcome since it depends solely on those third party libraries introducing standard compliant code.

- **Scalability Not Fully Tested:** The DSL should performs well on example models. Its performance and maintainability might not be validated on extremely large or complex input systems, which are typical in some industrial environments. This might not be a constraint if such models are later used for testing.

These limitations do not compromise the core contributions of this work, but they highlight areas for future development, such as broader configuration domains, wider user studies, and deeper toolchain integration.

3.9 Toward a Generator for System Integrators

RAMSES has proven to be a robust and effective tool for model-to-code generation, particularly within academic contexts and controlled environments. However, as embedded

systems development increasingly intersects with regulatory and industrial demands, the expectations placed on code generators have grown. Beyond correctness, they must now offer adaptability, traceability, and long-term maintainability to meet evolving project requirements.

Introducing a configuration language is not merely a feature, it's an architectural shift. It allows users to define their generation context without touching generator internals, reducing risk, increasing reuse, and enabling automation across product lines.

Development Plan

*This chapter details the development plan for the configuration language, its
implementation, testing, documentation, etc*

The implementation of the proposed DSL requires a methodical and objective approach
in terms of timing and task management. For that reason, a Gantt chart was made with
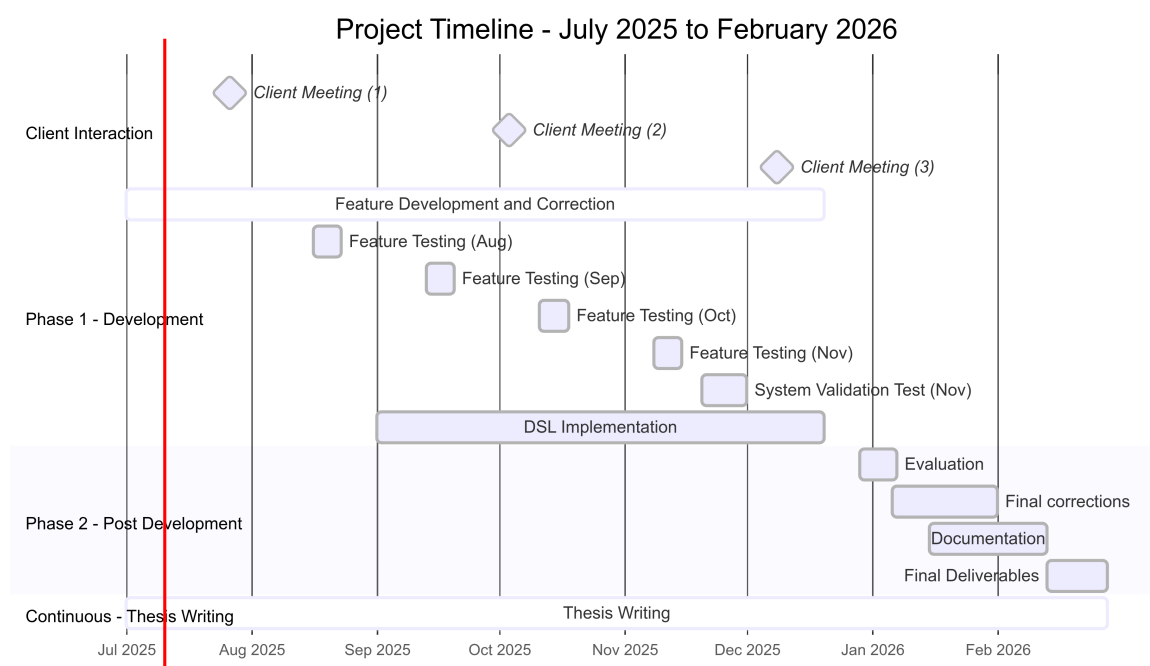the intent of guiding the development and ensuring the end product meets expectations.

Figure 4.1: Thesis planning Gantt chart

As we can see in Figure 4.1, the plan is separate in two distinct phases: Development
and Documentation. Those phases overlap with both client meetings and the writing of
the thesis, which is done in parallel with other tasks from start to finish.

4.1   Client Interaction

Client meetings are present to ensure that the features developed are useful and needed addition to the end user. The first and second meeting are mostly interviews directed to acquire quick feedback on the matter, while the third meeting will focus on the evaluation and testing of the mostly finished code generator configurator.

4.2   Development Phase

The development follows an iterative nature, which means that once a feature is developed, it is tested, documented and corrected when needed. Each month, full system wide testing, is also conducted to ensure multiple features work in harmony with each other. A system integration test is also performed near the end of the development phase to ensure that end users can use the system effectively and can achieve goals unachievable before. This is a critical point in the development since it will be used in addition to the third client meeting to build a system evaluation report.

These will be the types of testing used and their goals:

- **Unit Testing** during development, as new features get developed. Ensures new features meet their goals and catches early bugs.

- **Regression Testing** done monthly to make sure previous features are still working with the addition of new ones.

- **Integration Testing** done also monthly in addition to Regression Testing to test how different parts of the system work together.

- **Validation Testing** after development to validate that the features developed are in line with the requests of the end users and clients that will be testing the whole system.

Parallel to the code generator configuration feature development, a DSL will also be developed. Near the end of the development of a specific feature we need to also implement its actual configuration functionality via the DSL which will provide a way for end users to use the configurator, hence its parallel development method which also includes testing. Both the feature and DSL development go beyond the last evaluation and testing in order to ensure a safety gap for last minute features or corrections.

4.3   Documentation Phase

The final phase begins with the analysis of previously gathered evaluation results and insights on end user opinion, thought process and interaction when using the configuration language. Its a very important step since it will allow for quick overview on implementation

problems, bugs and other suggestions that might have not been thought of before. After that comes the documentation part for both users and future developers, since the configuration language will be very extensible. Documentation, which includes user manuals, developer guides and system documentation, will also be finished during development, however this interval mainly serves to make sure everything is documented properly.

The final part consists in ensuring every deliverable is correct and ready to send. It mostly include meetings and information exchange with every involved party to validate the work done, while not directed to development or corrections, it provides another safe window to make changes if strictly needed.

4.4   Final Considerations

The overall structure of the Gantt chart is clear and direct, although it provides a detailed and structured timeline, in practice it will serve mostly as a base and the development might end up with unforeseen requirements depending on the necessities of the parts involved, or technical challenges. The chart provides clarity and direction but remains adaptable to eventual changing project needs.

System Design and Realization

This chapter describes the technical implementation of the proposed configuration language for RAMSES. It details how the insights from industrial collaboration informed the architectural decisions and how each feature was implemented.

5.1 Industry Partner Meeting

The first meeting with an industry partner revealed to be very useful with many insights to note, particularly that a few features were highlighted and prioritized in order to maintain a good development speed and an overall better plan. Most of the previously thought of features (Identifiers, Traceability, Report, etc) were endorsed, however, a new feature was suggested by an expert Robot Operating System (ROS) developer: the separation of ROS specific code and C++ specific code, meaning that code that is specific to ROS can be isolated as it is very similar along the various generated files. This will essentially be done with an interface that will implement ROS code on C++ only files, which effectively serves as a separation of concerns.

> This meeting provided extremely valuable insight as the industry partner has been using ROS for over seven years, effectively distilling their experience into actionable knowledge.

5.2 Implementation Plan

In order to better understand what features will be implemented, when and why, the implementation will be done by layers, each layers will depend on things done previously on other layers, hence some features are developed before others.

Table 5.1: Feature dependency table

Feature	Depends On	Notes
Identifiers	–	Core to code structure
Comments	Identifiers	Cheap to implement early
Traceability	Identifiers, Comments	Hard to retrofit later
Report	Traceability	Uses trace data
Standard Compliance	Report, Traceability	Enforce compliance early
Dead Code Elimination	Traceability, Report	Needs stable generation logic
Memory Optimization	Dead Code Elim, Traceability	Impacts data structures directly
Node Interface	Memory Optimization	Enables ROS decoupling
Legacy Code Integration	Node Interface	Most architecture-dependent

Table 5.1 presents the planned order of feature implementation based on their dependencies. Each feature may require elements developed in previous steps, ensuring a coherent and manageable progression. This dependency structure prevents rework by implementing features that provide necessary infrastructure before higher-level functionalities.

5.2.1 Core and Metadata Layer

The Core and Metadata Layer (Table 5.2) includes foundational features that establish the basic code structure and metadata necessary for subsequent processing. Features such as identifiers and comments form the backbone of the generator, while traceability and reporting provide essential tools for debugging and verification. Their relatively low to medium complexity allows them to be developed early, enabling smooth integration of more advanced features later.

Table 5.2: Core and Metadata features and their complexity

Feature	Complexity	Notes
Identifiers	Low	Nearly complete.
Comments	Very Low	Can parallel identifier completion.
Traceability	Medium	Requires structured tagging throughout the generator.

Report	Low-Medium	Build once traceability is present; not deeply complex.

This first layer will serve as a base for other features to come since it has fairly straight-forward but very useful features.

5.2.2 Structural Optimization Layer

The Structural Optimization Layer (Table 5.3) focuses on improving code quality and efficiency. Standard compliance ensures generated code meets coding norms and best practices. Dead code elimination and memory optimization reduce code bloat and resource consumption but require stable metadata and analysis frameworks built in previous layers. These features have higher complexity, reflecting the need for careful analysis and manipulation of generated code.

Table 5.3: Structural Optimization Layer

Feature	Depends On	Notes
Standard Compliance	Report, Traceability	Defining and checking rules requires steady effort.
Dead Code Elimination	Traceability, Report	Non-trivial static analysis in model-to-code context.
Memory Optimization	Dead Code Elim, Traceability	Impacts core data structure generation. Error-prone.

The structural layer cements what was done before and adds to it with an optimization focused approach.

5.2.3 Architecture Layer

The Architecture Layer (Table 5.4) introduces features that abstract and modularize the generated code to support scalability and maintainability. The node interface abstraction separates platform-specific code (ROS dependencies) from core logic, facilitating reuse and easier updates. Legacy code integration is the most complex and architecture-dependent feature, involving interfacing with existing external codebases, which requires careful design and planning to avoid integration pitfalls.

Table 5.4: Architecture Layer

Feature	Depends On	Notes
Node Interface	Memory Optimization	Architecture-dependent. Needs careful planning.
Legacy Code Integration	Node Interface	Undefined scope and likely the most architecture-sensitive

The last layer of implementation has a higher degree of abstraction and requires a much more project architecture understanding than the previous layers, which is why it is done last and after every other feature layer is implemented. This will allow for the full focus to be on these core features that will have a much higher individual impact on code generation than the previous ones.

5.3 Implementation

As the implementation methodology was done iteratively and following the plan discussed in Section 5.2, the first feature to implement was the naming conventions.

5.3.1 Naming Conventions

The implementation of naming conventions was straightforward since the demo developed in Section 3.6 was a pretty big portion of the whole thing. Using Acceleo we are able to modify the properties coming from the model as they are added into the C++ file, in this case, the name is formatted according to the selected convention. The main addition to the prototype work was the introduction of dynamic naming conventions that depend on the selected language, for example: in C++, classes are declared in PascalCase, whereas variables are declared in snake_case, the following generated code illustrates that very well:

```
class ImageDisplay : public rclcpp::Node
{
public:

ImageDisplay(): Node("ImageDisplay")
{

  image_subscriber_sub_ = this->create_subscription(...);

  nd_spinner_sub_ = this->create_subscription(...)
```

As we can see, ImageDisplay and image_subscriber_sub_ are both identifiers but each one has a different format. In contrast, the previous implementation without configuration

was strictly be *Image_Display* and *Image_Subscriber_Sub_*. These names come directly from the AADL model as **Image_Display** and **Image_Subscriber**.

To summarize, the whole naming convention implementation consisted in:

- The ability to change identifier (class, variable, method) formats dynamically from a list of implemented options (camelCase, UPPERCASE, lowercase, snake_case and UPPER_SNAKE_CASE).

- Based on the selected language, assign different styles to different identifiers.

- Option to input a prefix or suffix in every identifier[1].

This implementation was very important since the remaining features used it to the fullest, hence why it needed to be the first.

5.3.2 Comments

The implementation of dynamic code comments allow the user to get much more context on the code while preserving the ability to view raw, non-commented code. The way it works is pretty straight forward, the user can toggle the comments on or off, comments are dynamic according to their component and only provide essential info.

```cpp
// Constructor: Initializes all components
ColorTracker(): Node("ColorTracker")
{
    // --------------- CtSubscriber Component ---------------
    ct_subscriber_sub_ = this->create_subscription<std_msgs::msg::Int32>("/
        10, std::bind(&ColorTracker::ct_subscriber_callback, this, std::pla

    // --------------- CtPublisher Component ---------------
    ct_publisher_pub_ = this->create_publisher<std_msgs::msg::Int32>("/colo
    ct_publisher_timer_ = this->create_wall_timer(500ms, std::bind(&ColorTr

    // --------------- Image_broadcaster Component ---------------
    image_broadcaster_pub_ = this->create_publisher<std_msgs::msg::Int32>("
    image_broadcaster_timer_ = this->create_wall_timer(500ms, std::bind(&Co
```

Figure 5.1: Example of generated code with comments

Figure 5.1 illustrates how the comments vary depending on the component, this also provides a degree of separation that helps developers understand where parts of the input model are being implemented.

More examples of code comments (with before and after) can be found in the Appendix A.1.

---

[1]This affix is not affected by the naming convention.

5.3.3 Traceability

The implementation of traceability is related with the report feature, however, traceability can exist without a report and a report can exist without traceability 5.3.4. The main idea of traceability is to relate the coding files directly to the model so we are able to verify that a certain file came from a certain part of the model.

The base construction of the trace file was done in JSON, this provides information while still maintaining a structure that can easily be ported elsewhere.

This is an example of the JSON produced by a traceability analysis:

```
{
 "system": "Robot_Vision_I_Instance",
 "generatedAt": "2025-07-28 16:26:49",
 "location": "C:/Users/aurel/git/codegen-config/fr.mem4csd.codegen.config/results",
 "components": {
  "Usb_Cam_Nd": {
   "file": "robot_vision_i_instance/src/Usb_Cam_Nd_process.cpp",
   "type": "process",
   "subcomponent": [
   {
    "name": "Image_Broadcaster",
    "qualifiedName": "Robot_Vision_I_Instance::Usb_Cam_Nd::Image_Broadcaster",
    "codeTrace": "@trace name='Image_Broadcaster'",
    "line": "15",
    "type": "thread",
    "properties": {
     "Dispatch_Protocol": "",
     "Period": ""
    }
   },
   {
    "name": "UsbSpinner",
    "qualifiedName": "Robot_Vision_I_Instance::Usb_Cam_Nd::UsbSpinner",
    "codeTrace": "@trace name='UsbSpinner'",
    "line" : "20",
    "type": "thread",
    "properties": {
     "Dispatch_Protocol": "",
     "Period": ""
    }
   }
   ]
```

```
},(...)
```

As we can see, it relates the model to the code in various forms, mainly code traces, qualified names, file and package locations. On top of that, it provides additional info that is useful for both developers and end users, developers can easily use this JSON as input for various tools (generating an alternate report from the one provided for example) and users can use it to trace parts of the code dedicated to a specific component or subcomponent using a combination of the file locations, code traces and line number.

The way the line number is detected is quite complex in practice but more straightforward in theory:

- When traceability is enabled, the code generates with traces, those traces are not language or tool specific, hence they don't get picked up by other tools or as language syntax.

- The same traces are generated in the JSON.

- Using the JSON and the traces corresponding to sub components in the larger component (which is represented by a C++ file) the C++ files are scavenged for traces, which, when found, get saved directly below the *codeTrace*, as the line correspondent to it.

In addition to this, traceability links were created using Capra to ensure that the JSON file can be directly related to the C++ files using the @*trace* indicative.

5.3.4   Report

The report was implemented with the goal of providing a clear, concise and structured way of visualizing exactly how the model translates to code. To do this, the report is generated in HyperText Markup Language (HTML) using Acceleo. This is important since it overcomes a few constraints, mainly:

- **PDF with Java:** The major issue that producing HTML code with Acceleo instead of a direct PDF with Java fixes is that producing PDFs with Java code, although more practical on the surface, has many limitations when it comes to designing and structuring PDFs. In addition, the verbose is very unintuitive and can quickly grow in complexity.

- **Extensibility:** Since HTML is a simple and well known language, adding details and making changes is straightforward when comparing it to other tools like LaTeX. In this case, the usage of LaTeX isn't really justified since it is more directed to academic circumstances and not fast paced development.

- **Styling Issues:** Using raw Cascading Style Sheets (CSS) in combination with HTML is no doubt the best choice when it comes to styling the report. Java has many limitations when it comes to styling and LaTeX is very strict and error prone. For a basic report such as the one generated, CSS styling is more than enough.

Once these constraints were no longer an issue, the development of the report was much more intuitive. The base was made with HTML and CSS, which enabled a high degree of custom styles while also simplifying the integration of useful information.

Since the first iteration of the report was done in HTML, the transition to PDF was mostly seamless, however, its implementation was done not by Java code, but via Workflow. This was at a later stage, a more detailed explanation can be found in Section **??**.

The report has many features, its high customization enables the creation of different parts when needed, hence omitting redundant or unwanted sections according to user needs. One of the core elements is the ROS specific option, which highlights ROS specific components that were generated or ported from the model. This allows ROS developers to have a much clearer idea of how parts of the input model were represented in the ROS part of the code, for instance, what thread corresponds to a Publisher[2] or Subscriber[3].

5.3.5 Code Quality Checker

A fine addition to the code generator was the introduction of a code quality checker powered by *Cppchecker*.

5.3.5.1 Cppchecker

Cppcheck is a static analysis tool for C and C++ code, it's designed to detect a wide range of programming errors, potential bugs, and code style issues without executing the program. *Cppcheck* analyzes the source code to identify problems such as memory leaks, uninitialized variables, unused functions, and violations of coding best practices. It's a lightweight, platform-independent, and highly configurable tool, which makes it suitable for integration into our code generation workflow.

Cppcheck has both an open source and a commercial version, in this case,

5.3.6 ROS and C++ separation

One of the underlining problems of the base generated code was the fact that ROS code and C++ code where closely coupled together. This is an issue since most of the C++ implementation is ROS independent, meaning that it could, in theory, be reused in contexts where ROS is not available or not needed. However, since the generated code is tight coupled with ROS, it forces the presence of ROS-specific dependencies even in parts of the system that do not require them.

---

[2]Publishers generate and publish data (sensor readings, etc).
[3]Subscribers listen for and process data sent by Publishers.

To combat this, a new approach to code generation was implemented. The main goal was to port the ROS code inside the C++ files onto another folder, adjacent to the remaining C++ files. This can be better understood visually, we part from this:

```
src/
  Color_Tracker_process.cpp
  Image_Display_2_process.cpp
```

And we split the files into their ROS and C++ counterparts, like so:

```
src/
  impl/
  Color_Tracker_logic.cpp
  Image_Display_2_logic.cpp
  ros/
  Color_Tracker_node.cpp
  Image_Display_2_node.cpp
```

With the files now separated, it is possible to reuse the C++ code present in the impl folder without the ROS dependencies. This also opens more the window to the inclusion of company specific libraries that can, if needed, be split into ROS and C++ only counterparts.

5.3.7   Legacy Code Integration

The main problem that legacy code integration solves is the following:

There are C++ libraries that need to be used by specific ROS components, these libraries should be included in the ROS project when needed.

To put it simply, in order to produce the functionality that the input model entices, certain legacy libraries need to be included, not just in the project, but called in the code aswell. The way it was done is clear:

- Firstly, libraries that the code will need based on model components used are listed.

- Then, the code for those libraries is copied to the project folder.

- Lastly, this newly copied code is called in certain places based on functional necessity.

6

SOFTWARE TESTING

This chapter documents the software testing done throughout the development of the thesis. It highlights how the specific tests where done, in what context and their results.

Intro intro intro?

6.1 Feature Testing

This type of testing is recurrent along the development, it serves to ensure that the features implemented are functioning as intended and that the previous version of the software has retained its functionality when new features were added.

6.2 Feature Testing

After the development of the first batch of features, the testing served to ensure that the previous version of the code generator was still working as intended and to validate the integration of the code generator configurator in the pipeline.

The report for the first volley of tests can be found in the Annex **??**, where it is explained the purpose of each individual test and the results they withheld.

The regression tests applied at this time were limited since the previous version of the software did not include any code generation customization, only the simple transition from model to code (M2T), which, after generating a ROS project in both the previous code generator and the at the time current one (with the first batch of features implemented), it was possible to verify that both generators still produced the same practical output, essentially, the code produced by both generators yields the same functionality.

The integration test part was done much more in depth since the goal was to intertwine the different new features and find out if they worked in perfect harmony as expected. The following tests were conducted:

- **TC-INT-01 – Identifier Consistency:** Verify that the identifier configuration set in the `config.generator` model is applied consistently in generated C++ code and reflected in the traceability file as well as in the report.

- **TC-INT-02 – Comment Preservation Across Modules:** Generate code from a model containing both single-line and multi-line comments, verifying that:

  1. Acceleo templates correctly output the comments.

  2. Java generator preserves formatting.

- **TC-INT-03 – Traceability Link Accuracy:** Confirm that model elements are mapped to the correct files and line numbers in generated code, and that these mappings are:

  1. Present in the JSON traceability output.

  2. Referenced in the HTML report.

- **TC-INT-04 – Report Generation Data Flow:** Metrics collected during generation (number of components, code quality warnings) are passed from the Java generator to the report module and rendered correctly in the final report.

- **TC-INT-05 – Code Quality & Report Synchronization:** Trigger a known style violation in the model, generate code, and verify that:

  1. The code quality checker detects it and outputs XML.

  2. The Java generator correctly embeds this warning into the final HTML reportx.

- **TC-INT-07 – Cross-Feature Stability:** Use a large, complex model with:

  - Custom identifiers

  - Extensive comments

  - Nested subcomponents

  - Known style violations

Verify that all features (identifier config, comments, traceability, reporting, quality checking) work together without breaking any module.

Bibliography

[1] Addison-Wesley, 2015 (cit. on p. 25).

[2] In: *Developing Safety-Critical Software* (2017-12), pp. 51–72. DOI: `10.1201/9781315218168-7` (cit. on p. 15).

[3] URL: `https://mem4csd.telecom-paristech.fr/blog/index.php/ramses/` (cit. on p. 5).

[4] URL: `https://www.mathworks.com/help/simulink/` (cit. on p. 5).

[5] URL: `https://www.mathworks.com/help/simulink/automotive-applications.html` (cit. on p. 6).

[6] URL: `https://wiki.eclipse.org/Acceleo/User_Guide` (cit. on pp. 9, 11).

[7] URL: `https://www.mathworks.com/help/simulink/ug/create-a-template-from-a-model.html` (cit. on p. 11).

[8] URL: `https://www.dspace.com/en/pub/home/support/kb/dsutil/kbtladd/tlutil.cfm` (cit. on p. 11).

[9] URL: `https://eclipse.dev/atl/documentation/old/ATL_Flyer_Normal_Version.pdf` (cit. on p. 14).

[10] URL: `https://google.github.io/styleguide/cppguide.html` (cit. on p. 21).

[11] URL: `https://peps.python.org/pep-0008/` (cit. on p. 21).

[12] URL: `https://www.eclipse.org/forums/index.php/t/1092427/` (cit. on p. 21).

[13] R. G. Bias and D. J. Mayhew. *Cost-justifying usability*. Academic Press, 1994 (cit. on p. 25).

[14] E. Borde et al. "Architecture models refinement for fine grain timing analysis of embedded systems". In: *2014 25nd IEEE International Symposium on Rapid System Prototyping* (2014-10), pp. 44–50. DOI: `10.1109/rsp.2014.6966691` (cit. on p. 2).

[15] M. Broy et al. "Engineering automotive software". In: *Proceedings of the IEEE* 95.2 (2007-02), pp. 356–373. DOI: `10.1109/jproc.2006.888386` (cit. on p. 1).

[16] M. Consortium. *MISRA C:2025 Guidelines for the use of the C language in critical systems*. The MISRA Consortium Limited, 2025. ɪsʙɴ: 978-1-911700-19-7 (cit. on pp. 7, 12, 15).

[17] M. Consortium and C. Tapp. *MISRA C++:2023 Guidelines for the use of C++17 in critical systems*. The MISRA Consortium Limited, 2023. ɪsʙɴ: 978-1911700104 (cit. on pp. 7, 12, 15).

[18] K. Czarnecki and S. Helsen. "Feature-based survey of Model Transformation Approaches". In: *IBM Systems Journal* 45.3 (2006), pp. 621–645. ᴅᴏɪ: 10.1147/sj.453.0621 (cit. on p. 2).

[19] R. Debouk. "Overview of the second edition of ISO 26262: Functional Safety—Road Vehicles". In: *Journal of System Safety* 55.1 (2019-03), pp. 13–21. ᴅᴏɪ: 10.56094/jss.v55i1.55 (cit. on p. 15).

[20] P. Feiler, B. Lewis, and S. Vestal. "The SAE Architecture Analysis and Design Language (Aadl) a standard for Engineering Performance Critical Systems". In: *2006 IEEE Conference on Computer-Aided Control Systems Design* (2006-10), pp. 1206–1211. ᴅᴏɪ: 10.1109/cacsd.2006.285483 (cit. on pp. 1, 4).

[21] R. France and B. Rumpe. "Model-driven development of complex software: A research roadmap". In: *Future of Software Engineering (FOSE '07)* (2007-05), pp. 37–54. ᴅᴏɪ: 10.1109/fose.2007.14 (cit. on p. 4).

[22] J. García-García et al. "NDT-Suite: A Methodological Tool Solution in the Model-Driven Engineering Paradigm". In: *Journal of Software Engineering and Applications* 7.4 (2014), pp. 206–217. ᴅᴏɪ: 10.4236/jsea.2014.74022 (cit. on p. 4).

[23] O. Khatib and B. Siciliano. *Springer Handbook of Robotics*. Springer International Publishing: Imprint: Springer, 2016 (cit. on p. 1).

[24] E. A. Lee. "Cyber Physical Systems: Design Challenges". In: *2008 11th IEEE International Symposium on Object and Component-Oriented Real-Time Distributed Computing (ISORC)*. 2008, pp. 363–369. ᴅᴏɪ: 10.1109/ISORC.2008.25 (cit. on p. 1).

[25] K. Mikova. *Why code flexibility is crucial in low-code development?* 2025-02. ᴜʀʟ: https://www.appbuilder.dev/blog/code-flexibility (cit. on p. 2).

[26] R. Mittal et al. "Solving the instance model-view update problem in AADL". In: *Proceedings of the 25th International Conference on Model Driven Engineering Languages and Systems* (2022-10), pp. 55–65. ᴅᴏɪ: 10.1145/3550355.3552396 (cit. on p. 14).

[27] *OpenAPI Generator Configuration*. Accessed: 2025-05-26. ᴜʀʟ: https://openapi-generator.tech/docs/configuration/ (cit. on p. 7).

[28] *OpenAPI Generator Plugins*. Accessed: 2025-05-26. ᴜʀʟ: https://openapi-generator.tech/docs/plugins/ (cit. on p. 7).

[29] *OpenAPI Generator Template Customization*. Accessed: 2025-05-26. ᴜʀʟ: https://openapi-generator.tech/docs/customization/ (cit. on p. 7).

[30]   *OpenAPI Generators.* Accessed: 2025-05-26. URL: https://openapi-generator.tech/docs/generators/ (cit. on p. 6).

[31]   OpenAPITools. *OpenAPITools/openapi-generator: Openapi generator allows generation of API client libraries (SDK Generation), server stubs, documentation and configuration automatically given an openapi Spec (V2, v3).* URL: https://github.com/OpenAPITools/openapi-generator (cit. on p. 6).

[32]   *OpenModelica - Open Source Modelica-based Modeling and Simulation Environment.* Accessed: 2025-05-27. URL: https://openmodelica.org/ (cit. on p. 6).

[33]   *OpenModelica Users Guide: Code Generation and Simulation.* Accessed: 2025-05-27. URL: https://openmodelica.org/doc/OpenModelicaUsersGuide/latest/ (cit. on p. 6).

[34]   *OpenModelica Users Guide: Performance Considerations.* Accessed: 2025-05-27. URL: https://openmodelica.org/doc/OpenModelicaUsersGuide/latest/profiler.html (cit. on p. 6).

[35]   R. Rajkumar et al. "44.1 Cyber-Physical Systems: The Next Computing Revolution". In: 2010-06, pp. 731–736. DOI: 10.1145/1837274.1837461 (cit. on p. 1).

[36]   D. Schmidt. "Guest Editor's Introduction: Model-Driven Engineering". In: *Computer* 39.2 (2006), pp. 25–31. DOI: 10.1109/MC.2006.58 (cit. on pp. 1, 4).

[37]   M. Voelter and S. Benz. *DSL Engineering: Designing, implementing and using domain-specific languages.* Dslbook.org, 2013 (cit. on p. 22).

[38]   E. Web. *Eclipse EMF.* 2025-03. URL: https://projects.eclipse.org/projects/modeling.emf.emf (cit. on p. 21).

A.1 Comment Configuration

A.1.1 Comment Configuration Example 1

```
1
2
3   #include "rclcpp/rclcpp.hpp"
4   #include "std_msgs/msg/int32.hpp"
5
6   using namespace std::chrono_literals;
7
8       class UsbCamNd : public rclcpp::Node
9       {
10          public:
11
12              UsbCamNd(): Node("UsbCamNd")
13              {
14
15                  image_broadcaster_pub_ = this->create_publisher<std_msgs::msg::Int32>("/usb_cam_nd/image_broadcaster", 10);
16                  image_broadcaster_timer_ = this->create_wall_timer(500ms, std::bind(&UsbCamNd::image_broadcaster_timer_callback, this));
17
18
19                  usb_spinner_sub_ = this->create_subscription<std_msgs::msg::Int32>("/usb_cam_nd/usbspinner",
20                      10, std::bind(&UsbCamNd::usb_spinner_callback, this, std::placeholders::_1));
21
22              }
```

Figure A.1: Example of generated code without comments

When code comments are toggled on, the code from Figure A.1 becomes the code from Figure A.2

45

```
comment_example1.png
```

Figure A.2: Example of generated code with comments