



**УНИВЕРЗИТЕТ У НОВОМ САДУ
ФАКУЛТЕТ ТЕХНИЧКИХ НАУКА**



Димитрије Ћук

**Од изворног С кода до извршне бинарне
слике - компилација и распоред у меморији
микроконтролера**

Дипломски рад

Нови Сад 2025.



УНИВЕРЗИТЕТ У НОВОМ САДУ • ФАКУЛТЕТ ТЕХНИЧКИХ НАУКА

21000 НОВИ САД, Трг Доситеја Обрадовића 6

КЉУЧНА ДОКУМЕНТАЦИЈСКА ИНФОРМАЦИЈА

Редни број, РБР:											
Идентификациони број, ИБР:											
Тип документације, ТД:	Монографска публикација										
Тип записа, ТЗ:	Текстуални штампани примерак										
Врста рада, ВР:	Дипломски рад										
Аутор, АУ:	Димитрије Ћук										
Ментор, МН:	проф. др Дарко Марчетић										
Наслов рада, НР:	Од изворног С кода до извршне бинарне слике - компилација и распоред у меморији микроконтролера										
Језик публикације, ЈП:	српски										
Језик извода, ЈИ:	српски										
Земља публикавања, ЗП:	Република Србија										
Уже географско подручје, УГП:	АП Војводина										
Година, ГО:	2024.										
Издавач, ИЗ:	Ауторски репринт										
Место и адреса, МА:	Факултет техничких наука, 21000 Нови Сад, Трг Доситеја Обрадовића 6										
Физички опис рада, ФО: (поглавља/страна/ цитата/табела/слика/графика/прилога)	(9/72/15/2/35/0/0)										
Научна област, НО:	Електротехничко и рачунарско инжењерство										
Научна дисциплина, НД:	Рачунарске науке и уграђени системи										
Предметна одредница/Кључне речи, ПО:	Меморија у микроконтролеру, компајлер и компајлирање, линкер и линковање										
УДК											
Чува се, ЧУ:	Библиотека ФТН, Трг Доситеја Обрадовића 6, Нови Сад										
Важна напомена, ВН:											
Извод, ИЗ:	<p>С обзиром да у микроконтролеру не постоји оперативни систем који управља меморијом, потребно је конфигурисати линкер кроз линкер директиву, односно линкер скрипту, на начин да уважи физичку расподелу меморије по врстама и адресама, на основу техничког листа, тј. спецификације коју произвођач доставља.</p> <p>Дигитално потписивање је процес који осигурава аутентичност и интегритет података коришћењем криптографије. Користи се асиметрична енкрипција која подразумева постојање приватног и јавног кључа.</p>										
Датум прихватања теме, ДП:	10.09.2024.										
Датум одбране, ДО:	27.09.2024.										
Чланови комисије, КО:	<table><tr><td>Председник:</td><td>др Владимир Поповић, доц., ФТН Нови Сад</td></tr><tr><td>Члан:</td><td>др Стеван Цветићанин, ван. проф., ФТН Нови Сад</td></tr><tr><td>Члан:</td><td></td></tr><tr><td>Члан:</td><td></td></tr><tr><td>Члан, ментор:</td><td>др Дарко Марчетић, ред. проф., ФТН Нови Сад</td></tr></table>	Председник:	др Владимир Поповић, доц., ФТН Нови Сад	Члан:	др Стеван Цветићанин, ван. проф., ФТН Нови Сад	Члан:		Члан:		Члан, ментор:	др Дарко Марчетић, ред. проф., ФТН Нови Сад
Председник:	др Владимир Поповић, доц., ФТН Нови Сад										
Члан:	др Стеван Цветићанин, ван. проф., ФТН Нови Сад										
Члан:											
Члан:											
Члан, ментор:	др Дарко Марчетић, ред. проф., ФТН Нови Сад										
	<table><tr><td>Потпис ментора</td></tr><tr><td></td></tr></table>	Потпис ментора									
Потпис ментора											

Образац **Q2.НА.06-05** - Издање 1




UNIVERSITY OF NOVI SAD • FACULTY OF TECHNICAL SCIENCES

21000 NOVI SAD, Trg Dositeja Obradovića 6

KEY WORDS DOCUMENTATION

Accession number, ANO :											
Identification number, INO :											
Document type, DT :	Monographic publication										
Type of record, TR :	Textual Printed Material										
Contents code, CC :	Bachelor thesis										
Author, AU :	Dimitrije Ćuk										
Mentor, MN :	prof. dr Darko Marčetić										
Title, TI :	Memory Mapping in a Microcontroller and the Use of Digital Signatures										
Language of text, LT :	Serbian										
Language of abstract, LA :	Serbian										
Country of publication, CP :	Republic of Serbia										
Locality of publication, LP :	AP of Vojvodina										
Publication year, PY :	2024.										
Publisher, PB :	Author's reprint										
Publication place, PP :	Faculty of technical sciences, 21000 Novi Sad, Trg Dositeja Obradovića 6										
Physical description, PD : (chapters/pages/ref./tables/pictures/graphs/appendixes)	(9/72/15/2/35/0/0)										
Scientific field, SF :	Electrical and computer engineering										
Scientific discipline, SD :	Computer science and embedded systems										
Subject/Key words, S/KW :	From Source C Code to Executable Binary Image – Compilation and Memory Layout in Microcontrollers										
UC											
Holding data, HD :	Library of Faculty of technical sciences, Trg Dositeja Obradovića 6, Novi Sad										
Note, N :											
Abstract, AB :	<p>Since there is no operating system in a microcontroller to manage memory, it is necessary to configure the linker through a linker directive or linker script in such a way that it takes into account the physical distribution of memory by type and address, based on the datasheet or specification provided by the manufacturer.</p> <p>Digital signing is a process that ensures the authenticity and integrity of data through cryptography. It uses asymmetric encryption, which involves the existence of both a private and a public key.</p>										
Accepted by the Scientific Board on, ASB :	10.09.2024.										
Defended on, DE :	27.09.2024.										
Defended Board, DB :	<table><tr><td>President:</td><td>dr Vladimir Popović, assist. prof., FTN Novi Sad</td></tr><tr><td>Member:</td><td>dr Stevan Cvetičanin, assoc. prof., FTN Novi Sad</td></tr><tr><td>Member:</td><td></td></tr><tr><td>Member:</td><td></td></tr><tr><td>Member, mentor:</td><td>dr Darko Marčetić, full prof., FTN Novi Sad</td></tr></table>	President:	dr Vladimir Popović, assist. prof., FTN Novi Sad	Member:	dr Stevan Cvetičanin, assoc. prof., FTN Novi Sad	Member:		Member:		Member, mentor:	dr Darko Marčetić, full prof., FTN Novi Sad
President:	dr Vladimir Popović, assist. prof., FTN Novi Sad										
Member:	dr Stevan Cvetičanin, assoc. prof., FTN Novi Sad										
Member:											
Member:											
Member, mentor:	dr Darko Marčetić, full prof., FTN Novi Sad										
	<table><tr><td>Menthor's sign</td></tr><tr><td></td></tr></table>	Menthor's sign									
Menthor's sign											

Образац Q2.HA.06-05 - Издање 1

	УНИВЕРЗИТЕТ У НОВОМ САДУ • ФАКУЛТЕТ ТЕХНИЧКИХ НАУКА	Број:
	21000 НОВИ САД, Трг Доситеја Обрадовића 6	012-40/1732
	ЗАДАТАК ЗА ДИПЛОМСКИ РАД	Датум: 10.09.2024.

(Податке уноси предметни наставник - ментор)

СТУДИЈСКИ ПРОГРАМ:	Е1 – енергетика, електроника, телекомуникације
РУКОВОДИЛАЦ СТУДИЈСКОГ ПРОГРАМА:	др Милан Сечујски

Студент:	Димитрије Ћук	Број индекса:	ЕЕ 3/2016
Област:	Рачунарске науке и уграђени системи		
Ментор:	др Дарко Марчетић		
НА ОСНОВУ ПОДНЕТЕ ПРИЈАВЕ, ПРИЛОЖЕНЕ ДОКУМЕНТАЦИЈЕ И ОДРЕДБИ СТАТУТА ФАКУЛТЕТА ИЗДАЈЕ СЕ ЗАДАТАК ЗА ДИПЛОМСКИ РАД, СА СЛЕДЕЋИМ ЕЛЕМЕНТИМА: <ul style="list-style-type: none"> - проблем – тема рада; - начин решавања проблема и начин практичне провере резултата рада, ако је таква провера неопходна; 			

НАСЛОВ ДИПЛОМСКОГ РАДА:

Од изворног С кода до извршне бинарне слике - компилација и распоред у меморији микроконтролера
--

ТЕКСТ ЗАДАТКА:

--

Руководилац студијског програма:	Ментор рада:
др Милан Сечујски	др Дарко Марчетић

Примерак за: <input type="radio"/> - Студента; <input type="radio"/> - Ментора
--

Садржај

1. Увод.....	5
1.1. Циљ и мотивација рада.....	5
1.2. Методологија и приступ	5
1.3. Релевантност теме.....	5
2. Улога C језика у програмирању микроконтролера	6
2.1. Историјска еволуција C језика у embedded окружењима.....	6
2.2. Кључне особине C језика у embedded контексту	7
2.3. Упоредна анализа C језика и алтернативних језика	7
3. Организација C изворног кода за embedded окружења.....	8
3.1. Основне компоненте embedded пројекта	9
3.2. Употреба CMSIS и HAL слојева.....	10
3.3. Стил програмирања и стандардизација.....	11
3.4. Приступ меморијски мапираним регистрима.....	12
3.5. Пример за Infineon TRAVEO™ T2G	14
4. Фазе компилације (превођења) у GCC.....	15
4.1. Препроцесирање.....	16
4.2. Компилација	17
4.3. Асемблирање.....	18
4.4. Линковање.....	19
4.5. Интеграција фаза компилације у оквиру GCC алатке	20
5. Формати резултујућих датотека	21
5.1. ELF формат	22
5.2. Intel HEX формат	23
5.3. RAW бинарни формат (.bin).....	24
6. Употреба GNU алата: objcopy, readelf, nm, size, objdump.....	26
6.1. objcopy	27
6.2. readelf.....	28
6.3. nm	29
6.4. size.....	30
6.5. objdump	31
6.6. Закључак.....	32
7. Линкерска скрипта: MEMORY и SECTIONS дефиниције	33
7.1. MEMORY дефиниција	34
7.2. SECTIONS расподела	35
8. Закључак.....	37

1. Увод

Програмски језик C доминира у развоју софтвера за микроконтролере захваљујући својој ефикасности и близини хардверу. Овај рад обрађује процес развоја једног C програма за 32-битни аутомобилски микроконтролер TRAVEO™ T2G (породица CYT3BB/4BB) заснован на ARM Cortex-M7 архитектури. Кроз логичку целину, рад покрива фазе писања и организације изворног кода, преко превођења (компилације) и линковања у извршну бинарну датотеку. Биће описан формат генерисаних датотека (ELF, HEX, BIN) и алати за њихову анализу (GNU binutils алати), као и структура линкерске скрипте прилагођена конкретној меморијској архитектури TRAVEO T2G чипа.

1.1. Циљ и мотивација рада

Циљ овог рада је да се систематски прикаже цео ток развоја уграђеног софтвера — од писања изворног C кода до добијања финалне бинарне слике која се програмира у меморију микроконтролера. Фокус је на компилационом процесу у оквиру GCC алатке, као и на детаљној анализи распореда меморијских секција кроз линкерску скрипту прилагођену конкретном микроконтролеру. У контексту растуће примене микроконтролера у аутомобилској индустрији, индустрији аутоматике и IoT систему, дубље разумевање унутрашњег функционисања овог процеса је од суштинске важности за пројектовање поузданих и безбедних система.

1.2. Методологија и приступ

Рад прати практичан инжењерски приступ: анализира се развој C програма за Infineon TRAVEO™ T2G микроконтролер (CYT4BB породица), који користи ARM Cortex-M7 језгро. Изложена су сва релевантна техничка средства: GCC компајлер и GNU binutils алати, структура startup кода и организација пројекта, као и конфигурација линкерске скрипте на основу спецификације меморијске мапе циљног хардвера. Кроз све фазе (препроцесирање, компилација, асемблирање, линковање), обрађен је начин на који се C код претвара у ELF датотеку, која се затим конвертује у HEX или BIN формат погодан за флешовање.

1.3. Релевантност теме

Обрада теме је уско везана за потребе реалног развоја софтвера у embedded окружењу, посебно у системима који раде без оперативног система (bare-metal). Правилна организација извора, разумевање формата резултујућих датотека и контрола над меморијским распоредом су предуслов за функционалан и безбедан рад микроконтролера у критичним апликацијама. Такође, разумевање ових аспеката представља основу за проширене теме као што су имплементација bootloader-a, сигурносна верификација кода (нпр. путем дигиталног потписа), и оптимизација потрошње ресурса.

2. Улога C језика у програмирању микроконтролера

Језик C је најзаступљенији у програмирању микроконтролера због своје флексибилности и ефикасности (мали оверхед). Под флексибилношћу се подразумева могућност директног приступа хардверским ресурсима, као што су меморијске адресе и регистри. Са друге стране, мали оверхед значи да компајлирани код заузима минимално меморије и омогућава брзо извршавање. Захваљујући овим особинама, C омогућава оптимално коришћење ограничених ресурса карактеристичних за уграђене системе.

Практично све – од малих контролера у уређајима до оперативних система – може бити написано у C-у због његове преносивости и способности да уз минималне наредбе пружи максималну контролу над хардвером. У домену уграђених система, C пружа низак ниво апстракције: омогућава директан приступ меморијским адресама и периферијама, управљање битовима и регистрима, као и прецизну контролу над временски критичним секцијама кода. За разлику од виших језика, компајлер C језика генерише ефикасан машински код који се извршава готово једнако брзо као ручно писани асемблер, чиме је погодан за примене у реалном времену (real-time). Истовремено, C код је знатно читљивији и одрживији од чистог асемблера, што је важно при тимском развоју софтвера. Захваљујући овим особинама, C (и његов надскуп C++) је постао стандард у развоју фирмвера за микроконтролере, омогућивши и преносивост кода између различитих архитектура. Другим речима, C обезбеђује **директан приступ хардверу** и **високе перформансе**, што су кључни захтеви у embedded систему.

2.1. Историјска еволуција C језика у embedded окружењима

Језик C је развијен раних 1970-их у Bell Labs-у за потребе оперативног система UNIX, али је због своје минималистичке и ефикасне структуре веома брзо нашао примену у embedded системима. Током развоја микроконтролера и појаве 8-битних и 16-битних архитектура, C је потиснуо асемблер као доминантни језик због боље читљивости и лакше преносивости. Данас, C представља основну полазну тачку за развој софтвера у реалновременским и ресурсно ограниченим системима, посебно захваљујући стандардизацији преко ISO/IEC 9899:2018 (C18) и спецификацијама као што је CMSIS (Cortex Microcontroller Software Interface Standard) од стране ARM-a.

2.2. Кључне особине С језика у embedded контексту

У контексту програмирања микроконтролера, С се издваја следећим особинама:

- **Директан рад са меморијом** — Показивачи омогућавају манипулацију на нивоу бајта и регистра, што је неопходно за рад са периферијама. Кроз `volatile` квалификатор могуће је обезбедити исправно понашање при асинхроним изменама вредности (нпр. од стране хардвера или прекида).
- **Фино управљање меморијским распоредом** — Коришћењем `__attribute__((section("секција")))`, програмер може утицати на то у коју меморијску регију ће одређена функција или променљива бити смештена, што је критично у систему без оперативног система.
- **Интеграција са асемблером** — Када је неопходна максимална оптимизација или директна манипулација регистрима, С омогућава уметање асемблерских инструкција (`__asm__`) или позиве на екстерне рутине, што га чини флексибилним алатом за нисконивоовски развој.
- **Прецизна контрола времена извршавања** — У `real-time` системима детерминистичност је од пресудне важности. С не садржи `runtime` механизме попут `garbage collector`-а, чиме обезбеђује временски предвидљиво понашање.

2.3. Упоредна анализа С језика и алтернативних језика

Иако се у `embedded` индустрији појављују и други језици (нпр. Rust, Ada, Python/MicroPython), ниједан не достиже ниво подршке и зрелости који има С. Python се користи углавном у образовне и прототипске сврхе, Rust још увек има ограничену подршку за специфичне архитектуре и `toolchain`-ове, док је Ada присутна углавном у високо-регулисаним индустријама (авионика, нуклеарна техника).

С се тако намеће као оптималан компромис између перформанси, контроле и одрживости. Он пружа довољну блискост хардверу за временски критичне апликације, а истовремено омогућава тимски развој, проверу стандарда као што су MISRA и лако повезивање са библиотекама и хардверским апстракционим слојевима.

3. Организација С изворног кода за embedded окружења

Развој софтвера за микроконтролере у програмском језику С захтева дисциплиновану и модуларну организацију изворног кода, посебно у условима где не постоји оперативни систем и где је програм одговоран за директно управљање хардвером — такозвано **bare-metal програмирање**. У овом контексту, коректна организација пројекта није само питање структуралне естетике, већ предуслов за исправно иницијализовање система, ефикасну меморијску расподелу и могућност проширивости и тестирања.

3.1. Основне компоненте embedded пројекта

Типичан C пројекат за микроконтролер обухвата следеће компоненте:

- `main.c` — Главна изворна јединица која садржи `main()` функцију и представља улазну тачку програма. У њој се обично врши иницијализација система (хардвера) и започиње такозвани `super-loop` или `RTOS scheduler`.
- **Модули/драјвери** — Функционално одвојене целине попут `gpio.c`, `uart.c`, `motor_control.c`, итд. Сваки модул обично има и одговарајућу заглавну датотеку (`.h`), која излаже јавни интерфејс модула.
- `startup.s` или `startup.c` — Код који се извршава одмах након ресета микроконтролера. Садржи векторску табелу прекида, `reset handler`, и иницијализацију стек показивача.
- `linker.ld` — Линкерска скрипта која дефинише начин расподеле кода и података у физичку меморију микроконтролера (Flash, SRAM и сл.).
- `system_*.c` — Фајл који садржи функције за иницијализацију системског такта и конфигурацију PLL/clock система, обично део HAL или CMSIS пакета.
- `Makefile` или `CMakeLists.txt` — Датотека за аутоматизацију компилације (у GCC контексту), или пројекат за IDE алат (нпр. IAR, Keil, STM32CubeIDE).

Оваква структура промовише **модуларност**, **одрживост** и **јединично тестирање**, што су кључне вредности у професионалном embedded развоју.

3.2. Употреба CMSIS и HAL слојева

У пракси, развојни процес се ослања на стандардизоване библиотеке:

- **CMSIS** (Cortex Microcontroller Software Interface Standard), развијен од ARM-а, дефинише структуре, симболе и функције за приступ системским и периферијским регистрима у Cortex-M архитектури. Пример су заглавља типа `core_cm7.h`, `system_device.h` и дефиниције као што је `SCB->VTOR`.
- **HAL (Hardware Abstraction Layer)** библиотеке, које пружају генерализоване функције за рад са периферијама (нпр. `HAL_UART_Transmit()`), смањују количину хардверски зависног кода. Иако уносе одређени оверхед, HAL слојеви убрзавају развој и олакшавају преносивост између различитих модела микроконтролера истог произвођача.

CMSIS обезбеђује формалну конзистентност, док HAL омогућава брзу продукцију прототипа. За перформансно критичне секције, HAL функције могу бити замењене директним приступом регистрима преко CMSIS симбола.

3.3. Стил програмирања и стандардизација

Због природе критичних апликација (аутомобилска индустрија, индустријска аутоматика, медицински уређаји), развој С кода у embedded систему често подлеже строгим стилским и безбедносним смерницама. Најзначајнији од њих је **MISRA C** стандард (Motor Industry Software Reliability Association), који дефинише правила за избегавање неодређеног понашања, нпр. забрана употребе динамичке алокације (malloc), неконтролисаних type cast-ова, и употребе goto.

Придржавање оваквих стандарда омогућава **формалну верификацију, аутоматизовану анализу кода** (нпр. алатима попут PC-lint или Coverity), и повећава поузданост и сигурност система. Поред MISRA-е, користе се и други стандарди као што су CERT C (сигурносне смернице) и ISO 26262 (функционална безбедност у аутомотив домену).

3.4. Приступ меморијски мапираним регистрима

Комуникација између процесора и хардвера у *embedded* системима реализује се кроз меморијски мапиране регистре, што подразумева директан упис у специфичне адресе у меморијском простору. Програмски језик С омогућава дефинисање структура које одговарају распореду регистра, уз употребу кључне речи **volatile**, која обавештава компајлер да не сме оптимизовати приступ тим променљивим, јер се њихове вредности могу мењати асинхроно — нпр. од стране прекидне рутине, хардверског тајмера или другог језгра процесора.

На пример:

```
typedef struct {
    volatile uint32_t IN;
    volatile uint32_t OUT;
    volatile uint32_t DIR;
    // ... остали регистри
} GPIO_TypeDef;
#define GPIO ((GPIO_TypeDef *) 0x50000000UL)
```

Горњи код показује дефиницију структуре која моделује GPIO регистре и симболичку адресу на коју она мапира. У том случају, приступ регистру **OUT** врши се на следећи начин:

- **GPIO->OUT = 0x1;**

Ова једноставна линија кода представља суштину рада са периферијама у *bare-metal* окружењу. Она подразумева:

- да **GPIO** представља показивач на структуру која моделује GPIO регистре;
- да оператор **->** омогућава приступ пољу **OUT** те структуре;
- да се вредност **0x1** уписује у регистар **OUT**, чиме се поставља логичка „1“ на пин број 0, док се остали пинови постављају на „0“.

Под условом да је претходно извршена конфигурација правца рада пинова (нпр. **GPIO->DIR |= 0x1;**), ова наредба резултује у појави високог логичког нивоа на излазном пину. У Assembly коду, оваква операција се мапира на једну инструкцију (**STR**) која уписује вредност у конкретну адресу у меморијском простору, чиме се постиже детерминистичко и брзо извршавање — особина кључна за реалновременске системе.

Захваљујући CMSIS библиотеци и заглављима произвођача, програмери више не морају да користе непрегледне изразе склоне грешкама попут:

```
*(volatile uint32_t *) (0x50000004) = 0x1;
```

већ могу користити симболичке, типски безбедне и читљиве форме попут `GPIO->OUT = 0x1 ;`, чиме се обезбеђује значајно већа одрживост, транспарентност и преносивост кода.

3.5. Пример за Infineon TRAVEO™ T2G

У случају микроконтролера TRAVEO T2G (нпр. из породице CYT4BB), произвођач Infineon испоручује почетни код (*startup assembly* или С датотеку) који садржи векторску табелу прекида и функцију за ресет. Такође, обезбеђују се CMSIS-заглавља специфична за циљани модел, као и HAL библиотеке за руковање периферијама. Програмер затим развија сопствени код у `main.c`, организује логичке модуле (нпр. `motor_control.c`, `sensors.c`), и укључује потребна заглавља.

Пројекат поред тога садржи и линкерску скрипту (`.ld` датотеку), као што је `TRAVEO_T2G.ld`, прилагођену меморијској мапи микроконтролера. Ова скрипта дефинише секције као што су `.text`, `.data`, `.bss`, `.stack`, и њихово постављање у Flash или RAM меморију. Тиме се омогућава да свака компонента програма буде смештена у предвиђено меморијско подручје током фазе линковања.

На овај начин, добром организацијом извора, доследним коришћењем CMSIS-а и HAL библиотека, као и строго дефинисаним фазама иницијализације и петље извршавања, обезбеђује се основа за стабилан, поуздан и преносив *embedded* систем, који може бити успешно преведен и повезан у потпуности током следеће фазе — компилације у оквиру GCC окружења.

4. Фазе компилације (превођења) у GCC

Превођење C програма у машински код одвија се кроз више дискретних фаза. GCC компајлер (GNU Compiler Collection) интерно дели процес на до четири корака: **препроцесирање**, **компилацију** (у ужем смислу), **асемблирање** и **линковање**, тим редоследом ([Overall Options \(Using the GNU Compiler Collection \(GCC\)\)](#)). Свака фаза има своју улогу у претварању изворног `.c` кода у извршну бинарну датотеку. Следи преглед ових фаза у табели 1.

Табела 1. Фазе превођења C програма уз GCC

Фаза	Алат (GCC позив)	Улаз	Изназ
Препроцесирање	<code>gcc -E</code>	<code>*.c</code> , <code>*.h</code> (изворник)	Препроцесирани код (<code>*.i</code>)
Компилација (C->ASM)	<code>gcc -S</code>	<code>*.i</code> (из претходног)	Асемблерски код (<code>*.s</code>)
Асемблирање	<code>gcc -c</code> или <code>as</code>	<code>*.s</code> (асемблерски код)	Објектни фајл (<code>*.o</code>)
Линковање	<code>gcc</code> или <code>ld</code>	<code>*.o</code> (+ библиотеке)	Извршна датотека (ELF)

4.1. Препроцесирање

Препроцесирање: Прва фаза у којој се извршава С препроцесор. Он обрађује директиве почевши знаком `#` – на пример, убацује садржај хедер датотека на место `#include` директива, проширује макрое дефинисане са `#define`, и условно уклања/укључује делове кода на основу `#ifdef` услова. Резултат ове фазе је *препроцесирани исходни код*, типично са екстензијом `.i` (или `.ii` за C++) који више не садржи препроцесорске директиве, већ само “чист” С код.

4.2. Компилација

Компилација (у ужем смислу): У овој фази GCC преводи препроцесирани C код у асемблерски код за циљну архитектуру. То значи да се синтакса и конструкције C језика преводe у низ асемблерских инструкција (нпр. ARM Cortex-M7 инструкције) које остварују еквивалентну функционалност. Излаз из ове фазе је `.s` датотека (асемблерски код у текстуалном облику). Ова фаза укључује и различите оптимизације које компајлер примењује (према подешеним опцијама, нпр. `-O2`) како би генерисани код био што ефикаснији.

4.3. Асемблирање

Асемблирање: Генерисани `.s` асемблерски код затим пролази кроз асемблер (саставни део GCC алата, нпр. `arm-none-eabi-as`), који га претвара у *релокативни објектни фајл* – машински код са нерешеним релокацијама и симболима. Ова датотека обично има екстензију `.o` и у формату је објектне датотеке (најчешће ELF формат, о чему ће бити речи касније). Објектни фајл садржи машинске инструкције за дати модул, али још увек није самостално извршна целина, јер адресе функција и података који се налазе у другим модулима нису још познате (остају као симболи које треба повезати).

4.4. Линковање

Линковање: Последња фаза је позив линкера (нпр. GNU ld) који узима један или више објектних фајлова (.o), као и евентуално предефинисане библиотеке (нпр. libc, или драјверске библиотеке), и **повезује** их у јединствену извршну датотеку. Линкер разрешава све међусобне референце – нпр. када функција у `main.o` зове функцију која је имплементирана у `uart.o`, линкер ће уписати исправну адресу те функције у машински код позива. Такође, линкер припаја и стандардни стартуп код (нпр. `crt0` за C) ако је део toolchain-а, мада у embedded окружењу стартуп и векторска табела обично долазе као засебан модул пројекта. Резултат линковања је извршна бинарна датотека у *ELF формату* (или сличном), са свим спојеним секцијама на одговарајућим меморијским адресама. Ова датотека је сада самосталан програм који се може учитати у меморију микроконтролера и покренути.

4.5. Интеграција фаза компилације у оквиру GCC алатке

Напоменимо да се у пракси већина ових корака обавља "у пролазу" помоћу исте `gcc` наредбе, јер GCC аутоматски позива препроцесор, па компајлер, асемблер и линкер. На пример, позив `arm-none-eabi-gcc -O2 -mcpu=cortex-m7 -o program.elf main.c uart.c startup.s -T linker.ld` ће обавити све кораке и произвести коначни `program.elf`. Ипак, корисно је разумети ове међукоре, јер алати омогућавају да се сваки корак изведе одвојено (нпр. опција `-save-temps` чува привремене `.i` и `.s` датотеке). GCC документација наглашава постојање наведене четири фазе и одговарајуће суфиксе/екстензије фајлова ([Overall Options \(Using the GNU Compiler Collection \(GCC\)\)](#)). Током компилације могу настати и помоћни фајлови као што је *листинг* (са мешовитим C кодом и асемблером, ако је затражено), али они нису нужни у даљем процесу.

Важно је истаћи да линкер за успешно повезивање за *embedded* мету мора знати распоред меморије циљног микроконтролера – ту ступа на снагу *линкерска скрипта* која описује меморијске регије (Flash, RAM) и како распоредити секције програма у њих. Линкерска скрипта је критична за добијање исправне бинарне слике програма и њена структура биће детаљно анализирана у посебном одељку.

5. Формати резултујућих датотека

Увод.

5.1. ELF формат

По завршеном линковању, добија се извршна датотека, најчешће у **ELF формату** (Executable and Linkable Format). ELF је стандардни бинарни формат који се користи на Unix/Linux системима за извршне датотеке, објектне модуле, па чак и библиотеке ([Executable and Linkable Format - Wikipedia](#)). Он је прихваћен и код cross-компајлера за микроконтролере јер је флексибилан и независан од архитектуре – подржава различите процесоре, ендијаност и величине адресног простора ([Executable and Linkable Format - Wikipedia](#)). За потребе embedded програмирања, ELF садржи све потребне информације о програму: машински код сегментиран у секције (.text, .data, .bss, итд.), али и симболичке табеле, таблице релокација, програмска заглавља са описом сегмената за извршавање, и опционе дебаг информације. ELF формат подржава двоструку анализу: према табели секција (*section header table*), која описује структуру изворног кода и симболе, или према табели сегмената (*program header table*), која описује начин учитавања у меморију при извршавању ([Executable and Linkable Format - Wikipedia](#)). У контексту микроконтролера, важнији је распоред секција, јер сегменти одговарају меморијским регијама у које ће секције бити смештене (Flash, RAM).

Иако ELF датотека садржи извршни код, она се обично **не програмира директно** у микроконтролер. Разлог је што ELF носи и метаподатке (нпр. симболе, одређене секције које нису потребне за сам рад програма) и није у формату који типични програмабилни хардвер очекује. Зато се из ELF-а изводе *чисти бинарни формати* погодни за флешовање. Најчешће се срећу три таква формата у embedded свету: **Intel HEX**, **RAW BIN** (сиров бинарни фајл), и **Motorola S-Record (S19)** формат.

5.2. Intel HEX формат

Intel HEX формат је текстуални формат у ASCII нотацији који представља садржај меморије у хексадецималном облику. Датотека се састоји од више линија, где свака линија представља један *рекорд* са одређеним бројем бајтова, њиховом адресом у меморији и контролном сумом ([Intel HEX - Wikipedia](#)). Конкретно, свака линија почиње двотачком :, затим следи бајт бројача (колико бајтова података та линија носи), па 16-битна почетна адреса, бајт типа записа (нпр. 00 за податке, 01 за крај датотеке, 04 за проширену адресу код већих адресних простора итд.), затим сами подаци (парови хекс цифара), и на крају једна контролна сума за проверу тачности ([Intel HEX - Wikipedia](#)). Овај формат је веома погодан јер је читљив и садржи адресе – нпр. ако програм није континуиран у меморији, HEX фајл може имати "рупе" у адресама између линија. Програматори (алати за флешовање) читају HEX датотеку линију по линију и уписују бајтове на наведене адресе у флеш меморију микроконтролера ([Intel HEX - Wikipedia](#)). Intel HEX је историјски настао 1970-их за потребе учитавања програма са папирне траке у Intel MCS systems, али је и данас широко коришћен због једноставности и поузданости провере (свака линија носи своју контролну суму) ([Intel HEX - Wikipedia](#)). Генерисање HEX фајла се обично ради алатом **objcopy**, о чему ће бити речи касније.

5.3. RAW бинарни формат (.bin)

RAW бинарни формат (.bin) је најједноставнији могући формат – низ бајтова идентичан бајтовима који треба да се упишу у меморију, без икакве додатне структуре или информација. Сирови бинарни фајл представља *меморијски дамп* програма, обично тачно оних секција које се налазе у непрекидном опсегу адреса. При конверзији ELF-а у .bin, одбацују се сви симболи, заглавља и вишак информација, и добија се само секвенца бајтова која одговара садржају флеша (и евентуално других меморија ако се спајају у један фајл). GNU `objcopy` алат омогућава ову конверзију: на пример команда `arm-none-eabi-objcopy -O binary program.elf program.bin` узима ELF и ствара .bin фајл. Према документацији, када се `objcopy` користи за генерисање raw binary датотеке, он ефективно производи меморијски дамп укупног садржаја ELF-а – од најнижег до највишег адресног бајта садржаног у ELF-у – одбацујући све симболе и релокације ([objcopy \(GNU Binary Utilities\)](#)). Битно је напоменути да .bin не носи информацију о томе на коју адресу ти бајтови треба да се упишу; претпоставља се подразумевани почетак (нпр. код већине микроконтролера почетак флеша је или 0x00000000 или нека позната базна адреса). Због тога, .bin формат се углавном користи када се читава слика програма ставља на почетак флеш меморије. Ако је потребно флешовати програм који не почиње од 0 или ако програм обухвата више одвојених меморијских области, Intel HEX је погоднији, јер носи адресе.

У пракси, многи произвођачи алата и IDE-ови (нпр. KEIL uVision, IAR EWARM, GCC toolchain) омогућавају генерисање HEX или BIN датотека из ELF-а. Неки *debugger*-и и *програматори* могу чак директно учитати ELF (користећи информације из ELF заглавља о сегментима за читавање). Ипак, имајући у виду да HEX и BIN представљају стандард у размену фирмвер слика (нпр. HEX за надоградњу софтвера у сервису, или BIN за брзо читавање преко bootloader-а), важно је разумети њихову структуру и разлике.

Motorola S-Record (S19)

Поред Intel HEX-а, чест је и **Motorola S-Record (S19)** формат – сличан ASCII хекс запису линија. Алати **objcopy** са опцијом **-O srec** може генерисати S-Record фајл ([objcopy \(GNU Binary Utilities\)](#)). Разлика је углавном у синтакси линија (S-Record линије почињу са 'S' и имају мало другачију организацију адресних поља). Пошто је у питању алтернативни формат, нећемо детаљно разматрати његову структуру, али вреди споменути да алат **srec_cat** (део SRecord пакета) може манипулисати и HEX и S19 фајловима.

6. Употреба GNU алата: objcopy, readelf, nm, size, objdump

При раду на озбиљним embedded пројектима, корисно је познавати алате за анализу и конверзију објектних и извршних датотека. GCC toolchain долази са низом **GNU binutils** алата који омогућавају увид у ELF садржај, конверзију формата, дисасемблирање, мерење величине и друго. У наставку су описани најважнији алати и њихова примена у контексту програмирања микроконтролера.

6.1. objcopy

objcopy: Алат за копирање и конверзију формата објектних датотека. Користи се за генерисање HEX или BIN фајла из ELF-а. Примери: `objcopy -O ihex program.elf program.hex` (створи Intel HEX из ELF-а), `objcopy -O binary program.elf program.bin` (створи raw бинарни дамп). `objcopy` може и издвајати поједине секције, уклањати debug информације опцијом `-S`, итд. Базиран је на BFD библиотеци, тако да подржава бројне формате и аутоматски препознаје улазни и излазни формат ([objcopy \(GNU Binary Utilities\)](#)) ([objcopy \(GNU Binary Utilities\)](#)). Треба бити опрезан: при конверзији у `.bin`, ако ELF садржи "празнину" (нпр. секције које нису континуалне у меморији), `objcopy` ће ту празнину испунити нулама у `.bin` фајлу (јер прави континуални дамп од најнижег до највишег адресног бајта). Стога `.bin` може бити већи него износ корисног кода ако постоје велике неиницијализоване секције на високим адресама.

6.2. readelf

readelf: Алат за читање ELF датотека – даје детаљан увид у ELF заглавља, секције, сегменте, симболе, итд. На пример, `readelf -h program.elf` приказује опште заглавље (тип, машина, ендијаност, улазну тачку...), `readelf -S program.elf` листа секције (са именима, величинама, стартним адресама у фајлу и у меморији), а `readelf -s program.elf` листа симбол таблицу (са именима функција/променљивих и њиховим адресама или офсетовима). У анализи линкер скрипте и меморијског распореда, `readelf -S` је посебно користан да видимо где су `.text`, `.data`, `.bss` и друге секције смештене и колике су. **Пример:** излаз `readelf -S` може показати да је `.text` секција величине 0x1000 бајтова на адреси 0x10000000 (у флешу), `.data` величине 0x100 бајтова на адреси 0x08040000 (RAM), са LMA (Load Memory Address) у флешу, итд. Ово нам јасно говори распоред по меморијским регијама.

6.3. nm

nm: Листирање симбола (из објектних или извршних фајлова). **nm program.elf** ће исцртати листу свих симбола (функција, глобалних променљивих) које постоје у програму, са њиховим адресама и ознаком типа (Т=текст/функција, D=иницијализовани податак, B=неиницијализовани податак *bss*, и сл.). Ово је корисно кад желимо да знамо на којој адреси се налази одређена функција или променљива након линковања. На пример, можемо проверити да ли је глобална променљива мапирана у RAM (биће означена са **B** или **D** и имаће адресу у опсегу RAM меморије), или да ли је функција у флешу (ознака **T** са адресом у опсегу флеша). **nm** је посебно драгоцен за грубу проверу исправности линкер скрипте – нпр. да ли симбол `_estack` (врх стека) има очекивану вредност, да ли су неки симболи преклопљени итд.

6.4. size

size: Приказује резиме величина секција у извршној датотеци. Обично се позива као `size program.elf` и избацује три колоне: величину `.text` (код + константни подаци у флешу), `.data` (иницијални подаци који ће бити учитани у RAM) и `.bss` (неиницијализовани подаци који ће заузети RAM), као и укупан збир. Ово је врло прегледно да се види колики је „отисак“ програма у флешу и RAM-у. На пример, output може бити:

text	data	bss	dec	hex	filename
4528	128	256	4912	1330	program.elf

што значи да код заузима ~4528 бајтова флеша, статички иницијализовани подаци 128 бајтова RAM-а (плус још толико у флешу за њихове почетне вредности), а `.bss` (нпр. глобалне променљиве иницијализоване на 0) заузимају 256 бајтова RAM-а. Ови бројеви су битни у планирању да ли програм стаје у меморију микроконтролера. (Напомена: `size` уз опцију `-A` или `--format=SysV` даје детаљнији приказ по именованим секцијама.)

6.5. objdump

objdump: Многофункционални алат за испис садржаја објектних датотека. Може да прикаже хексадецимални *dump* секција (`objdump -s program.elf`), али најкориснија функционалност је **дисасемблирање** машинског кода у читљив асемблер.

Наредба `objdump -d -M reg-names-std program.elf` произвешће асемблерски листинг свих секција које имају код (нпр. `.text`, `.init`) са људски читљивим именима регистара. Ово је изузетно корисно за дебаговање ниског нивоа – можемо видети тачно које је инструкције компајлер генерисао из нашег С кода, што помаже у оптимизацији или трагању за баговима на ниском нивоу.

- `objdump -t program.elf` исписује таблицу симбола (слично `nm`).
- `objdump -x program.elf` исписује пуна ELF заглавља, секције, сегменте, симболе (комбинација информација, мање читљива од специјализованих алата попут `readelf` или `nm`).

Углавном, **objdump** је згодан за брзи увид у садржај бинарног кода – било у хекс или у асемблерском облику.

6.6. Закључак

Набројани алати покривају најважније аспекте: конверзију формата (`objcopy`, `srec_cat`), анализу садржаја (`readelf`, `nm`, `objdump`) и величину (`size`). У типичном развојном циклусу, након добијања `program.elf`, програмер може покренути `size` да провери заузеће меморије, `objdump -d` ако сумња у неку оптимизацију компајлера, или `nm` да пронађе адресу битне променљиве за дебаговање. При припреми HEX-а за програмирање, користи се `objcopy` или `srec_cat`. На тај начин, ови алати представљају продужетак функционалности самог компајлера и линкера, дајући увид "испод хаубе" готовог програма.

7. Линкерска скрипта: MEMORY и SECTIONS дефиниције

Линкерска скрипта или линкер директива (`.ld` датотека) је суштински део embedded пројекта – она диктира како ће линкер распоредити преведене секције програма у меморијски адресни простор микроконтролера. Пошто микроконтролер има одређене величине и адресе флеш и RAM меморије, потребно је линкер обавестити где може смештати код, а где податке. GNU ld линкер има свој интерни језик за скрипте који омогућава опис меморијских региона и расподелу секција.

7.1. MEMORY дефиниција

У скрипти се најпре дефинишу именовани *меморијски региони*. Пример за неки Cortex-M7 микроконтролер може бити:

```
MEMORY
{
    FLASH (rx)  : ORIGIN = 0x10000000, LENGTH = 4096K
    WORK_FLASH (rx) : ORIGIN = 0x14000000, LENGTH = 256K
    SRAM (rwx)   : ORIGIN = 0x28000000, LENGTH = 768K
}
```

Овим смо линкеру дали до знања да постоје три региона: **FLASH** који почиње на адреси 0x1000_0000 величине 4 MB, додатни **WORK_FLASH** од 256 KB, и **SRAM** од 768 KB почев од 0x2800_0000. Сваки регион има атрибуте: **r** (читљив), **w** (уписив), **x** (извршан). Флеш је обично означен са **rx** (читање + извршавање, није уписив у току рада програма), RAM са **rwx** (читање/писање, извршавање ако се код копира у RAM). Ове MEMORY декларације омогућавају линкеру да разуме расположиви адресни простор и да пријави грешку ако одређена секција премашује додељени меморијски блок. На пример, ако код постане превелик (прелази 4MB флеша), линкер ће јавити да **FLASH** регион нема довољно простора. MEMORY секција дакле описује распон и намену меморијских блокова циљног система ([The most thoroughly commented linker script \(probably\) - Stargirl \(Thea\) Flowers](#)).

7.2. SECTIONS расподела

Након дефинисања меморије, линкерска скрипта садржи **SECTIONS** блок, који је срце скрипте – ту се наводи како се свакој излазној секцији ELF-а додељује меморијски регион и адреса. Пример делимичне SECTIONS скрипте за ARM:

```
SECTIONS {
    .text :
    {
        _stext = .;           /* почетак .text секције */
        KEEP(*(.isr_vector)) /* векторска табела, ако је дефинисана у секцији
.isr_vector */
        *(.text*)             /* све .text секције из свих објеката */
        *(.rodata*)           /* све .rodata (константе) секције */
        _etext = .;           /* крај .text/.rodata секције */
    } > FLASH                /* овај блок иде у FLASH меморију */

    .data : AT (ADDR(FLASH) + SIZEOF(.text)) /* AT(...) каже да се почетни
садржај .data налази након .text у FLASH */
    {
        _sdata = .;          /* почетак .data у RAM */
        *(.data*)            /* све иницијализоване статичке променљиве */
        _edata = .;
    } > SRAM                 /* овај блок се мапира у SRAM на извршавање */

    .bss (NOLOAD) :
    {
        _sbss = .;           /* почетак .bss */
        *(.bss*)
        *(COMMON)
        _ebss = .;
    } > SRAM                 /* .bss је у SRAM, NOLOAD јер нема почетних
података у флешу */

    /* ... остале секције (стек, heap, секције за C++ EH, итд.) ... */
}
```

Овај пример илуструје кључне концепте. Прво, **.text** (са **.rodata**) је смештен у **FLASH** регион. У њему се чак *принудно чува* (**KEEP**) векторска табела ако је дефинисана у посебној секцији **.isr_vector** (многи startup кодови стављају векторску таблицу у засебну секцију на почетак флеша). Затим **.text** обухвата сав програмски код и неизменљиве податке. Символи као **_stext** и **_etext** су обележивачи које можемо касније користити (нпр. **_etext** указује на крај садржаја који се налази у флешу, што је уједно почетак података за копирање у RAM).

Део за **.data** је интересантан јер демонстрира одвајање *LOAD адресе* и *RUN адресе*: секција **.data** ће током извршавања бити у SRAM (због **> SRAM**), али је наведено **AT (ADDR(FLASH) + SIZEOF(.text))**. Ово каже линкеру да резервише простор у FLASH-у за иницијалне вредности **.data** секције, одмах након **.text** блока.

Практично, то значи да ће у ELF-у `.data` имати два различита почетка: `VMA` (Virtual Memory Address) у RAM (нпр. `0x2800_0000` ако ту почиње SRAM) и `LMA` (Load Memory Address) у флешу (нпр. `0x1000_1234` ако `.text` заузима до `0x1000_1233`). Линкер ће поред ELF секције уписати и тај садржај (иницијалне бајтове) на `LMA` адресу. Потом, стартап код микроконтролера копира те бајтове из флеша на одговарајућу `VMA` адресу у SRAM током покретања програма. На овај начин, глобалне променљиве које имају иницијалне вредности (нпр. `int count = 5;`) налазе се у `.data` секцији: након ресета, у RAM-у на адреси `_sdata` треба да се нађе вредност 5, коју је програм негде морао сачувати – управо у флешу, у склопу `.data` `LMA`. Директива `AT()` у линкер скрипти се користи да раздвоји те адресе и веома је битна за исправно покретање програма ([The most thoroughly commented linker script \(probably\) - Stargirl \(Thea\) Flowers](#)).

Секција `.bss` у примеру има атрибут `NOLOAD` – то значи да за њу не постоји садржај у флешу; она представља блок RAM меморије коју треба испунити нулама. Линкерски симболи `_sbss` и `_ebss` означавају почетак и крај `.bss`. Важно је и да су сви `COMMON` симболи (неиницијализоване глобалне променљиве које нису експлицитно у `.bss`) такође стављени у `.bss` (директива `*(COMMON)`). Стартап код ће једноставно обрисати (испунити нулама) овај RAM опсег при покретању.

Поред ових, линкер скрипта често резервише простор за стек (нпр. дефинише симбол `_estack` на крај SRAM-а), за `heap` (ако се користи), и укључује специјалне секције за C++ ако су присутне (нпр. `.init_array` за конструкторе објеката). У случају ARM Cortex-M, потребне су и секције `.ARM.exidx` и `.ARM.extab` за информације о изузецима (корисно ако се користе одређене библиотеке и функције или C++ бацање изузетака – `throw exception`). Они се обично смештају на крај `.text` сегмента. Укључујући све те ставке, линкерска скрипта обезбеђује да у излазном ELF-у стоје исправно дефинисане адресе и границе секција.

За пример TRAVEO T2G микроконтролера, линкерска скрипта би била прилагођена тачним величинама меморија тог чипа. На основу **меморијске архитектуре**, могли бисмо дефинисати MEMORY са `FLASH (rx) : ORIGIN = 0x10000000, LENGTH = 4160K` (главна флеш меморија од ~4.16 MB), евентуално `WORKFLASH (rx) : ORIGIN = 0x14000000, LENGTH = 256K`, `SRAM (rwx) : ORIGIN = 0x28000000, LENGTH = 768K`, и можда посебно `ITCM` и `DTCM` меморије ако би се користиле (16KB + 16KB по језгру). За једноставност, претпоставимо да сав код иде у главни FLASH, а сав RAM садржај у главни SRAM. Линкерска скрипта би онда одразила тај модел: `.text`, `.rodata`, евентуално векторска табела и `startup` код у FLASH; `.data` `LMA` у FLASH али `VMA` у SRAM; `.bss` у SRAM.

Закључак о линкер скрипти: Она повезује свет C кода са физичком меморијом хардвера. MEMORY секција описује где може шта да иде ([The most thoroughly commented linker script \(probably\) - Stargirl \(Thea\) Flowers](#)), а `SECTIONS` секција како да се садржај распореди ([The most thoroughly commented linker script \(probably\) - Stargirl \(Thea\) Flowers](#)). За типичан Cortex-M пројекат, већина програмера користи унапред припремљену скрипту (било од произвођача или генерисану алатом), али разумевање исте је кључно при решавању проблема као што су: преливање меморије, смештање специфичних функција у одређени регион (нпр. у посебан сегмент који се може ажурирати независно), прављење боотлоадера/апликација поделе, итд.

8. Закључак