#### Instruction Sets and Code Generation

**Eelco Visser** 



CS4200 | Compiler Construction | November 25, 2021

#### This Lecture

#### **Operational Semantics**

of ChocoPy

#### **Machine Architecture**

- components of a (virtual) machine

#### **RISC-V Instruction Set**

- instructions, registers, conventions

## Code Generation by Term Transformation

from source AST to target AST

#### **Compilation Schemas**

- how do source language constructs map to machine code

# Operational Semantics

#### Operational Semantics

#### 6 Operational semantics

This section contains the formal operational semantics for the ChocoPy language.

The operational semantics define how every definition, statement, or expression in a ChocoPy program should be evaluated in a given context.

#### Literals

$$\overline{G,E,S \vdash \mathtt{None} : None,S,\_} \quad [\mathtt{NONE}]$$
 
$$\overline{G,E,S \vdash \mathtt{False} : bool(false),S,\_} \quad [\mathtt{BOOL\text{-}FALSE}]$$
 
$$\overline{G,E,S \vdash \mathtt{True} : bool(true),S,\_} \quad [\mathtt{BOOL\text{-}TRUE}]$$
 
$$\frac{i \text{ is an integer literal}}{G,E,S \vdash i : int(i),S,\_} \quad [\mathtt{INT}]$$
 
$$s \text{ is a string literal}$$
 
$$\frac{n \text{ is the length of the string } s}{G,E,S \vdash s : str(n,s),S,\_} \quad [\mathtt{STR}]$$

#### **Expression Statement**

$$\frac{G, E, S \vdash e : v, S', \_}{G, E, S \vdash e : \_, S', \_}$$
 [EXPR-STMT]

#### Arithmetic Expressions

$$G, E, S \vdash e : int(i_1), S_1, \_$$

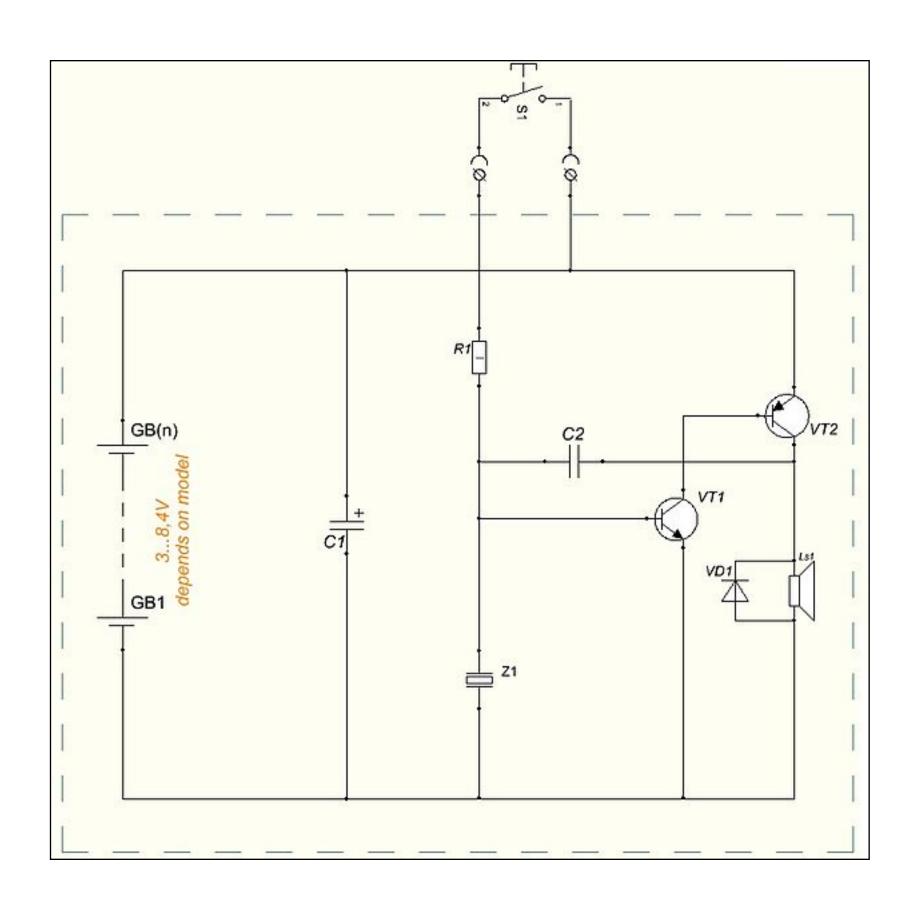
$$v = int(-i_1)$$

$$G, E, S \vdash \neg e : v, S_1, \_$$
[NEGATE]

$$G, E, S \vdash e_1 : int(i_1), S_1, _ G, E, S_1 \vdash e_2 : int(i_2), S_2, _ op \in \{+, -, *, //, \%\}$$
 $op \in \{//, \%\} \Rightarrow i_2 \neq 0$ 
 $v = int(i_1 \ op \ i_2)$ 
 $G, E, S \vdash e_1 \ op \ e_2 : v, S_2, _-$ 
[ARITH]

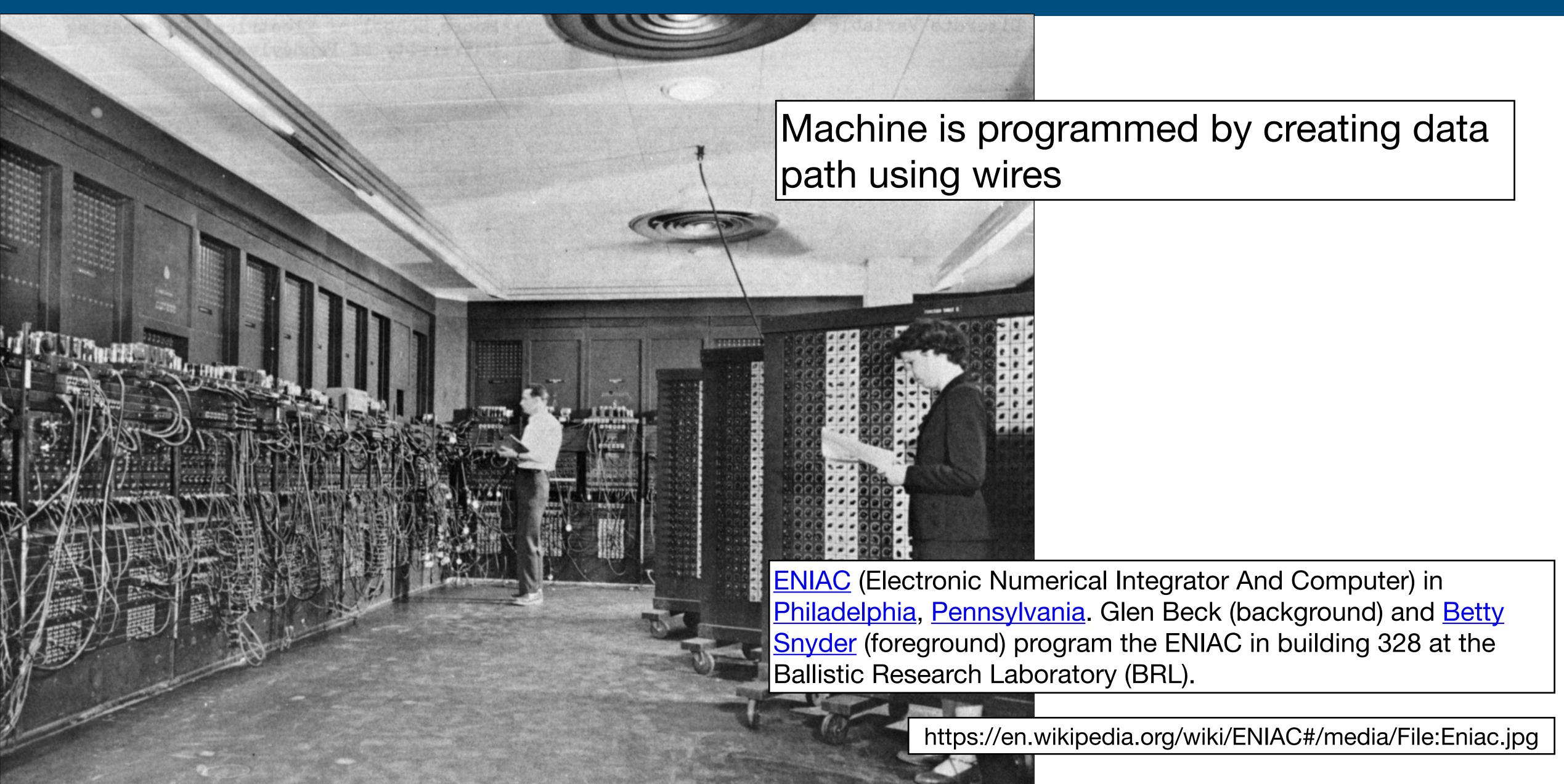
## Machine Architecture

## Hard-Wired Programs



Fixed to perform one computation from input to output

#### Programmable Machines



## Stored-Program Computer (Von Neumann Architecture)

## Central Processing Unit

- Processor registers
- Arithmetic logic unit

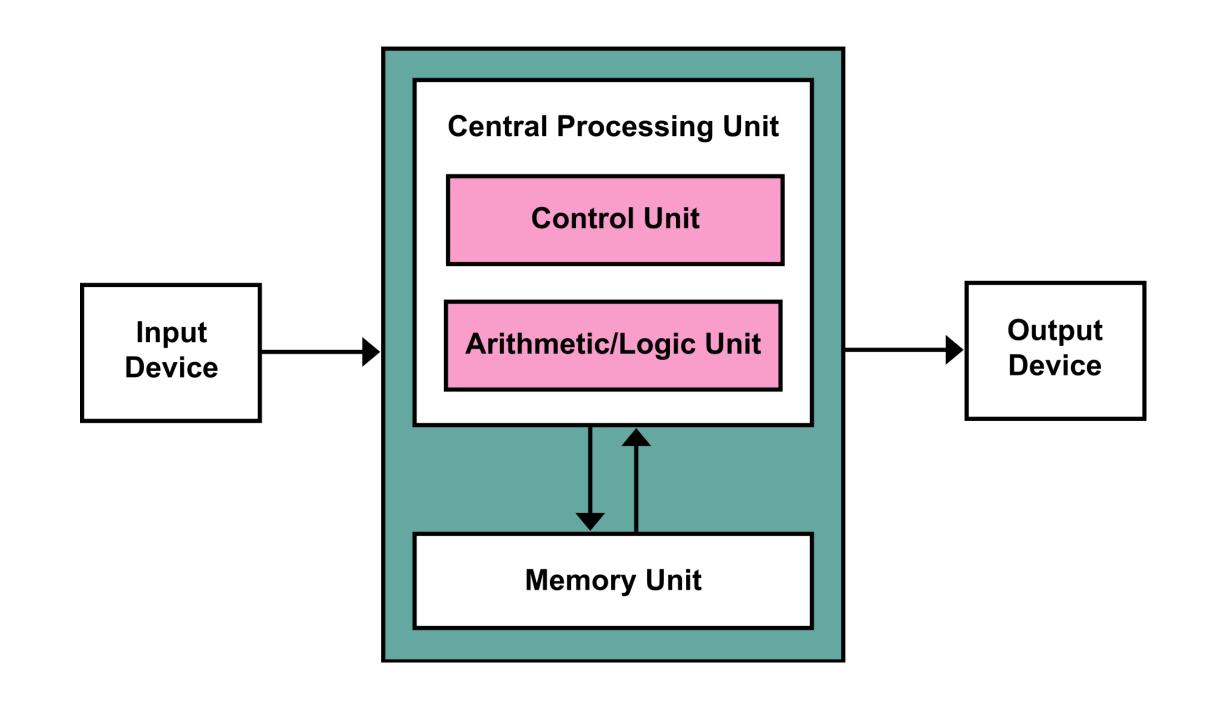
## Main Memory

- Stores data and instructions

## **External Storage**

Persistent storage of data

## Input/Output



#### State

#### Machine state

- data stored in memory
- memory hierarchy: registers, RAM, disk, network, ...

## Imperative program

- computation is series of changes to memory
- basic operations on memory (increment register)
- controlling such operations (jump, return address, ...)
- control represented by state (program counter, stack, ...)

```
mov AX [1]

mov CX AX

L: dec CX
mul CX
cmp CX 1
ja L

mov [2] AX
```

```
mov AX [1]

read memory

mov CX AX

L: dec CX
mul CX
cmp CX 1
ja L

mov [2] AX
```

```
mov AX [1] read memory

mov CX AX

L: dec CX
mul CX
cmp CX 1
ja L

mov [2] AX write memory
```

```
mov AX [1] read memory

mov CX AX

L: dec CX

mul CX

cmp CX 1

ja L

mov [2] AX write memory
```

```
mov AX [1] read memory

mov CX AX

L: dec CX
mul CX
cmp CX 1
ja L jump

mov [2] AX write memory
```

```
.method static public m(I)I
        iload 1
         ifne else
         iconst_1
         ireturn
  else: iload 1
         dup
        iconst_1
         isub
        invokestatic Math/m(I)I
         imul
         ireturn
```

```
.method static public m(I)I
         iload 1
         ifne else
         iconst_1
         ireturn
        iload 1
  else:
                          read memory
         dup
        iconst_1
         isub
         invokestatic Math/m(I)I
         imul
         ireturn
```

```
.method static public m(I)I
         iload 1
         ifne else
         iconst_1
         ireturn
         iload 1
  else:
                          read memory
         dup
         iconst_1
         isub
                          calculation
         invokestatic Math/m(I)I
         imul
         ireturn
```

```
.method static public m(I)I
         iload 1
         ifne else
                          jump
         iconst_1
         ireturn
         iload 1
  else:
                          read memory
         dup
         iconst_1
         isub
                          calculation
         invokestatic Math/m(I)I
         imul
         ireturn
```

#### Memory & Control Abstractions

## Memory abstractions

- variables: abstract over data storage
- expressions: combine data into new data
- assignment: abstract over storage operations

#### Control-flow abstractions

- structured control-flow: abstract over unstructured jumps
- 'go to statement considered harmful' Edgser Dijkstra, 1968

```
int f = 1
int x = 5
int s = f + x

while (x > 1) {
   f = x * f;
   x = x - 1
}
```

```
int f = 1
int x = 5
int s = f + x

while (x > 1) {
   f = x * f;
   x = x - 1
}
```

```
int f = 1
int x = 5
int s = f + x

while (x > 1) {
   f = x * f;
   x = x - 1
}
```

```
int f = 1
int x = 5
int s = f + x

while (x > 1) {
  f = x * f;
  x = x - 1
}
assignment
```

```
int f = 1
int x = 5
int s = f + x

while (x > 1) {
  f = x * f;
  x = x - 1
}
control flow
assignment
}
```

#### Control-flow abstraction

- Procedure: named unit of computation
- Procedure call: jump to unit of computation and return

#### Control-flow abstraction

- Procedure: named unit of computation
- Procedure call: jump to unit of computation and return

## Memory abstraction

- Formal parameter: the name of the parameter
- Actual parameter: value that is passed to procedure
- Local variable: temporary memory

#### Control-flow abstraction

- Procedure: named unit of computation
- Procedure call: jump to unit of computation and return

## Memory abstraction

- Formal parameter: the name of the parameter
- Actual parameter: value that is passed to procedure
- Local variable: temporary memory

#### Recursion

- Procedure may (indirectly) call itself
- Consequence?

## RISC-V Instruction Set

#### Concrete Syntax

```
.globl main
main:
                                           # Initialize heap size (in multiples of 4KB)
  lui a0, 8192
  add s11, s11, a0
                                           # Save heap size
  jal heap.init
                                           # Call heap.init routine
                                           # Initialize heap pointer
  mv gp, a0
                                           # Set beginning of heap
 mv s10, gp
                                           # Set end of heap (= start of heap + heap size)
  add s11, s10, s11
                                           # No normal return from main program.
  mv ra, zero
  mv fp, zero
                                           # No preceding frame.
                                           # Top saved FP is 0.
  mv fp, zero
                                           # No function return from top level.
  mv ra, zero
                                           # Reserve space for stack frame.
  addi sp, sp, -@..main.size
  sw ra, @..main.size-4(sp)
                                           # return address
  sw fp, @..main.size-8(sp)
                                           # control link
  addi fp, sp, @..main.size
                                           # New fp is at old SP.
  jal initchars
                                           # Initialize one-character strings.
  li a0, 1
                                           # Load boolean literal: true
  beqz a0, label_1
                                           # Operator and: short-circuit left operand
  li a0, 0
                                           # Load boolean literal: false
  seqz a0, a0
                                           # Logical not
label_1:
                                           # Done evaluating operator: and
  .equiv @..main.size, 16
label_0:
                                           # End of program
                                           # Code for ecall: exit
 li a0, 10
  ecall
```

#### Syntax Definition (\*)

```
// RV32I - Base
// Math
                                               {case-insensitive}
Instruction.Add = <add <ID>, <ID>, <ID>>
Instruction.Addi = <addi <ID>, <ID>, <IntOrID>> {case-insensitive}
Instruction.And = <and <ID>, <ID>, <ID>>
                                               {case-insensitive}
Instruction.Andi = <andi <ID>, <ID>, <IntOrID>> {case-insensitive}
Instruction.Or = <or <ID>, <ID>, <ID>>
                                               {case-insensitive}
Instruction.Ori = <ori <ID>, <ID>, <IntOrID>> {case-insensitive}
Instruction.Xor = <xor <ID>, <ID>, <ID>>
                                               {case-insensitive}
Instruction.Xori = <xori <ID>, <ID>, <IntOrID>> {case-insensitive}
Instruction.Sub = <sub <ID>, <ID>, <ID>>
                                               {case-insensitive}
// Branches
Instruction.Beg = <beg <ID>, <ID>, <IntOrID>> {case-insensitive}
Instruction.Bne = <bne <ID>, <ID>, <IntOrID>> {case-insensitive}
Instruction.Blt = <blt <ID>, <ID>, <IntOrID>> {case-insensitive}
Instruction.Bge = <bge <ID>, <ID>, <IntOrID>> {case-insensitive}
Instruction.Bltu = <bltu <ID>, <ID>, <IntOrID>> {case-insensitive}
Instruction.Bgeu = <bgeu <ID>, <ID>, <IntOrID>> {case-insensitive}
// Misc.
Instruction.Ecall = <ecall>
Instruction.Lui = <lui <ID>, <IntOrID>> {case-insensitive}
Instruction.Auipc = <auipc <ID>, <IntOrID>> {case-insensitive}
// Jumps
Instruction.Jal = <jal <ID>, <IntOrID>>
                                               {case-insensitive}
Instruction.Jalr = <jalr <ID>, <ID>, <IntOrID>> {case-insensitive}
```

#### Abstract Syntax Signature (\*)

```
module signatures/RV32IM-sig
imports signatures/Common-sig
signature
  sorts Start Line Label Statement Pseudodirective Instruction IntOrID
  constructors
                           : List(Line) \rightarrow Start
    Program
                           : Statement → Line
                           : Label → Line
    Label
                           : ID \rightarrow Label
                           : INT \rightarrow IntOrID
                           : ID \rightarrow IntOrID
                           : Pseudodirective → Statement
                           : Instruction → Statement
    PSData
                           : Pseudodirective
    PSText
                           : Pseudodirective
                           : STRING → Pseudodirective
    PSString
    PSAsciiz
                          : STRING → Pseudodirective
    PSWord
                           : List(IntOrID) → Pseudodirective
                           : INT → Pseudodirective
    PSSpace
                           : ID * ID * ID → Instruction
    Add
    Addi
                           : ID * ID * IntOrID → Instruction
                           : ID * ID * ID → Instruction
    And
                           : ID * ID * IntOrID → Instruction
    Andi
    0r
                           : ID * ID * ID → Instruction
                           : ID * ID * IntOrID → Instruction
    Ori
```

#### RISC-V Assembly Programmer's Manual

#### **Load Immediate**

The following example shows the li pseudo instruction which is used to load immediate values:

```
.equ CONSTANT, 0xdeadbeef
li a0, CONSTANT
```

Which, for RV32I, generates the following assembler output, as seen by objdump:

### RISC-V Assembly Programmer's Manual

#### **Load Address**

The following example shows the la pseudo instruction which is used to load symbol addresses:

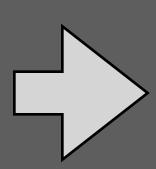
```
la a0, msg + 1
```

Which generates the following assembler output and relocations for non-PIC as seen by objdump:

And generates the following assembler output and relocations for PIC as seen by objdump:

### From Concrete Syntax to Abstract Syntax (\*)

```
.text
li a0, 1
li a1, 15
ecall
li a0, 10
ecall
```

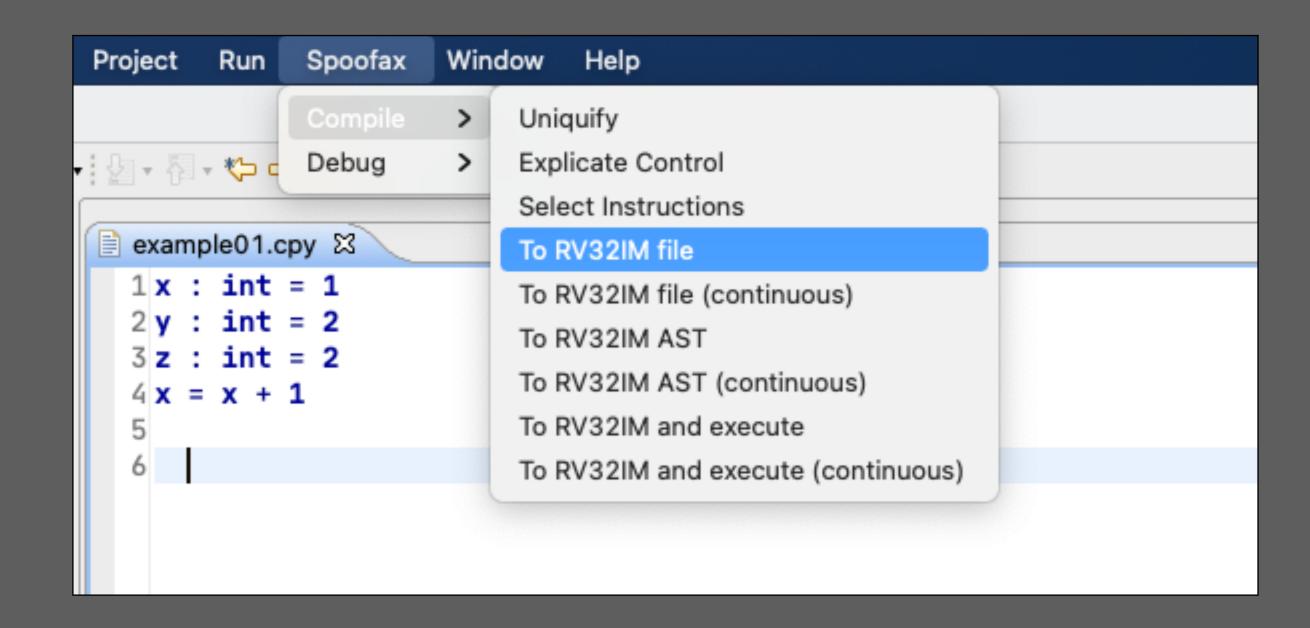


```
Program(
   [ PSText()
   , Li("a0", "1")
   , Li("a1", "15")
   , Ecall()
   , Li("a0", "10")
   , Ecall()
   ]
)
```

# Code Generation by Term Transformation

### Compilation Menu: chocopy.cfg

```
editor-context-menu [
 menu "Compile" [
   command-action {
      command-def = uniquifyToAstCommand
      execution-type = Once
      required-enclosing-resource-types = [Project]
    command-action {
      command-def = explicateControlToAstCommand
      execution-type = Once
      required-enclosing-resource-types = [Project]
   command-action {
      command-def = selectInstructionsToAstCommand
      execution-type = Once
      required-enclosing-resource-types = [Project]
    command-action {
      command-def = toFileCommand
      execution-type = Once
      required-enclosing-resource-types = [Project]
    command-action {
     command-def = toFileCommand
      execution-type = Continuous
      required-enclosing-resource-types = [Project]
```



### Invoking the Compiler: chocopy.cfg

```
let toFile = task-def mb.chocopy.show.ShowCompileToRv32ImFile
let toFileCommand = command-def {
  task-def = toFile
  display-name = "To RV32IM file"
  parameters = [
    rootDirectory = parameter {
      type = java mb.resource.hierarchical.ResourcePath
      argument-providers = [EnclosingContext(Project)]
    file = parameter {
      type = java mb.resource.hierarchical.ResourcePath
      argument-providers = [Context(File)]
```

### Invoking the Compiler: Java

```
package mb.chocopy.show;
|import ...
@ChocopyScope
public class ShowCompileToRv32ImFile implements TaskDef<ShowCompileToRv32ImFile.Args, CommandFeedback> {
    @SuppressWarnings("serial")
                                    public static final class Args implements Serializable {
        public final ResourcePath rootDirectory;
        public final ResourcePath file;
        public Args(ResourcePath rootDirectory, ResourcePath file) {
            this.rootDirectory = rootDirectory;
            this.file = file;
        @Override public boolean equals(@Nullable Object o) {
            if(this = o) return true;
            if(o = null || getClass() \neq o.getClass()) return false;
            Args input = (Args)o;
            if(!rootDirectory.equals(input.rootDirectory)) return false;
            return file.equals(input.file);
        @Override public int hashCode() {
            int result = rootDirectory.hashCode();
            result = 31 * result + file.hashCode();
            return result;
        @Override public String toString() {
            return "ShowCompileToRv32ImFile$Args{" +
                                                                   @Override public CommandFeedback exec(ExecContext context, Args args) throws IOException {
                "rootDirectory=" + rootDirectory +
                                                                    final ResourcePath file = args.file;
                ", file=" + file +
                                                                          return context.require(compileToRv32ImText.createSupplier(analyzeFile.createSupplier(new ChocopyAnalyzeFile.Input(args.rootDirectory,
                '}';
                                                                                                                                                                                              args.file)))).mapThrow
                                                                         text \rightarrow {
                                                                          final ResourcePath outputFilePath = file.appendToLeafExtension("rv32im");
                                                                          final WritableResource outputFile = context.getWritableResource(outputFilePath);
    private final ChocopyAnalyzeFile analyzeFile;
                                                                          outputFile.writeString(text);
    private final CompileToRv32ImText compileToRv32ImText;
                                                                          context.provide(outputFile);
                                                                          return CommandFeedback.of(ShowFeedback.showFile(outputFilePath));
    @Inject public ShowCompileToRv32ImFile(
               ChocopyAnalyzeFile analyzeFile,
                                                                          e → CommandFeedback.ofTryExtractMessagesFrom(e, file)
               CompileToRv32ImText compileToRv32ImText
                                                                          );
       this.analyzeFile = analyzeFile;
       this.compileToRv32ImText = compileToRv32ImText;
                                                                   @Override public String getId() {
                                                                          return getClass().getName();
```

### The Compiler Pipeline

```
rules

compile-to-rv32im-ast :: Program → RProgram

compile-to-rv32im-ast =
   compile-cpy-to-cir
   ; compile-cir-to-rv32im
   ; compile-rv32im
```

```
rules

compile-cpy-to-cir :: Program → CProgram

compile-cpy-to-cir =
   explicate-types
   ; desugar
   ; uniquify
   ; remove-complex-operands
   ; explicate-control
```

```
compile-cir-to-rv32im :: CProgram → RProgram
compile-cir-to-rv32im =
  select-instructions-cprogram
```

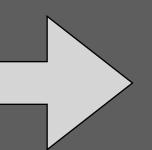
```
compile-rv32im :: RProgram → RProgram

compile-rv32im =
  assign-homes
; patch-instructions
```

## Uniquify

### Uniquify

```
x : int = 1
y : int = 2
z : int = 2
x = x + 1
z = x + y
```



```
Program(
  [ VarDef(TypedVar("x3", Type("int")), Int("1"))
   VarDef(TypedVar("y3", Type("int")), Int("2"))
   VarDef(TypedVar("z1", Type("int")), Int("2"))
 [ Assign(
      [Target(Var("x3"))]
     Add(Var("x3"), Int("1"))
  , Assign(
      [Target(Var("z1"))]
     Add(Var("x3"), Var("y3"))
```

### Uniquify

```
rules

declare-new-name :: ID → ID

declare-new-name :
    x1 → x2
    with <newname> x1 ⇒ x2
    with rules( Rename : x1 → x2 )

rename :: string → string
rename :
    x1 → x2
    with <stx-get-ast-analysis> x1 ⇒ analysis
    with <stx-get-ast-ref(|analysis)
        ; if is-list then Hd end> x1 ⇒ d
    where <Rename>d ⇒ x2
```

```
rules
  uniquify
                 :: Program → Program
  uniquify-gen :: ? \rightarrow ?
  uniquify-def :: Definition → Definition
  uniquify-ref :: Var → Var
  uniquify =
    uniquify-gen
  uniquify-gen =
    topdown(try(is(Definition); uniquify-def))
    ; topdown(try(is(Var); uniquify-ref))
  uniquify-def :
    VarDef(TypedVar(x1, t), e1) \rightarrow VarDef(TypedVar(x2, t), e1)
    with <declare-new-name> x1 \Rightarrow x2
  uniquify-ref :
    Var(x1) \rightarrow Var(x2)
    where <rename> x1 \Rightarrow x2
```

### C-IR

### 'C' Intermediate Representation

```
signature
  sorts CID CINT CProgram CBlock CLabel CTail CStmt
       CType CExp CAtom CVar
 constructors
                    : string → CID
                    : string \rightarrow CINT
                    : List(CBlock) → CProgram
   CProgram
                    : CLabel * CTail → CBlock
   CBlock
   CLabel : CID → CLabel
   CReturn
                    : CExp \rightarrow CTail
   CReturnNone
                    : CTail
                    : CStmt * CTail → CTail
   CSeq
   CVarDec
                    : CVar * CType * CExp → CStmt
   CAssign
                    : CVar * CExp → CStmt
   CIntT
                    : CType
                    : CAtom \rightarrow CExp
   CRead
                    : CExp
   CMin
                    : CAtom \rightarrow CExp
                    : CAtom ★ CAtom → CExp
   CAdd
                    : CAtom ★ CAtom → CExp
   CMul
   CDiv
                    : CAtom ★ CAtom → CExp
   CInt
                    : CINT \rightarrow CAtom
                    : CVar → CAtom
                    : CID \rightarrow CVar
   CVar
```

### Explicate Control: Generating a Control-Flow Graph

```
rules // control-flow graph

add-cfg-node :: CBlock → CBlock
all-cfg-nodes :: List(CBlock) → List(CBlock)

add-cfg-node =
?block
; rules( CFGNode :+ _ → block )

all-cfg-nodes =
<br/>
<br/>
all-cfg-nodes =
<br/>
<br/>
<br/>
cbagof-CFGNode <+ ![]>()
```

```
rules

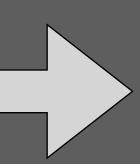
explicate-control :: Program → CProgram

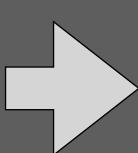
explicate-control :
    Program(defs, stms) → CProgram([CBlock(CLabel("Main"), tail2) | blocks])
    with <explicate-tail-seq> stms ⇒ tail1
    with <explicate-defs(|tail1)> defs ⇒ tail2
    with <all-cfg-nodes>[] ⇒ blocks
```

### Instruction Selection

### Select Instructions: Example

```
z : int = 3
z + 1
```





```
RProgram(
    [ RPSData()
    , RPSText()
    , RLabel("Main")
    , RLocal(RVar("z3"), RIntT())
    , RLi(RVar("z3"), RInt("3"))
    , RAddi(RReg("a0"), RVar("z3"), RInt("1"))
    ]
)
```

### Select Instructions: Programs

```
rules select-instructions-cprogram :: CProgram → RProgram

select-instructions-cprogram :
    CProgram(blocks) → RProgram(<concat>[
        [RPSData()],
        [RPSText()],
        instrs
    ])
    with <mapconcat(select-instrs-block(|"a0"))> blocks ⇒ instrs
```

```
rules select-instrs-block(|string) :: CBlock → List(RLine)

select-instrs-block(|r) :
    CBlock(CLabel(lbl), tail) → [RLabel(<cid-to-string>lbl) | instrs]
    with <select-instrs-tail(|RReg("a0"))> tail ⇒ instrs

rules select-instrs-tail(|RArg) :: CTail → List(RLine)

select-instrs-tail(|r) :
    CReturn(exp) → instrs
    with <select-instrs-exp(|r)> exp ⇒ instrs

select-instrs-tail(|r) :
    CReturnNone() → []
```

### Select Instructions: Programs

```
rules select-instructions-cprogram :: CProgram → RProgram

select-instructions-cprogram :
    CProgram(blocks) → RProgram(<concat>[
        [RPSData()],
        [RPSText()],
        instrs
    ])
    with <mapconcat(select-instrs-block(|"a0"))> blocks ⇒ instrs
```

```
rules select-instrs-block(|string) :: CBlock → List(RLine)

select-instrs-block(|r) :
    CBlock(CLabel(lbl), tail) → [RLabel(<cid-to-string>lbl) | instrs]
    with <select-instrs-tail(|RReg("a0"))> tail ⇒ instrs

rules select-instrs-tail(|RArg) :: CTail → List(RLine)

select-instrs-tail(|r) :
    CReturn(exp) → instrs
    with <select-instrs-exp(|r)> exp ⇒ instrs

select-instrs-tail(|r) :
    CReturnNone() → []
```

### Select Instructions: Expressions

```
rules select-instrs-exp(|RArg) :: CExp → List(RLine)

select-instrs-exp(|x) :
   CInt(i) → [RLi(x, <cint-to-rint>i)]

select-instrs-exp(|x) :
   CVar(y) → [RMv(x, RVar(<cid-to-string>y))]

select-instrs-exp(|x) :
   CAdd(y@CVar(_), z@CVar(_)) → [RAdd(x, <cvar-to-rvar>y, <cvar-to-rvar>z)]
```

### Compilation Schemas

### Abstract From Implementation Details

```
|[i]|_{r,regs} \Rightarrow li r, i
```

```
||[e + i]|_{r,regs} \rightarrow |[e]|_{r,regs}
addi r, r, i
```

```
|[\ e1 + e2\ ]|_{r_1,r_2,r_{egs}} \Rightarrow |[\ e1\ ]|_{r_1,r_2,r_{egs}} \\ |[\ e2\ ]|_{r_2,r_{egs}} \\ |[\ add\ r1,\ r1,\ r2]|
```

Except where otherwise noted, this work is licensed under

