# Introduction

The stackl environment is intended as a virtual environment that can be used for teaching operating systems classes. The environment allows students to program in C and be able to write interrupt service routines, syscalls, and other operating system code without the need for real hardware.

# stacklc: the stackl compiler

The stacklc compiler compiles a subset of the C programming language. The following are the current restrictions:

1. No separate compilation
2. The switch statement is not implemented
3. The only array type supported is char
4. The only pointer type supported is char
5. Pointer dereference is not supported. Instead of \*ptr, use ptr[0]
6. void is not supported
7. static is not supported

## Compiler modes

The compiler runs in two modes: boot and regular. Boot mode is selected with the boot command line option. In regular mode, the compiler includes the file <startup.h> prior to compiling any other source code. It is assumed that the startup code will supply a startup\_\_ function that will call main() and then exit (halt the CPU).

In boot mode, the compiler does not imply any include file. The compiled source must include definitions for the startup\_\_, systrap, and interrupt functions.

When run in regular mode (-boot is not specified) the compiler includes a file by the name of startup.h. This file typically defines the special functions listed in the next section. The function startup\_\_ is expected to call main() and then exit.

## Special functions

The following functions have special meaning. These functions must be defined in every compilation unit.

### startup\_\_

This function must initialize the execution environment and then, typically, calls main() and then exit(). The prototype for startup\_\_ is as follows:

int startup\_\_();

### systrap

If the resulting .sl file is loaded as the boot file, then this function will be executed every time the TRAP opcode executes. This function must use an asm(RTI\_OP) to return. The contents of the stack at the time of the TRAP will determine the arguments for this function. Typically, the TRAP function gets executed by a syscall function, and the arguments for systrap are the same as for the syscall function.

If the arguments to this function are used, then this function must not have any local variables.

### interrupt

This function is the interrupt service routine. It gets called any time there is a hardware interrupt. This function must use asm(RTI\_OP) to return.

## Generating assembly code

Assembly instructions can be generated using the asm keyword. The asm and asm2 keywords behave like functions with the following prototype:

int asm(opcode, ...);

int asm2(opcode, opcode2, ...);

The optional parameters will be pushed onto the stack, then the opcode value(s) will be emitted to the instruction stream. The asm2 form is useful for two word instructions such as PUSH\_OP or JMPUSER\_OP.

If the asm keywords are used in an assignment statement, then the value at the top of the stack will be popped and stored in the lval of the assignment.

# stackl machine

The stackl machine is a stack based machine. As such, there are no general purpose registers. All operations make use of the stack. Typically, operands are popped off the stack and the result of an operation is pushed back onto the stack.

There are two categories of special purpose registers: execution and memory management. The execution registers are used for general instructions. The memory management instructions are used to manage memory when multiple processes are running on the machine. The registers are described in the following sections.

## Execution Registers

### Instruction Pointer (IP)

The instruction pointer contains the address of the next instruction to be executed.

### Stack Pointer (SP)

The stack pointer contains the address of the first unused memory location at the top of the stack. The stack always contains word aligned data. The stack grows to increasing memory addresses.

Pseudocode for pushing an item onto the stack is as follows:

Memory[SP] = value;

SP++;

Pseudo-code for popping an item from the stack is as follows:

SP--;

value = Memory[SP];

### Frame Pointer (FP)

The frame pointer is used for function calls. It points to the location of the first local variable in a called function. The opcodes that include VAR in the opcode name reference the FP to locate the variable.

## Memory Management Registers

When operating in user mode, memory accesses are relative to the Base Pointer. When not in user mode, the Base Pointer and Limit Pointer are ignored, and all addresses are absolute.

### Base Pointer (BP)

The base pointer contains the address of the start of memory for the current process. All memory addresses are relative to the base pointer. In other words, if a process wants to access memory at address N, the actual address is BP+N. This is true whether the memory operation is fetching data or instructions.

### Limit Pointer (LP)

The limit pointer contains the value just above the highest memory that a process can access. In other words, legal memory access are those in the range of BP .. LP-1. Any access outside this range will cause a machine check.

## Stack Frames

The following diagram illustrates a stack frame. Higher addresses are higher in the diagram.



The caller must push the parameters onto the stack prior to executing the CALL instruction. Parameters must be pushed from right to left. In other words, a call to

foo(a,b,c)

would push c then b then a.

## Input/Output

The CPU has two instructions for doing input/output the OUTS instruction prints the NULL terminated string whose address is on the top of the stack. This is a blocking instruction: the next instruction will not execute until the IO operation is complete.

The INP instruction provides a non-blocking means of doing input. The value on the top of the stack is assumed to be the address of an io\_blk\_t. The IO block has the following form:

typedef struct

{

int op;

int param1;

int param2;

} io\_blk\_t;

The op is used to specify the desired operation and indicate the completion of the operation. The two param values are operation dependent.

When the operation is finished the most significant bit of the op field is set (resulting in a negative value).

The following operations are defined:

GETL\_CALL 6 This operation will perform a C gets() function. The address of the buffer is in param1. There is no attempt to prevent buffer overflow.

GETS\_CALL 5 This operation will perform the following C function:

scanf("%s", param1);

where param1 is assumed to be the address of the buffer.

GETI\_CALL 7 This operation will perform the following C function:

scanf("%d", param1);

where param1 is assumed to be the address of an integer.

EXEC\_CALL 8 This operation will call the loader to load a file into memory. The name of the file is in a string pointed to by param1. The high memory address of the load will be stored in param2 upon completion. The BP and LP registers should be set prior to initiating this operation.

Since these are non-blocking operations, the initiator needs to poll the op field waiting for it to turn negative to determine when the operation is complete.

## Loading programs

Programs can be loaded into memory using the INP instruction with the EXEC\_CALL operation. The loader will load the code starting at the address indicated by the BP. The loader will adjust addresses so that they are relative to BP. After loading, the param2 of the io\_blk will contain the address of the first free word above the loaded file. This value can be used to set the SP and FP for executing the code.

The following procedure should be used to load code:

set the BP and LP to bound the memory available to the program

initiate the load using the INP instruction

wait for the load to complete

Clear the interrupt mode bit in the FLAG register (if set)

Set the FP and SP to the value in param2

Use the JMPUSER instruction to jump to address 8

The code for doing this is as follows:

// set BP to start\_addr

asm(POPREG\_OP, start\_addr);

asm(BP\_REG);

// initiate and wait for INP instruction

asm(SETMODE\_OP); // clear interrupt bit

asm(0);

asm(DUP\_OP, io\_blk.param2);

asm(POPREG\_OP);

asm(FP\_REG); // update FP

asm(POPREG\_OP);

asm(SP\_REG); // update SP

asm(JMPUSER\_OP);

asm(8); // jump to addr 8 in user mode

Following the update to the FP, no VAR instructions should execute because they would reference the new FP. Following the update to the SP, no stack operations should execute because they will reference the new stack.

# Devices

This section discusses the IO devices available on the stackl machine.

## Programmed IO Terminal (pio\_term)

This device is used for doing programmed IO from/to the console. The device can operate in polled mode or in interrupt mode.

### Registers

The following registers are available for this device. These are byte-wide registers and should only be accessed using byte wide instructions. If a word sized instruction is used to read or write these registers, the behavior is undefined.

#### RDR – Read Data Register

This register is at address 0x0E000000. This is a read only register.

Data input from the console will appear in this register. Every time a new value appears in this register, the RECV bit in the IIR register will be set. The bit will be cleared when the data is read from the RDR.

#### XDR – Xmit Data Register

This register is at address 0x0E000000. This is a write only register.

Data written to this register will be sent to the console. Writing to this register will clear the XMIT bit in the IIR. The XMIT bit in the IIR will be set when the device is ready to transmit another byte.

#### IER – Interrupt Enable Register

This register is at address 0x0E000001. This is a read/write register.

This register controls what events will trigger an interrupt. Interrupts will be triggered if the bit corresponding to the event is set in this register. Bit 0 controls the READ interrupt, bit 1 controls the XMIT interrupt.

#### IIR – Interrupt Information Register

This register is at address 0x0E000002. This is a read only register. Reading the register will clear it.

This register indicates what caused the interrupt. The following bits are used:

0x01 ATTN: This device needs attention

0x02 XMIT: The device finished transmitting a byte

0x04 RECV: The device received a new character

# Interpreter Architecture

## Memory

In order to facilitate memory management, all memory references are made through the following four functions:

Get\_Word Returns a word-sized value. The address must be word aligned.

Set\_Word Sets a word-sized value. The address must be word aligned

Get\_Byte Returns a byte-sized value. There are no alignment requirements on the address.

Set\_Byte Sets a word-sized value. There are no alignment requirements on the address.

If the address specified to these functions is outside the range of main memory, then the address will be interpreted as the address of a memory mapped IO register.

There are two functions for managing the machine state. They get/set all the registers. These functions are Get\_Machine\_State and Set\_Machine\_State.

## Executable file format and the Loader

To be specified

## Adding IO devices to the interpreter

To be specified

# Instruction Reference

This section defines the opcodes for the stackl machine. All opcodes occupy a word.

For binary operations, the left operand is always pushed onto the stack first.

Some opcodes take a parameter from the instruction stream (that is, they are two word instructions). These instructions are indicated by <param> in the opcode name where param is replaced by the name of the parameter. The parameter can be found at Memory[IP+1].

In the descriptions of the opcodes, memory offsets are given in word-sized increments. In other words, “IP+1” means the word following the IP, not the byte following the IP. Offsets relative to the IP assume the IP is pointing at the opcode.

In the descriptions of the effects of the opcodes, the operators have the meaning defined by the C language.

The numerical value of specific opcodes should not re relied upon because they are subject to change in future versions of stackl.

## ADJSP\_OP <amount>

This opcode is used to adjust the stack pointer. This opcode has the following effect:

SP += <amount>

The <amount> can be a positive or negative number. The <amount> is in bytes, not words.

## AND\_OP

This opcode performs a logical AND. This instruction has the following effect:

Memory[SP-2] = Memory[SP-2] && Memory[SP-1]

SP--

## CALL\_OP <addr>

This opcode is used for making function calls. This opcode has the following effect:

push IP+2 // return address

push FP

SP += 2

FP = SP

IP = <addr>

## CALLI\_OP

This opcode is reserved and should not be used

## DIVIDE\_OP

This opcode performs a divide. This instruction has the following effect:

Memory[SP-2] = Memory[SP-2] / Memory[SP-1]

SP--

## DUP\_OP

This opcode duplicates the value at the top of the stack. This instruction has the following effect:

Memory[SP] = Memory[SP-1]

SP++

## EQ\_OP

This opcode performs an equivalence check. This instruction has the following effect:

Memory[SP-2] = (Memory[SP-2] == Memory[SP-1])

SP--

## GE\_OP

This opcode performs a greater than or equal to check. This instruction has the following effect:

Memory[SP-2] = (Memory[SP-2] >= Memory[SP-1])

SP--

## GT\_OP

This opcode performs a greater than check. This instruction has the following effect:

Memory[SP-2] = (Memory[SP-2] > Memory[SP-1])

SP--

## HALT\_OP

This opcode will cause the processor to stop executing instructions.

## ILLEGAL\_OP

This opcode will cause an illegal instruction exception and the simulator will exit.

## INP\_OP

This opcode will initiate an IO operation. The value at the top of the stack must be the address of an io\_blk. See the section on IO processing for a description of the IO blocks and their function.

## JMPUSER\_OP <dest>

This opcode causes the processor state to switch to user mode and then jump to an address in the user modes address space. This instruction has the following effect:

FL\_USER\_MODE bit is set in FLAG register

IP = <dest>

## JUMPE\_OP <dest>

This is the conditional jump instruction. It will jump if the value on the top of the stack is equal to zero. It has the following effect:

SP--

val = Memory[SP]

if (val == 0)

IP = <dest>

else

IP += 2

## JUMP\_OP <dest>

This is the jump instruction. It has the following effect:

IP = <dest>

## LAND\_OP

This opcode performs a logical (bitwise) AND operation. This instruction has the following effect:

Memory[SP-2] = (Memory[SP-2] & Memory[SP-1])

SP--

## LE\_OP

This opcode performs a less than or equal to check. This instruction has the following effect:

Memory[SP-2] = (Memory[SP-2] <= Memory[SP-1])

SP--

## LOR\_OP

This opcode performs a logical (bitwise) OR operation. This instruction has the following effect:

Memory[SP-2] = (Memory[SP-2] | Memory[SP-1])

SP--

## LT\_OP

This opcode performs a less than check. This instruction has the following effect:

Memory[SP-2] = (Memory[SP-2] < Memory[SP-1])

SP--

## LXOR\_OP

This opcode performs a logical (bitwise) XOR operation. This instruction has the following effect:

Memory[SP-2] = (Memory[SP-2] ^ Memory[SP-1])

SP--

## MINUS\_OP

This opcode performs a subtraction. This instruction has the following effect:

Memory[SP-2] = Memory[SP-2] - Memory[SP-1]

SP--

## MOD\_OP

This opcode performs a modulus. This instruction has the following effect:

Memory[SP-2] = Memory[SP-2] % Memory[SP-1]

SP--

## NEG\_OP

This opcode negates the value on the top of the stack. This instruction has the following effect:

Memory[SP-1] = - Memory[SP-1]

## NE\_OP

This opcode performs a not equal to check. This instruction has the following effect:

Memory[SP-2] = (Memory[SP-2] != Memory[SP-1])

SP--

## NOP

This opcode performs no operation.

## OR\_OP

This opcode performs a logical OR. This instruction has the following effect:

Memory[SP-2] = Memory[SP-2] || Memory[SP-1]

SP--

## OUTS\_OP

This opcode sends output to the console. The value at the top of the stack is assumed to be a the address of a NULL terminated string. The string at that address will be sent to the console. This is a blocking operation. The CPU will not execute another instruction until after the IO is complete.

## PLUS\_OP

This opcode performs an addition. This instruction has the following effect:

Memory[SP-2] = Memory[SP-2] + Memory[SP-1]

SP--

## POPCVARIND\_OP

This opcode pops a value off the stack and stores it in a byte-sized variable. The offset to the variable is taken from the stack. Note that a word-sized value is popped, but the destination is a single byte. To use this opcode, the value should be pushed on the stack and then the offset. The opcode has the following effect:

Memory[Memory[SP-1]] = Memory[SP-2]

## POPCVAR\_OP <offset>

This opcode pops a value off the stack and stores it in a byte variable. The offset to the variable is taken from the stack. Note that a word-sized value is popped, but the destination is a single byte.

The opcode has the following effect:

SP--

Memory[FP+offset] = Memory[SP]

## POPREG\_OP <reg num>

This opcode pops the value from the top of the stack into the specified register. The registers are numbered as follows:

0 BP register

1 LP register

2 IP register

3 SP register

4 FP register

5 FLAG register

## POPVAR\_OP <offset>

This opcode pops a value off the stack and stores it in the word indicated by FP+<offset>. The opcode has the following effect:

SP--

Memory[FP+offset] = Memory[SP]

## POP\_OP

This opcode pops the stack. The value that is popped in not saved. This opcode has the following effect:

SP--

## PUSHCVARIND\_OP

This opcode pushes a byte onto the stack. The offset to the byte comes from the stack. A single byte is pushed onto the stack but the SP is updated by a word-sized amount. The value of the unused three bytes is unspecified. To use this opcode, push the address of the value to be pushed onto the stack. This opcode has the following effect:

Memory[SP-1] = Memory[ Memory[SP-1] ]

SP -= 2

## PUSHCVAR\_OP <offset>

This opcode pushes a byte onto the stack. The byte is located at FP+<offset>. To preserve stack alignment, the SP is increased by a word-size amount. The value of the unused bytes is unspecified.

This opcode has the following effect:

Memory[SP] = Memory[FP+<offset>]

SP++

## PUSHFP\_OP

This opcode pushes the frame pointer onto the stack. This opcode has the following effect:

Memory[SP] = FP

SP++

## PUSHREG\_OP <reg num>

This opcode pushes the value of the specified register onto the stack. The registers are numbered as follows:

0 BP register

1 LP register

2 IP register

3 SP register

4 FP register

5 FLAG register

## PUSHVAR\_OP <offset>

This opcode pushes a word onto the stack. The word is located at FP+<offset>. This opcode has the following effect:

Memory[SP] = Memory[FP+<offset>]

SP++

## PUSH\_OP <value>

This opcode pushes a value onto the stack. This opcode has the following effect:

Memory[SP] = <value>

SP++

## RETURNV\_OP

This opcode is used to return from a function when the return function returns a value. The value to be returned is assumed to be at the top of the stack at the beginning of the instruction, and it is left on the top of the stack at the end of the function (after execution returns to the caller).

This opcode has the following effect:

temp = Memory[SP-1]

SP = FP-1

IP = Memory[FP-2]

FP = Memory[FP-1]

Memory[SP-1] = temp

## RETURN\_OP

This opcode is used to return from a void function. This opcode has the following effect:

SP = FP-1

IP = Memory[FP-2]

FP = Memory[FP-1]

## RTI\_OP

This opcode is used to return from interrupt mode. Both the interrupt and systrap functions should use this opcode to return.

## SETMODE\_OP <mode>

This opcode changes the FLAG register. The instruction has the following effect:

FLAG = <mode>

IP += 2

## SHIFTL\_OP

This opcode performs a bit shift left. This instruction has the following effect:

Memory[SP-2] = Memory[SP-2] << Memory[SP-1]

SP--

## SHIFTR\_OP

This opcode performs a bit shift right. This instruction has the following effect:

Memory[SP-2] = Memory[SP-2] >> Memory[SP-1]

SP--

## SWAP\_OP

This opcode performs a multiply. This instruction has the following effect:

Memory[SP-2] = Memory[SP-2] \* Memory[SP-1]

SP--

## TIMES\_OP

This opcode performs a multiply. This instruction has the following effect:

Memory[SP-2] = Memory[SP-2] \* Memory[SP-1]

SP--

## TRAPTOC\_OP

This opcode is reserved and should not be used.

## TRAP\_OP

This opcode performs a trap. It has the effect of calling the systrap() function. See the syscall section of this document for more details.

# stacklc grammar

The stacklc grammar is defined below:

program: global\_decls

block: open stmts close

| '{' '}'

open: '{'

close: '}'

decls: decls decl

| decl

decl: var\_decl ';'

| struct\_decl ';'

| UNSUPPORTED

| error ';'

var\_decl: TYPE\_ID IDENTIFIER

| CHAR '\*' IDENTIFIER

| CHAR IDENTIFIER arrayspec

struct\_decl: TYPEDEF STRUCT open decls close IDENTIFIER

global\_decls: global\_decls global\_decl

| global\_decl

global\_decl: func\_decl

| struct\_decl

| CONST TYPE\_ID IDENTIFIER '=' INT\_VAL ';'

| DEFINE IDENTIFIER INT\_VAL

| DEFINE IDENTIFIER '-' INT\_VAL

| UNSUPPORTED {

func\_decl: func\_header ';'

| func\_header '{' stmts '}'

| func\_header '{' '}'

func\_header: func\_prefix paramsspec ')'

| func\_prefix ')'

func\_prefix: TYPE\_ID IDENTIFIER '('

func\_prefix: CHAR '\*' IDENTIFIER '('

paramsspec: paramsspec',' paramspec

| paramspec

paramspec: var\_decl

arrayspec: '[' INT\_VAL ']'

stmts: stmts stmt

| stmt

stmt: decl

| ';'

| IF '(' ccomp ')' stmt ELSE stmt

| IF '(' ccomp ')' stmt

| FOR '(' assign ';' ccomp ';' assign ')' stmt

| WHILE '(' ccomp ')' stmt

| assign ';'

| func\_call ';'

| block

| RETURN expr ';'

| asm\_stmt ';'

assign: lval '=' expr

| lval '=' string\_lit

| lval '=' asm\_stmt

| lval PLUS\_EQ expr

| lval MINUS\_EQ expr

| lval TIMES\_EQ expr

| lval DIVIDE\_EQ expr

| lval INC

| INC lval

| lval DEC

| DEC lval

asm\_stmt : ASM '(' INT\_VAL ')'

| ASM '(' INT\_VAL ',' params ')'

func\_call: IDENTIFIER '(' params ')'

| IDENTIFIER '(' ')'

varref: varref '.' varpart

| varpart

varpart: IDENTIFIER arrayval

lval: varref

arrayval: arrayval '[' expr ']'

| /\* empty \*/

params: params',' param

| param

param: expr

| string\_lit

| '&' varref

ccomp: ccomp OR comp

| ccomp AND comp

| comp

comp: comp EQ expr

| comp NE expr

| comp GT expr

| comp LT expr

| comp LE expr

| comp GE expr

| expr

expr: expr '+' term

| expr '-' term

| term

term: term '\*' value

| term '/' value

| term '%' value

| value

value: fact

| '-' fact

fact: '(' ccomp ')'

| INT\_VAL

| varref

| func\_call

| SIZE\_OF '(' IDENTIFIER ')'

| SIZE\_OF '(' TYPE\_ID ')'

| SIZE\_OF '(' CHAR ')'

| SIZE\_OF IDENTIFIER

string\_lit: STRING\_LIT