CC64 Language Reference

Table of Contents

Overview

Language Representation
CC64 has more ways to represent constructs than standard C does. There are functionally redundant keywords (for instance ‘until’) that allow expressing meaning in a more variable manner. Hopefully, this is a human readble improvement over C.

CC64 does not require expressions to be surrounded with ‘(‘ and ‘)’ in all cases if it can be determined by the compiler what is part of the expression. It is recommended to use the brackets however. For instance the ‘if’ statement may be written as:

```
if a < 10 then dosomething();
```

CC64 supports an extended ‘C’ language compiler. CC64 is able to compile most C language programs with little or no modification required. In addition to the standard ‘C’ language CC64 adds the following:

run-time type identification (via typenum())
exception handling (via try/throw/catch)
function prolog / epilog control
multiple case constants eg. case ‘1’,’2’,’3’:
assembler code (asm)
pascal calling conventions (pascal)
no calling conventions (nocall / naked)
inline code
coroutines
additional loop constructs (until, loop, forever)
true/false are defined as 1 and 0 respectively
thread storage class
structure alignment control
firstcall blocks
block naming
branch prediction hints
bit slicing

Compiler Options

Option	Description
-fno-exceptions	This option tells the compiler not to generate code for processing exceptions. It results in smaller code, however the try/catch mechanism will no longer work.
-fpoll[n]	This option tells the compiler to generate code to poll for interrupts periodically.
-o[pxrcl]	This option disables optimizations done by the compiler causing really poor code to be generated. p – this disables the peephole optimization step x – this disables optimization of expressions (constants) r – this disables the allocation of register variables and common subexpression elimination. Also turns off constant optimizations. c – this disables optimizations done during code generation l – this disables loop invariant optimization -o by itself disables all optimizations done by the compiler
-os	Optimize for size. This will disable optimizations that increase code size such as loop inversions.
-w	This option disables wchar_t as a keyword. This keyword is sometimes #defined rather than being built into some compilers.
-S	generate assembly code with source code in comments.

The following additions have been made:

typenum(<type>)

allow run-time type identification. It returns a hash code for the type specified. It works the same way the sizeof() operator works, but it returns a code for the type, rather than the types size.

CC64 supports a simple try/throw/catch mechanism. A catch statement without a variable declaration catches all exceptions.

```
try { <statement> }
catch(var decl) {
}
catch(var decl)
{
}
catch {
}
```

Types:

A byte is one byte (8 bits) in size.
A char is two bytes (16 bits) in size.
An int is eight bytes (64 bits) wide.
An short int is four bytes (32 bits) wide
Pointers are eight bytes (64 bits) wide.

__attribute__

__attribute__ defines attributes associated with functions. Currently the only defined attribute is __no_temps which indicates to the compiler that the function does not use any temporary registers. This allows the compiler to omit code to save and restore temporaries around function calls. This is used primarily for functions defined in assembly language.

```
Example:
	extern signed byte KeybdGetStatus() __attribute__(__no_temps);
	extern byte KeybdGetScancode() __attribute__(__no_temps);
```

__check

__check causes the compiler to output a bounds checking instruction. The bounds expression must be of the format shown in the example.

```
Example:
	__check (hMbx; 0; 1024);
```

The first expression must be greater than or equal to the second expression and less than the third expression or a processor check interrupt will be invoked. Note any valid expressions may be used.

__leaf

In some cases, the compiler can not tell if a function is a leaf function, for instance in prototypes of external functions. The __leaf keyword indicates to the compiler that function is a leaf function and uses the link register to perform a return.

__mulf

__mulf causes the compiler to output a fast, single cycle multiply instruction. The fast multiply instruction is limited to 24 x 16 bits.

```
Example:
	ndx = __mulf (row, 56);
```

align()

The align keyword is used to specify structure alignment in memory. For example, the following structure will be aligned on 64-byte boundaries even though the structure itself is smaller in size.

struct my_struct align(64) {

```
    byte name[40];
}
```

Place the align keyword just before the opening brace of a structure or union declaration.

Note that specifying the structure alignment overrides the compiler's capability to automatically determine structure alignment. Care must be taken to specify a structure alignment that is at least the size of the structure.

Taking the size of a structure with an alignment specified returns the alignment.

and

'and' is defined as a keyword and is a synonym for '&&'. It can make code a little more readable.
Example:

```
if (a and b) {
}
```

'and' resolves to a Boolean value of true (1) or false (0).

asm [__leafs]

The asm keyboard allows assembler code to be placed in a 'C' function. The compiler does not process the block of assembler code, It simply copies it verbatim to the output. Global variables may be referenced by name by following the compiler convention of adding an '_' to the name. Stack arguments have to be specifically addressed referenced to the fp register. Register arguments can use the register directly.

```
pascal void SetRunningTCB(hTCB ht)
{
    asm {
        lw      tr,32[fp]      ; this references the ht variable
        asli    tr,tr,#10
        add     tr,tr,#_tcbs           ; this is a global variable reference
    }
}
```

The __leafs keyword indicates that the assembler code contains leafs (calls to other functions). Using the __leafs keyword causes the compiler to emit code to save and restore the subroutine linkage register.

```
// ------------------------------------------------------------------------
----
// Set an IRQ vector
```

```
// ------------------------------------------------------------------------
----

pascal void set_vector(unsigned int vecno, unsigned int rout)
{
     if (vecno > 255) return;
     if (rout == 0) return;
        asm __leafs {
                     lw            r2,32[fp]
                     lw            r1,40[fp]
                     call          set_vector
        }
}
```

case

Case statement may have more than one case constant specified by separating the constants with commas.

CC64:

```
switch (option) {
case 1,2,3,4:
        printf(“option 1-4);
case 5:
        printf(“option 5”);
}
```

Standard C:

```
switch (option) {
case 1:
case 2:
case 3:
case 4:
        printf(“option 1-4);
case 5:
        printf(“option 5”);
}
```

The compiler will make use of a jump table if there are enough cases and the density of the cases is greater than 33%. The table must be at least 33% populated. Otherwise, the compiler reverts to using a series of branches in a binary tree pattern. If there are just a few cases (<4) then a linear series of branch testing is used.

If the case options are powers of two, the compiler may generate more efficient code using BBS and BBC instructions. BBS means branch-on-bit-set.

catch (<type>)

The catch statement “catches” a specific type of object used for exception handling. A catch handler corresponds to object type used in the throw statement. If the thrown object is not of a type caught by a local catch handler, then a search for the correct catch handler will continue at a more outer level.
The type may be specified as an ellipsis which will then match any type, but the object thrown is not available to the catch handler. The type may also be omitted which will also match any type.
If none of the catches match the error object then catch statements at more outer levels are checked.
When exception processing is enabled the compiler always generates a default catch for each function. The default catch merely returns up the try/catch chain. The default catch will perform the same cleanup as would be activated by a return statement.

```
try {
  < some code>
}
catch (int erc) {
   < process error code>
}
catch {
   < catch anything else>
}
```

class

CC64 features a simple class keyword which may be used to implement classes. A class is very similar to a struct except that class methods may be declared to be part of the class. Classes in CC64 can have only single inheritance. Methods may be overloaded with different parameters.

coroutine(<expression>)

The coroutine keyword allows the implementation of coroutines using CC64. It accepts an expression which is a value to initialize the stack pointer to. Each coroutine must have its own stack.

```
int Astack[1024];
int Bstack[1024];

coroutine B();

int coroutine(Astack + 1024) A()
{
       forever {
              yield B();
       }
       return (0);
}
```

```
void coroutine(Bstack + 1024) B()
{
	forever {
		yield A();
	}
}

// One of the coroutines should be started with an ordinary function
// call.

void main()
{
	A();
}
```

delete

delete calls the run-time function __delete() to delete an object allocated by the new operator. __delete() takes a pointer to the object which was returned by new() and deallocates the object from the heap. If an object is deleted it is immediately deallocated and is not garbage collected. Delete does not call an object destructor. The object should be destroyed before using delete.

enum

The stride may be specified for the enumeration by following the enum keyword with the parenthesized stride value. The value must be a constant. If not specified the enumeration will increment by one.

In the following example the enumeration will decrement by 1 the value for each enumerated constant. so BADARG is equal to -1. This may be useful in cases where functions return a negative value indicating error or a positive value indicating proper operation.

```
enum(-1) {
        OKAY = 0,
        BADARG
}
```

The following will increment the enumeration value by 32.

```
enum (0x20) {
        ErrorClass0 = 0,
        ErrorClass1,
        ErrorClass2
}
```

This may be useful to define constants for bit-fields. The following example shows usage for a bit-field at bit position 7.

```
enum (0x80) {
        GFX_POINT = 0,
        GFX_LINE,    // will equal 0x080
        GFX_RECT     // will equal 0x100
}
```

enum may also indicate a powers-of enumeration as in:

```
enum (2^) { SelectA, SelectB, SelectC }
```

In which case SelectA would be equal to one, SelectB equal to two, and SelectC equal to four.

The powers-of enumeration causes the enumeration to multiply by the enumeration value for each step instead of adding. Different power series can be defined by setting one of the enumeration values to an arbitrary value. From that point on the values will multiply. So, “enum (2^) { A, B = 3, C }; will assign the values 1, 6, and 12 to A, B, C.

Enumeration values are only 16-bit.

epilog

The epilog keyword identifies a block of code to be executed as the function epilog code. An epilog block maybe placed anywhere in a function, but the compiler will output it at the function’s return point.

```
nocall myfunction()
{
        // other code
          epilog asm {
              // do some epilog work here, eg. setup return values
          }
}
```

firstcall

The firstcall keyword defines a statement that is to be executed only once the first time a function is called.

```
firstcall {
        printf("this prints the first time.");
}
```

The compiler automatically generates a static variable in the data segment that controls the firstcall block. The firstcall statement is equivalent to:

```
static char first=1;
if (first) {
   first = 0;
   <other statements>
}
```

The if statement may also precede the firstcall for readability.

forever

Forever is a loop construct that allows writing an unconditional loop. A forever loop should have an alternate means of exiting. For instance, a break statement will exit the loop.

```
forever {
        printf("this prints forever.");
}
```

inline

The inline keyword may be applied to a function declaration to cause the compiler to emit the function "inline" with other code. Every time the inline function is called, the code for the function is replicated inline. It is best to use register parameters to inline code.

if

If statements can accept a branch predictor hint. The hint must be a constant value determined at compile time. The syntax adds an options second expression ';' into the expression clause as shown below.

```
if (a < 10; 1)    // predict taken all the time
…
```

nocall / naked

The nocall or naked keyword causes the compiler to omit all the conventional stack operations required to call a function. (Omits function prologue and epilogue code) It's use is primarily to allow inline assembler code to handle function calling conventions instead of allowing the compiler to handle the calling convention. The naked keyword may also be applied to the switch() statement to cause the compiler to omit bounds checking on the switch.
A naked function also omits the default exception handler.

```
nocall myfunction()
{
        asm {
        }
}
```

new

The new operator generates a call to the run-time library function __new(). __new() will allocate storage for the object on the heap using malloc(). The object is given a 64-byte header and the object state is set to new. Objects whose state is new will not be deleted automatically by the garbage collector. They must be deleted using the delete keyword. new may be preceded by the keyword 'auto' which indicates that automatic garbage collection should be used for the object. In this case the objects state is set to auto new.
new() does not call the object's constructor.

or

'or' is defined as a keyword and is a synonym for '||'. It can make code a little more readable.
Example:

```
if (a or b) {
}
```

pascal

The pascal keyword causes the compiler to use the pascal calling convention rather than the usual C calling convention. For the pascal calling convention, function arguments are popped off the stack by the called routine. This may allow slightly faster and smaller code in some circumstances.

```
pascal char myfunction(int arg1, int arg2)
{
}
```

Note the default calling convention for the compiler is pascal.

prolog

The prolog keyword identifies a block of code to be executed as the function prolog. A prolog block may be placed anywhere in a function, but the compiler will output it at the function's entry point.

```
nocall myfunction()
{
        prolog asm {
            // do some prolog work here, eg. setup stack parameters
        }
}
```

Prolog code is not as highly optimized as other code as the prolog is generally placed before register variables are saved. That means it cannot make use of register variables.

register

The compiler will not use registers to pass arguments to functions unless specifically instructed to do so. The register keyword is used for this. The compiler automatically uses registers for temporaries and other variables where possible. Using the register keyword on anything other than an argument is likely to be ignored. Note the compiler allocates storage space on the stack for variables even if they are in registers. This storage space is often unused.

switch

The naked keyword may be applied to the switch() statement to cause the compiler to omit bounds checking. Normally the compiler will check the switch variable to ensure that it's within the range of the defined case values. With a naked switch the compiler assumes that the switch value is between the minimum and maximum case value in the switch statement. Naked switches result in faster code, but results are undefined if the switch is out of range. For a naked switch if the switch value isn't valid then the program will likely crash. So use with caution.

Regular switch:

```
;             switch(btn) {
              lw     r3,-32[bp]
              ldi    r4,#1
              ldi    r5,#9
              chk    r3,r4,r5,BIOSMain_13
              sub    r3,r3,#1
              shl    r3,r3,#3
              lw     r3,BIOSMain_19[r3]
              jal    r0,0[r3]
```

Naked Switch:

```
;             switch(btn; naked) {
              lw     r3,-32[bp]
              sub    r3,r3,#1
              shl    r3,r3,#3
              lw     r3,BIOSMain_19[r3]
              jal    r0,0[r3]
```

Note that if the minimum case value is zero then the code may omit the subtract of the minimum value making the switch slightly faster.

then

'Then' is defined as a keyword. It's only purpose is to make code more readable. It may be used with 'if' statements in which case it is ignored.

throw

Throw acts in a similar fashion to the return statement. Throw returns to the latest catch handler. The latest catch handler does not have to be defined in the current routine or a previous routine. Throwing an exception will walk backwards up the stack to the most recently defined catch handler. Unlike c++ throw does not automatically destroy objects created in the subroutine or method.

thread

The ‘thread’ keyword may be applied in variable declarations to indicate that a variable is thread-local. Thread local variables are treated like static declarations by the compiler, except that the variable’s storage is allocated in the thread-local-storage segment (tls).

thread int varname;

try { }

Try defines a try – catch block. A block of statements followed by a series of catch statements. Try causes the compiler to output code to point to the catch block of the try statement. This pointer will be used by subsequent throw statements.

When exception handling is enabled every function has a default exception handler that merely returns up to the next higher exception handler.

typenum()

Typenum() works like the sizeof() operator, but it returns a hashcode representing the type, rather than the size of the type. Typenum() can be used to identify types at run-time.

```
struct tag { int i; };

main()
{
        int n;

        n = typenum(struct tag);
}
```

The compiler numbers the types it encounters in a program, up to 10,000 types are supported. Pointers to types add 10,000 to the hash number for each level of pointer. Program should not depend on specific hash numbers as the hash numbers of type may vary between different versions of the compiler. The compiler uses a 15-bit hash number which may be encoded into a three character ascii string.

until

Until is a loop construct that allows writing a loop that continues until a condition is true. Until and while are almost the same except that until waits for the inverted condition. It may have better semantics in some circumstances to use until instead of inverting a while condition. while(!x) is better represented as until(x)

```
x = 0;
until (x==10) {
        printf("this prints 10 times.");
        x = x + 1;
}
```

using

The using keyword is used to activate features of CC64. In particular a name mangler used for classes can be activated by using the phrase ‘using name mangler;’ Without activating the name mangler all class methods have global scope.

name mangler

The name mangler generates unique names for methods so that there is no name clash for overloaded or derived methods. A hash string representing the type, method parameter and return value types is added to the method name.

__cdecl

Specifies that the default calling convention is the C calling convention

pascal

Specifies that the default calling convention is the Paacal calling convention.

real names

Disables name mangling.

yield <expression>

The yield keyword transfers control to the specified coroutine. The yield expression should point to a coroutine or results are undefined. The expression may include pointers to coroutines in addition to directly specifying the routine.

&&&

The &&& operator indicates to the compiler that a safe optimization is to generate code that executes both sides of the operator then uses an ‘and’ operation to determine the result. This may eliminate branches.

|||

The ||| operator indicates to the compiler that a safe optimization is to generate code that executes both sides of the operator then uses an ‘or’ operation to determine the result. This may eliminate branches.

??

The ?? ternary operator indicates to the compiler that a safe optimization is to generate code that executes both sides of the operator then select between the result. This may eliminate branches.

Character Constants

String constants may be pre-pended with one of ‘B’, ‘W’, ‘T’, or ‘O’ to indicate the size of encoded characters. ‘B’ = byte, ‘W’ = wyde (16-bit), ‘T’ = tetra(32-bit), and ‘O’ = octa (64-bit).

```
Example: The following string is encoded as bytes:
B”? = display help”
```

String constants may also be placed inline with code using the letter ‘I’ as a prefix. Inlined string constants are automatically 16-bit encoded.

Block Naming

The compiler supports named compound statement blocks. To name a compound statement follow the opening brace with a colon then the name.

```
void SomeFunc()
{
    while (x) {: x_name
      <other statements>
    }
}
```

An eventual goal for the compiler is to have the break statement be able to identify which block statement to break out of.

Calling Convention

CC64 uses the pascal calling convention by default. However, for functions with variable argument lists the variable portion of the list must be popped off the stack by the caller, as the called routine does not know how many variable arguments there are. The called routine will only pop the fixed portion of the argument list. The ’C’ calling convention may be specified using the keyword __cdecl.

Bit Slicing

The compiler allows bits slices of primitive types to be taken. Specific bits of a variable may be selected using the indexing operator[]. In the following example bits 26 to 31 of the variable ab are updated, then bits 6 to 10 is returned with 21 subtracted from the value.

```
int main(int a, int b, int c)
{
  int ab;
  ab[31:26] = a + b + c[15:5];
  return (ab[10:6]-21);
}
```

The indicated bits must specify the high bit followed by a colon then the low bit. Bits are numbered from 0 to 63.

Bit slicing aids the compiler in determining the use of bit field instructions that may be available on the processor. Bit manipulations can also be done with bitwise operators (~, &, |, ^, <<, and >>).

Array Handling Differences from ‘C’

The following is “in the works”. It may or may not work.

Arrays may be passed by value using the standard declaration of an array as a parameter. In ‘C’ arrays are always passed by reference.

In CC64:

```
SomeFn(int ary[50]) {
}
```

Declares a function that accepts an array of 50 integers passed by value. Declaring the function the same way in ‘C’ results in a reference to the array being passed to the function rather than the array values.

In order to pass an array by reference in CC64 the pointer indicator ‘*’ must be used as in the following:

```
SomeFn(int *ary) {
}
```

It is not recommended to pass large arrays or structures around in a program by value as program performance may be adversely affected. Passing aggregate types by value causes the compiler to output code to copy the values. The alternative, passing references around is significantly faster.

Exception handling

When exception handling is enabled, CC64 generates a default exception handler for each function. The action of the default handler is merely to return to the next higher exception handler. If an exception handler is not coded for the function then the default handler will be in effect. During function prolog, the address of any exception handler is stored in the stack frame at 16[$FP]. When a throw() operation occurs the address of the exception handler is loaded from 16[$FP] and jumped to. Throw() loads the exception type into register $a1 and the exception value into $a0. The catch handlers check the exception type in $a1 against the type of exception they handle. If there is no match, the next catch handler tests the value. If none of the catch handlers match the exception type then the default handler which returns up the exception chain is jumped to.

This causes an unhandled exception to unwind the stack just as a return would, then return to the caller's exception handler address rather than the normal return address.

To receive hardware exceptions the operating system must be notified of which exceptions are desired. A 256-bit bit mask is used to track which exceptions the application will respond to. This bitmap is stored in the applications ACB. When a selected exception occurs the OS forces the cause code into $a0, and the "exception" type into $a1 then causes the application to resume execution at the exception handler the next time the application is active.

Because of the simplicity of the exception handling mechanism objects created in the function are not automatically destroyed. That means it's necessary to keep track of which objects got created and destroy them in the catch handler.

Garbage Collection

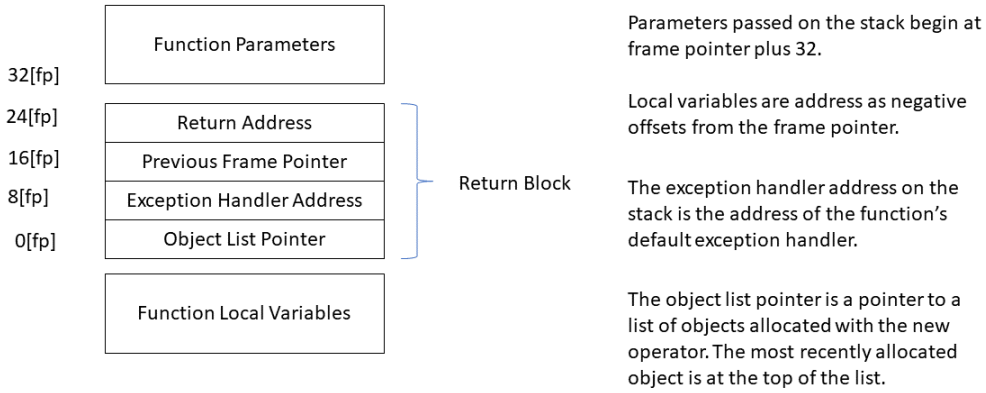
If a function or method uses the new operator, then a call is made to the run-time library function __AddGarbage() when the function returns. The __AddGarbage() function moves objects off the function's object list onto the garbage collector's list.

Limitations

The CC64 compiler has some built in limitations, most of the limits are far beyond what would be required for ordinary use:

The maximum number of function parameters: 20
The maximum number of symbols in a translation unit: 32,000.
The maximum number of functions in a translation unit: 3,000
The maximum number of different types allowed: 32,000
The maximum number of labels in a translation unit: 65,000.
The maximum number of dimensions that may be specified for an array is 19.
The maximum number of different expressions in a function: 450.
Number of basic blocks in a function: 10,000 max.

Return Block



- Deleting objects in the reverse order that they were allocated in will give the best performance, because then the top object on the object list can be deleted without having to search the object list.
- When the function exits any objects remaining on the object list are re-listed on the garbage collection list.