#### Albert Ludwigs Universität Freiburg

TECHNISCHE FAKULTÄT

### PicoC-Compiler

# Übersetzung einer Untermenge von C in den Befehlssatz der RETI-CPU

BACHELORARBEIT

 $Abgabedatum: 28^{th}$  April 2022

 $\begin{array}{c} Author: \\ \text{J\"{u}rgen Mattheis} \end{array}$ 

Gutachter: Prof. Dr. Scholl

Betreung: M.Sc. Seufert

Eine Bachelorarbeit am Lehrstuhl für Betriebssysteme

ERKLÄRUNG
ERRLARONS
Hiermit erkläre ich, dass ich diese Abschlussarbeit selbständig verfasst habe, keine anderen
als die angegebenen Quellen/Hilfsmittel verwendet habe und alle Stellen, die wörtlich oder
sinngemäß aus veröffentlichten Schriften entnommen wurden, als solche kenntlich gemacht
habe. Darüber hinaus erkläre ich, dass diese Abschlussarbeit nicht, auch nicht
auszugsweise, bereits für eine andere Prüfung angefertigt wurde.

# Inhaltsverzeichnis

L	Implementierung			
	1.1	Archit	ktur	
	1.2	Lexika	ische Analyse	10
		1.2.1	Verwendung von Lark	10
		1.2.2	Basic Parser	1
	1.3	Synta	tische Analyse	1
		1.3.1	Verwendung von Lark	1
		1.3.2	Umsetzung von Präzidenz	12
		1.3.3	Derivation Tree Generierung	13
			1.3.3.1 Early Parser	13
		1.3.4	Derivation Tree Vereinfachung	13
		1.3.5	Abstrakt Syntax Tree Generierung	13
			1.3.5.1 PicoC Nodes	13
			1.3.5.2 RETI Nodes	13
			1.3.5.3 Kompositionen mit besonderer Bedeutung	13
			1.3.5.4 Abstrakte Syntax	

# Abbildungsverzeichnis

1.1	Cross-Compiler Kompiliervorgang ausgeschrieben
1.2	Cross-Compiler Kompiliervorgang Kurzform
1.3	Architektur mit allen Passes ausgeschrieben

Codeverzeichnis			

# Tabellenverzeichnis

Definitionsverzeichnis		

## Grammatikverzeichnis

1.2.1 Konkrette Syntax des Lexers in EBNF
1.3.1 Konkrette Syntax des Parsers in EBNF, Teil 1
1.3.2 Konkrette Syntax des Parsers in EBNF, Teil 2
1.3.3 Abstrakte Syntax für $L_{PiocC}$
1.3.4 Abstrakte Syntax für $L_{PicoC\_Blocks}$
1.3.5 Abstrakte Syntax für $L_{PicoC\_Mon}$
1.3.6 Abstrakte Syntax für $L_{RETI\_Blocks}$
1.3.7 Abstrakte Syntax für $L_{RETI\_Patch}$
1.3.8 Abstrakte Syntax für $L_{RETI}$
$1.2.2 \lambda$ calculus syntax
1.2.3 Advanced capabilities of grammar.sty

# 1 Implementierung

#### 1.1 Architektur

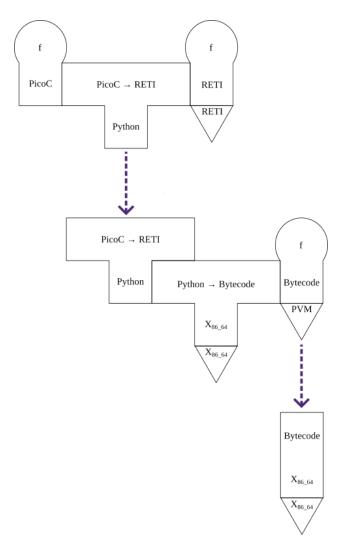
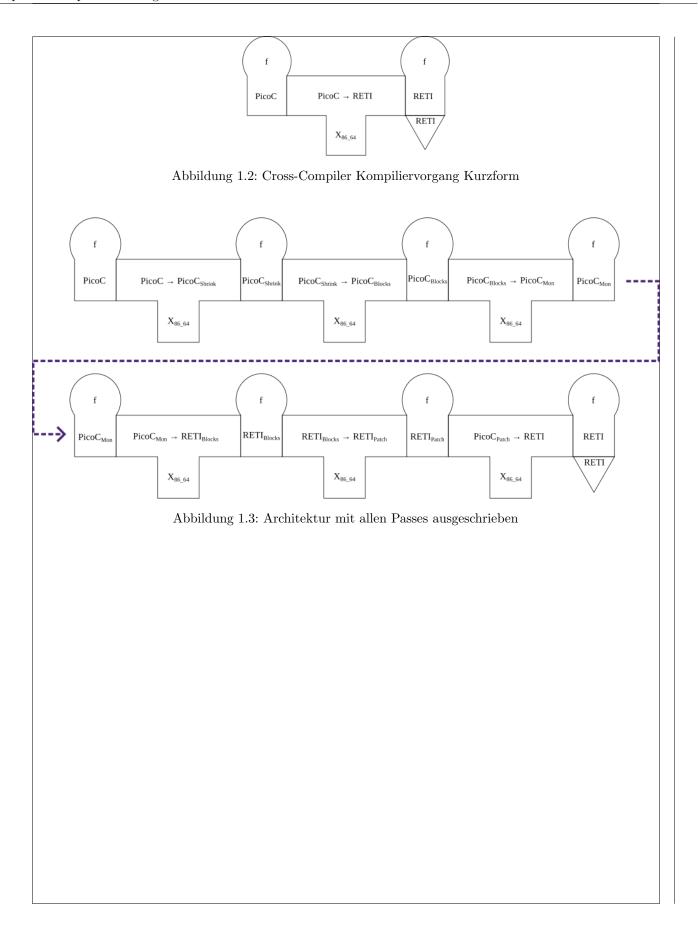


Abbildung 1.1: Cross-Compiler Kompiliervorgang ausgeschrieben



#### 1.2 Lexikalische Analyse

#### 1.2.1 Verwendung von Lark

```
"//" /[\wedge n]*/ | "/*" /(. | n)*?/ "*/"
COMMENT
                                                                    L\_Comment
                     ::=
                         "//""_{-}"?"#"/[\wedge \setminus n]*/
RETI\_COMMENT.2
                     ::=
DIG\_NO\_0
                         "1"
                                "2"
                                       "3"
                                              "4"
                                                    "5"
                                                                    L_Arith
                     ::=
                         "6"
                                "7"
                                       "8"
                                              "9"
DIG\_WITH\_0
                         "0"
                                DIG\_NO\_0
                     ::=
NUM
                         "0" | DIG_NO_0DIG_WITH_0*
                     ::=
                         "ບ"…"่∼ "
ASCII\_CHAR
                     ::=
                     ::= "'" ASCII\_CHAR"'"
CHAR
FILENAME
                     ::= ASCII\_CHAR + ".picoc"
LETTER
                     ::= "a".."z" | "A".."Z"
                     ::= (LETTER \mid """)
NAME
                             (LETTER — DIG_WITH_0 — "_")*
                     ::= NAME \mid INT\_NAME \mid CHAR\_NAME
name
                         VOID\_NAME
                         " \sim "
NOT
                     ::=
                     ::= "&"
REF\_AND
                     ::= SUB\_MINUS \mid LOGIC\_NOT \mid NOT
un\_op
                         MUL\_DEREF\_PNTR \mid REF\_AND
MUL\_DEREF\_PNTR
                         "*"
                     ::=
                         "/"
DIV
                     ::=
                         "%"
MOD
                     ::=
                     ::= MUL\_DEREF\_PNTR \mid DIV \mid MOD
prec1\_op
ADD
                     ::=
                         "+"
                     ::= "-"
SUB\_MINUS
                     ::= ADD \mid SUB\_MINUS
prec2\_op
                     ::= "<"
LT
                                                                    L_{-}Logic
                         "<="
LTE
                     ::=
                     ::= ">"
GT
GTE
                     ::= ">="
                     ::= LT | LTE | GT | GTE
rel\_op
                         "=="
EQ
                     ::=
                     ::= "!="
NEQ
                     ::= EQ \mid NEQ
eq\_op
                         "!"
LOGIC_NOT
                     ::=
INT\_DT.2
                     ::=
                         "int"
INT\_NAME.3
                         "int" (LETTER | DIG_WITH_0 | "_")+ L_Assign_Alloc
                     ::=
                         "char"
CHAR\_DT.2
                     ::=
                         "char" (LETTER \mid DIG\_WITH\_0 \mid "_")+
CHAR\_NAME.3
                     ::=
VOID\_DT.2
                     ::= "void"
VOID\_NAME.3
                     ::= "void" (LETTER | DIG_WITH_0 | "_")+
prim_{-}dt
                     ::= INT\_DT \mid CHAR\_DT \mid VOID\_DT
```

Grammar 1.2.1: Konkrette Syntax des Lexers in EBNF

#### 1.2.2 Basic Parser

#### 1.3 Syntaktische Analyse

#### 1.3.1 Verwendung von Lark

#### In 1.3.1

prim_exp post_exp un_exp	::= ::=   ::=	name   NUM   CHAR   "("logic_or")" array_subscr   struct_attr   fun_call input_exp   print_exp   prim_exp un_opun_exp   post_exp	$L\_Arith + L\_Array + L\_Pntr + L\_Struct + L\_Fun$
input_exp print_exp arith_prec1 arith_prec2 arith_and arith_oplus arith_or	::= ::= ::= ::= ::=	"input""("")"  "print""("logic_or")"  arith_prec1 prec1_op un_exp   un_exp  arith_prec2 prec2_op arith_prec1   arith_prec1  arith_and "&" arith_prec2   arith_prec2  arith_oplus "^" arith_and   arith_and  arith_or " " arith_oplus   arith_oplus	$L\_Arith$
rel_exp eq_exp logic_and logic_or	::= ::= ::=	rel_exp rel_op arith_or   arith_or eq_exp eq_oprel_exp   rel_exp logic_and "&&" eq_exp   eq_exp logic_or "  " logic_and   logic_and	$L\_Logic$
type_spec alloc assign_stmt initializer init_stmt const_init_stmt	::= ::= ::= ::=	<pre>prim_dt   struct_spec type_spec pntr_decl un_exp "=" logic_or";" logic_or   array_init   struct_init alloc "=" initializer";" "const" type_spec name "=" NUM";"</pre>	$L\_Assign\_Alloc$
$pntr\_deg$ $pntr\_decl$	::=	"*"*  pntr_deg array_decl   array_decl	$L\_Pntr$
array_dims array_decl array_init array_subscr	::= ::= ::=	("["NUM"]")* name array_dims   "("pntr_decl")"array_dims "{"initializer("," initializer) *"}" post_exp"["logic_or"]"	$L\_Array$
struct_spec struct_params struct_decl struct_init	::= ::= ::=	"struct" name (alloc";")+  "struct" name "{"struct_params"}"  "{""."name"="initializer  ("," "."name"="initializer)*"}"	$L\_Struct$
$struct\_attr$	::=	$post\_exp"."name$	
$\begin{array}{c} if\_stmt \\ if\_else\_stmt \end{array}$	::= ::=	"if""("logic_or")" exec_part "if""("logic_or")" exec_part "else" exec_part	$L_{-}If_{-}Else$
while_stmt do_while_stmt	::=	"while""("logic_or")" exec_part "do" exec_part "while""("logic_or")"";"	$L_{-}Loop$

Grammar 1.3.1: Konkrette Syntax des Parsers in EBNF, Teil 1

```
alloc";"
decl\_exp\_stmt
                                                                                                L_Stmt
                   ::=
decl\_direct\_stmt
                         assign_stmt | init_stmt | const_init_stmt
                   ::=
decl\_part
                         decl\_exp\_stmt \mid decl\_direct\_stmt \mid RETI\_COMMENT
                   ::=
                         "{"exec\_part*"}"
compound\_stmt
                   ::=
                         logic_or";"
exec\_exp\_stmt
                   ::=
exec\_direct\_stmt
                   ::=
                        if\_stmt \mid if\_else\_stmt \mid while\_stmt \mid do\_while\_stmt
                         assign\_stmt \quad | \quad fun\_return\_stmt
exec\_part
                         compound\_stmt \mid exec\_exp\_stmt \mid exec\_direct\_stmt
                   ::=
                         RETI\_COMMENT
                     decl\_exec\_stmts
                         decl\_part * exec\_part *
                   ::=
                         [logic\_or("," logic\_or)*]
                                                                                                L_Fun
fun\_args
                   ::=
fun\_call
                         name" ("fun_args")"
                   ::=
                         "return" [logic_or]";"
fun\_return\_stmt
                   ::=
                         [alloc("," alloc)*]
fun\_params
                   ::=
fun\_decl
                         type_spec pntr_deg name"("fun_params")"
                   ::=
                         type_spec_pntr_deg_name"("fun_params")" "{"decl_exec_stmts"}"
fun_{-}def
                         (struct\_decl \mid
                                          fun_decl)";" | fun_def
decl\_def
                                                                                                L_File
                   ::=
                         decl\_def*
decls\_defs
                         FILENAME\ decls\_defs
file
                   ::=
```

Grammar 1.3.2: Konkrette Syntax des Parsers in EBNF, Teil 2

#### 1.3.2 Umsetzung von Präzidenz

Die PicoC Programmiersprache hat dieselben Präzidenzregeln implementiert, wie die Programmiersprache  $\mathbb{C}^1$ . Die Präzidenzregeln von PicoC sind in Tabelle 1.1 aufgelistet.

Präzidenz	Operator	Beschreibung	Assoziativität	
1	a()	Funktionsaufruf	Links, dann rechts $\rightarrow$	
	a[]	Indexzugriff		
	a.b	Attributzugriff		
2	-a	Unäres Minus		
	!a ~a	Logisches NOT und Bitweise NOT	Rechts, dann links $\leftarrow$	
	*a &a	Dereferenz und Referenz, auch	reents, dann miks —	
		Adresse-von		
3	a*b a/b a%b	Multiplikation, Division und Modulo		
4	a+b a-b	Addition und Subtraktion		
5	a <b a="" a<="b">b a&gt;=b</b>	Kleiner, Kleiner Gleich, Größer,		
		Größer gleich		
6	a==b a!=b	Gleichheit und Ungleichheit	Links, dann rechts $\rightarrow$	
7	a&b	Bitweise UND	Links, daim recits →	
8	a^b	Bitweise XOR (exclusive or)		
9	a b	Bitweise ODER (inclusive or)		
10	a&&b	Logiches UND		
11	a  b	Logisches ODER		
12	a=b	Zuweisung	Rechts, dann links $\leftarrow$	
13	a,b	Komma	Links, dann rechts $\rightarrow$	

Tabelle 1.1: Präzidenzregeln von PicoC

<sup>1</sup>C Operator Precedence - cppreference.com.

Derivation Tree Generierung 1.3.31.3.3.1 Early Parser 1.3.4Derivation Tree Vereinfachung 1.3.5Abstrakt Syntax Tree Generierung 1.3.5.1 PicoC Nodes 1.3.5.2RETI Nodes 1.3.5.3Kompositionen mit besonderer Bedeutung 1.3.5.4 Abstrakte Syntax

```
Minus()
                                              Not()
                                                                                                                 L_Arith
un\_op
                   ::=
bin\_op
                   ::=
                           Add()
                                          Sub()
                                                        Mul() \mid Div() \mid
                                                                                        Mod()
                                            And() \mid Or()
                           Oplus()
                                                                  Char(str)
                           Name(str) \mid Num(str)
exp
                           BinOp(\langle exp \rangle, \langle bin\_op \rangle, \langle exp \rangle)
                           UnOp(\langle un\_op \rangle, \langle exp \rangle) \mid Call(Name('input'), None)
                           Alloc(\langle type\_qual \rangle, \langle dataype \rangle, Name(str))
exp\_stmts
                  ::=
                           Call(Name('print'), \langle exp \rangle)
un\_op
                          LogicNot()
                                                                                                                 L\_Logic
                   ::=
                                   |NEq()|Lt()|LtE()|Gt()|GtE()
rel
                   ::=
                           Eq()
                           LogicAnd() | LogicOr()
bin\_op
                   ::=
                           Atom(\langle exp \rangle, \langle rel \rangle, \langle exp \rangle)
exp
                           ToBool(\langle exp \rangle)
                          Const() \mid Writeable()
                                                                                                                 L\_Assign\_Alloc
type\_qual
                   ::=
datatype
                           IntType() \mid CharType() \mid VoidType()
                   ::=
                           Alloc(\langle type\_qual \rangle, \langle dataype \rangle, Name(str))
assign\_lhs
                   ::=
exp\_stmts
                           Alloc(\langle type\_qual \rangle, \langle dataype \rangle, Name(str))
                   ::=
stmt
                           Assign(\langle assign\_lhs \rangle, \langle exp \rangle)
                   ::=
                           Exp(\langle exp\_stmts \rangle)
datatype
                   ::=
                           PntrDecl(Num(str), \langle datatype \rangle)
                                                                                                                 L_{-}Pntr
deref\_loc
                           Ref(\langle ref\_loc \rangle) \mid \langle ref\_loc \rangle
                   ::=
ref\_loc
                           Name(str)
                   ::=
                           Deref(\langle deref\_loc \rangle, \langle exp \rangle)
                           Subscr(\langle deref\_loc \rangle, \langle exp \rangle)
                           Attr(\langle ref\_loc \rangle, Name(str))
                          Deref(\langle deref\_loc \rangle, \langle exp \rangle)
exp
                   ::=
                           Ref(\langle ref\_loc \rangle)
                           ArrayDecl(Num(str)+, \langle datatype \rangle)
datatype
                                                                                                                 L\_Array
                   ::=
                           Subscr(\langle deref\_loc \rangle, \langle exp \rangle)
                                                                       Array(\langle exp \rangle +)
exp
                   ::=
                           StructSpec(Name(str))
                                                                                                                 L\_Struct
datatype
                   ::=
                           Attr(\langle ref\_loc \rangle, Name(str))
exp
                   ::=
                           Struct(Assign(Name(str), \langle exp \rangle) +)
decl\_def
                           StructDecl(Name(str),
                   ::=
                                 Alloc(Writeable(), \langle datatype \rangle, Name(str)) +)
                           If(\langle exp \rangle, \langle stmt \rangle *)
                                                                                                                 L_If_Else
stmt
                   ::=
                           IfElse(\langle exp \rangle, \langle stmt \rangle *, \langle stmt \rangle *)
                           While(\langle exp \rangle, \langle stmt \rangle *)
                                                                                                                 L_{-}Loop
stmt
                   ::=
                           DoWhile(\langle exp \rangle, \langle stmt \rangle *)
                           Call(Name(str), \langle exp \rangle *)
                                                                                                                 L_Fun
                   ::=
exp
                           Call(Name(str), \langle exp \rangle *)
exp\_stmts
                   ::=
                           Return(\langle exp \rangle)
stmt
                   ::=
decl\_def
                           FunDecl(\langle datatype \rangle, Name(str),
                   ::=
                                 Alloc(Writeable(), \langle datatype \rangle, Name(str))*)
                           FunDef(\langle datatype \rangle, Name(str),
                                 Alloc(Writeable(), \langle datatype \rangle, Name(str))*, \langle stmt \rangle*)
                                                                                                                 L_{-}File
file
                  ::=
                           File(Name(str), \langle decl\_def \rangle *)
```

Grammar 1.3.3: Abstrakte Syntax für  $L_{PiocC}$ 

```
FunDef(\langle datatype \rangle, Name(str),
                                                                                                        L_Fun
         decl\_def
                      ::=
                                  Alloc(Writeable(), \langle datatype \rangle, Name(str))*, \langle block \rangle*)
                             Block(Name(str), \langle stmt \rangle *)
         block
                      ::=
                                                                                                        L\_Blocks
                             Goto(Name(str)) \mid NewStackframe(Name(), Goto(str))
         stmt
                      ::=
                             RemoveStackframe() \mid SetScope(Name(str))
                             SingleLineComment(str, str)
                               Grammar 1.3.4: Abstrakte Syntax für L_{PicoC\_Blocks}
                         Tmp(Num(str))
                                                  StackRead(Num(str))
ref\_loc
                                                                                                         L\_Assign\_Alloc
                  ::=
                         StackWrite(Num(str))
                                                       | GlobalRead(Num(str))|
                         GlobalWrite(Num(str))
                         \langle exp \rangle \mid Pos(Num(str), Num(str))
error\_data
                  ::=
                         Stack(Num(str)) \mid Ref(\langle ref_loc \rangle, \langle datatype \rangle, \langle error_data \rangle)
exp
                  ::=
stmt
                  ::=
                         Exp(\langle exp \rangle)
                         Assign(Alloc(Writeable(), StructSpec(Name(str)), Name(str)),
                              Struct(Assign(Name(str), \langle exp \rangle) +, \langle datatype \rangle))
                         Assign(Alloc(Writeable(), ArrayDecl(Num(str)+, \langle datatype \rangle),
                              Name(str), Array(\langle exp \rangle +, \langle datatype \rangle))
symbol\_table
                         SymbolTable(\langle symbol \rangle)
                                                                                                          L_Symbol_Table
                  ::=
symbol
                         Symbol(\langle type_qual \rangle, \langle datatype \rangle, \langle name \rangle, \langle val \rangle, \langle pos \rangle, \langle size \rangle)
                  ::=
                         Empty()
type\_qual
                  ::=
                                          SelfDefined()
datatype
                         BuiltIn()
                  ::=
name
                         Name(str)
                  ::=
                         Num(str)
                                        |Empty()
val
                  ::=
                         Pos(Num(str), Num(str)) \mid Empty()
pos
                  ::=
                         Num(str)
size
                                           Empty()
                  ::=
                                Grammar 1.3.5: Abstrakte Syntax für L_{PicoC\ Mon}
```

```
Program(Name(str), \langle block \rangle *)
                                                                                                      L\_Program
program
                    ::=
                           Goto(str)
exp\_stmts
                    ::=
                                                                                                      L_Blocks
                           Num(str)
instrs\_before
                    ::=
num\_instrs
                           Num(str)
                    ::=
                           Block(Name(str), \langle instr\rangle *, \langle instrs\_before \rangle, \langle num\_instrs \rangle)
block
                    ::=
instr
                           Goto(Name(str))
                    ::=
```

Grammar 1.3.6: Abstrakte Syntax für  $L_{RETI\_Blocks}$ 

```
Exit(Num(str))
stmt
```

Grammar 1.3.7: Abstrakte Syntax für  $L_{RETI\_Patch}$ 

```
ACC() \mid IN1() \mid IN2() \mid PC() \mid SP() \mid BAF()
                                                                                                       L\_Program
reg
                   CS() \mid DS()
               Reg(\langle reg \rangle) \mid Num(str)
             ::=
arg
                   Eq() \mid NEq() \mid Lt() \mid LtE() \mid Gt() \mid GtE()
rel
                   Always() \mid NOp()
                   Add() \mid Addi() \mid Sub() \mid Subi() \mid Mult()
op
             ::=
                   Multi() \mid Div() \mid Divi()
                   Mod() | Modi() | Oplus() | Oplusi() | Or()
                   Ori() \mid And() \mid Andi()
                   Load() \mid Loadin() \mid Loadi()
                   Store() \mid Storein() \mid Move() \\ Instr(\langle op \rangle, \langle arg \rangle +) \mid Jump(\langle rel \rangle, Num(str)) \mid Int(Num(str))
instr
             ::=
                   RTI() \mid Call(Name('print'), \langle reg \rangle) \mid Call(Name('input'), \langle reg \rangle)
                   SingleLineComment(str, str)
                   Program(Name(str), \langle instr \rangle *)
program
            ::=
```

Grammar 1.3.8: Abstrakte Syntax für  ${\cal L}_{RETI}$ 

```
T ::= V Variable \ | (TT) Application \ | \lambda V \cdot T Abstraction \ V ::= x, y, ... Variables
```

Grammar 1.2.2:  $\lambda$  calculus syntax

```
egin{array}{c|ccccc} A & ::= & \mathcal{T} & | & \mathcal{V} & Multiple \ option \ on \ a \ single \ line \ & \mathcal{A} & Highlighted \ form \ & | & \mathcal{B} & | & \mathcal{C} & Downplayed \ form \ & | & \mathcal{A} & | & \mathcal{B} & Emphasize \ part \ of \ the \ line \ \end{array}
```

Grammar 1.2.3: Advanced capabilities of grammar.sty

Literatur
Online
• C Operator Precedence - cppreference.com. URL: https://en.cppreference.com/w/c/language/operator_precedence (besucht am 27.04.2022).