The University of Nottingham

SCHOOL OF COMPUTER SCIENCE

A LEVEL 3 MODULE, AUTUMN SEMESTER 2018–2019

COMPILERS

Time allowed TWO hours

Candidates may complete the front cover of their answer book and sign their desk card but must NOT write anything else until the start of the examination period is announced.

Answer ALL THREE questions

No calculators are permitted in this examination.

Dictionaries are not allowed with one exception. Those whose first language is not English may use a standard translation dictionary to translate between that language and English provided that neither language is the subject of this examination. Subject-specific translation directories are not permitted.

No electronic devices capable of storing and retrieving text, including electronic dictionaries, may be used.

DO NOT turn examination paper over until instructed to do so

ADDITIONAL MATERIAL: Appendix A, Appendix B

INFORMATION FOR INVIGILATORS: none

Question 1
See Appendix A for the MiniTriangle grammars relevant to this question.

(a) The following is a Haskell datatype definition for representing the abstract syntax of a selection of MiniTriangle commands. The type Expression represents the abstract syntax of expressions.

```
data Command = CmdAssign Expression Expression
             | CmdIf Expression Command Command
             | CmdWhile Expression Command
             | CmdSeq [Command]
```

A Happy parser specification dealing with commands and sequences of commands is given below. The semantic actions for constructing an abstract syntax tree (AST) have been left out (indicated by a boxed number, like [3]). Complete the specification by providing semantic actions for constructing an AST. The type of the semantic values of the non-terminals var_expression and expression is Expression.

```
commands :: { [Command] }
commands : command                { [1] }
         | command ';' commands   { [2] }

command :: { Command }
command
    : var_expression ':=' expression          { [3] }
    | IF expression THEN command ELSE command { [4] }
    | WHILE expression DO command             { [5] }
    | BEGIN commands END                      { [6] }
```

(6)

(b) Suppose we wish to extend MiniTriangle with a for-loop (a new command). The following two code fragments illustrate the idea:

- `for i from 1 to 10 do x[i] := i * i`
- `for j from 2 * m to n step -2 do sum := sum + j`

The for-loop has the following semantics. The expressions defining the start, end, and step are evaluated exactly once. The loop variable is initialised to the value given by the expression following the keyword from. The loop body is then repeated 0 or more times, incrementing (if positive step size) or decrementing (if negative step size) the loop variable after each execution of the body until the value of the loop variable is greater (positive step size) or smaller (negative step size) than the value of the expression following the keyword to. Note that the step size is optional.

If left out, it should default to 1. Thus, in the first example, `i` will assume the values 1, 2, . . . , 10 in that order, with the loop body `x[i] := i * i` executed once for each assignment.

(i) Extend the MiniTriangle lexical and concrete syntax with new productions defining the syntax of the `for`-loop. Pick the syntactic categories for the constituent parts with care: your extended grammars should be reasonably general, and in particular general enough to accept both examples above. (4)

(ii) Extend the type `Command` with a new constructor for representing `for`-loops. Then show how to extend the Happy parser specification so that the new construct is accepted and a corresponding AST gets constructed. You may assume that all extensions related to the lexical syntax, including extending the scanner, have already been carried out. (5)

(c) Write the case(s) of a code-generation function *execute* for generating code for the `for`-loop, targetting the Triangle Abstract Machine (TAM). See appendix B for a specification of the TAM instructions. The code generation function should be specified through *code templates* in the style used in the lectures. Thus, for the case without the optional step size, something along the lines

$$execute\ n\ [\![\, \texttt{for}\ E_{\mathrm{x}}\ \texttt{from}\ E_{\mathrm{f}}\ \texttt{to}\ E_{\mathrm{t}}\ \texttt{do}\ C\,]\!] = \ldots$$

where n is the current stack depth.

Assume a code-generation function *evaluate* (which does not need the current stack depth as expressions do not introduce new variables) for generating code for expressions, leaving the value of the expression on the top of the stack. Assume further that calling *evaluate* on the expression corresponding to the loop variable generates code that leaves the *address* of the variable on the stack (for use by instructions such as `LOADI` and `STOREI`). Call *execute* recursively for commands. Generation of fresh labels need not be considered; it suffices that labels are distinct within each case of the code function. (Also, there is no need to consider environments for mapping identifiers to addresses etc.) Take care to only generate code for the body once. (10)

Question 2

(a) Consider the following expression language:

e	$\rightarrow$		*expressions:*
	$\mid$	n	*natural numbers, $n \in \mathbb{N}$*
	$\mid$	x	*variables, $x \in \mathrm{Name}$*
	$\mid$	$e \texttt{ + } e$	*addition*
	$\mid$	$e \texttt{ - } e$	*subtraction*
	$\mid$	$e \texttt{ * } e$	*multiplication*
	$\mid$	$e \texttt{ = } e$	*equality test*
	$\mid$	$\texttt{if}\ e\ \texttt{then}\ e\ \texttt{else}\ e$	*conditional*
	$\mid$	$\texttt{let var}\ x = e\ \texttt{in}\ e$	*variable definition*
	$\mid$	$\texttt{let fun}\ f\,(x{:}t){:}t = e\ \texttt{in}\ e$	*function definition*
	$\mid$	$e(e)$	*function application*

where Name is the set of variable names. The types are given by the following grammar:

t	$\rightarrow$		*types:*
	$\mid$	$\texttt{Nat}$	*natural numbers*
	$\mid$	$\texttt{Bool}$	*Booleans*
	$\mid$	$t \rightarrow t$	*function (arrow) type*

The ternary relation $\Gamma \vdash e : t$ says that expression e has type t in the typing context Γ. It is defined by the following typing rules:

$$\Gamma \vdash n : \texttt{Nat} \qquad \text{(T-NAT)}$$

$$\frac{x : t \in \Gamma}{\Gamma \vdash x : t} \qquad \text{(T-VAR)}$$

$$\frac{\Gamma \vdash e_1 : \texttt{Nat} \quad \Gamma \vdash e_2 : \texttt{Nat}}{\Gamma \vdash e_1 \texttt{ + } e_2 : \texttt{Nat}} \qquad \text{(T-ADD)}$$

$$\frac{\Gamma \vdash e_1 : \texttt{Nat} \quad \Gamma \vdash e_2 : \texttt{Nat}}{\Gamma \vdash e_1 \texttt{ - } e_2 : \texttt{Nat}} \qquad \text{(T-SUB)}$$

$$\frac{\Gamma \vdash e_1 : \texttt{Nat} \quad \Gamma \vdash e_2 : \texttt{Nat}}{\Gamma \vdash e_1 \texttt{ * } e_2 : \texttt{Nat}} \qquad \text{(T-MUL)}$$

$$\frac{\Gamma \vdash e_1 : \texttt{Nat} \quad \Gamma \vdash e_2 : \texttt{Nat}}{\Gamma \vdash e_1 \texttt{ = } e_2 : \texttt{Bool}} \qquad \text{(T-EQ)}$$

$$\frac{\Gamma \vdash e_1 : \texttt{Bool} \quad \Gamma \vdash e_2 : t \quad \Gamma \vdash e_3 : t}{\Gamma \vdash \texttt{if}\ e_1\ \texttt{then}\ e_2\ \texttt{else}\ e_3 : t} \quad \text{(T-COND)}$$

$$\frac{\Gamma \vdash e_1 : t_1 \quad \Gamma,\ x : t_1 \vdash e_2 : t_2}{\Gamma \vdash \texttt{let var}\ x\ =\ e_1\ \texttt{in}\ e_2 : t_2} \quad \text{(T-LETVAR)}$$

$$\frac{\Gamma,\ f : t_{11} \to t_{12},\ x : t_{11} \vdash e_1 : t_{12} \quad \Gamma,\ f : t_{11} \to t_{12} \vdash e_2 : t_2}{\Gamma \vdash \texttt{let fun}\ f(x{:}t_{11}){:}t_{12}\ =\ e_1\ \texttt{in}\ e_2 : t_2} \quad \text{(T-LETFUN)}$$

$$\frac{\Gamma \vdash e_1 : t_2 \to t_1 \quad \Gamma \vdash e_2 : t_2}{\Gamma \vdash e_1(e_2) : t_1} \quad \text{(T-APP)}$$

A typing context, Γ in the rules above, is a comma-separated sequence of variable-name and type pairs, such as

$$\texttt{x} : \texttt{Nat},\ \texttt{y} : \texttt{Bool},\ \texttt{z} : \texttt{Nat}$$

or empty, denoted $\emptyset$. Typing contexts are extended on the right, e.g. $\Gamma,\ \texttt{z} : \texttt{Nat}$, the membership predicate is denoted by $\in$, and lookup is from right to left, ensuring recent bindings hide earlier ones.

Use the typing rules given above to formally derive the type of the following (well-typed) expressions in the empty environment ($\emptyset$). Your proof should be in the form of a *proof tree*.

(i)
```
let var x = 1 + 7 in x * x
```
(4)

(ii)
```
let fun fac(n : Nat) : Nat =
    if n = 0 then 1 else n * fac(n - 1)
in
    fac(7)
```
(9)

(b) Suppose we wish to extend MiniTriangle with a command `break`:

Command	$\to$	...	...
	|	`break` <u>*IntegerLiteral*</u>	CmdBreak

See Appendix A for the abstract syntax for the remaining MiniTriangle commands. The intended semantics of `break` n, where $n \geq 1$, is to terminate the innermost n loops, with the execution continuing immediately after the nth loop. It should be a static error if there are fewer than n loops enclosing a command `break` n or if $n < 1$. Define, using inference rules, a binary relation *Well Enclosed* on numbers and commands characterising the static correctness of commands in this sense. *Hint:* Think of the number as a form of context keeping track of the number of enclosing loops. (12)

Question 3

(a) Transform the following code fragment into *static single assignment* (SSA) form:

```
a := 1;
b := 17;
i := 0;
while i < n do begin
    c := a + i;
    i := i + 1;
    a := c
end;
b := b + a
```

(10)

(b) This question concerns *register allocation by graph colouring*. Consider the following assembly code fragment for a typical register machine:

```
        load    R0, 1
        load    R1, 0
loop:   mul     R2, R0, R0
        mul     R3, R0, R0
        mul     R4, R3, R0
        add     R5, R2, R4
        add     R1, R1, R5
        load    R6, 1
        add     R0, R0, R6
        load    R7, 10
        cmp     R0, R7
        ble     loop
```

The load instruction stores a numeric constant into the designated register. Arithmetic instructions with three register arguments perform the arithmetic operation on the two last registers and store the result into the first. The instruction ble is a conditional branch (jump) instruction.

(i) Draw the *interference graph* for the above code fragment. It should have one node for each of the eight registers being used. (6)

(ii) "Colour" the interference graph using as few colours as possible such that no two adjacent nodes have the same colour. Use this result to carry out register allocation for the above code fragment by associating each colour with one register. Your answer should include the coloured graph and the final version of the code using a minimal number of registers. (9)

Appendix A: MiniTriangle Grammars

This appendix contains the grammars for the MiniTriangle lexical, concrete, and abstract syntax. The following typographical conventions are used to distinguish between terminals and non-terminals:

- nonterminals are written like *this*
- terminals are written like `this`
- terminals with *variable spelling* and special symbols are written like <u>*this*</u>

MiniTriangle Lexical Syntax:

Program → (*Token* | *Separator*)*

Token → *Keyword* | *Identifier* | *IntegerLiteral* | *Operator*
| `,` | `;` | `:` | `:=` | `=` | `(` | `)` | `[` | `]` | <u>*eot*</u>

Keyword → `begin` | `const` | `do` | `else` | `end` | `fun` | `if` | `in`
| `let` | `out` | `proc` | `then` | `var` | `while`

Identifier → *Letter* | *Identifier Letter* | *Identifier Digit*
except *Keyword*

IntegerLiteral → *Digit* | *IntegerLiteral Digit*

Operator → `^` | `*` | `/` | `+` | `-` | `<` | `<=` | `==` | `!=` | `>=` | `>` | `&&` | `||` | `!`

Letter → `A` | `B` | ... | `Z` | `a` | `b` | ... | `z`

Digit → `0` | `1` | `2` | `3` | `4` | `5` | `6` | `7` | `8` | `9`

Separator → *Comment* | <u>*space*</u> | <u>*eol*</u>

Comment → `//` (any character except <u>*eol*</u>)* <u>*eol*</u>

MiniTriangle Concrete Syntax:

Program → *Command*

Commands → *Command*
| *Command* `;` *Commands*

Command → *VarExpression* `:=` *Expression*
| *VarExpression* `(` *Expressions* `)`
| `if` *Expression* `then` *Command* `else` *Command*
| `while` *Expression* `do` *Command*
| `let` *Declarations* `in` *Command*
| `begin` *Commands* `end`

Expressions → ϵ
| *Expressions₁*

Expressions₁ → *Expression*
| *Expression* `,` *Expressions₁*

Expression → *PrimaryExpression*
| *Expression* *BinaryOperator* *Expression*

PrimaryExpression → <u>*IntegerLiteral*</u>
| *VarExpression*
| *UnaryOperator* *PrimaryExpression*
| *VarExpression* `(` *Expressions* `)`
| `[` *Expressions* `]`
| `(` *Expression* `)`

VarExpression → <u>*Identifier*</u>
| *VarExpression* `[` *Expression* `]`

BinaryOperator → `^` | `*` | `/` | `+` | `-` | `<` | `<=` | `==` | `!=` | `>=` | `>` | `&&` | `||`

UnaryOperator → `-` | `!`

```
Declarations  →  Declaration
              |  Declaration ; Declarations

Declaration   →  const Identifier : TypeDenoter = Expression
              |  var Identifier : TypeDenoter
              |  var Identifier : TypeDenoter := Expression
              |  fun Identifier ( ArgDecls ) : TypeDenoter = Expression
              |  proc Identifier ( ArgDecls ) Command

ArgDecls      →  ϵ
              |  ArgDecls1

ArgDecls1     →  ArgDecl
              |  ArgDecl , ArgDecls1

ArgDecl       →  Identifier : TypeDenoter
              |  in Identifier : TypeDenoter
              |  out Identifier : TypeDenoter
              |  var Identifier : TypeDenoter

TypeDenoter   →  Identifier
              |  TypeDenoter [ IntegerLiteral ]
```

Note that the productions for *Expression* make the grammar as stated above ambiguous. Operator precedence and associativity for the *binary* operators as defined in the following table are used to disambiguate:

<table>
<tr><th>Operator</th><th>Precedence</th><th>Associativity</th></tr>
<tr><td>^</td><td>1</td><td>right</td></tr>
<tr><td>* /</td><td>2</td><td>left</td></tr>
<tr><td>+ -</td><td>3</td><td>left</td></tr>
<tr><td>< <= == != >= ></td><td>4</td><td>non</td></tr>
<tr><td>&&</td><td>5</td><td>left</td></tr>
<tr><td>||</td><td>6</td><td>left</td></tr>
</table>

A precedence level of 1 means the highest precedence, 2 means second highest, and so on.

MiniTriangle Abstract Syntax: $\underline{\mathit{Name}} = \underline{\mathit{Identifier}} \cup \underline{\mathit{Operator}}$.

Program	→	*Command*	Program
Command	→	*Expression* `:=` *Expression*	CmdAssign
	|	*Expression* `(` *Expression** `)`	CmdCall
	|	`begin` *Command** `end`	CmdSeq
	|	`if` *Expression* `then` *Command* `else` *Command*	CmdIf
	|	`while` *Expression* `do` *Command*	CmdWhile
	|	`let` *Declaration** `in` *Command*	CmdLet
Expression	→	$\underline{\mathit{IntegerLiteral}}$	ExpLitInt
	|	$\underline{\mathit{Name}}$	ExpVar
	|	*Expression* `(` *Expression** `)`	ExpApp
	|	`[` *Expression** `]`	ExpAry
	|	*Expression* `[` *Expression* `]`	ExpIx
Declaration	→	`const` $\underline{\mathit{Name}}$ `:` *TypeDenoter* `=` *Expression*	DeclConst
	|	`var` $\underline{\mathit{Name}}$ `:` *TypeDenoter* (`:=` *Expression* $\mid \epsilon$)	DeclVar
	|	`fun` $\underline{\mathit{Name}}$ `(` *ArgDecl** `)` `:` *TypeDenoter* `=` *Expression*	DeclFun
	|	`proc` $\underline{\mathit{Name}}$ `(` *ArgDecl** `)` *Command*	DeclProc
ArgDecl	→	*ArgMode* $\underline{\mathit{Name}}$ `:` *TypeDenoter*	ArgDecl
ArgMode	→	ϵ	ByValue
	|	`in`	ByRefIn
	|	`out`	ByRefOut
	|	`var`	ByRefVar
TypeDenoter	→	$\underline{\mathit{Name}}$	TDBaseType
	→	*TypeDenoter* `[` $\underline{\mathit{IntegerLiteral}}$ `]`	TDArray

Appendix B: Triangle Abstract Machine (TAM) Instructions

Meta variable	Meaning
a	Address: one of the forms specified by table below when part of an instruction, specific stack address when on the stack
b	Boolean value ($\mathrm{false} = 0$ or $\mathrm{true} = 1$)
ca	Code address; address to routine in the code segment
d	Displacement; i.e., offset w.r.t. address in register or on the stack
l	Label name
m, n, p	Integer
x, y, z	Any kind of stack data
x^n	Vector of n items, $n \geq 0$, here any kind

Address form	Description
`[SB + ` d`]` `[SB - ` d`]`	Address given by contents of register `SB` (Stack Base) $+/-$ displacement d
`[LB + ` d`]` `[LB - ` d`]`	Address given by contents of register `LB` (Local Base) $+/-$ displacement d
`[ST + ` d`]` `[ST - ` d`]`	Address given by contents of register `ST` (Stack Top) $+/-$ displacement d

Instruction	Stack effect	Description
	Label	
`LABEL` l	—	Pseudo instruction: symbolic location
	Load and store	
`LOADL` n	$\ldots \Rightarrow n, \ldots$	Push literal integer n onto stack
`LOADCA` l	$\ldots \Rightarrow \mathrm{addr}(l), \ldots$	Push address of label l (code segment) onto stack
`LOAD` a	$\ldots \Rightarrow [a], \ldots$	Push contents at address a onto stack
`LOADA` a	$\ldots \Rightarrow a, \ldots$	Push address a onto stack
`LOADI` d	$a, \ldots \Rightarrow [a+d], \ldots$	Load indirectly; push contents at address $a+d$ onto stack
`STORE` a	$n, \ldots \Rightarrow \ldots$	Pop value n from stack and store at address a
`STOREI` d	$a, n, \ldots \Rightarrow \ldots$	Store indirectly; store n at address $a+d$

Instruction	Stack effect	Description
	Block operations	
LOADLB m n	$\ldots \Rightarrow m^n, \ldots$	Push block of n literal integers m onto stack
LOADIB n	$a, \ldots \Rightarrow [a+(n-1)], \ldots, [a+0], \ldots$	Load block of size n indirectly
STOREIB n	$a, x^n, \ldots \Rightarrow \ldots$	Store block of size n indirectly
POP m n	$x^m, y^n, \ldots \Rightarrow x^m, \ldots$	Pop n values below top m values
	Arithmetic operations	
ADD	$n_2, n_1, \ldots \Rightarrow n_1 + n_2, \ldots$	Add n_1 and n_2, replacing n_1 and n_2 with the sum
SUB	$n_2, n_1, \ldots \Rightarrow n_1 - n_2, \ldots$	Subtract n_2 from n_1, replacing n_1 and n_2 with the difference
MUL	$n_2, n_1, \ldots \Rightarrow n_1 \cdot n_2, \ldots$	Multiply n_1 by n_2, replacing n_1 and n_2 with the product
DIV	$n_2, n_1, \ldots \Rightarrow n_1 / n_2, \ldots$	Divide n_1 by n_2, replacing n_1 and n_2 with the (integer) quotient
NEG	$n, \ldots \Rightarrow -n, \ldots$	Negate n, replacing n with the result
	Comparison & logical operations (false $= 0$, true $= 1$)	
LSS	$n_2, n_1, \ldots \Rightarrow n_1 < n_2, \ldots$	Check if n_1 is smaller than n_2, replacing n_1 and n_2 with the Boolean result
EQL	$n_2, n_1, \ldots \Rightarrow n_1 = n_2, \ldots$	Check if n_1 is equal to n_2, replacing n_1 and n_2 with the Boolean result
GTR	$n_2, n_1, \ldots \Rightarrow n_1 > n_2, \ldots$	Check if n_1 is greater than n_2, replacing n_1 and n_2 with the Boolean result
AND	$b_2, b_1, \ldots \Rightarrow b_1 \wedge b_2, \ldots$	Logical conjunction of b_1 and b_2, replacing b_1 and b_2 with the Boolean result
OR	$b_2, b_1, \ldots \Rightarrow b_1 \vee b_2, \ldots$	Logical disjunction of b_1 and b_2, replacing b_1 and b_2 with the Boolean result
NOT	$b, \ldots \Rightarrow \neg b, \ldots$	Logical negation of b, replacing b with the result

Instruction	Stack effect	Description
	Control transfer	
JUMP l	—	Jump unconditionally to location identified by label l
JUMPIFZ l	$n, \ldots \Rightarrow \ldots$	Jump to location identified by label l if $n = 0$ (i.e., n is false)
JUMPIFNZ l	$n, \ldots \Rightarrow \ldots$	Jump to location identified by label l if $n \neq 0$ (i.e., n is true)
CALL l	$\ldots \Rightarrow$ PC $+ 1$, LB, 0, $\ldots$	Call global subroutine at location l: Activation record set up by pushing static link (0 for global level), dynamic link (value of LB), and return address (PC$+1$, address of instruction after the call instruction) onto the stack; PC $= l$; LB $=$ start of activation record (address of static link)
CALLI	ca, sl, $\ldots \Rightarrow$ PC $+ 1$, LB, sl, $\ldots$	Call subroutine indirectly: address of routine (ca) and static link to use (sl) on top of the stack; activation record and new PC and LB as for CALL
RETURN m n	x^m, y^p, ra, olb, sl, y^n, $\ldots \Rightarrow x^m$, $\ldots$	Return from subroutine, replacing activation record by result, jumping to return address (PC $= ra$), and restoring the old local base (LB $= olb$)
	Input/Output	
PUTINT	$n, \ldots \Rightarrow \ldots$	Print n to the terminal as a decimal integer
PUTCHR	$n, \ldots \Rightarrow \ldots$	Print the character with character code n to the terminal
GETINT	$\ldots \Rightarrow n, \ldots$	Read decimal integer n from the terminal and push onto the stack
GETCHR	$\ldots \Rightarrow n, \ldots$	Read character from the terminal and push its character code n onto the stack
	TAM Control	
HALT	—	Stop execution and halt the machine