

IBM XL C/C++ for AIX, V12.1



Optimization and Programming Guide

Version 12.1

IBM XL C/C++ for AIX, V12.1



Optimization and Programming Guide

Version 12.1

Note

Before using this information and the product it supports, read the information in "Notices" on page 163.

First edition

This edition applies to IBM XL C/C++ for AIX, V12.1 (Program 5765-J02; 5725-C72) and to all subsequent releases and modifications until otherwise indicated in new editions. Make sure you are using the correct edition for the level of the product.

© Copyright IBM Corporation 1996, 2012.

US Government Users Restricted Rights – Use, duplication or disclosure restricted by GSA ADP Schedule Contract with IBM Corp.

Contents

About this information	vii
Who should read this information.	vii
How to use this information	vii
How this information is organized	vii
Conventions.	viii
Related information.	xi
IBM XL C/C++ information.	xii
Standards and specifications	xiii
Other IBM information	xiv
Other information	xiv
Technical support	xiv
How to send your comments	xiv

Chapter 1. Using 32-bit and 64-bit modes	1
Assigning long values	2
Assigning constant values to long variables	2
Bit-shifting long values	3
Assigning pointers	3
Aligning aggregate data	4
Calling Fortran code.	4

Chapter 2. Using XL C/C++ with Fortran	5
Identifiers	5
Corresponding data types	5
Character and aggregate data.	6
Function calls and parameter passing	7
Pointers to functions.	7
Sample program: C/C++ calling Fortran	7

Chapter 3. Aligning data.	9
Using alignment modes.	9
Alignment of aggregates	11
Alignment of bit fields.	14
Using alignment modifiers	15
Guidelines for determining alignment of scalar variables	17
Guidelines for determining alignment of aggregate variables	18

Chapter 4. Handling floating-point operations	19
Floating-point formats.	19
Handling multiply-add operations.	19
Compiling for strict IEEE conformance	20
Handling floating-point constant folding and rounding	20
Matching compile-time and runtime rounding modes	21
Rounding modes and standard library functions	22
Handling floating-point exceptions	23
Compiling a decimal floating-point program	23

Chapter 5. Using memory heaps.	25
---	-----------

Managing memory with multiple heaps	25
Functions for managing user-created heaps.	26
Creating a heap	27
Expanding a heap	28
Using a heap	29
Getting information about a heap	30
Closing and destroying a heap	30
Changing the default heap used in a program.	31
Compiling and linking a program with user-created heaps	31
Examples of creating and using user heaps.	31
Debugging memory heaps	36
Functions for checking memory heaps	37
Functions for debugging memory heaps.	38
Using memory allocation fill patterns.	39
Skipping heap checking	39
Using stack traces	40

Chapter 6. Using C++ constructors	41
Using delegating constructors (C++0x)	41

Chapter 7. Using C++ templates	43
Using the -qtempinc compiler option	44
Example of -qtempinc	45
Regenerating the template instantiation file.	47
Using -qtempinc with shared libraries	48
Using the -qtemplateregistry compiler option	48
Recompiling related compilation units	49
Switching from -qtempinc to -qtemplateregistry	49
Using explicit instantiation declarations (C++0x)	49

Chapter 8. Constructing a library	51
Compiling and linking a library	51
Compiling a static library.	51
Compiling a shared library	51
Exporting symbols with the CreateExportList utility	53
Linking a library to an application.	54
Linking a shared library to another shared library	54
Initializing static objects in libraries (C++)	54
Assigning priorities to objects	55
Order of object initialization across libraries	57
Dynamically loading a shared library.	59
Loading and initializing a module with the loadAndInit function	60
Terminating and unloading a module with the terminateAndUnload function	61

Chapter 9. Replacing operator new and operator delete in applications that use shared libraries (C++)	63
--	-----------

Chapter 10. Using the C++ utilities.	65
Demangling compiled C++ names.	65

Demangling compiled C++ names with <code>c++filt</code>	65
Demangling compiled C++ names with the demangle class library.	66
Creating a shared library with the <code>makeC++SharedLib</code> utility	68
Linking with the <code>linkxlc</code> utility.	70

Chapter 11. Optimizing your applications 71

Distinguishing between optimization and tuning.	71
Steps in the optimization process	72
Basic optimization	72
Optimizing at level 0	72
Optimizing at level 2	73
Advanced optimization	74
Optimizing at level 3	75
An intermediate step: adding <code>-qhot</code> suboptions at level 3	76
Optimizing at level 4	76
Optimizing at level 5	77
Tuning for your system architecture	78
Getting the most out of target machine options	79
Using high-order loop analysis and transformations	80
Getting the most out of <code>-qhot</code>	81
Using shared-memory parallelism (SMP)	82
Getting the most out of <code>-qsmp</code>	83
Using interprocedural analysis	83
Getting the most from <code>-qipa</code>	85
Using profile-directed feedback.	86
Viewing profiling information with <code>showpdf</code>	89
Object level profile-directed feedback.	92
Using compiler reports to diagnose optimization opportunities.	93
Parsing compiler reports with development tools	95
Other optimization options	95

Chapter 12. Debugging optimized code 99

Understanding different results in optimized programs.	100
Debugging in the presence of optimization	100
Using <code>-qoptdebug</code> to help debug optimized programs.	101

Chapter 13. Coding your application to improve performance. 105

Finding faster input/output techniques.	105
Reducing function-call overhead	105
Using delegating constructors (C++0x)	107
Using template explicit instantiation declarations (C++0x)	107
Managing memory efficiently	108
Optimizing variables	108
Manipulating strings efficiently	109
Optimizing expressions and program logic	110
Optimizing operations in 64-bit mode	110
Tracing functions in your code.	111
Using rvalue references (C++0x)	115

Chapter 14. Using the high performance libraries 119

Using the Mathematical Acceleration Subsystem libraries (MASS)	119
Using the scalar library	120
Using the vector libraries	122
Using the SIMD library for POWER7	128
Compiling and linking a program with MASS	131
Using the Basic Linear Algebra Subprograms – BLAS	132
BLAS function syntax	132
Linking the <code>libxlopt</code> library.	134

Chapter 15. Parallelizing your programs 135

Countable loops	136
Enabling automatic parallelization	137
Using IBM SMP directives (C only)	137
Using OpenMP directives	138
Shared and private variables in a parallel environment.	140
Reduction operations in parallelized loops.	141

Chapter 16. Selecting the standard allocation method to suit performance (C++) 143

Chapter 17. Ensuring thread safety (C++) 145

Ensuring thread safety of template objects.	145
Ensuring thread safety of stream objects	145

Chapter 18. Memory debug library functions 147

Memory allocation debug functions	147
<code>_debug_calloc</code> — Allocate and initialize memory	147
<code>_debug_free</code> — Free allocated memory	148
<code>_debug_heapmin</code> — Free unused memory in the default heap.	149
<code>_debug_malloc</code> — Allocate memory	150
<code>_debug_ucalloc</code> — Reserve and initialize memory from a user-created heap	151
<code>_debug_uheapmin</code> — Free unused memory in a user-created heap	152
<code>_debug_umalloc</code> — Reserve memory blocks from a user-created heap	153
<code>_debug_realloc</code> — Reallocate memory block	154
String handling debug functions	155
<code>_debug_memcpy</code> — Copy bytes	155
<code>_debug_memset</code> — Set bytes to value	156
<code>_debug_strcat</code> — Concatenate strings	157
<code>_debug_strcpy</code> — Copy strings	158
<code>_debug_strncat</code> — Concatenate strings	159
<code>_debug_strncpy</code> — Copy strings	160
<code>_debug_strnset</code> — Set characters in a string	161
<code>_debug_strset</code> — Set characters in a string.	161

Notices 163

Trademarks and service marks 165

Index 167

About this information

This guide discusses advanced topics related to the use of the IBM® XL C/C++ for AIX®, V12.1 compiler, with a particular focus on program portability and optimization. The guide provides both reference information and practical tips for getting the most out of the compiler's capabilities, through recommended programming practices and compilation procedures.

Who should read this information

This document is addressed to programmers building complex applications, who already have experience compiling with XL C/C++, and would like to take further advantage of the compiler's capabilities for program optimization and tuning, support for advanced programming language features, and add-on tools and utilities.

How to use this information

This document uses a "task-oriented" approach to presenting the topics, by concentrating on a specific programming or compilation problem in each section. Each topic contains extensive cross-references to the relevant sections of the reference guides in the IBM XL C/C++ for AIX, V12.1 documentation set, which provide detailed descriptions of compiler options and pragmas, and specific language extensions.

How this information is organized

This guide includes these topics:

- Chapter 1, "Using 32-bit and 64-bit modes," on page 1 discusses common problems that arise when porting existing 32-bit applications to 64-bit mode, and provides recommendations for avoiding these problems.
- Chapter 2, "Using XL C/C++ with Fortran," on page 5 discusses considerations for calling Fortran code from XL C/C++ programs.
- Chapter 3, "Aligning data," on page 9 discusses the different compiler options available for controlling the alignment of data in aggregates, such as structures and classes, on all platforms.
- Chapter 4, "Handling floating-point operations," on page 19 discusses options available for controlling the way floating-point operations are handled by the compiler.
- Chapter 5, "Using memory heaps," on page 25 discusses compiler library functions for heap memory management, including using custom memory heaps, and validating and debugging heap memory.
- Chapter 6, "Using C++ constructors," on page 41 discusses delegating constructors that can concentrate common initializations in one constructor.
- Chapter 7, "Using C++ templates," on page 43 discusses the different options for compiling programs that include C++ templates.
- Chapter 8, "Constructing a library," on page 51 discusses how to compile and link static and shared libraries, and how to specify the initialization order of static objects in C++ programs.

- Chapter 9, “Replacing operator new and operator delete in applications that use shared libraries (C++),” on page 63 discusses how to use a user defined operator new() and operator delete() for the shared libraries.
- Chapter 10, “Using the C++ utilities,” on page 65 discusses some additional utilities shipped with XL C/C++, for demangling compiled symbol names, creating shared libraries, and linking C++ modules.
- Chapter 11, “Optimizing your applications,” on page 71 discusses the various options provided by the compiler for optimizing your programs, and provides recommendations for use of the different options.
- Chapter 12, “Debugging optimized code,” on page 99 discusses the potential usability problems of the optimized programs and the options that can be used to debug the optimized code.
- Chapter 13, “Coding your application to improve performance,” on page 105 discusses recommended programming practices and coding techniques for enhancing program performance and compatibility with the compiler’s optimization capabilities.
- Chapter 14, “Using the high performance libraries,” on page 119 discusses two performance libraries that are shipped with XL C/C++: the Mathematical Acceleration Subsystem (MASS), which contains tuned versions of standard math library functions; and the Basic Linear Algebra Subprograms (BLAS), which contains basic functions for matrix multiplication.
- Chapter 15, “Parallelizing your programs,” on page 135 provides an overview of the different options offered by the XL C/C++ for creating multi-threaded programs, including IBM SMP and OpenMP language constructs.
- Chapter 16, “Selecting the standard allocation method to suit performance (C++),” on page 143 discusses how to select the allocation method used by the standard allocator of the C++ Standard Library to suit the performance needs of the application.
- Chapter 17, “Ensuring thread safety (C++),” on page 145 discusses thread-safety issues related to C++ class libraries, including input/output streams, and standard templates.
- Chapter 18, “Memory debug library functions,” on page 147 provides a reference listing and examples of all compiler debug memory library functions.

Conventions

Typographical conventions

The following table explains the typographical conventions used in the IBM XL C/C++ for AIX, V12.1 information.

Table 1. Typographical conventions

Typeface	Indicates	Example
bold	Lowercase commands, executable names, compiler options, and directives.	The compiler provides basic invocation commands, xlc and xlc (xlc++), along with several other compiler invocation commands to support various C/C++ language levels and compilation environments.
<i>italics</i>	Parameters or variables whose actual names or values are to be supplied by the user. Italics are also used to introduce new terms.	Make sure that you update the <i>size</i> parameter if you return more than the <i>size</i> requested.











Table 1. Typographical conventions (continued)

Typeface	Indicates	Example
<u>underlining</u>	The default setting of a parameter of a compiler option or directive.	nomaf <u>maf</u>
monospace	Programming keywords and library functions, compiler builtins, examples of program code, command strings, or user-defined names.	To compile and optimize myprogram.c, enter: xlc myprogram.c -O3.

Qualifying elements (icons)

Most features described in this information apply to both C and C++ languages. In descriptions of language elements where a feature is exclusive to one language, or where functionality differs between languages, this information uses icons to delineate segments of text as follows:

Table 2. Qualifying elements

Qualifier/Icon	Meaning
C only, or C only begins   C only ends	The text describes a feature that is supported in the C language only; or describes behavior that is specific to the C language.
C++ only, or C++ only begins   C++ only ends	The text describes a feature that is supported in the C++ language only; or describes behavior that is specific to the C++ language.
IBM extension begins   IBM extension ends	The text describes a feature that is an IBM extension to the standard language specifications.
C1X, or C1X begins   C1X ends	The text describes a feature that is introduced into standard C as part of C1X.
C++0x, or C++0x begins   C++0x ends	The text describes a feature that is introduced into standard C++ as part of C++0x.

Syntax diagrams

Throughout this information, diagrams illustrate XL C/C++ syntax. This section will help you to interpret and use those diagrams.

- Read the syntax diagrams from left to right, from top to bottom, following the path of the line.

The \blacktriangleright — symbol indicates the beginning of a command, directive, or statement.

The — \blacktriangleright symbol indicates that the command, directive, or statement syntax is continued on the next line.

The \blacktriangleright — symbol indicates that a command, directive, or statement is continued from the previous line.

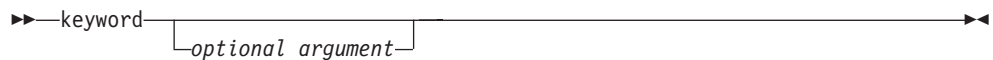
The — \blacktriangleleft symbol indicates the end of a command, directive, or statement.

Fragments, which are diagrams of syntactical units other than complete commands, directives, or statements, start with the \mid — symbol and end with the — \mid symbol.

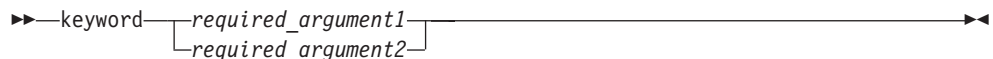
- Required items are shown on the horizontal line (the main path):



- Optional items are shown below the main path:



- If you can choose from two or more items, they are shown vertically, in a stack. If you *must* choose one of the items, one item of the stack is shown on the main path.



If choosing one of the items is optional, the entire stack is shown below the main path.



- An arrow returning to the left above the main line (a repeat arrow) indicates that you can make more than one choice from the stacked items or repeat an item. The separator character, if it is other than a blank, is also indicated:



- The item that is the default is shown above the main path.

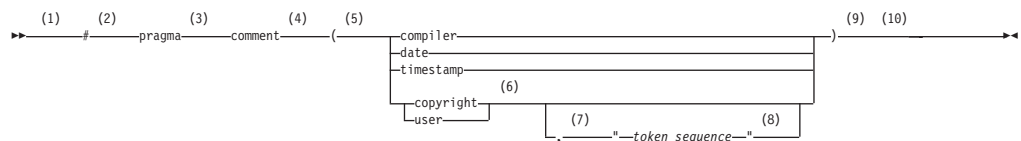


- Keywords are shown in nonitalic letters and should be entered exactly as shown.

- Variables are shown in italicized lowercase letters. They represent user-supplied names or values.
- If punctuation marks, parentheses, arithmetic operators, or other such symbols are shown, you must enter them as part of the syntax.

Sample syntax diagram

The following syntax diagram example shows the syntax for the **#pragma comment** directive.



Notes:

- 1 This is the start of the syntax diagram.
- 2 The symbol # must appear first.
- 3 The keyword pragma must appear following the # symbol.
- 4 The name of the pragma comment must appear following the keyword pragma.
- 5 An opening parenthesis must be present.
- 6 The comment type must be entered only as one of the types indicated: compiler, date, timestamp, copyright, or user.
- 7 A comma must appear between the comment type copyright or user, and an optional character string.
- 8 A character string must follow the comma. The character string must be enclosed in double quotation marks.
- 9 A closing parenthesis is required.
- 10 This is the end of the syntax diagram.

The following examples of the **#pragma comment** directive are syntactically correct according to the diagram shown above:

```
#pragma comment(date)
#pragma comment(user)
#pragma comment(copyright,"This text will appear in the module")
```

Examples in this information

The examples in this information, except where otherwise noted, are coded in a simple style that does not try to conserve storage, check for errors, achieve fast performance, or demonstrate all possible methods to achieve a specific result.

The examples for installation information are labelled as either *Example* or *Basic example*. *Basic examples* are intended to document a procedure as it would be performed during a basic, or default, installation; these need little or no modification.

Related information

The following sections provide related information for XL C/C++:

IBM XL C/C++ information

XL C/C++ provides product information in the following formats:

- README files
 README files contain late-breaking information, including changes and corrections to the product information. README files are located by default in the XL C/C++ directory and in the root directory of the installation CD.
- Installable man pages
 Man pages are provided for the compiler invocations and all command-line utilities provided with the product. Instructions for installing and accessing the man pages are provided in the *IBM XL C/C++ for AIX, V12.1 Installation Guide*.
- Information center
 The information center of searchable HTML files can be launched on a network and accessed remotely or locally. Instructions for installing and accessing the online information center are provided in the *IBM XL C/C++ for AIX, V12.1 Installation Guide*.
 The information center is viewable on the web at <http://publib.boulder.ibm.com/infocenter/comphelp/v121v141/index.jsp>.
- PDF documents
 PDF documents are located by default in the `/usr/vacpp/doc/LANG/pdf/` directory, where *LANG* is one of `en_US`, `zh_CN`, or `ja_JP`. The PDF files are also available on the web at <http://www.ibm.com/software/awdtools/xlcpp/aix/library/>.

The following files comprise the full set of XL C/C++ product information:

Table 3. XL C/C++ PDF files

Document title	PDF file name	Description
<i>IBM XL C/C++ for AIX, V12.1 Installation Guide, GC14-7329-00</i>	install.pdf	Contains information for installing XL C/C++ and configuring your environment for basic compilation and program execution.
<i>Getting Started with IBM XL C/C++ for AIX, V12.1, SC14-7328-00</i>	getstart.pdf	Contains an introduction to the XL C/C++ product, with information on setting up and configuring your environment, compiling and linking programs, and troubleshooting compilation errors.
<i>IBM XL C/C++ for AIX, V12.1 Compiler Reference, SC14-7330-00</i>	compiler.pdf	Contains information about the various compiler options, pragmas, macros, environment variables, and built-in functions, including those used for parallel processing.
<i>IBM XL C/C++ for AIX, V12.1 Language Reference, SC14-7331-00</i>	langref.pdf	Contains information about the C and C++ programming languages, as supported by IBM, including language extensions for portability and conformance to nonproprietary standards.
<i>IBM XL C/C++ for AIX, V12.1 Optimization and Programming Guide, SC14-7332-00</i>	progguide.pdf	Contains information on advanced programming topics, such as application porting, interlanguage calls with Fortran code, library development, application optimization and parallelization, and the XL C/C++ high-performance libraries.
<i>Standard C++ Library Reference, SC14-7333-00</i>	stdlib.pdf	Contains reference information about the standard C++ runtime libraries and headers.

Table 3. XL C/C++ PDF files (continued)

Document title	PDF file name	Description
C/C++ Legacy Class Libraries Reference, SC09-7652-00	legacy.pdf	Contains reference information about the USL I/O Stream Library and the Complex Mathematics Library.

To read a PDF file, use the Adobe Reader. If you do not have the Adobe Reader, you can download it (subject to license terms) from the Adobe website at <http://www.adobe.com>.

More information related to XL C/C++ including IBM Redbooks® publications, white papers, tutorials, and other articles, is available on the web at:

<http://www.ibm.com/software/awdtools/xlcpp/aix/library/>

For more information about boosting performance, productivity, and portability, see the C/C++ café at <http://www.ibm.com/software/rational/cafe/community/ccpp>.

Standards and specifications

XL C/C++ is designed to support the following standards and specifications. You can refer to these standards for precise definitions of some of the features found in this information.

- *Information Technology - Programming languages - C, ISO/IEC 9899:1990*, also known as C89.
- *Information Technology - Programming languages - C, ISO/IEC 9899:1999*, also known as C99.
- *Information Technology - Programming languages - C++, ISO/IEC 14882:1998*, also known as C++98.
- *Information Technology - Programming languages - C++, ISO/IEC 14882:2003(E)*, also known as *Standard C++*.
- *Information Technology - Programming languages - Extensions for the programming language C to support new character data types, ISO/IEC DTR 19769*. This draft technical report has been accepted by the C standards committee, and is available at <http://www.open-std.org/JTC1/SC22/WG14/www/docs/n1040.pdf>.
- *Draft Technical Report on C++ Library Extensions, ISO/IEC DTR 19768*. This draft technical report has been submitted to the C++ standards committee, and is available at <http://www.open-std.org/JTC1/SC22/WG21/docs/papers/2005/n1836.pdf>.
- *AltiVec Technology Programming Interface Manual, Motorola Inc*. This specification for vector data types, to support vector processing technology, is available at http://www.freescale.com/files/32bit/doc/ref_manual/ALTIVECPIM.pdf.
- *Information Technology - Programming Languages - Extension for the programming language C to support decimal floating-point arithmetic, ISO/IEC WDTR 24732*. This draft technical report has been submitted to the C standards committee, and is available at <http://www.open-std.org/JTC1/SC22/WG14/www/docs/n1176.pdf>.
- *Decimal Types for C++: Draft 4* <http://www.open-std.org/jtc1/sc22/wg21/docs/papers/2006/n1977.html>
- *ANSI/IEEE Standard for Binary Floating-Point Arithmetic, ANSI/IEEE Std 754-1985*.

- *OpenMP Application Program Interface Version 3.1*, available at <http://www.openmp.org>

Other IBM information

- *Parallel Environment for AIX: Operation and Use*
- The IBM Systems Information Center, at <http://publib.boulder.ibm.com/infocenter/systems/index.jsp?topic=/com.ibm.aix.doc/doc/base/aixparent.htm> is a resource for AIX information.

You can find the following books for your specific AIX system:

- *AIX Commands Reference, Volumes 1 - 6*
- *Technical Reference: Base Operating System and Extensions, Volumes 1 & 2*
- *AIX National Language Support Guide and Reference*
- *AIX General Programming Concepts: Writing and Debugging Programs*
- *AIX Assembler Language Reference*
- *ESSL for AIX V5.1/ESSL for Linux on POWER V5.1 Guide and Reference* available at the Engineering and Scientific Subroutine Library (ESSL) and Parallel ESSL web page.

Other information

- *Using the GNU Compiler Collection* available at <http://gcc.gnu.org/onlinedocs>

Technical support

Additional technical support is available from the XL C/C++ Support page at <http://www.ibm.com/software/awdtools/xlcpp/aix/support/>. This page provides a portal with search capabilities to a large selection of Technotes and other support information.

If you cannot find what you need, you can send email to compinfo@ca.ibm.com.

For the latest information about XL C/C++, visit the product information site at <http://www.ibm.com/software/awdtools/xlcpp/aix/>.

How to send your comments

Your feedback is important in helping to provide accurate and high-quality information. If you have any comments about this information or any other XL C/C++ information, send your comments by email to compinfo@ca.ibm.com.

Be sure to include the name of the information, the part number of the information, the version of XL C/C++, and, if applicable, the specific location of the text you are commenting on (for example, a page number or table number).

Chapter 1. Using 32-bit and 64-bit modes

You can use the XL C/C++ compiler to develop either 32-bit or 64-bit applications. To do so, specify **-q32** (the default) or **-q64**, respectively, during compilation. Alternatively, you can set the *OBJECT_MODE* environment variable to 32 or 64 at compile time. If both *OBJECT_MODE* and **-q32/-q64** are specified, **-q32/-q64** takes precedence.

However, porting existing applications from 32-bit to 64-bit mode can lead to a number of problems, mostly related to the differences in C/C++ long and pointer data type sizes and alignment between the two modes. The following table summarizes these differences.

Table 4. Size and alignment of data types in 32-bit and 64-bit modes

Data type	32-bit mode		64-bit mode	
	Size	Alignment	Size	Alignment
long, signed long, unsigned long	4 bytes	4-byte boundaries	8 bytes	8-byte boundaries
pointer	4 bytes	4-byte boundaries	8 bytes	8-byte boundaries
size_t (defined in the header file <stddef>)	4 bytes	4-byte boundaries	8 bytes	8-byte boundaries
ptrdiff_t (defined in the header file <stddef>)	4 bytes	4-byte boundaries	8 bytes	8-byte boundaries

The following sections discuss some of the common pitfalls implied by these differences, as well as recommended programming practices to help you avoid most of these issues:

- “Assigning long values” on page 2
- “Assigning pointers” on page 3
- “Aligning aggregate data” on page 4
- “Calling Fortran code” on page 4

When compiling in 32-bit or 64-bit mode, you can use the **-qwarn64** option to help diagnose some issues related to porting applications. In either mode, the compiler immediately issues a warning if undesirable results, such as truncation or data loss, will occur when the program is executed.

For suggestions on improving performance in 64-bit mode, see “Optimize operations in 64-bit mode”.

Related information in the *XL C/C++ Compiler Reference*



-q32, -q64



-qwarn64



Compile-time and link-time environment variables

Assigning long values

The limits of long type integers defined in the `limits.h` standard library header file are different in 32-bit and 64-bit modes, as shown in the following table.

Table 5. Constant limits of long integers in 32-bit and 64-bit modes

Symbolic constant	Mode	Value	Hexadecimal	Decimal
LONG_MIN (smallest signed long)	32-bit	$-(2^{31})$	0x80000000L	-2,147,483,648
	64-bit	$-(2^{63})$	0x8000000000000000L	-9,223,372,036,854,775,808
LONG_MAX (largest signed long)	32-bit	$2^{31}-1$	0x7FFFFFFFL	+2,147,483,647
	64-bit	$2^{63}-1$	0x7FFFFFFFFFFFFFFFL	+9,223,372,036,854,775,807
ULONG_MAX (largest unsigned long)	32-bit	$2^{32}-1$	0xFFFFFFFFUL	+4,294,967,295
	64-bit	$2^{64}-1$	0xFFFFFFFFFFFFFFFFUL	+18,446,744,073,709,551,615

Implications of these differences are:

- Assigning a long value to a double variable can cause loss of accuracy.
- Assigning constant values to long variables can lead to unexpected results. This issue is explored in more detail in “Assigning constant values to long variables.”
- Bit-shifting long values will produce different results, as described in “Bit-shifting long values” on page 3.
- Using `int` and long types interchangeably in expressions will lead to implicit conversion through promotions, demotions, assignments, and argument passing, and can result in truncation of significant digits, sign shifting, or unexpected results, without warning. These operations can impact performance.

In situations where a long value can overflow when assigned to other variables or passed to functions, you must:

- Avoid implicit type conversion by using explicit type casting to change types.
- Ensure that all functions that return long types are properly prototyped.
- Ensure that long parameters can be accepted by the functions to which they are being passed.

Assigning constant values to long variables

Although type identification of constants follows explicit rules in C and C++, many programs use hexadecimal or unsuffixed constants as “typeless” variables and rely on a two’s complement representation to exceed the limits permitted on a 32-bit system. As these large values are likely to be extended into a 64-bit long type in 64-bit mode, unexpected results can occur, generally at boundary areas such as:

- `constant >= UINT_MAX`
- `constant < INT_MIN`
- `constant > INT_MAX`

Some examples of unexpected boundary side effects are listed in the following table.

Table 6. Unexpected boundary results of constants assigned to long types

Constant assigned to long	Equivalent value	32 bit mode	64 bit mode
-2,147,483,649	INT_MIN-1	+2,147,483,647	-2,147,483,649
+2,147,483,648	INT_MAX+1	-2,147,483,648	+2,147,483,648
+4,294,967,726	UINT_MAX+1	0	+4,294,967,296
0xFFFFFFFF	UINT_MAX	-1	+4,294,967,295
0x100000000	UINT_MAX+1	0	+4,294,967,296
0xFFFFFFFFFFFFFFFF	ULONG_MAX	-1	-1

Unsuffixes constants can lead to type ambiguities that can affect other parts of your program, such as when the results of `sizeof` operations are assigned to variables. For example, in 32-bit mode, the compiler types a number like 4294967295 (`UINT_MAX`) as an unsigned long and `sizeof` returns 4 bytes. In 64-bit mode, this same number becomes a signed long and `sizeof` returns 8 bytes. Similar problems occur when passing constants directly to functions.

You can avoid these problems by using the suffixes `L` (for long constants), `UL` (for unsigned long constants), `LL` (for long long constants), or `ULL` (for unsigned long long constants) to explicitly type all constants that have the potential of affecting assignment or expression evaluation in other parts of your program. In the example cited in the preceding paragraph, suffixing the number as `4294967295U` forces the compiler to always recognize the constant as an unsigned int in 32-bit or 64-bit mode. These suffixes can also be applied to hexadecimal constants.

Bit-shifting long values

Left-bit-shifting long values produces different results in 32-bit and 64-bit modes. The examples in Table 7 show the effects of performing a bit-shift on long constants, using the following code segment:

```
long l=valueL<<1;
```

Table 7. Results of bit-shifting long values

Initial value	Symbolic constant	Value after bit shift by one bit	
		32-bit mode	64-bit mode
0x7FFFFFFFL	INT_MAX	0xFFFFFFFFE	0x00000000FFFFFFFFE
0x80000000L	INT_MIN	0x00000000	0x0000000100000000
0xFFFFFFFFL	UINT_MAX	0xFFFFFFFFE	0x00000001FFFFFFFFE

In 32-bit mode, `0xFFFFFFFFE` is negative. In 64 bit mode, `0x00000000FFFFFFFFE` and `0x00000001FFFFFFFFE` are both positive.

Assigning pointers

In 64-bit mode, pointers and `int` types are no longer the same size. The implications of this are:

- Exchanging pointers and `int` types causes segmentation faults.
- Passing pointers to a function expecting an `int` type results in truncation.
- Functions that return a pointer, but are not explicitly prototyped as such, return an `int` instead and truncate the resulting pointer, as illustrated in the following example.

To avoid these types of problems:

- Prototype any functions that return a pointer, where possible by using the appropriate header file.
- Be sure that the type of parameter you are passing in a function (pointer or int) call matches the type expected by the function being called.
- For applications that treat pointers as an integer type, use type long or unsigned long in either 32-bit or 64-bit mode.

Aligning aggregate data

Normally, structures are aligned according to the most strictly aligned member in both 32-bit and 64-bit modes. However, since long types and pointers change size and alignment in 64-bit, the alignment of a structure's strictest member can change, resulting in changes to the alignment of the structure itself.

Structures that contain pointers or long types cannot be shared between 32-bit and 64-bit applications. Unions that attempt to share long and int types, or overlay pointers onto int types can change the alignment. In general, you need to check all but the simplest structures for alignment and size dependencies.

In 64-bit mode, member values in a structure passed by value to a va_arg argument might not be accessed properly if the size of the structure is not a multiple of 8-bytes.

For detailed information on aligning data structures, including structures that contain bit fields, see Chapter 3, "Aligning data," on page 9.

Calling Fortran code

A significant number of applications use C, C++, and Fortran together, by calling each other or sharing files. It is currently easier to modify data sizes and types on the C side than on the Fortran side of such applications. The following table lists C and C++ types and the equivalent Fortran types in the different modes.

Table 8. Equivalent C/C++ and Fortran data types

C/C++ type	Fortran type	
	32-bit	64-bit
signed int	INTEGER	INTEGER
signed long	INTEGER	INTEGER*8
unsigned long	LOGICAL	LOGICAL*8
pointer	INTEGER	INTEGER*8
		integer POINTER (8 bytes)

Related information:

Chapter 2, "Using XL C/C++ with Fortran," on page 5

Chapter 2. Using XL C/C++ with Fortran

With XL C/C++, you can call functions written in Fortran from your C and C++ programs. This section discusses some programming considerations for calling Fortran code in the following areas:

- “Identifiers”
- “Corresponding data types”
- “Character and aggregate data” on page 6
- “Function calls and parameter passing” on page 7
- “Pointers to functions” on page 7
- “Sample program: C/C++ calling Fortran” on page 7 provides an example of a C program which calls a Fortran subroutine.

Related information:

“Calling Fortran code” on page 4

Identifiers

C++ functions callable from Fortran should be declared with extern "C" to avoid name mangling. For details, see the appropriate section about options and conventions for mixing Fortran with C/C++ code in the Fortran optimization and programming guide.

You need to follow these recommendations when writing C and C++ code to call functions written in Fortran:

- Avoid using uppercase letters in identifiers. Although XL Fortran folds external identifiers to lowercase by default, the Fortran compiler can be set to distinguish external names by case.
- Avoid using long identifier names. The maximum number of significant characters in XL Fortran identifiers is 250¹.

Note:

1. The Fortran 90 and 95 language standards require identifiers to be no more than 31 characters; the Fortran 2003 standard requires identifiers to be no more than 63 characters.
-

Corresponding data types

The following table shows the correspondence between the data types available in C/C++ and Fortran. Several data types in C have no equivalent representation in Fortran. Do not use them when programming for interlanguage calls.

Table 9. Correspondence of data types among C, C++ and Fortran

C and C++ data types	Fortran data types
bool (C++)_Bool (C)	LOGICAL(1)
char	CHARACTER
signed char	INTEGER*1
unsigned char	LOGICAL*1
signed short int	INTEGER*2

Table 9. Correspondence of data types among C, C++ and Fortran (continued)

C and C++ data types	Fortran data types
unsigned short int	LOGICAL*2
signed long int	INTEGER*4
unsigned long int	LOGICAL*4
signed long long int	INTEGER*8
unsigned long long int	LOGICAL*8
float	REAL REAL*4
double	REAL*8 DOUBLE PRECISION
long double (default)	REAL*8 DOUBLE PRECISION
long double (with -qlongdouble or -qdbl128)	REAL*16
float _Complex	COMPLEX*8 or COMPLEX(4)
double _Complex	COMPLEX*16 or COMPLEX(8)
long double _Complex (default)	COMPLEX*16 or COMPLEX(8)
long double _Complex(with -qlongdouble or -qdbl128)	COMPLEX*32 or COMPLEX(16)
structure or union	derived type
enumeration	INTEGER*4
char[n]	CHARACTER*n
array pointer to type, or type []	Dimensioned variable (transposed)
pointer to function	Functional parameter
structure (with -qalign=packed)	Sequence derived type

Related information in the *XL C/C++ Compiler Reference*



-qdbl128, -qlongdouble



-qalign

Character and aggregate data

Most numeric data types have counterparts across C/C++ and Fortran. However, character and aggregate data types require special treatment:

- C character strings are delimited by a '\0' character. In Fortran, all character variables and expressions have a length that is determined at compile time. Whenever Fortran passes a string argument to another routine, it appends a hidden argument that provides the length of the string argument. This length argument must be explicitly declared in C. The C code should not assume a null terminator; the supplied or declared length should always be used.
- An n-element C/C++ array is indexed with 0...n-1, whereas an n-element Fortran array is typically indexed with 1...n. In addition, Fortran supports user-specified bounds while C/C++ does not.
- C stores array elements in row-major order (array elements in the same row occupy adjacent memory locations). Fortran stores array elements in ascending storage units in column-major order (array elements in the same column occupy adjacent memory locations). Table 10 on page 7 shows how a two-dimensional array declared by A[3][2] in C and by A(3,2) in Fortran, is stored:

Table 10. Storage of a two-dimensional array

Storage unit	C and C++ element name	Fortran element name
Lowest	A[0][0]	A(1,1)
	A[0][1]	A(2,1)
	A[1][0]	A(3,1)
	A[1][1]	A(1,2)
Highest	A[2][0]	A(2,2)
	A[2][1]	A(3,2)

- In general, for a multidimensional array, if you list the elements of the array in the order they are laid out in memory, a row-major array will be such that the rightmost index varies fastest, while a column-major array will be such that the leftmost index varies fastest.

Function calls and parameter passing

Functions must be prototyped identically in both C/C++ and Fortran.

In C, by default, all function arguments are passed by value, and the called function receives a copy of the value passed to it. In Fortran, by default, arguments are passed by reference, and the called function receives the address of the value passed to it. You can use the Fortran %VAL built-in function or the VALUE attribute to pass by value. Refer to the *XL Fortran Language Reference* for more information.

For call-by-reference (as in Fortran), the address of the parameter is passed in a register. When passing parameters by reference, if you write C or C++ functions that call a program written in Fortran, all arguments must be pointers, or scalars with the address operator.

For more information about interlanguage calls to functions or routines, see "Interlanguage calls" in the *XL Fortran Optimization and programming guide*.

Pointers to functions

A function pointer is a data type whose value is a function address. In Fortran, a dummy argument that appears in an EXTERNAL statement is a function pointer. Function pointers are supported in contexts such as the target of a call statement or an actual argument of such a statement.

Sample program: C/C++ calling Fortran

The following example illustrates how program units written in different languages can be combined to create a single program. It also demonstrates parameter passing between C/C++ and Fortran subroutines with different data types as arguments.

```
#include <stdio.h>
extern double add(int *, double [], int *, double []);

double ar1[4]={1.0, 2.0, 3.0, 4.0};
double ar2[4]={5.0, 6.0, 7.0, 8.0};

main()
{
```

```

int x, y;
double z;

x = 3;
y = 3;

z = add(&x, ar1, &y, ar2); /* Call Fortran add routine */
/* Note: Fortran indexes arrays 1..n */
/* C indexes arrays 0..(n-1) */

printf("The sum of %1.0f and %1.0f is %2.0f \n",
ar1[x-1], ar2[y-1], z);
}

```

The Fortran subroutine is:

C Fortran function add.f - for C/C++ interlanguage call example

C Compile separately, then link to C/C++ program

```

REAL*8 FUNCTION ADD (A, B, C, D)
REAL*8 B,D
INTEGER*4 A,C
DIMENSION B(4), D(4)
ADD = B(A) + D(C)
RETURN
END

```

Chapter 3. Aligning data

XL C/C++ provides many mechanisms for specifying data alignment at the levels of individual variables, members of aggregates, entire aggregates, and entire compilation units. If you are porting applications between different platforms, or between 32-bit and 64-bit modes, you need to take into account the differences between alignment settings available in the different environments, to prevent possible data corruption and deterioration in performance. In particular, vector types have special alignment requirements which, if not followed, can produce incorrect results. For more information, see the *AltiVec Technology Programming Interface Manual*.

XLC provides alignment modes and alignment modifiers for specifying data alignment. Using alignment modes, you can set alignment defaults for all data types for a compilation unit (or subsection of a compilation unit), by specifying a predefined suboption.

Using alignment modifiers, you can set the alignment for specific variables or data types within a compilation unit, by specifying the exact number of bytes that should be used for the alignment.


“Using alignment modes” discusses the default alignment modes for all data types on the different platforms and addressing models; the suboptions and pragmas you can use to change or override the defaults; and rules for the alignment modes for simple variables, aggregates, and bit fields. This section also provides examples of aggregate layouts based on the different alignment modes.

“Using alignment modifiers” on page 15 discusses the different specifiers, pragmas, and attributes you can use in your source code to override the alignment mode currently in effect, for specific variable declarations. It also provides the rules governing the precedence of alignment modes and modifiers during compilation.

Related information in the XL C/C++ Compiler Reference

 **-qaltivec**

Related external information

 AltiVec Technology Programming Interface Manual, available at http://www.freescale.com/files/32bit/doc/ref_manual/ALTIVECPIM.pdf

Using alignment modes

Each data type supported by XL C/C++ is aligned along byte boundaries according to platform-specific default alignment modes. On AIX, the default alignment mode is **power** or **full**, which are equivalent.

You can change the default alignment mode, by using any of the following mechanisms:

- Set the alignment mode for all variables in a single file or multiple files during compilation

To use this approach, you specify the **-qalign** compiler option during compilation, with one of the suboptions listed in Table 11 on page 10.

- Set the alignment mode for all variables in a section of source code

To use this approach, you specify the `#pragma align` or `#pragma options align` directives in the source files, with one of the suboptions listed in Table 11. Each directive changes the alignment mode in effect for all variables that follow the directive until another directive is encountered, or until the end of the compilation unit.

Each of the valid alignment modes is defined in Table 11, which provides the alignment value, in bytes, for scalar variables, for all data types. Where there are differences between 32-bit and 64-bit modes, these are indicated. Also, where there are differences between the first (scalar) member of an aggregate and subsequent members of the aggregate, these are indicated.

Table 11. Alignment settings (values given in bytes)

Data type	Storage	Alignment setting				
		natural	power, full	mac68k, twobyte ³	bit_packed ²	packed ²
_Bool (C), bool (C++) (32-bit mode)	1	1	1	1	1	1
_Bool (C), bool (C++) (64-bit mode)	1	1	1	not supported	1	1
char, signed char, unsigned char	1	1	1	1	1	1
wchar_t (32-bit mode)	2	2	2	2	1	1
wchar_t (64-bit mode)	4	4	4	not supported	1	1
int, unsigned int	4	4	4	2	1	1
short int, unsigned short int	2	2	2	2	1	1
long int, unsigned long int (32-bit mode)	4	4	4	2	1	1
long int, unsigned long int (64-bit mode)	8	8	8	not supported	1	1
_Decimal32	4	4	4	2	1	1
_Decimal64	8	8	8	2	1	1
_Decimal128	16	16	16	2	1	1
long long	8	8	8	2	1	1
float	4	4	4	2	1	1
double	8	8	see note ¹	2	1	1
long double	8	8	see note ¹	2	1	1
long double with <code>-qldbl128</code>	16	16	see note ¹	2	1	1
pointer (32-bit mode)	4	4	4	2	1	1
pointer (64-bit mode)	8	8	8	not supported	1	1
vector types	16	16	16	16	1	1

Table 11. Alignment settings (values given in bytes) (continued)

Data type	Storage	Alignment setting			
		natural	power, full	mac68k, twobyte ³	bit_packed ²
Notes:					
1. In aggregates, the first member of this data type is aligned according to its natural alignment value; subsequent members of the aggregate are aligned on 4-byte boundaries.					
2. The packed alignment will not pack bit-field members at the bit level; use the <code>bit_packed</code> alignment if you want to pack bit fields at the bit level.					
3. For mac68k alignment, if the aggregate does not contain a vector member, the alignment is 2 bytes. If an aggregate contains a vector member, then the alignment is the largest alignment of all of its members.					

If you are working with aggregates containing double, long long, or long double data types, use the **natural** mode for highest performance, as each member of the aggregate is aligned according to its natural alignment value. If you generate data with an application on one platform and read the data with an application on another platform, it is recommended that you use the **bit_packed** mode, which results in equivalent data alignment on all platforms.

Notes:

- Vectors in a bit-packed structure may not be correctly aligned unless you take extra action to ensure their alignment.
- Vectors might suffer from alignment issues if they are accessed via heap-allocated storage or through pointer arithmetic. For example, `double __align(16) my_array[1000]` is 16-byte aligned while `my_array[1]` is not. How `my_array[i]` is aligned is determined by the value of `i`.

“Alignment of aggregates” discusses the rules for the alignment of entire aggregates and provide examples of aggregate layouts. “Alignment of bit fields” on page 14 discusses additional rules and considerations for the use and alignment of bit fields, and provides an example of bit-packed alignment.

Related information in the XL C/C++ Compiler Reference



-qalign




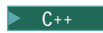
-qldbl128, -qlongdouble



#pragma options

Alignment of aggregates

The data contained in Table 11 on page 10 (in “Using alignment modes” on page 9) apply to scalar variables, and variables that are members of aggregates such as structures, unions, and classes. The following rules apply to aggregate variables, namely structures, unions or classes, as a whole (in the absence of any modifiers):

- For all alignment modes, the size of an aggregate is the smallest multiple of its alignment value that can encompass all of the members of the aggregate.
-  Empty aggregates are assigned a size of 0 bytes. As a result, two distinct variables might have the same address.
-  Empty aggregates are assigned a size of 1 byte. Note that static data members do not participate in the alignment or size of an aggregate; therefore a structure or class containing only a single static data member has a size of 1 byte.


- For all alignment modes except **mac68k**, the alignment of an aggregate is equal to the largest alignment value of any of its members. With the exception of packed alignment modes, members whose natural alignment is smaller than that of their aggregate's alignment are padded with empty bytes.
- For **mac68k** alignment, if the aggregate does not contain a vector member, the alignment is 2 bytes. If an aggregate contains a vector member, then the alignment is the largest alignment of all of its members.
- Aligned aggregates can be nested, and the alignment rules applicable to each nested aggregate are determined by the alignment mode that is in effect when a nested aggregate is declared.

The following table shows some examples of the size of an aggregate according to alignment mode.

Table 12. Alignment and aggregate size

Example	Size of aggregate		
	-qalign=power	-qalign=natural	-qalign=packed
struct Struct1 { double a1; char a2; };	16 bytes (The member with the largest alignment requirement is a1; therefore, a2 is padded with 7 bytes.)	16 bytes (The member with the largest alignment requirement is a1; therefore, a2 is padded with 7 bytes.)	9 bytes (Each member is packed to its natural alignment; no padding is added.)
struct Struct2 { char buf[15]; };	15 bytes	15 bytes	15 bytes
struct Struct3 { char c1; double c2; };	12 bytes (The member with the largest alignment requirement is c2; however, because it is a double and is not the first member, the 4-byte alignment rule applies. c1 is padded with 3 bytes.)	16 bytes (The member with the largest alignment requirement is c2; therefore, c1 is padded with 7 bytes.)	9 bytes (Each member is packed to its natural alignment; no padding is added.)

Notes:

-  The C++ compiler might generate extra fields for classes that contain base classes or virtual functions. Objects of these types might not conform to the usual mappings for aggregates.
- The alignment of an aggregate must be the same in all compilation units. For example, if the declaration of an aggregate is in a header file and you include that header file into two distinct compilations units, choose the same alignment mode for both compilations units.

For rules on the alignment of aggregates containing bit fields, see “Alignment of bit fields” on page 14.

Alignment examples

The following examples use these symbols to show padding and boundaries:

p = padding

| = halfword (2-byte) boundary

: = byte boundary

Mac68K example

```
#pragma options align=mac68k
struct B {
    char a;
    double b;
};
#pragma options align=reset
```

The size of B is 10 bytes. The alignment of B is 2 bytes. The layout of B is:
|a:p|b:b|b:b|b:b|b:b|

Packed example

```
#pragma options align=bit_packed
struct {
    char a;
    double b;
} B;
#pragma options align=reset
```

The size of B is 9 bytes. The layout of B is:
|a:b|b:b|b:b|b:b|b:

Nested aggregate example

```
#pragma options align=mac68k
struct A {
    char a;
    #pragma options align=power
    struct B {
        int b;
        char c;
    } B1; // <-- B1 laid out using power alignment rules
    #pragma options align=reset // <-- has no effect on A or B,
                                // but on subsequent structs
    char d;
};
#pragma options align=reset
```

The size of A is 12 bytes. The alignment of A is 2 bytes. The layout of A is:
|a:p|b:b|b:b|c:p|p:p|d:p|

C++ derived class example

In 32-bit mode:

```
#pragma options align=natural

class A {
    double _a;
} sa;

class C : public A {
public:
    virtual void f() {}
private:
    char* name;
} sc;
```

The size of sc is 24 bytes. The alignment of sc is 8 bytes. The layout of sc is:
|a:a|a:a|a:a|a:a|f:f|f:f|p:p|p:p|n:n|n:n|p:p|p:p|

Alignment of bit fields

You can declare a bit field as a `_Bool` (C), `bool` (C++), `char`, `signed char`, `unsigned char`, `short`, `unsigned short`, `int`, `unsigned int`, `long`, `unsigned long`, `long long`, or `unsigned long long` data type. The alignment of a bit field depends on its base type and the compilation mode (32-bit or 64-bit).

Note: `long long` and `unsigned long long` are not available for C on AIX.

C In the C language, you can specify bit fields as `char` or `short` instead of `int`, but XL C/C++ maps them as if they were `unsigned int`. The length of a bit field cannot exceed the length of its base type. In extended mode, you can use the `sizeof` operator on a bit field. The `sizeof` operator on a bit field always returns 4.

C++ The length of a bit field can exceed the length of its base type, but the remaining bits are used to pad the field, and do not actually store any value.

However, alignment rules for aggregates containing bit fields are different depending on the alignment mode in effect. These rules are described below.

Rules for natural alignment

- A zero-length bit field pads to the next alignment boundary of its base declared type. This causes the next member to begin on a 4-byte boundary for all types except `long` in 64-bit mode, which moves the next member to the next 8-byte boundary. Padding does not occur if the previous member's memory layout ended on the appropriate boundary.
- **C** An aggregate that contains only zero-length bit fields has a length of 0 bytes and an alignment of 4 bytes.
- **C++** An aggregate that contains only zero-length bit fields has a length of 4 or 8 bytes, depending on the declared type of the bit field and the compilation mode (32-bit or 64-bit).

Rules for power alignment

- Aggregates containing bit fields are 4-byte (word) aligned.
- Bit fields are packed into the current word. If a bit field would cross a word boundary, it starts at the next word boundary.
- A bit field of length zero causes the bit field that immediately follows it to be aligned at the next word boundary, or 8 bytes, depending on the declared type and the compilation mode. If the zero-length bit field is at a word boundary, the next bit field starts at this boundary.
- **C** An aggregate that contains only zero-length bit fields has a length of 0 bytes.
- **C++** An aggregate that contains only zero-length bit fields has the length of 1 byte.

Rules for Mac68K alignment

- Bit fields are packed into a word and are aligned on a 2-byte boundary.
- Bit fields that would cross a word boundary are moved to the *next* halfword boundary even if they are already starting on a halfword boundary. (The bit field can still end up crossing a word boundary.)

- A bit field of length zero forces the next member (even if it is not a bit field) to start at the *next* halfword boundary even if the zero-length bit field is currently at a halfword boundary.
- An aggregate containing nothing but zero-length bit fields has a length, in bytes, of two times the number of zero-length bit fields.
- For unions, there is one special case: unions whose largest element is a bit field of length 16 or less have a size of 2 bytes. If the length of the bit field is greater than 16, the size of the union is 4 bytes.

Rules for bit-packed alignment

- Bit fields have an alignment of 1 byte, and are packed with no default padding between bit fields.
- A zero-length bit field causes the next member to start at the next byte boundary. If the zero-length bit field is already at a byte boundary, the next member starts at this boundary. A non-bit field member that follows a bit field is aligned on the next byte boundary.

Example of bit-packed alignment

```
#pragma options align=bit_packed
struct {
    int a : 8;
    int b : 10;
    int c : 12;
    int d : 4;
    int e : 3;
    int : 0;
    int f : 1;
    char g;
} A;
```

```
pragma options align=reset
```

The size of A is 7 bytes. The alignment of A is 1 byte. The layout of A is:

Member name	Byte offset	Bit offset
a	0	0
b	1	0
c	2	2
d	3	6
e	4	2
f	5	0
g	6	0

Using alignment modifiers

XL C/C++ also provides alignment modifiers, with which you can exercise even finer-grained control over alignment, at the level of declaration or definition of individual variables. Available modifiers are:

#pragma pack(...)

Valid application:

The entire aggregate (as a whole) immediately following the directive.

Note: on AIX **#pragma pack** does not apply to bit-field union members.

Effect: Sets the maximum alignment of the members of the aggregate to which it applies, to a specific number of bytes. Also allows a bit-field to cross a container boundary. Used to reduce the effective alignment of the selected aggregate.

Valid values:

n: where *n* is 1, 2, 4, 8, or 16. That is, structure members are aligned on *n*-byte boundaries or on their natural alignment boundary, whichever is less. *nopack*: disables packing. *pop*: removes the previous value added with **#pragma pack**. **Note:** empty brackets has the same functionality as *pop*.

__attribute__((aligned(*n*)))

Valid application:

As a variable attribute, it applies to a single aggregate (as a whole), namely a structure, union, or class; or to an individual member of an aggregate.¹ As a type attribute, it applies to all aggregates declared of that type. If it is applied to a typedef declaration, it applies to all instances of that type.²

Effect:

Sets the minimum alignment of the specified variable (or variables), to a specific number of bytes. Typically used to increase the effective alignment of the selected variables.

Valid values:

n must be a positive power of 2, or NIL. NIL can be specified as either **__attribute__((aligned()))** or **__attribute__((aligned))**; this is the same as specifying the maximum system alignment (16 bytes on all UNIX platforms).

__attribute__((packed))

Valid application:

As a variable attribute, it applies to simple variables, or individual members of an aggregate, namely a structure or class¹. As a type attribute, it applies to all members of all aggregates declared of that type.

Effect: Sets the maximum alignment of the selected variable, or variables, to which it applies, to the smallest possible alignment value, namely one byte for a variable and one bit for a bit field.

__align(*n*)

Effect: Sets the minimum alignment of the variable or aggregate to which it applies to a specific number of bytes; also effectively increases the amount of storage occupied by the variable. Used to increase the effective alignment of the selected variables.

Valid application:

Applies to simple static (or global) variables or to aggregates as a whole, rather than to individual members of aggregates, unless these are also aggregates.

Valid values:

n must be a positive power of 2. XL C/C++ also allows you to specify a value greater than the system maximum.

Notes:

- In a comma-separated list of variables in a declaration, if the modifier is placed at the beginning of the declaration, it applies to all the variables in the declaration. Otherwise, it applies only to the variable immediately preceding it.

- Depending on the placement of the modifier in the declaration of a struct, it can apply to the definition of the type, and hence applies to all instances of that type; or it can apply to only a single instance of the type. For details, see *Type Attributes* in the *XL C/C++ Language Reference*.

When you use alignment modifiers, the interactions between modifiers and modes, and between multiple modifiers, can become complex. The following sections outline precedence guidelines for alignment modifiers, for the following types of variables:

- Simple, or scalar, variables, including members of aggregates (structures or classes) and user-defined types created by typedef statements
- Aggregate variables (structures or classes)


Related information in the *XL C/C++ Compiler Reference*

 #pragma pack

Related information in the *XL C/C++ Language Reference*

 The aligned type attribute (IBM extension)

 The packed type attribute (IBM extension)

 The __align type qualifier (IBM extension)

 Type attributes (IBM extension)

 The aligned variable attribute (IBM extension)

 The packed variable attribute (IBM extension)

Guidelines for determining alignment of scalar variables

The following formulas use a "top-down" approach to determining the alignment, given the presence of alignment modifiers, for both non-embedded (stand-alone) scalar variables and embedded scalars (variables declared as members of an aggregate):

Alignment of variable = maximum(*effective type alignment* , *modified alignment value*)

where *effective type alignment* = maximum(maximum(aligned type attribute value, __align specifier value) , minimum(*type alignment*, packed type attribute value))

and *modified alignment value* = maximum(aligned variable attribute value, packed variable attribute value)

and where *type alignment* is the alignment mode currently in effect when the variable is declared, or the alignment value applied to a type in a typedef statement.

In addition, for embedded variables, which can be modified by the #pragma pack directive, the following rule applies:

Alignment of variable = maximum(#pragma pack value, maximum(*effective type alignment* , *modified alignment value*))

Note: If a type attribute and a variable attribute of the same kind are both specified in a declaration, the second attribute is ignored.

Guidelines for determining alignment of aggregate variables

The following formulas determine the alignment for aggregate variables, namely structures, unions, and classes:

Alignment of variable = maximum(*effective type alignment* , *modified alignment value*)

where *effective type alignment* = maximum(maximum(aligned type attribute value, `__align` specifier value) , minimum(*aggregate type alignment*, packed type attribute value))

and *modified alignment value* = maximum (aligned variable attribute value , packed variable attribute value)

and where *aggregate type alignment* = maximum (alignment of all members)

Note: If a type attribute and a variable attribute of the same kind are both specified in a declaration, the second attribute is ignored.

Chapter 4. Handling floating-point operations

The following sections provide reference information, portability considerations, and suggested procedures for using compiler options to manage floating-point operations:

- “Floating-point formats”
- “Handling multiply-add operations”
- “Compiling for strict IEEE conformance” on page 20
- “Handling floating-point constant folding and rounding” on page 20
- “Handling floating-point exceptions” on page 23

Floating-point formats

XL C/C++ supports the following binary floating-point formats:


- 32-bit single precision, with an approximate absolute normalized range of 0 and 10^{-38} to 10^{+38} and precision of about 7 decimal digits
- 64-bit double precision, with an approximate absolute normalized range of 0 and 10^{-308} to 10^{+308} and precision of about 16 decimal digits
- 128-bit extended precision, with slightly greater range than double-precision values, and with a precision of about 32 decimal digits

Note that the `long double` type may represent either double-precision or extended-precision values, depending on the setting of the `-qldbl128` compiler option. The default is 128 bits. For compatibility with older compilations, you can use `-qnohdl128` if you need `long double` to be 64 bits.

Beginning in V9.0, on selected hardware and operating system levels, the compiler also supports the following decimal floating-point formats:

- 32-bit single precision, with an approximate range of 10^{-101} to 10^{+90} and precision of 7 decimal digits
- 64-bit double precision, with an approximate range of 10^{-398} to 10^{+369} and precision of 16 decimal digits
- 128-bit extended precision, with an approximate range of 10^{-6176} to 10^{+6111} , and with a precision of 34 decimal digits

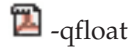
Related information in the *XL C/C++ Compiler Reference*

 `-qldbl128, -qlongdouble`

Handling multiply-add operations

By default, the compiler generates a single non-IEEE 754 compatible multiply-add instruction for binary floating-point expressions such as $a+b*c$, partly because one instruction is faster than two. Because no rounding occurs between the multiply and add operations, this may also produce a more precise result. However, the increased precision might lead to different results from those obtained in other environments, and may cause $x*y-x*y$ to produce a nonzero result. To avoid these issues, you can suppress the generation of multiply-add instructions by using the `-qfloat=nomaf` option.

Note: Decimal floating-point does not use multiply-add instructions



Compiling for strict IEEE conformance

By default, XL C/C++ follows most, but not all of the rules in the IEEE standard. If you compile with the **-qnostrict** option, which is enabled by default at optimization level **-O3** or higher, some IEEE floating-point rules are violated in ways that can improve performance but might affect program correctness. To avoid this issue, and to compile for strict compliance with the IEEE standard, use the following options:

- Use the **-qfloat=nomaf** compiler option.
- If the program changes the rounding mode at runtime, use the **-qfloat=rrm** option.
- If the data or program code contains signaling NaN values (NaNs), use the **-qfloat=nans** option. (A signaling NaN is different from a quiet NaN; you must explicitly code it into the program or data or create it by using the **-qinitauto** compiler option.)
- If you compile with **-O3**, **-O4**, or **-O5**, include the option **-qstrict** after it.

Related information:

“Advanced optimization” on page 74

Related information in the *XL C/C++ Compiler Reference*



Handling floating-point constant folding and rounding

By default, the compiler replaces most operations involving constant operands with their result at compile time. This process is known as constant folding. Additional folding opportunities might occur with optimization or with the **-qnostrict** option. The result of a floating-point operation folded at compile time normally produces the same result as that obtained at execution time, except in the following cases:


- The compile-time rounding mode is different from the execution-time rounding mode. By default, both are round-to-nearest; however, if your program changes the execution-time rounding mode, to avoid differing results, do either of the following operations:
 - Change the compile-time rounding mode to match the execution-time mode, by compiling with the appropriate **-y** option. For more information and an example, see “Matching compile-time and runtime rounding modes” on page 21.
 - Suppress folding, by compiling with the **-qfloat=nofold** option.
- Expressions like $a+b*c$ are partially or fully evaluated at compile time. The results might be different from those produced at execution time, because $b*c$ might be rounded before being added to a , while the runtime multiply-add instruction does not use any intermediate rounding. To avoid differing results, do either of the following operations:

- Suppress the use of multiply-add instructions, by compiling with the **-qfloat=nomaf** option.
- Suppress folding, by compiling with the **-qfloat=nofold** option.
- An operation produces an infinite or NaN result. Compile-time folding prevents execution-time detection of an exception, even if you compile with the **-qflttrap** option. To avoid missing these exceptions, suppress folding with the **-qfloat=nofold** option.

Related information:

“Handling floating-point exceptions” on page 23

Related information in the *XL C/C++ Compiler Reference*

 -qfloat

 -qstrict

 -qflttrap

Matching compile-time and runtime rounding modes

The default rounding mode used at compile time and run time is round-to-nearest, ties to even. If your program changes the rounding mode at run time, the results of a floating-point calculation might be slightly different from those that are obtained at compile time. The following example illustrates this:¹

```
#include <float.h>
#include <fenv.h>
#include <stdio.h>

int main ( )
{
    volatile double one = 1.f, three = 3.f; /* volatiles are not folded */
    double one_third;

    one_third = 1. / 3.; /* folded */
    printf ("1/3 with compile-time rounding = %.17f\n", one_third);

    fesetround (FE_TOWARDZERO);
    one_third = one / three; /* not folded */
    fesetround (FE_TONEAREST);2
    printf ("1/3 with execution-time rounding to zero = %.17f\n", one_third);

    fesetround (FE_TONEAREST);
    one_third = one / three; /* not folded */
    fesetround (FE_TONEAREST);2
    printf ("1/3 with execution-time rounding to nearest = %.17f\n", one_third);

    fesetround (FE_UPWARD);
    one_third = one / three; /* not folded */
    fesetround (FE_TONEAREST);2
    printf ("1/3 with execution-time rounding to +infinity = %.17f\n", one_third);

    fesetround (FE_DOWNWARD);
    one_third = one / three; /* not folded */
    fesetround (FE_TONEAREST);2
    printf ("1/3 with execution-time rounding to -infinity = %.17f\n", one_third);

    return 0;
}
```

Notes:

1. On AIX, this example must be linked with the system math library, `libm`, to obtain the functions and macros declared in the `fenv.h` header file.

2. See “Rounding modes and standard library functions” for an explanation of the resetting of the round mode before the call to printf.

When compiled with the default options, this code produces the following results:

```
1/3 with compile-time rounding = 0.3333333333333331
1/3 with execution-time rounding to zero = 0.3333333333333331
1/3 with execution-time rounding to nearest = 0.3333333333333331
1/3 with execution-time rounding to +infinity = 0.3333333333333337
1/3 with execution-time rounding to -infinity = 0.3333333333333331
```

Because the fourth computation changes the rounding mode to round-to-infinity, the results are slightly different from the first computation, which is performed at compile time, using round-to-nearest. If you do not use the **-qfloat=nofold** option to suppress all compile-time folding of floating-point computations, it is recommended that you use the **-y** compiler option with the appropriate suboption to match compile-time and runtime rounding modes. In the previous example, compiling with **-yp** (round-to-infinity) produces the following result for the first computation:

```
1/3 with compile-time rounding = 0.3333333333333337
```

In general, if the rounding mode is changed to +infinity or -infinity, or to any decimal floating-point only rounding mode, it is recommended that you also use the **-qfloat=rrm** option.

Related information in the XL C/C++ Compiler Reference

 [-qfloat](#)

 [-y](#)

Rounding modes and standard library functions

On AIX, C and C++ input/output and conversion functions apply the rounding mode in effect to the values that are input or output by the function. These functions include printf, scanf, atof, and ftoa, as well as the C++ input and output operators (>> and <<) on objects like cin and cout.

For example, if the current rounding mode is round-to-infinity, the printf function will apply that rounding mode to the floating-point digit string value it prints, in addition to the rounding that was already performed on a calculation. The following example illustrates this:

```
#include <float.h>
#include <fenv.h>
#include <stdio.h>

int main( )
{
    volatile double one = 1.f, three = 3.f; /* volatiles are not folded*/
    double one_third;

    fesetround (FE_UPWARD);
    one_third = one / three; /* not folded */
    printf ("1/3 with execution-time rounding to +infinity = %.17f\n", one_third);

    fesetround (FE_UPWARD);
    one_third = one / three; /* not folded */
    fesetround (FE_TONEAREST);
    printf ("1/3 with execution-time rounding to +infinity = %.17f\n", one_third);

    return 0;
}
```

When compiled with the default options, this code produces the following results:

```
1/3 with execution-time rounding to +infinity = 0.3333333333333338  
1/3 with execution-time rounding to -infinity = 0.3333333333333337
```

In the first calculation, the value returned is rounded upward to 0.3333333333333337, but the `printf` function rounds this value upward again, to print out 0.3333333333333338. The solution to this problem, which is used in the second calculation, is to reset the rounding mode to round-to-nearest just before the call to the library function is made.

Handling floating-point exceptions


By default, invalid operations such as division by zero, division by infinity, overflow, and underflow are ignored at run time. However, you can use the **-qflttrap** option to detect these types of exceptions. In addition, you can add suitable support code to your program to make program execution continue after an exception occurs, and to modify the results of operations causing exceptions.

Because, however, floating-point computations involving constants are usually folded at compile time, the potential exceptions that would be produced at runtime will not occur. To ensure that the **-qflttrap** option traps all runtime floating-point exceptions, consider using the **-qfloat=nofold** option to suppress all compile-time folding.

If you use the AIX operating system functions to enable hardware trapping on floating-point exceptions, use the **-qfloat=fenv** option to inform the compiler that exceptions might occur.

Related information in the *XL C/C++ Compiler Reference*

 [-qfloat](#)

 [-qflttrap](#)

Compiling a decimal floating-point program

If you are using decimal floating-point formats in your programs, use the **-qdfp** option when you compile them. For example, to compile the following Hello World program `dfp_hello.c`, the compiler invocation is:

```
xlc -qdfp dfp_hello.c  
#include <stdio.h>  
#include <float.h>  
int main() {  
    printf("Hello DFP World\n");  
    printf("DEC32_MAX = %Hf\n", DEC32_MAX);  
    float f = 12.34df;  
    printf("12.34df as a float = %f\n", f);  
}
```

Related information in the *XL C/C++ Compiler Reference*

 [-qdfp](#)

Chapter 5. Using memory heaps

In addition to the memory management functions defined by ANSI, XL C/C++ provides enhanced versions of memory management functions that can help you improve program performance and debug your programs. With these functions, you can:

- Allocate memory from multiple, custom-defined pools of memory, known as user-created heaps.
- Debug memory problems in the default runtime heap.
- Debug memory problems in user-created heaps.

All the versions of the memory management functions actually work in the same way. They differ only in the heap from which they allocate, and in whether they save information to help you debug memory problems. The memory allocated by all of these functions is suitably aligned for storing any type of object.

“Managing memory with multiple heaps” discusses the advantages of using multiple, user-created heaps; summarizes the functions available to manage user-created heaps; provides procedures for creating, expanding, using, and destroying user-defined heaps; and provides examples of programs that create user heaps using both regular and shared memory.

“Debugging memory heaps” on page 36 discusses the functions available for checking and debugging the default and user-created heaps.

Managing memory with multiple heaps

You can use XL C/C++ to create and manipulate your own memory heaps, either in place of or in addition to the default XL C/C++ runtime heap.

You can create heaps of regular memory or shared memory, and you can have any number of heaps of any type. The only limit is the space available on your operating system (your machine's memory and swapper size, minus the memory required by other running applications). You can also change the default runtime heap to a heap that you have created.

Using your own heaps is optional, and your applications can work well using the default memory management provided (and used by) the XL C/C++ runtime library. However, using multiple heaps can be more efficient and can help you improve your program's performance and reduce wasted memory for a number of reasons:

- When you allocate from a single heap, you can end up with memory blocks on different pages of memory. For example, you might have a linked list that allocates memory each time you add a node to the list. If you allocate memory for other data in between adding nodes, the memory blocks for the nodes could end up on many different pages. To access the data in the list, the system might have to swap many pages, which can significantly slow your program.

With multiple heaps, you can specify the heap from which you want to allocate. For example, you might create a heap specifically for a linked list. The list's memory blocks and the data they contain would remain close together on fewer pages, which reduces the amount of swapping required.

- In multithreaded applications, only one thread can access the heap at a time to ensure memory is safely allocated and freed. For example, if thread 1 is allocating memory, and thread 2 has a call to free, thread 2 must wait until thread 1 has finished its allocation before it can access the heap. Again, this can slow down performance, especially if your program does a lot of memory operations.

If you create a separate heap for each thread, you can allocate from them concurrently, eliminating both the waiting period and the overhead required to serialize access to the heap.

- With a single heap, you must explicitly free each block that you allocate. If you have a linked list that allocates memory for each node, you have to traverse the entire list and free each block individually, which can take some time.

If you create a separate heap only for that linked list, you can destroy it with a single call and free all the memory at once.

- When you have only one heap, all components share it (including the XL C/C++ runtime library, vendor libraries, and your own code). If one component corrupts the heap, another component might fail. You might have trouble discovering the cause of the problem and where the heap was damaged.

With multiple heaps, you can create a separate heap for each component, so if one damages the heap (for example, by using a freed pointer), the others can continue unaffected. You also know where to look to correct the problem.

Functions for managing user-created heaps

The `libhu.a` library provides a set of functions with which you can manage user-created heaps. These functions are all prefixed by `_u` (for "user" heaps), and they are declared in the header file `umalloc.h`. The following table summarizes the functions available for creating and managing user-defined heaps.

Table 13. Functions for managing memory heaps

Default heap function	Corresponding user-created heap function	Description
n/a	<code>_ucreate</code>	Creates a heap. Described in "Creating a heap" on page 27.
n/a	<code>_uopen</code>	Opens a heap for use by a process. Described in "Using a heap" on page 29.
n/a	<code>_ustats</code>	Provides information about a heap. Described in "Getting information about a heap" on page 30.
n/a	<code>_uaddmem</code>	Adds memory blocks to a heap. Described in "Expanding a heap" on page 28.
n/a	<code>_uclose</code>	Closes a heap from further use by a process. Described in "Closing and destroying a heap" on page 30.
n/a	<code>_udestroy</code>	Destroys a heap. Described in "Closing and destroying a heap" on page 30.
<code>calloc</code>	<code>_ucalloc</code>	Allocates and initializes memory from a heap you have created. Described in "Using a heap" on page 29.
<code>malloc</code>	<code>_umalloc</code>	Allocates memory from a heap you have created. Described in "Using a heap" on page 29.

Table 13. Functions for managing memory heaps (continued)

Default heap function	Corresponding user-created heap function	Description
<code>_heapmin</code>	<code>_uheapmin</code>	Returns unused memory to the system. Described in “Closing and destroying a heap” on page 30.
n/a	<code>_udefault</code>	Changes the default runtime heap to a user-created heap. Described in “Changing the default heap used in a program” on page 31.

Note: There are no user-created heap versions of `realloc` or `free`. These standard functions always determine the heap from which memory is allocated, and can be used with both user-created and default memory heaps.

Creating a heap

You can create a fixed-size heap, or a dynamically-sized heap. With a fixed-size heap, the initial block of memory must be large enough to satisfy all allocation requests made to it. With a dynamically-sized heap, the heap can expand and contract as your program needs demand.

Creating a fixed-size heap

When you create a fixed-size heap, you first allocate a block of memory large enough to hold the heap and to hold internal information required to manage the heap, and you assign it a handle. For example:

```
Heap_t fixedHeap; /* this is the "heap handle" */
/* get memory for internal info plus 5000 bytes for the heap */
static char block[_HEAP_MIN_SIZE + 5000];
```

The internal information requires a minimum set of bytes, specified by the `_HEAP_MIN_SIZE` macro (defined in `umalloc.h`). You can add the amount of memory your program requires to this value to determine the size of the block you need to get. Once the block is fully allocated, further allocation requests to the heap will fail.

After you have allocated a block of memory, you create the heap with `_ucreate`, and specify the type of memory for the heap, regular or shared. For example:

```
fixedHeap = _ucreate(block, (_HEAP_MIN_SIZE+5000), /* block to use */
                    !_BLOCK_CLEAN, /* memory is not set to 0 */
                    _HEAP_REGULAR, /* regular memory */
                    NULL, NULL); /* functions for expanding and shrinking
                                a dynamically-sized heap */
```

The `!_BLOCK_CLEAN` parameter indicates that the memory in the block has not been initialized to 0. If it were set to 0 (for example, by `memset`), you would specify `_BLOCK_CLEAN`. The `calloc` and `_ucalloc` functions use this information to improve their efficiency; if the memory is already initialized to 0, they don't need to initialize it.

The fourth parameter indicates the type of memory the heap contains: regular (`_HEAP_REGULAR`) or shared (`_HEAP_SHARED`).

Use `_HEAP_REGULAR` for regular memory. Most programs use regular memory. This is the type provided by the default run-time heap. Use `_HEAP_SHARED` for shared memory. Heaps of shared memory can be shared between processes or applications.

For a fixed-size heap, the last two parameters are always `NULL`.

Creating a dynamically-sized heap

With the XL C/C++ default heap, when not enough storage is available to fulfill a `malloc` request, the runtime environment gets additional storage from the system. Similarly, when you minimize the heap with `_heapmin` or when your program ends, the runtime environment returns the memory to the operating system.

When you create an expandable heap, you provide your own functions to do this work, which you can name however you choose. You specify pointers to these functions as the last two parameters to `_ucreate` (instead of the `NULL` pointers you use to create a fixed-size heap). For example:

```
Heap_t growHeap;
static char block[_HEAP_MIN_SIZE]; /* get block */

growHeap = _ucreate(block, _HEAP_MIN_SIZE, /* starting block */
                   !_BLOCK_CLEAN,      /* memory not set to 0 */
                   _HEAP_REGULAR,      /* regular memory */
                   expandHeap,          /* function to expand heap */
                   shrinkHeap);        /* function to shrink heap */
```

Note: You can use the same expand and shrink functions for more than one heap, as long as the heaps use the same type of memory and your functions are not written specifically for one heap.

Expanding a heap

To increase the size of a heap, you add blocks of memory to it by doing the following:

- For fixed-size or dynamically-sized heaps, calling the `_uaddmem` function.
- For dynamically-sized heaps only, writing a function that expands the heap, and that can be called automatically by the system if necessary, whenever you allocate memory from the heap.

Adding blocks of memory to a heap

You can add blocks of memory to a fixed-size or dynamically-sized heap with `_uaddmem`. This can be useful if you have a large amount of memory that is allocated conditionally. Like the starting block, you must first allocate memory for a block of memory. This block will be added to the current heap, so make sure the block you add is of the same type of memory as the heap to which you are adding it. For example, to add 64K to `fixedHeap`:

```
static char newblock[65536];

_uaddmem(fixedHeap, /* heap to add to */
         newblock, 65536, /* block to add */
         _BLOCK_CLEAN); /* sets memory to 0 */
```

Note: For every block of memory you add, a small number of bytes from it are used to store internal information. To reduce the total amount of overhead, it is better to add a few large blocks of memory than many small blocks.

Writing a heap-expanding function

When you call `_umalloc` (or a similar function) for a dynamically-sized heap, `_umalloc` tries to allocate the memory from the initial block you provided to `_ucreate`. If not enough memory is there, it then calls the heap-expanding function you specified as a parameter to `_ucreate`. Your function then gets more memory from the operating system and adds it to the heap. It is up to you how you do this.

Your function must have the following prototype:

```
void *(*functionName)(Heap_t uh, size_t *size, int *clean);
```

Where *functionName* identifies the function (you can name it however you want), *uh* is the heap to be expanded, and *size* is the size of the allocation request passed by `_umalloc`. You probably want to return enough memory at a time to satisfy several allocations; otherwise every subsequent allocation has to call your heap-expanding function, reducing your program's execution speed. Make sure that you update the *size* parameter if you return more than the *size* requested.

Your function must also set the *clean* parameter to either `_BLOCK_CLEAN`, to indicate the memory has been set to 0, or `!_BLOCK_CLEAN`, to indicate that the memory has not been initialized.

The following fragment shows an example of a heap-expanding function:

```
static void *expandHeap(Heap_t uh, size_t *length, int *clean)
{
    char *newblock;
    /* round the size up to a multiple of 64K * /
    *length = (*length / 65536) * 65536 + 65536;

    *clean = _BLOCK_CLEAN; /* mark the block as "clean" */
    return(newblock);      /* return new memory block */
}
```

Using a heap

Once you have created a heap, you can open it for use by calling `_uopen`:

```
_uopen(fixedHeap);
```

This opens the heap for that particular process; if the heap is shared, each process that uses the heap needs its own call to `_uopen`.

You can then allocate and free memory from your own heap just as you would from the default heap. To allocate memory, use `_ucalloc` or `_umalloc`. These functions work just like `calloc` and `malloc`, except you specify the heap to use as well as the size of block that you want. For example, to allocate 1000 bytes from `fixedHeap`:

```
void *up;
up = _umalloc(fixedHeap, 1000);
```

To reallocate and free memory, use the regular `realloc` and `free` functions. Both of these functions always check the heap from which the memory was allocated, so you don't need to specify the heap to use. For example, the `realloc` and `free` calls in the following code fragment look exactly the same for both the default heap and your heap:

```

void *p, *up;
p = malloc(1000); /* allocate 1000 bytes from default heap */
up = _umalloc(fixedHeap, 1000); /* allocate 1000 from fixedHeap */

realloc(p, 2000); /* reallocate from default heap */
realloc(up, 100); /* reallocate from fixedHeap */

free(p); /* free memory back to default heap */
free(up); /* free memory back to fixedHeap */

```

When you call any heap function, make sure the heap you specify is valid. If the heap is not valid, the behavior of the heap functions is undefined.

Getting information about a heap

You can determine the heap from which any object was allocated by calling `_mheap`. You can also get information about the heap itself by calling `_ustats`, which tells you:

- The amount of memory the heap holds (excluding memory used for overhead)
- The amount of memory currently allocated from the heap
- The type of memory in the heap
- The size of the largest contiguous piece of memory available from the heap

Closing and destroying a heap

When a process has finished using the heap, close it with `_uclose`. Once you have closed the heap in a process, that process can no longer allocate from or return memory to that heap. If other processes share the heap, they can still use it until you close it in each of them. Performing operations on a heap after you have closed it causes undefined behavior.

To destroy a heap, do the following:

- For a fixed-size heap, call `_udestroy`. If blocks of memory are still allocated somewhere, you can force the destruction. Destroying a heap removes it entirely even if it was shared by other processes. Again, performing operations on a heap after you have destroyed it causes undefined behavior.
- For a dynamically-sized heap, call `_uheapmin` to coalesce the heap (return all blocks in the heap that are totally free to the system), or `_udestroy` to destroy it. Both of these functions call your heap-shrinking function. (See the following section.)

After you destroy a heap, it is up to you to return the memory for the heap (the initial block of memory you supplied to `_ucreate` and any other blocks added by `_uaddmem`) to the system.

Writing the heap-shrinking function

When you call `_uheapmin` or `_udestroy` to coalesce or destroy a dynamically-sized heap, these functions call your heap-shrinking function to return the memory to the system. It is up to you how you implement this function.

Your function must have the following prototype:

```
void (*functionName)(Heap_t uh, void *block, size_t size);
```

Where *functionName* identifies the function (you can name it however you want), *uh* identifies the heap to be shrunk. The pointer *block* and its *size* are passed to

your function by `_uheapmin` or `_udestroy`. Your function must return the memory pointed to by `block` to the system. For example:

```
static void shrinkHeap(Heap_t uh, void *block, size_t size)
{
    free(block);
    return;
}
```

Changing the default heap used in a program

The regular memory management functions (`malloc` and so on) always use the current default heap for that thread. The initial default heap for all XL C/C++ applications is the runtime heap provided by XL C/C++. However, you can make your own heap the default by calling `_udefault`. Then all calls to the regular memory management functions allocate memory from your heap instead of the default runtime heap.

The default heap changes only for the thread where you call `_udefault`. You can use a different default heap for each thread of your program if you choose. This is useful when you want a component (such as a vendor library) to use a heap other than the XL C/C++ default heap, but you cannot actually alter the source code to use heap-specific calls. For example, if you set the default heap to a shared heap and then call a library function that calls `malloc`, the library allocates storage in shared memory

Because `_udefault` returns the current default heap, you can save the return value and later use it to restore the default heap you replaced. You can also change the default back to the XL C/C++ default runtime heap by calling `_udefault` and specifying the `_RUNTIME_HEAP` macro (defined in `umalloc.h`). You can also use this macro with any of the heap-specific functions to explicitly allocate from the default runtime heap.

Compiling and linking a program with user-created heaps

To compile an application that calls any of the user-created heap functions (prefixed by `_u`), specify `hu` on the `-l` linker option. For example, if the `libhu.a` library is installed in the default directory, you could specify:

```
xlc prog.c -o progf -lhu
```

Examples of creating and using user heaps

Example of a user heap with regular memory

The program below shows how you might create and use a heap that uses regular memory.

```
#include <stdlib.h>
#include <stdio.h>
#include <umalloc.h>

static void *get_fn(Heap_t usrheap, size_t *length, int *clean)
{
    void *p;
    /* Round up to the next chunk size */
    *length = ((*length) / 65536) * 65536 + 65536;
    *clean = _BLOCK_CLEAN;
    p = calloc(*length,1);
    return (p);
}

static void release_fn(Heap_t usrheap, void *p, size_t size)
```



```

    {
        free( p );
        return;
    }

int main(void)
{
    void    *initial_block;
    long    rc;
    Heap_t  myheap;
    char    *ptr;
    int     initial_sz;

    /* Get initial area to start heap */
    initial_sz = 65536;
    initial_block = malloc(initial_sz);
    if(initial_block == NULL) return (1);

    /* create a user heap */
    myheap = _ucreate(initial_block, initial_sz, _BLOCK_CLEAN,
                     _HEAP_REGULAR, get_fn, release_fn);
    if (myheap == NULL) return(2);

    /* allocate from user heap and cause it to grow */
    ptr = _umalloc(myheap, 100000);
    _ufree(ptr);

    /* destroy user heap */
    if (_udestroy(myheap, _FORCE)) return(3);

    /* return initial block used to create heap */

    free(initial_block);
    return 0;
}

```

Example of a shared user heap – parent process

The following program shows how you might implement a heap shared between a parent and several child processes. This program shows the parent process, which creates the shared heap. First the main program calls the `init` function to allocate shared memory from the operating system (using `CreateFileMapping`) and name the memory so that other processes can use it by name. The `init` function then creates and opens the heap. The loop in the main program performs operations on the heap, and also starts other processes. The program then calls the `term` function to close and destroy the heap.

```

#include <umalloc.h>
#include <stdio.h>
#include <stdlib.h>
#include <string.h>

#define PAGING_FILE  0xFFFFFFFF
#define MEMORY_SIZE  65536
#define BASE_MEM     (VOID*)0x01000000

static HANDLE hFile;    /* Handle to memory file          */
static void*  hMap;    /* Handle to allocated memory      */

typedef struct mem_info {
    void * pBase;
    Heap_t pHeap;
} MEM_INFO_T;

/*-----*/

```



```

/* inithp:                                                                 */
/* Function to create and open the heap with a named shared memory object */
/*-----*/
static Heap_t inithp(size_t heap_size)
{
    MEM_INFO_T info;          /* Info structure          */

    /* Allocate shared memory from the system by creating a shared memory */
    /* pool basing it out of the system paging (swapper) file.          */

    hFile = CreateFileMapping( (HANDLE) PAGING_FILE, NULL, PAGE_READWRITE,
                               0, heap_size + sizeof(Heap_t),
                               "MYNAME_SHAREMEM" );

    if (hFile == NULL) {
        return NULL;
    }

    /* Map the file to this process' address space, starting at an address */
    /* that should also be available in child processe(s)                  */

    hMap = MapViewOfFileEx( hFile, FILE_MAP_WRITE, 0, 0, 0, BASE_MEM );

    info.pBase = hMap;
    if (info.pBase == NULL) {
        return NULL;
    }

    /* Create a fixed sized heap. Put the heap handle as well as the      */
    /* base heap address at the beginning of the shared memory.          */

    info.pHeap = _ucreate((char *)info.pBase + sizeof(info),
                          heap_size - sizeof(info),
                          !_BLOCK_CLEAN, _HEAP_SHARED | _HEAP_REGULAR, NULL, NULL);

    if (info.pBase == NULL) {
        return NULL;
    }

    memcpy(info.pBase, info, sizeof(info));

    if (_uopen(info.pHeap)) {      /* Open heap and check result      */
        return NULL;
    }

    return info.pHeap;
}

/*-----*/
/* termhp:                                                                 */
/* Function to close and destroy the heap                                  */
/*-----*/
static int termhp(Heap_t uheap)
{
    if (_uclose(uheap))           /* close heap                      */
        return 1;
    if (_udestroy(uheap, _FORCE)) /* force destruction of heap      */
        return 1;

    UnmapViewOfFile(hMap);       /* return memory to system        */
    CloseHandle(hFile);

    return 0;
}

/*-----*/
/* main:                                                                    */

```

```

/* Main function to test creating, writing to and destroying a shared heap.
/*-----*/
int main(void)
{
    int i, rc;                /* Index and return code */
    Heap_t uheap;            /* heap to create */
    char *p;                 /* for allocating from heap */

    /*
    /* call init function to create and open the heap
    /*-----*/

    uheap = inithp(MEMORY_SIZE);
    if (uheap == NULL)        /* check for success */
        return 1;            /* if failure, return non zero */

    /*
    /* perform operations on uheap
    /*-----*/
    for (i = 1; i <= 5; i++)
    {
        p = _umalloc(uheap, 10); /* allocate from uheap */
        if (p == NULL)
            return 1;
        memset(p, 'M', _msize(p)); /* set all bytes in p to 'M' */
        p = realloc(p,50);        /* reallocate from uheap */
        if (p == NULL)
            return 1;
        memset(p, 'R', _msize(p)); /* set all bytes in p to 'R' */
    }

    /*
    /* Start a second process which accesses the heap
    /*-----*/
    if (system("memshr2.exe"))
        return 1;

    /*
    /* Take a look at the memory that we just wrote to. Note that memshr.c
    /* and memshr2.c should have been compiled specifying the
    /* alloc(debug[, yes]) flag.
    /*-----*/
    #ifdef DEBUG
        _udump_allocated(uheap, -1);
    #endif

    /*
    /* call term function to close and destroy the heap
    /*-----*/
    rc = termhp(uheap);

    #ifdef DEBUG
        printf("memshr ending... rc = %d\n", rc);
    #endif

    return rc;
}

```

Example of a shared user heap - child process

The following program shows the process started by the loop in the parent process. This process uses `OpenFileMapping` to access the shared memory by name, then extracts the heap handle for the heap created by the parent process. The

process then opens the heap, makes it the default heap, and performs some operations on it in the loop. After the loop, the process replaces the old default heap, closes the user heap, and ends.

```

#include <umalloc.h>
#include <stdio.h>
#include <stdlib.h>
#include <string.h>

static HANDLE hFile;          /* Handle to memory file          */
static void* hMap;           /* Handle to allocated memory    */

typedef struct mem_info {
    void * pBase;
    Heap_t pHeap;
} MEM_INFO_T;

/*-----*/
/* inithp: Subprocess Version          */
/* Function to create and open the heap with a named shared memory object */
/*-----*/
static Heap_t inithp(void)
{
    MEM_INFO_T info;          /* Info structure          */

    /* Open the shared memory file by name. The file is based on the
    /* system paging (swapper) file.          */

    hFile = OpenFileMapping(FILE_MAP_WRITE, FALSE, "MYNAME_SHAREMEM");

    if (hFile == NULL) {
        return NULL;
    }

    /* Figure out where to map this file by looking at the address in the
    /* shared memory where the memory was mapped in the parent process.  */

    hMap = MapViewOfFile( hFile, FILE_MAP_WRITE, 0, 0, sizeof(info) );

    if (hMap == NULL) {
        return NULL;
    }

    /* Extract the heap and base memory address from shared memory          */

    memcpy(info, hMap, sizeof(info));
    UnmapViewOfFile(hMap);

    hMap = MapViewOfFileEx( hFile, FILE_MAP_WRITE, 0, 0, 0, info.pBase );

    if (_uopen(info.pHeap)) {          /* Open heap and check result          */
        return NULL;
    }

    return info.pHeap;
}

/*-----*/
/* termhp:          */
/* Function to close my view of the heap          */
/*-----*/
static int termhp(Heap_t uheap)
{
    if (_uclose(uheap))          /* close heap          */
        return 1;
}

```

```

    UnmapViewOfFile(hMap);                /* return memory to system */
    CloseHandle(hFile);

    return 0;
}

/*-----*/
/* main:                                     */
/* Main function to test creating, writing to and destroying a shared */
/* heap.                                     */
/*-----*/
int main(void)
{
    int rc, i;                            /* for return code, loop iteration */
    Heap_t uheap, oldheap;                /* heap to create, old default heap */
    char *p;                              /* for allocating from the heap */

    /*                                     */
    /* Get the heap storage from the shared memory */
    /*                                     */
    uheap = inithp();
    if (uheap == NULL)
        return 1;

    /*                                     */
    /* Register uheap as default runtime heap, save old default */
    /*                                     */
    oldheap = _udefault(uheap);
    if (oldheap == NULL) {
        return termhp(uheap);
    }

    /*                                     */
    /* Perform operations on uheap */
    /*                                     */
    for (i = 1; i <= 5; i++)
    {
        p = malloc(10); /* malloc uses default heap, which is now uheap*/
        memset(p, 'M', _msize(p));
    }

    /*                                     */
    /* Replace original default heap and check result */
    /*                                     */
    if (uheap != _udefault(oldheap)) {
        return termhp(uheap);
    }

    /*                                     */
    /* Close my views of the heap */
    /*                                     */
    rc = termhp(uheap);

    #ifdef DEBUG
        printf("Returning from memshr2 rc = %d\n", rc);
    #endif
    return rc;
}

```

Debugging memory heaps

XL C/C++ provides two sets of functions for debugging memory problems:

- Heap-checking functions similar to those provided by other compilers. (Described in “Functions for checking memory heaps” on page 37.)

- Debug versions of all memory management functions. (Described in “Functions for debugging memory heaps” on page 38.)

Both sets of debugging functions have their benefits and drawbacks. The one you choose to use depends on your program, your problems, and your preference.

The heap-checking functions perform more general checks on the heap at specific points in your program. You have greater control over where the checks occur. The heap-checking functions also provide compatibility with other compilers that offer these functions. You only have to rebuild the modules that contain the heap-checking calls. However, you have to change your source code to include these calls, which you will probably want to remove in your final code. Also, the heap-checking functions only tell you if the heap is consistent or not; they do not provide the details that the debug memory management functions do.

On the other hand, the debug memory management functions provide detailed information about all allocation requests you make with them in your program. You don't need to change any code to use the debug versions; you need only specify the **-qheapdebug** option.

A recommended approach is to add calls to heap-checking functions in places you suspect possible memory problems. If the heap turns out to be corrupted, you can rebuild with **-qheapdebug**.

Regardless of which debugging functions you choose, your program requires additional memory to maintain internal information for these functions. If you are using fixed-size heaps, you might have to increase the heap size in order to use the debugging functions.

Related information:

Chapter 18, “Memory debug library functions,” on page 147

Related information in the *XL C/C++ Compiler Reference*

 **-qheapdebug**

Functions for checking memory heaps

The header file `umalloc.h` declares a set of functions for validating user-created heaps. These functions are not controlled by a compiler option, so you can use them in your program at any time. Regular versions of these functions, without the `_u` prefix, are also available for checking the default heap. The heap-checking functions are summarized in the following table.

Table 14. Functions for checking memory heaps

Default heap function	User-created heap function	Description
<code>_heapchk</code>	<code>_uheapchk</code>	Checks the entire heap for minimal consistency.
<code>_heapset</code>	<code>_uheapset</code>	Checks the free memory in the heap for minimal consistency, and sets the free memory in the heap to a value you specify.
<code>_heap_walk</code>	<code>_uheap_walk</code>	Traverses the heap and provides information about each allocated or freed object to a callback function that you provide.

To compile an application that calls the user-created heap functions, see “Compiling and linking a program with user-created heaps” on page 31.

Functions for debugging memory heaps

Debug versions are available for both regular memory management functions and user-defined heap memory management functions. Each debug version performs the same function as its non-debug counterpart, and you can use them for any type of heap, including shared memory. Each call you make to a debug function also automatically checks the heap by calling `_heap_check` (described below), and provides information, including file name and line number, that you can use to debug memory problems. The names of the user-defined debug versions are prefixed by `_debug_u` (for example, `_debug_umalloc`), and they are defined in `umalloc.h`.

For a complete list and details about all of the debug memory management functions, see *Memory debug library functions*.

Table 15. Functions for debugging memory heaps

Default heap function	Corresponding user-created heap function
<code>_debug_calloc</code>	<code>_debug_ucalloc</code>
<code>_debug_malloc</code>	<code>_debug_umalloc</code>
<code>_debug_heapmin</code>	<code>_debug_uheapmin</code>
<code>_debug_realloc</code>	n/a
<code>_debug_free</code>	n/a

To use these debug versions, you can do either of the following:

- In your source code, prefix any of the default or user-defined-heap memory management functions with `_debug_`.
- If you do not want to make changes to the source code, simply compile with the `-qheapdebug` option. This option maps all calls to memory management functions to their debug version counterparts. To prevent a call from being mapped, parenthesize the function name.

To compile an application that calls the user-created heap functions, see “Compiling and linking a program with user-created heaps” on page 31.

Notes:

1. When the `-qheapdebug` option is specified, code is generated to *pre-initialize* the local variables for all functions. This makes it much more likely that uninitialized local variables will be found during the normal debug cycle rather than much later (usually when the code is optimized).
2. Do not use the `-brtl` option with `-qheapdebug`.
3. You should place a `#pragma strings` (readonly) directive at the top of each source file that will call debug functions, or in a common header file that each includes. This directive is not essential, but it ensures that the file name passed to the debug functions cannot be overwritten, and that only one copy of the file name string is included in the object module.

Additional functions for debugging memory heaps

Three additional debug memory management functions do not have regular counterparts. They are summarized in the following table.

Table 16. Additional functions for debugging memory heaps

Default heap function	Corresponding user-created heap function	Description
<code>_dump_allocated</code>	<code>_udump_allocated</code>	Prints information to stderr about each memory block currently allocated by the debug functions.
<code>_dump_allocated_delta</code>	<code>_udump_allocated_delta</code>	Prints information to file descriptor 2 about each memory block allocated by the debug functions since the last call to <code>_dump_allocated</code> or <code>_dump_allocated_delta</code> .
<code>_heap_check</code>	<code>_uheap_check</code>	Checks all memory blocks allocated or freed by the debug functions to make sure that no overwriting has occurred outside the bounds of allocated blocks or in a free memory block.

The `_heap_check` function is automatically called by the debug functions; you can also call this function explicitly. You can then use `_dump_allocated` or `_dump_allocated_delta` to display information about currently allocated memory blocks. You must explicitly call these functions.

Related information:

Chapter 18, “Memory debug library functions,” on page 147

Related information in the XL C/C++ Compiler Reference

 `-brtl`

 `-qheapdebug`

 `-qro / #pragma strings`

Using memory allocation fill patterns

Some debug functions set all the memory they allocate to a specified fill pattern. This lets you easily locate areas in memory that your program uses.

The `debug_malloc`, `debug_realloc`, and `debug_umalloc` functions set allocated memory to a default repeating `0xAA` fill pattern. To enable this fill pattern, export the `HD_FILL` environment variable.

The `debug_free` function sets all free memory to a repeating `0xFB` fill pattern.

Skipping heap checking

Each debug function calls `_heap_check` (or `_uheap_check`) to check the heap. Although this is useful, it can also increase your program's memory requirements and decrease its execution speed.

To reduce the overhead of checking the heap on every debug memory management function, you can use the `HD_SKIP` environment variable to control how often the functions check the heap. You will not need to do this for most of your applications unless the application is extremely memory intensive.

Set `HD_SKIP` like any other environment variable. The syntax for `HD_SKIP` is:

```
set HD_SKIP=increment, [start]
```

where:

increment

Specifies the number of debug function calls to skip between performing heap checks.

start Specifies the number debug function calls to skip before starting heap checks.

Note: The comma separating the parameters is optional.

For example, if you specify:

```
set HD_SKIP=10
```

then every tenth debug memory function call performs a heap check. If you specify:

```
set HD_SKIP=5,100
```

then after 100 debug memory function calls, only every fifth call performs a heap check.

When you use the *start* parameter to start skipping heap checks, you are trading off heap checks that are done implicitly against program execution speed. You should therefore start with a small increment (like 5) and slowly increase until the application is usable.

Using stack traces

Stack contents are traced for each allocated memory object. If the contents of an object's stack change, the traced contents are dumped.

The trace size is controlled by the HD_STACK environment variable. If this variable is not set, the compiler assumes a stack size of 10. To disable stack tracing, set the HD_STACK environment variable to 0.

Chapter 6. Using C++ constructors

► C++0x

Before C++0x, common initialization in multiple constructors of the same class could not be concentrated in one place in a robust, maintainable manner. A basic approach can solve this problem:

Using delegating constructors:

With the delegating constructors feature, you can concentrate common initializations in one constructor, which can make program more readable and maintainable. Delegating constructors help reduce the code size and collective size of the object files. For more information, see "Using delegating constructors (C++0x)."

Related information in the XL C/C++ Compiler Reference

 -qlanglvl

Using delegating constructors (C++0x)

Note: C++0x is a new version of the C++ programming language standard. IBM continues to develop and implement the features of the new standard. The implementation of the language level is based on IBM's interpretation of the standard. Until IBM's implementation of all the features of the C++0x standard is complete, including the support of a new C++ standard library, the implementation may change from release to release. IBM makes no attempt to maintain compatibility, in source, binary, or listings and other compiler interfaces, with earlier releases of IBM's implementation of the new features of the C++0x standard and therefore they should not be relied on as a stable programming interface.

Syntactically, *delegating constructors* and *target constructors* present the same interface as other constructors, see "Delegating constructors (C++0x)" in the *XL C/C++ Language Reference*.

Consider the following points when you use the delegating constructors feature:

- Call the target constructor implementation in such a way that virtual bases, direct nonvirtual bases, class members and additional ABI artifacts are initialized by target constructor as appropriate.
- Respond to the exception thrown in the body of a delegating constructor by calling the destructor implementation on the object that is constructed through the target constructor. The destructor implementation must be called in such a way that it calls the destructors of subobjects as appropriate. In particular, it must call the destructors for virtual base classes if the virtual base classes are created through the target constructor.
- Perform proper construction and destruction when initializing static objects with delegating constructors and on termination of a program that does such initialization.
- When an exception is thrown, a corresponding destructor must be called. Otherwise, virtual bases may have their destructors called more than once or not at all. With a delegating constructor, the call to the target constructor does not necessarily match a specific destructor implementation.

- The feature has minimal impact on compile time and run time performance. However, use of default arguments with an existing constructor is recommended in place of a delegating constructor where possible. Without inlining and interprocedural analysis, run time performance may degrade because of function call overhead and increased opacity.

Related information in the *XL C/C++ Compiler Reference*

 **-qlanglvl**

Related information in the *XL C/C++ Language Reference*

 Delegating constructors (C++0x)

Chapter 7. Using C++ templates

In C++, you can use a template to declare a set of related:

- Classes (including structures)
- Functions
- Static data members of template classes

Reducing redundant template instantiations

Within an application, you can instantiate the same template multiple times with the same arguments or with different arguments. If you use the same arguments, the repeated instantiations are redundant. These redundant instantiations increase compilation time, increase the size of the executable, and deliver no benefit.

There are several basic approaches to the problem of redundant instantiations:

Handling redundancy during linking

The size increase of the final executable might be small enough that it does not justify changing the way you compile your program or modify the source file. Most linkers have some form of garbage collection functionality. On AIX, the linker performs garbage collection well, especially when you use the `-qfuncsect` option. If you use `-qmplinst=always` or `-qmplinst=auto` without using `-qtemplateregistry` or `-qtempinc`, no compile time management of redundant instantiations is done. In this case, you can use the `-qfuncsect` option to reduce the executable size. For details, see `-qfuncsect` in the *XL C/C++ Compiler Reference*.

Controlling implicit instantiation in the source code

Concentrating implicit instantiations of a specialization: Organize your source code so that object files contain fewer instances of each required instantiation and fewer unused instantiations. This is the least usable approach, because you must know where each template is defined, which instantiations are used, and where to declare an explicit instantiation for each instantiation.

C++0x **Using explicit instantiation declarations:** With the explicit instantiation declarations feature, you can suppress the implicit instantiation of a template specialization or its members. This helps reduce the collective size of the object files. It might also reduce the size of the final executable if the suppressed symbol definitions are meant to be found in a shared library, or if the system linker is unable to always remove additional definitions of a symbol. For more information, see “Using explicit instantiation declarations (C++0x)” on page 49.

Note: If you want to control implicit instantiation in the source code, or use explicit instantiation declarations, you can use the `-qmplinst=none` or `-qmplinst=noinlines` option to prevent accidental implicit instantiations from occurring.

Having the compiler store instantiation information in a registry

Use the `-qtemplateregistry` compiler option. Information about each template instantiation is stored in a template registry. If the compiler is asked to instantiate the same template again with the same arguments, it

points to the instantiation in the first object file instead. This approach is described in “Using the `-qtemplateregistry` compiler option” on page 48.


Having the compiler store instantiations in a template include directory

Use the `-qtempinc` compiler option. If the template definition and implementation files have the required structure, each template instantiation is stored in a template include directory. If the compiler is asked to instantiate the same template again with the same arguments, it uses the stored version instead. The source file created in the template include directory is compiled during the link step recursively until all instantiations are done. This approach is described in “Using the `-qtempinc` compiler option.”

Notes:

- The `-qtempinc` and `-qtemplateregistry` compiler options are mutually exclusive.
- `-qtemplateregistry` is a better approach than `-qtempinc` for the following reasons:
 - `-qtemplateregistry` provides better benefits than `-qtempinc`.
 - `-qtemplateregistry` does not require modifications to the header files.

The compiler generates code for an implicit instantiation unless one of the following conditions is true:

- You use either `-qtmplinst=none` or `-qtmplinst=noinlines`.
- You use `-qtmplinst=auto`, which is the default suboption of `-qtmplinst` with `-qnotemplateregistry`.
- You use `-qtmplinst=auto` with `-qtempinc` and the template source that is organized to use `-qtempinc`.
-  An explicit instantiation declaration for that instantiation is in the current translation unit.

Related information in the *XL C/C++ Compiler Reference*

 `-qfuncsect`

 `-qtempinc` (C++ only)

 `-qtemplateregistry` (C++ only)

 `-qtmplinst` (C++ only)

 `-qlanglvl`

Using the `-qtempinc` compiler option

To use `-qtempinc`, you must structure your application as follows:

- Declare your class templates and function templates in template header files, with a `.h` extension.
- For each template declaration file, create a template implementation file. This file must have the same file name as the template declaration file and an extension of `.c` or `.t`, or the name must be specified in a **`#pragma implementation`** directive. For a class template, the implementation file defines the member functions and static data members. For a function template, the implementation file defines the function.
- In your source program, specify an `#include` directive for each template declaration file.

- Optionally, to ensure that your code is applicable for both **-qtempinc** and **-qnotempinc** compilations, in each template declaration file, conditionally include the corresponding template implementation file if the `__TEMPINC__` macro is *not* defined. (This macro is automatically defined when you use the **-qtempinc** compilation option.) This produces the following results:
 - Whenever you compile with **-qnotempinc**, the template implementation file is included.
 - Whenever you compile with **-qtempinc**, the compiler does not include the template implementation file. Instead, the compiler looks for a file with the same name as the template implementation file and extension `.c` the first time it needs a particular instantiation. If the compiler subsequently needs the same instantiation, it uses the copy stored in the template include directory.

Note: You can also use **-qtemplateregistry** that provides more benefits than **-qtempinc**, and does not require modifications to your source files. For details, see "**-qtemplateregistry (C++ only)**" in the *XL C/C++ Compiler Reference*.

Related information in the XL C/C++ Compiler Reference



-qtempinc (C++ only)



-qtemplateregistry (C++ only)



-qtmplinst (C++ only)



#pragma implementation (C++ only)

Example of **-qtempinc**

This example includes the following source files:

- A template declaration file: `stack.h`.
- The corresponding template implementation file: `stack.c`.
- A function prototype: `stackops.h` (not a function template).
- The corresponding function implementation file: `stackops.cpp`.
- The main program source file: `stackadd.cpp`.

In this example:

- Both source files include the template declaration file `stack.h`.
- Both source files include the function prototype `stackops.h`.
- The template declaration file conditionally includes the template implementation file `stack.c` if the program is compiled with **-qnotempinc**.

Template declaration file: `stack.h`

You must follow these steps if you want **-qtempinc** to manage the instantiations for the templates in `stack.h`:

1. Take the two template implementation definitions that start with `template <class Item, int size>` in the following example and place them in a `.c` or `.t` file.
2. Include `#ifndef __TEMPINC__` in the `stack.h` file.

Note: This header file compiles with **-qtempinc** but does not remove redundant instantiations. This is because the template code is not organized for **-qtempinc**.

Consider the following example that compiles successfully but does not use **-qtempinc** to manage implicit instantiations :

```
#ifndef STACK_H
#define STACK_H

template <class Item, int size> class Stack {
public:
    void push(Item item); // Push operator
    Item pop();          // Pop operator
    int isEmpty(){
        return (top==0); // Returns true if empty, otherwise false
    }
    Stack() { top = 0; } // Constructor defined inline
private:
    Item stack[size];   // The stack of items
    int top;            // Index to top of stack
};

template <class Item, int size>
void Stack<Item,size>::push(Item item) {
    if (top >= size) throw size;
    stack[top++] = item;
}

template <class Item, int size>
Item Stack<Item,size>::pop() {
    if (top <= 0) throw size;
    Item item = stack[--top];
    return(item);
}

#endif
```

Here is the revised example that compiles successfully with **-qtempinc**:

```
#ifndef STACK_H
#define STACK_H
#ifdef __TEMPINC__

template <class Item, int size> class Stack {
public:
    void push(Item item); // Push operator
    Item pop();          // Pop operator
    int isEmpty(){
        return (top==0); // Returns true if empty, otherwise false
    }
    Stack() { top = 0; } // Constructor defined inline
private:
    Item stack[size];   // The stack of items
    int top;            // Index to top of stack
};

template <class Item, int size>
void Stack<Item,size>::push(Item item) {
    if (top >= size) throw size;
    stack[top++] = item;
}

template <class Item, int size>
Item Stack<Item,size>::pop() {
    if (top <= 0) throw size;
    Item item = stack[--top];
    return(item);
}

#endif
```

Function declaration file: stackops.h

This header file contains the prototype for the add function, which is used in both stackadd.cpp and stackops.cpp.

```
#ifndef STACKOPS_H
#define STACKOPS_H
#include "stack.h"
void add(Stack<int, 50>& s);
#endif
```

Function implementation file: stackops.cpp

This file provides the implementation of the add function, which is called from the main program.

```
#include "stack.h"
#include "stackops.h"

void add(Stack<int, 50>& s) {
    int tot = s.pop() + s.pop();
    s.push(tot);
    return;
}
```

Main program file: stackadd.cpp

This file creates a Stack object.

```
#include <iostream>
#include "stack.h"
#include "stackops.h"

main() {
    Stack<int, 50> s;           // create a stack of ints
    int left=10, right=20;
    int sum;

    s.push(left);             // push 10 on the stack
    s.push(right);           // push 20 on the stack
    add(s);                   // pop the 2 numbers off the stack
                                // and push the sum onto the stack
    sum = s.pop();           // pop the sum off the stack

    cout << "The sum of: " << left << " and: " << right << " is: " << sum << endl;

    return(0);
}
```

Regenerating the template instantiation file

The compiler builds a template instantiation file in the TEMPINC directory corresponding to each template implementation file. With each compilation, the compiler can add information to the file but it never removes information from the file.

As you develop your program, you might remove template function references or reorganize your program so that the template instantiation files become obsolete. You can periodically delete the TEMPINC destination and recompile your program.

Using `-qtempinc` with shared libraries

In a traditional application development environment, different applications can share both source files and compiled files. When you use templates, applications can share source files but cannot share compiled files.

If you use `-qtempinc`:

- Each application must have its own TEMPINC destination.
- You must compile all of the source files for the application, even if some of the files have already been compiled for another application.

Using the `-qtemplateregistry` compiler option

The template registry uses a "first-come first-served" algorithm:

- When a compiler performs an implicit instantiation for the first time, it is instantiated in the compilation unit in which it occurs.
- When another compilation unit performs the same implicit instantiation, it is not instantiated. Thus, only one copy is generated for the entire program.

The instantiation information is stored in a template registry file. You must use the same template registry file for the entire program. Two programs cannot share a template registry file.

The default file name for the template registry file is `templateregistry`, but you can specify any other valid file name to override this default. When cleaning your program build environment before starting a fresh or scratch build, you must delete the registry file along with the old object files.

You can perform multiple compilations in parallel using the same template registry file with minimal impact on compile time.

When you recompile your program, the information in the template registry file is also used to determine whether a recompilation of a source file might introduce link errors because of missing template instantiations. If the following conditions are true, the compiler will schedule the recompilation of one or more source files when you recompile a source file:

- The source file instantiated a template that other source files instantiated.
- The source file was chosen by the template registry to actually instantiate the template.
- The source file no longer instantiates the template.

When the preceding conditions are all true, the compiler chooses another source file to instantiate the template in. That file is scheduled for recompilation during the link step. If you happen to recompile a source file that is scheduled to be recompiled during the link step, the scheduled recompilation is cancelled.

You can use `-qnotemplatercompile` to disable the scheduled recompilation during the link step. For details, see "`-qtemplatercompile (C++ only)`" in the *XL C/C++ Compiler Reference*.

Related information in the *XL C/C++ Compiler Reference*

 `-qtempinc` (C++ only)

 `-qtemplatercompile` (C++ only)

 `-qtemplateregistry` (C++ only)

Recompiling related compilation units

If two compilation units, A and B, reference the same instantiation, the `-qtemplateregistry` compiler option has the following effect:

- If you compile A first, the object file for A contains the code for the instantiation.
- When you later compile B, the object file for B does not contain the code for the instantiation because object A already does.
- If you later change A so that it no longer references this instantiation, the reference in object B would produce an unresolved symbol error. When you recompile A, the compiler detects this problem and handles it as follows:
 - If the `-qtemplaterecompile` compiler option is in effect, the compiler automatically recompiles B during the link step, using the same compiler options that were specified for A. (Note, however, that if you use separate compilation and linkage steps, you need to include the compilation options in the link step to ensure the correct compilation of B.)
 - If the `-qnotemplaterecompile` compiler option is in effect, the compiler issues a warning and you must manually recompile B.

Related information in the *XL C/C++ Compiler Reference*

 `-qtemplateregistry` (C++ only)

 `-qtemplaterecompile` (C++ only)

Switching from `-qtempinc` to `-qtemplateregistry`

Because the `-qtemplateregistry` compiler option does not impose any restrictions on the file structure of your application, it has less administrative overhead than `-qtempinc`. You can make the switch as follows:



- If your application compiles successfully with both `-qtempinc` and `-qnotempinc`, you do not need to make any changes.
- If your application compiles successfully with `-qtempinc` but not with `-qnotempinc`, you must change it so that it will compile successfully with `-qnotempinc`. In each template definition file, conditionally include the corresponding template implementation file if the `__TEMPINC__` macro is not defined. This is illustrated in “Example of `-qtempinc`” on page 45.

Using explicit instantiation declarations (C++0x)

Note: C++0x is a new version of the C++ programming language standard. IBM continues to develop and implement the features of the new standard. The implementation of the language level is based on IBM's interpretation of the standard. Until IBM's implementation of all the features of the C++0x standard is complete, including the support of a new C++ standard library, the implementation may change from release to release. IBM makes no attempt to maintain compatibility, in source, binary, or listings and other compiler interfaces, with earlier releases of IBM's implementation of the new features of the C++0x standard and therefore they should not be relied on as a stable programming interface.

Syntactically, an *explicit instantiation declaration* is an *explicit instantiation definition* preceded by the extern keyword, see “Explicit instantiation (C++ only)” in the *XL C/C++ Language Reference*.

Consider the following points when you use the explicit instantiation declarations feature:

-  An explicit instantiation declaration of a class template specialization does not cause implicit instantiation of said specialization.
- In a translation unit, if a user-defined inline function is subject to an explicit instantiation declaration and not subject to an explicit instantiation definition:
 - Implicit instantiation of said function occurs no matter whether it is inlined or not.
 -  No out-of-line copy of the function is generated in that translation unit no matter whether the compiler option **-qkeepinlines** is specified or not.

Note: This does not limit the behavior for functions that are implicitly generated by the compiler. Implicitly declared special members such as the default constructor, copy constructor, destructor and copy assignment operator are inline and the compiler might instantiate them. In particular, out-of-line copies might be generated.

- The degradation of the amount of inlining achieved on functions that are not inline and are subject to explicit instantiation declarations might occur.
- When a non-pure virtual member function is subject to an explicit instantiation declaration, either directly or through its class, the virtual member function must be subject to an explicit instantiation definition somewhere in the entire program. Otherwise, an unresolved symbol error might result at link time.
- When implicit instantiation of a class template specialization is allowed, the user program must be written as if the implicit instantiation of all virtual member functions of that class specialization occurs. Otherwise, an unresolved symbol error for a virtual member function might result at link time.
- When implicit instantiation of a class template specialization is allowed and the specialization is subject to an explicit instantiation declaration, the class template specialization must be subject to an explicit instantiation definition somewhere in the user program. Otherwise, an unresolved symbol error might result at link time.

Related information in the *XL C/C++ Compiler Reference*

 **-qtempinc** (C++ only)

 **#pragma implementation** (C++ only)

 **-qlanglvl**

Related information in the *XL C/C++ Language Reference*

 **Explicit instantiation** (C++ only)

Chapter 8. Constructing a library

You can include static and shared libraries in your C and C++ applications.

“Compiling and linking a library” describes how to compile your source files into object files for inclusion in a library, how to link a library into the main program, and how to link one library into another.

“Initializing static objects in libraries (C++)” on page 54 describes how to use priorities to control the order of initialization of objects across multiple files in a C++ application.


“Dynamically loading a shared library” on page 59 describes two functions you can use in your application code to load, initialize, unload, and terminate a C++ shared library at runtime.

Related information:

 [Objects and libraries](#)

Compiling and linking a library

Related information:

 [Diagnosing link-time errors](#)

[Dynamic and static linking](#)

Compiling a static library

To compile a static (unshared) library:

1. Compile each source file into an object file, with no linking. For example:

```
xlc -c test.c example.c
```
2. Use the `ar` command to add the generated object files to an archive library file. For example:

```
ar -rv libex.a test.o example.o
```

Compiling a shared library

For use with dynamic linking

To compile a shared library that uses dynamic linking:

1. Compile each source file into an object file, with no linking. For example:

```
xlc -c test1.c -o test1.o
```
2. Optional: Create an export file listing the global symbols to be exported, by doing one of the following:
 - Use the **CreateExportList** utility, described in “Exporting symbols with the CreateExportList utility” on page 53.
 - Use the **-qexpfile** compiler option with the **-qmkshrobj** option. The **-qexpfile** option saves all exported symbols from a list of given object files in a designated file. For example:

```
xlc -qmkshrobj -qexpfile=exportlist test1.o test2.o
```
 - Manually create an export file using a text editor. You can edit an export file to include or exclude global symbols from the target shared library.


3. Use the **-qmkshrobj** option to create a shared library from the generated object files.
 - If you created an export file in step 2, use the **-bE** linker option to use your global symbol export list. If you do not specify a **-bE** option, all symbols are exported.
 - If you are creating a shared library from C++ object files, you can also assign an initialization priority to the shared library, as described in “Assigning priorities to objects” on page 55.

For example:

```
xlc -qmkshrobj -o mySharedObject.o test1.o test2.o -bE:exportlist
```

Notes:

- The default name of the shared object is **shr.o**, unless you use the **-o** option to specify another name.
- Exporting some functions (such as `restf#` where `#` is a number) might cause incorrect execution.

 Users without XL C/C++ installed can use the **makeC++SharedLib** utility described in “Creating a shared library with the `makeC++SharedLib` utility” on page 68 to create a shared library from C++ object files. This method is not recommended for compiler users. The **-qmkshrobj** option is preferred because it has several advantages, including the ability to automatically handle C++ template instantiation, and compatibility with the **-O5** optimization option.



4. Optional: Use the AIX **ar** command to produce an archive library file from multiple shared or static objects. For example:


```
ar -rv libtest.a mySharedObject.o myStaticObject.o
```
5. Link the shared library to the main application, as described in “Linking a library to an application” on page 54.

For use with runtime linking

To create a shared library that uses runtime linking:

1. Follow steps 1 and 2 in the procedure described above.
2. Use the **-G** option to create a shared library from the generated object files, and to enable runtime linking with applications that support it.
 - If you created an export file, use the **-bE** linker option to use your global symbol export list. If you do not specify a **-bE** option, all symbols are exported.
 - If you are creating a shared library from C++ object files, you can also assign an initialization priority to the shared library, as described in “Assigning priorities to objects” on page 55.

For example:


```
xlc -G -o libtest.so test1.o test2.o -bE:exportlist
```

3. Link the shared library to the main application using the **-brtl** option, as described in “Linking a library to an application” on page 54.

Dynamic loading of a shared library

Shared libraries built for either dynamic or runtime linking can be dynamically loaded. See the AIX documentation for more information about using the dynamic loading routines:






- dlopen
- dlclose
- dlerror
- loadAndInit
- loadbind
- loadquery
- terminateAndUnload

 If you want the system to perform static initialization when dynamically loading a shared library, use the load and unload functions described in “Dynamically loading a shared library” on page 59.

Related external information

- *AIX Linking and Loading Mechanisms*
- AIX Linking 101
- AIX Linking 102
- **ar** and **ld** in the *AIX Commands Reference, Volumes 1 - 6*
- Shared Objects and Runtime Linking

Related information in the *XL C/C++ Compiler Reference*

-  -qexpfile
-  -qmkshrobj
-  -O, -qoptimize
-  -G
-  -brtl

Exporting symbols with the CreateExportList utility

CreateExportList is a shell script that creates a file containing a list of all the global symbols found in a given set of object files. Note that this command is run automatically when you use the **-qmkshrobj** option, unless you specify an alternative export file with the **-qexpfile** command.

The syntax of the **CreateExportList** command is as follows:

```

▶▶ CreateExportList [_r] exp_list [-f file_list] [obj_files] [-w] [-x 32 64]

```

You can specify one or more of the following options:

-r If specified, template prefixes are pruned. The resource file symbol (`__rsrc`) is not added to the resource list.

exp_list

The name of a file that will contain a list of global symbols found in the object files. This file is overwritten each time the **CreateExportList** command is run.

-f*file_list*

The name of a file that contains a list of object file names.

obj_files

One or more names of object files.

- w** Excludes weak symbols from the export list.
- X32** Generates names from 32-bit object files in the input list specified by *-f file_list* or *obj_files*. This is the default.
- X64** Generates names from 64-bit object files in the input list specified by *-f file_list* or *obj_files*.

The **CreateExportList** command creates an empty list if any of the following are true:

- No object files are specified by either *-f file_list* or *obj_files*.
- The file specified by the *-f file_list* parameter is empty.

Linking a library to an application

You can use the same command string to link a static or shared library to your main program. For example:

```
xlc -o myprogram main.c -Ldirectory -ltest
```

where *directory* is the path to the directory containing the library *libtest.a*.

If your library uses runtime linking, add the **-brtl** option to the command:

```
xlc -brtl -o myprogram main.c -Ldirectory -ltest
```

By using the **-l** option, you instruct the linker to search in the directory specified via the **-L** option for *libtest.so*; if it is not found, the linker searches for *libtest.a*. For additional linkage options, including options that modify the default behavior, see the AIX **ld** documentation (<http://publib.boulder.ibm.com/infocenter/aix/v7r1/index.jsp?topic=/com.ibm.aix.cmds/doc/aixcmds3/ld.htm>).

Related information in the XL C/C++ Compiler Reference

 **-l**

 **-L**

 **-brtl**

Linking a shared library to another shared library

Just as you link modules into an application, you can create dependencies between shared libraries by linking them together. For example:

```
xlc -qmkshrobj -o mylib.so myfile.o -Ldirectory -ltest
```

Related information in the XL C/C++ Compiler Reference

 **-qmkshrobj**

 **-L**

Initializing static objects in libraries (C++)

The C++ language definition specifies that, before the `main` function in a C++ program is executed, all objects with constructors, from all the files included in the program must be properly constructed. Although the language definition specifies the order of initialization for these objects within a file (which follows the order in which they are declared), it does not specify the order of initialization for these

objects across files and libraries. You might want to specify the initialization order of static objects declared in various files and libraries in your program.

To specify an initialization order for objects, you assign relative priority numbers to objects. The mechanisms by which you can specify priorities for entire files or objects within files are discussed in “Assigning priorities to objects.” The mechanisms by which you can control the initialization order of objects across modules are discussed in “Order of object initialization across libraries” on page 57.

Related information:

“Assigning priorities to objects”

“Order of object initialization across libraries” on page 57

Assigning priorities to objects

You can assign a priority number to objects and files within a single library, and the objects will be initialized at run time according to the order of priority. However, because of the differences in the way modules are loaded and objects initialized on the different platforms, the levels at which you can assign priorities vary among the different platforms, as follows:

Set the priority level for an entire file

To use this approach, you specify the **-qpriority** compiler option during compilation. By default, all objects within a single file are assigned the same priority level, and are initialized in the order in which they are declared, and terminated in reverse declaration order.

Set the priority level for objects within a file

To use this approach, you include **#pragma priority** directives in the source files. Each **#pragma priority** directive sets the priority level for all objects that follow it, until another pragma directive is specified. Within a file, the first **#pragma priority** directive must have a higher priority number than the number specified in the **-qpriority** option (if it is used), and subsequent **#pragma priority** directives must have increasing numbers. While the relative priority of objects within a single file will remain the order in which they are declared, the pragma directives will affect the order in which objects are initialized across files. The objects are initialized according to their priority, and terminated in reverse priority order.

On AIX only, you can additionally set the priority of an entire shared library, by using the **priority** suboption of the **-qmksprobj** compiler option. As loading and initialization on AIX occur as separate processes, priority numbers assigned to files (or to objects within files) are entirely independent of priority numbers assigned to libraries, and do not need to follow any sequence.

Using priority numbers

Priority numbers can range from -2147483648 to 2147483647. However, numbers from -2147483648 to -2147482625 are reserved for system use. The smallest priority number that you can specify, -2147482624, is initialized first. The largest priority number, 2147483647, is initialized last. If you do not specify a priority level, the default priority is 0 (zero).

The examples below show how to specify the priority of objects within a single file, and across two files. “Order of object initialization across libraries” on page 57 provides detailed information on the order of initialization of objects.

Example of object initialization within a file

The following example shows how to specify the priority for several objects within a source file.

```
...
#pragma priority(2000) //Following objects constructed with priority 2000
...

static Base a ;

House b ;
...
#pragma priority(3000) //Following objects constructed with priority 3000
...

Barn c ;
...
#pragma priority(2500) // Error - priority number must be larger
                        // than preceding number (3000)
...
#pragma priority(4000) //Following objects constructed with priority 4000
...

Garage d ;
...
```

Example of object initialization across multiple files

The following example describes the initialization order for objects in two files, farm.C and zoo.C. Both files are contained in the same shared module, and use the **-qpriority** compiler option and **#pragma priority** directives.

```
farm.C -qpriority=1000                                zoo.C -qpriority=2000
...
Dog a ;
Dog b ;
...
#pragma priority(6000)
...
Cat c ;
Cow d ;
...
#pragma priority(7000)
Mouse e ;
...

...
Bear m ;
...
#pragma priority(5000)
...
Zebra n ;
Snake s ;
...
#pragma priority(8000)
Frog f ;
...
```

At runtime, the objects in these files are initialized in the following order:

Sequence	Object	Priority value	Comment
1	Dog a	1000	Takes option priority (1000).
2	Dog b	1000	Follows with the same priority.
3	Bear m	2000	Takes option priority (2000).
4	Zebra n	5000	Takes pragma priority (5000).
5	Snake s	5000	Follows with same priority.
6	Cat c	6000	Next priority number.
7	Cow d	6000	Follows with same priority.

Sequence	Object	Priority value	Comment
8	Mouse e	7000	Next priority number.
9	Frog f	8000	Next priority number (initialized last).

Related information in the *XL C/C++ Compiler Reference*

 `-qpriority / #pragma priority` (C++ only)

 `-qmkshrobj`

Order of object initialization across libraries

At run time, once all modules in an application have been loaded, the modules are initialized in their order of priority (the executable program containing the main function is always assigned a priority of 0). When objects are initialized within a library, the order of initialization follows the rules outlined in “Assigning priorities to objects” on page 55. If objects do not have priorities assigned, or have the same priorities, object files are initialized in random order, and the objects within the files are initialized according to their declaration order. Objects are terminated in reverse order of their construction.

For objects assigned the same priorities, if you want to control their initialization order, you can use the `-Wm` option to do so. `-Wm` with the `-c` suboption specifies that object files with the same priority are to be initialized in link order — where link order is the order in which the files were given on the command line during linking into the library — and the static objects within the files are initialized according to their declaration order. `-Wm` with the `-r` suboption specifies that the object files with the same priority are to be initialized in reverse link order.

Example of object initialization across libraries

In this example, the following modules are used:

- `main.out`, the executable containing the main function
- `libS1` and `libS2`, two shared libraries
- `libS3` and `libS4`, two shared libraries that are dependencies of `libS1`
- `libS5` and `libS6`, two shared libraries that are dependencies of `libS2`

The source files are compiled into object files with the following command strings:

```
x1C -qpriority=101 -c fileA.C -o fileA.o
x1C -qpriority=150 -c fileB.C -o fileB.o
x1C -c fileC.C -o fileC.o
x1C -c fileD.C -o fileD.o
x1C -c fileE.C -o fileE.o
x1C -c fileF.C -o fileF.o
x1C -qpriority=300 -c fileG.C -o fileG.o
x1C -qpriority=200 -c fileH.C -o fileH.o
x1C -qpriority=500 -c fileI.C -o fileI.o
x1C -c fileJ.C -o fileJ.o
x1C -c fileK.C -o fileK.o
x1C -qpriority=600 -c fileL.C -o fileL.o
```

The dependent libraries are created with the following command strings:

```

x1C -qmkshrobj=50 -o libS3.a fileE.o fileF.o
x1C -qmkshrobj=-600 -o libS4.a fileG.o fileH.o
x1C -qmkshrobj=-200 -o libS5.a fileI.o fileJ.o
x1C -qmkshrobj=-150 -o libS6.a fileK.o fileL.o

```

The dependent libraries are linked with their parent libraries using the following command strings:

```

x1C -qmkshrobj=-300 -o libS1.a fileA.o fileB.o -L. -lS3 -lS4
x1C -qmkshrobj=100 -o libS2.a fileC.o fileD.o -L. -lS5 -lS6

```

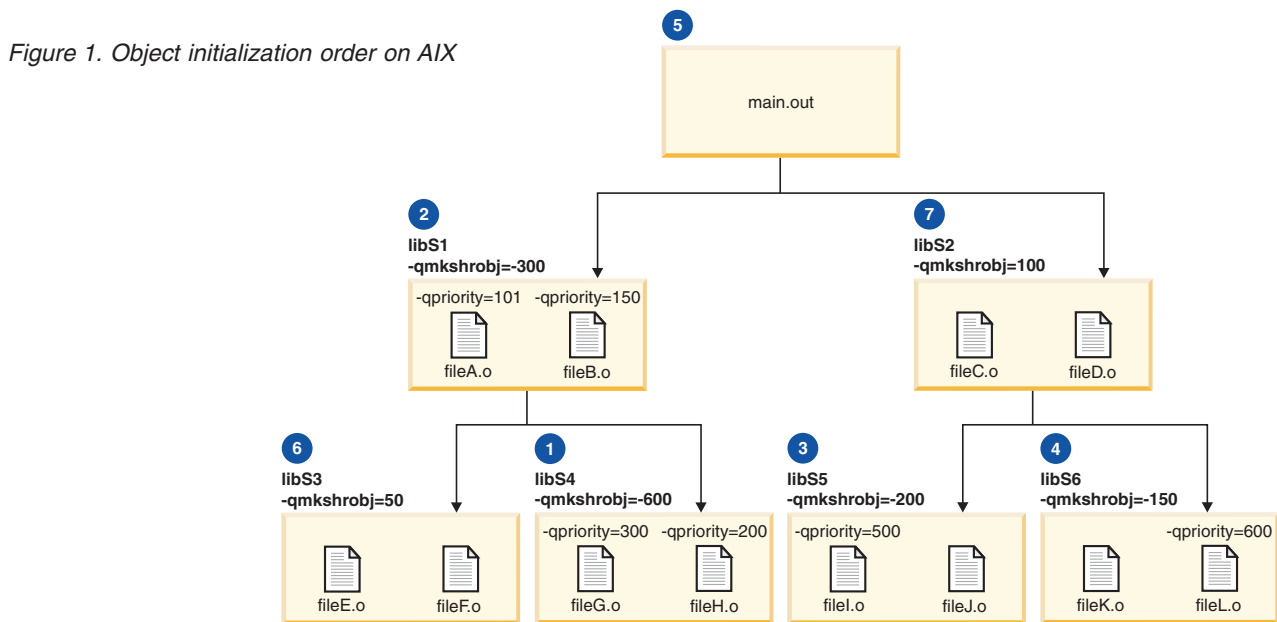
The parent libraries are linked with the main program with the following command string:

```

x1C main.C -o main.out -L. -R. -lS1 -lS2

```

The following diagram shows the initialization order of the objects in the shared libraries.



First, the shared libraries are initialized, in the following order:

Sequence	Object	Priority value	Comment
1	libS4	-600	Initialized first (lowest priority number).
2	libS1	-300	Initialized next (next priority number).
3	libS5	-200	Initialized next (next priority number).
4	libS6	-150	Initialized next (next priority number).
5	main.out	0	Initialized next (next priority number). The main program always has a priority of 0.
6	libS3	50	Initialized next (next priority number).
7	libS2	100	Initialized last (next priority number).

Then, each of the files is initialized, in the following order:

Sequence	Object	Priority value	Comment
8	fileH	200	Initialized first (contained in libS4; lowest priority number).
9	fileG	300	Initialized next (contained in libS4; next priority number).
10	fileA	101	Initialized next (contained in libS1; lowest priority number).
11	fileB	150	Initialized next (contained in libS1; next priority number).
12	fileJ	0	Initialized next (contained in libS5; lowest priority number).
13	fileI	500	Initialized next (contained in libS5; next priority number).
14	fileK	0	Initialized next (contained in libS6; lowest priority number).
15	fileL	600	Initialized next (contained in libS6; next priority number).
16			Objects in main are initialized according to their priority.
17	fileE, fileF	0, 0	Initialized next, in random order (contained in libS3; same priority number).
18	fileC, fileD	0, 0	Initialized next, in random order (contained in libS2; same priority number).

Related information in the *XL C/C++ Compiler Reference*

 -qmkshrobj

 -W

Dynamically loading a shared library

If you want to programmatically control the loading and initialization of C++ objects contained in shared libraries, you can use two functions provided by XL C/C++: `loadAndInit` and `terminateAndUnload`. These functions are declared in the header file `load.h`, and you can call them from the main program to load, initialize, terminate, and unload any named shared library. These functions work in the same way, take the same parameters, and return the same values and error codes as the AIX `load` and `unload` routines, respectively, but they additionally perform initialization of C++ objects. See the `load` and `unload` routines in the *Technical Reference: Base Operating System and Extensions, Volumes 1 & 2* for more information.

Note: For portability, you might wish to use the POSIX `dlopen` and `dclose` functions, which also perform initialization and termination, and interact correctly with `loadAndInit` and `terminateAndUnload`. For more information on `dlopen` and `dclose`, see the *Technical Reference: Base Operating System and Extensions, Volumes 1 & 2*.

Related information:

“Loading and initializing a module with the loadAndInit function”

“Terminating and unloading a module with the terminateAndUnload function” on page 61

Loading and initializing a module with the loadAndInit function

The loadAndInit function takes the same parameters and returns the same values and error codes as the load routine. See the load routine in the *Technical Reference: Base Operating System and Extensions, Volumes 1 & 2* for more information.

Format

```
#include <load.h>
int (*loadAndInit(char *FilePath, unsigned int Flags, char *LibraryPath))();
```

Description

The loadAndInit function calls the AIX load routine to load the specified module (shared library) into the calling process's address space. If the shared library is loaded successfully, any C++ initialization is performed. The loadAndInit function ensures that a shared library is only initialized once, even if dlopen is used to load the library too. Subsequent loads of the same shared library will not perform any initialization of the shared library.

If loading a shared library results in other shared libraries being loaded, the initialization for those shared libraries will also be performed (if it has not been previously). If loading a shared library results in the initialization of multiple shared libraries, the order of initialization is determined by the priority assigned to the shared libraries when they were built. Shared libraries with the same priority are initialized in random order.

To terminate and unload the shared library, use the terminateAndUnload function, described below.

Do not reference symbols in the C++ initialization that need to be resolved by a call to the AIX loadbind routine, since the loadbind routine normally is not called until after the loadAndInit function returns.

Parameters

FilePath

Points to the name of the shared library being loaded, or to the member of an archive. If you specify a relative or full path name (that is, a name containing one or more / characters), the file is used directly, and no search of directories specified in the *LibraryPath* is performed. If you specify a base name (that is, a name containing no / characters), a search is performed of the directory you specify in the *LibraryPath* parameter (see below).

Flags Modifies the behavior of loadAndInit. If no special behavior is required, set the value to 0 (or 1). The possible flags are:

L_LIBPATH_EXEC

Specifies that the library path used at program execution time be prepended to any library path specified in the loadAndInit call. You should use this flag.

L_NOAUTODEFER

Specifies that any deferred imports must be explicitly resolved by the use of the `loadbind` routine.

L_LOADMEMBER

Specifies that the *FilePath* is the name of a member in an archive. The format is *archivename.a(member)*.

LibraryPath

Points to the default library search path.

Return values

Upon successful completion, the `loadAndInit` function returns the pointer to function for the entry point (or data section) of the shared library.

If the `loadAndInit` function fails, a null pointer is returned, the module is not loaded or initialized, and the `errno` global variable is set to indicate the error.

Terminating and unloading a module with the `terminateAndUnload` function

The `terminateAndUnload` function takes the same parameters and returns the same values and error codes as the `unload` routine. See the `unload` routine in *Technical Reference: Base Operating System and Extensions, Volumes 1 & 2* for more information.

Format

```
#include <load.h>
int terminateAndUnload(int (*FunctionPointer)());
```

Description

The `terminateAndUnload` function performs any C++ termination that is required and unloads the module (shared library). The function pointer returned by the `loadAndInit` routine is used as the parameter for the `terminateAndUnload` function. If this is the last time the shared library is being unloaded, any C++ termination is performed for this shared library and any other shared libraries that are being unloaded for the last time as well. The `terminateAndUnload` function ensures that the shared library is only terminated once, even if `dlopen` is used to unload the library too. The order of termination is the reverse order of initialization performed by the `loadAndInit` function. If any uncaught exceptions occur during the C++ termination, the termination is stopped and the shared library is unloaded.

If the `loadAndInit` function is called more times for a shared library than `terminateAndUnload`, the shared library will never have the C++ termination performed. If you rely on the C++ termination being performed at the time the `terminateAndUnload` function is called, ensure the number of calls to the `terminateAndUnload` function matches the number of calls to the `loadAndInit` function. If any shared libraries loaded with the `loadAndInit` function are still in use when the program exits, the C++ termination is performed.

If the `terminateAndUnload` function is used to unload shared libraries not loaded with the `loadAndInit` function, no termination will be performed.

Parameters

FunctionPointer

Specifies the name of the function returned by the `loadAndInit` function.

Return values

Successful completion of the `terminateAndUnload` function returns a value of 0, even if the C++ termination was not performed and the shared library was not unloaded because the shared library was still in use.

If the `terminateAndUnload` function fails, it returns a value of -1 and sets `errno` to indicate the error.

Chapter 9. Replacing operator new and operator delete in applications that use shared libraries (C++)

You can define your own versions of operator new() and operator delete() in C++ applications. In applications that use shared libraries, it may be useful for the shared library to use a user defined operator new() and operator delete() in the main application executable. You may want to do this if you want more control over memory management than if you use the default calls to these operators in the C++ Runtime Library libC.a. Enabling this facility in your applications requires using the runtime linking option **-brtl**, creating an export list with the mangled names for the operators you are defining, and building your applications with the correct link time option so that calls to operator new() and operator delete() are replaceable. The mangled names indicated by the export list are then available to the runtime environment so that libraries loaded at run time use your versions of operator new() and operator delete().

Follow these steps:

1. Write the code that defines your own versions of operator new() or operator delete(). For example, this program shows operator new() being defined:

```
#include <new>
#include <cstdio>
#include <cstdlib>
void* operator new(unsigned long x) {
    printf("operator new %ld\n", x);
    return malloc(x);
}

int main() {
    return 5;
}
```

2. Create an export list that contains the mangled name symbols for the operator you are defining. For new() and delete(), there are a limited number of name mangling possibilities when compiling with x1C. For example, depending on the exception handling specified with the **-qlanglvl=newexcp** option, different mangled names will be used. See Table 17 on page 64 for the list of possible mangled names.

As an aid to creating an export list, compile without linking the code that has your operator definitions; use the **nm** command on your object file to display the symbols the compiler is using in your object; then refer to Table 17 on page 64 to find the matching symbols. For example:

- a. Compile without linking:

```
x1C -c my_app.C
```

Creates my_app.o.

- b. Use the **nm** command to display the symbols in new.o

```
nm -epC my_app.o
```

The **nm** command displays a listing similar to this:

```
__nw__FU1      T      0
TOC            d      56
__nw__FU1      D      60      12
__nw__FU1      d      56      4
```

`__nw__FUI` is a valid symbol listed in Table 17. Add this symbol to your export list.

3. Link your application and use the `-bE` option to specify the export list you created that contains the mangled names for the operators you are defining. Also, specify the `-brtl` option so that the application uses runtime linking. For example:

```
xlc my_app.o -bE:my_app.exp -brtl
```

Where `my_app.exp` is the export file that you created in step 2.

Table 17. Mangled names for operator `new()`, operator `delete()`, vector `new`, and vector `delete`

	Mangled names
Operator <code>new</code> and vector <code>new</code> names when compiling with <code>-qlanglvl=nonewexp</code>	<ul style="list-style-type: none"> • <code>__nw__FUI</code> • <code>__nw__FUIPv</code> • <code>__nw__FUIRCQ2_3std9nothrow_t</code> • <code>__vn__FUI</code> • <code>__vn__FUIPv</code> • <code>__vn__FUIRCQ2_3std9nothrow_t</code>
Operator <code>new</code> and vector <code>new</code> names when compiling with <code>-qlanglvl=newexp</code>	<ul style="list-style-type: none"> • <code>__snw__FUI</code> • <code>__svn__FUI</code>
Operator <code>delete</code> names	<ul style="list-style-type: none"> • <code>__dl__FPv</code> • <code>__dl__FPvRCQ2_3std9nothrow_t</code> • <code>__dl__FPvT1</code> • <code>__dl__FPvUI</code> • <code>__vd__FPv</code> • <code>__vd__FPvRCQ2_3std9nothrow_t</code> • <code>__vd__FPvUI</code>

Related information in the *XL C/C++ Compiler Reference*

 `-qlanglvl`

 `-b`

 `-brtl`

Chapter 10. Using the C++ utilities

XL C/C++ ships with a set of additional utilities you can use for managing your C++ applications:

- A filter for demangling compiled symbol names in object files. Described in “Demangling compiled C++ names with `c++filt`.”
- A library of classes for demangling and manipulating mangled names. Described in “Demangling compiled C++ names with the demangle class library” on page 66.
- A distributable shell script for creating shared libraries from library files. Described in “Creating a shared library with the `makeC++SharedLib` utility” on page 68.
- A distributable shell script for linking C++ object files and archives. Described in “Linking with the `linkxlc` utility” on page 70.

Demangling compiled C++ names

When XL C/C++ compiles a C++ program, it encodes (mangles) all function names and certain other identifiers to include type and scoping information. The name mangling is necessary to accommodate overloading of C++ functions and operators. The linker uses these mangled names to resolve duplicate symbols and ensure type-safe linkage. These mangled names appear in the object files and final executable file.

Tools that can manipulate the files, the AIX dump utility for example, have only the mangled names and not the original source-code names, and present the mangled name in their output. This output might be undesirable because the names are no longer recognizable.

Two utilities convert the mangled names to their original source code names:

`c++filt` A filter that demangles (decodes) mangled names.

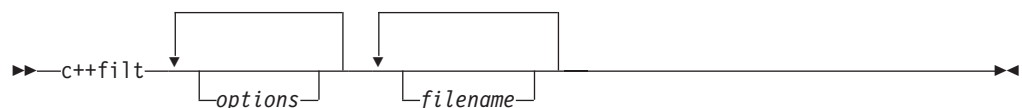
`demangle.h`

A class library that you can use to develop tools to manipulate mangled names.

Demangling compiled C++ names with `c++filt`

The `c++filt` utility is a filter that copies characters from file names or standard input to standard output, replacing all mangled names with their corresponding demangled names. You can use the filter directly with file name arguments, and the filter outputs the demangled names of all mangled names in the files; or you can use a shell command that inputs text, such as specific mangled names, and pipe it to the filter, so that the filter provides the demangled names of the names you specified.

`c++filt` command syntax



You can specify one or more of the following options:

- m** Produces a symbol map, containing a side-by-side listing of demangled names in the left column and their corresponding mangled names in the right column.
- s** Produces a continuous listing of each demangled name followed immediately by its mangled name.
- w *width*** Prints demangled names in fields *width* characters wide. If the name is shorter than *width*, it is padded on the right with blanks; if longer, it is truncated to *width*.
- C** Demangles standalone class names, such as Q2_1X1Y.
- S** Demangles special compiler-generated symbol names, such as __vft1X (represents a virtual function).

filename

Is the name of the file containing the mangled names you want to demangle. You can specify more than one file name.

For example, the following command shows the symbols contained in an object file `functions.o`, producing a side-by-side listing of the mangled and demangled names with a field width of 40 characters:

```
c++filt -m -w 40 functions.o
```

The output is displayed as follows:

C++ Symbol Mapping demangled:	Mangled:
Average::insertValue(double)	insertValue__7AverageFd
Average::getCount()	getCount__7AverageFv
Average::getTotal()	getTotal__7AverageFv
Average::getAverage()	getAverage__7AverageFv

The following command shows the demangled name immediately followed by the mangled name:

```
echo getAverage__7AverageFv | c++filt -s
```

The output is displayed as follows:

```
Average::getAverage()getAverage__7AverageFv
```

Demangling compiled C++ names with the demangle class library

The demangle class library contains a small class hierarchy that client programs can use to demangle names and examine the resulting parts of the name. It also provides a C-language interface for use in C programs. Although it is a C++ library, it uses no external C++ features, so you can link it directly to C programs. The demangle library is included as part of `libC.a`, and is automatically linked, when required, if `libC.a` is linked.

The header file declares a base class, `Name`, and a member function, `Demangle`, that takes a mangled name as a parameter, and returns the corresponding demangled name. The header file declares four additional subclasses, which each contain member functions that allow you to get additional information about the name. These classes are:

ClassName

Can be used to query names of independent or nested classes.

FunctionName

Can be used to query names of functions.

MemberVarName

Can be used to query names of member variables.

MemberFunctionName

Can be used to query names of member functions.

For each of these classes, functions are defined to provide you with information about the name. For example, for function names, a set of functions are defined that return the following information:

Kind Returns the type of the name being queried (that is, class, function, member variable, or member function).

Text Returns the fully qualified original text of the function.

Rootname

Returns the unqualified original name of the function.

Arguments

Returns the original text of the parameter list.

Scope Returns the original text of the function's qualifiers.

IsConst/IsVolatile/IsStatic

Returns true/false for these type qualifiers or storage class specifiers.

To demangle a name (represented as a character array), create a dynamic instance of the `Name` class, providing the character string to the class's constructor. For example, if the compiler mangled `X::f(int)` to the mangled name `f__1XFi`, in order to demangle the name, use the following code:

```
char *rest;
Name *name = Demangle("f__1XFi", rest) ;
```

If the supplied character string is not a name that requires demangling, because the original name was not mangled, the `Demangle` function returns `NULL`.

Once your program has constructed an instance of class `Name`, the program could query the instance to find out what kind of name it is, using the `Kind` method. Using the example of the mangled name `f__1XFi`, the following code:

```
name->Kind()
```

returns `MemberFunction`.

Based on the kind of name returned, the program might ask for the text of the different parts of the name, or the text of the entire name. The following table shows examples, still assuming the mangled name `f__1XFi`.

To return...	...use this code:	Result
The name of the function's qualifier	<code>((MemberFunctionName *)name)->Scope()->Text()</code>	X
The unqualified name of the function	<code>((MemberFunctionName *)name)->RootName()</code>	f

To return...	...use this code:	Result
The fully qualified name of the function	((MemberFunctionName *)name)->Text()	X::f(int)

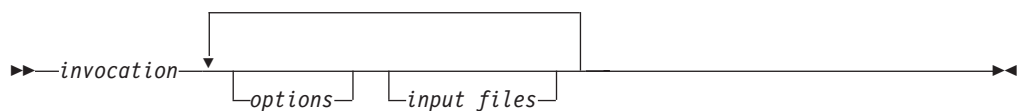
For further details about the demangle library and the C++ interface, see the comments in the library's header file, `/usr/vacpp/include/demangle.h`.

Creating a shared library with the `makeC++SharedLib` utility

`makeC++SharedLib` is a shell script that links C++ `.o` and `.a` files. It can be redistributed and used by someone who does not have XL C/C++ installed.

It is recommended that you use the `-qmkshrobj` compiler option instead of the `makeC++SharedLib` command. Among the advantages to using this option are the automatic handling of link-time C++ template instantiation (using either the template include directory or the template registry), and compatibility with the `-O5` option.

`makeC++SharedLib` command syntax



invocation

Is the command, preceded by the path. The following commands are provided:

- `makeC++SharedLib`
- `makeC++SharedLib_r`
- `makeC++SharedLib_r7`
- `makeC++SharedLib128`

You can specify one or more of the following options:

`-o shared_file.o`

The name of the file that will hold the shared file information. The default is `shr.o`.

`-b`

Uses the `-b` binder options of the `ld` command.

`-Llib_dir`

Uses the `-L` option of the `ld` command to add the directory `lib_dir` to the list of directories to be searched for unresolved symbols.

`-llibrary`

Adds `library` to the list of libraries to be searched for unresolved symbols.

`-p priority`

Specifies the priority level for the file. `priority` can be any number from `-214782623` (highest priority-initialized first) to `214783647` (lowest priority-initialized last). Numbers from `-214783648` to `-214782624` are reserved for system use. For more information, see "Assigning priorities to objects" on page 55.

- I** *import_list*
Uses the **-bI** option of the **ld** command to resolve the list of symbols in the file *import_list* that can be resolved by the binder.
- E** *export_list*
Uses the **-bE** option of the **ld** command to export the external symbols in the *export_list* file. If you do not specify **-E export_list**, a list of all global symbols is generated.
- e** *file* Saves in *file* the list computed by **-E export_list**.
- n** *name*
Sets the entry name for the shared executable to *name*. This is equivalent to using the command **ld -e name**.
- w** Excludes weak symbols from being exported.
- X** *mode*
Specifies the type of object file **makeC++SharedLib** should create. The mode must be either **32**, which processes only 32-bit object files, or **64**, which processes only 64-bit object files. The default is to process 32-bit object files (ignore 64-bit objects). You can also set the mode with the **OBJECT_MODE** environment variable. For example, **OBJECT_MODE=64** causes **makeC++SharedLib** to process any 64-bit objects and ignore 32-bit objects. The **-X** flag overrides the **OBJECT_MODE** variable.



The input files can be any of the following:

- file.o* Is an object file to be put into the shared library.
- file.a* Is an archive file to be put into the shared library.

The following table shows equivalent options between **makeC++SharedLib** and **-qmkshrobj**:

makeC++SharedLib options	-qmkshrobj and related options
-p <i>nmn</i>	-qmkshrobj=nmn
-e <i>file_name</i>	-qexpfile=file_name
-E <i>export_file</i>	-bE:export_file
-I <i>export_file</i>	-bI:export_file
-x	-qnolib
-x 32	-q32
-x 64	-q64
-n <i>entry_point</i>	-e entry_point
-w	-qnoweakexp

Related information in the XL C/C++ Compiler Reference

-  **-qmkshrobj**
-  **-O, -qoptimize**

Linking with the linkx1C utility

linkx1C is a small shell script that links C++ .o and .a files. It can be redistributed, and then used if you do not have XL C/C++ installed.

linkx1C supports the following subset of the x1C compiler options:

- **-q32** (build a 32-bit application)
- **-q64** (build a 64-bit application)
- **-b** (pass linker options to ld)
- **-f** (pass a list of object files to ld)
- **-l** (pass a library to ld)
- **-L** (pass a library path to ld)
- **-o** (specify the output file)
- **-s** (strip output)
- **-qtwolink** (enable two-step linking)

linkx1C does not support the following compiler options:











- **-G**
- **-p**
- **-pg**

linkx1C accepts and ignores all other compiler options.

Unlike x1C, **linkx1C** does not specify any runtime libraries. You must specify these libraries yourself. For example, x1C a.o becomes:

```
linkx1C a.o -L/usr/lpp/vacpp/lib -lC -lm -lc
```

Related information in the *XL C/C++ Compiler Reference*

-  **-q32, -q64**
-  **-b**
-  **-f**
-  **-l**
-  **-L**
-  **-o**
-  **-s**
-  **-qtwolink** (C++ only)
-  **-G**
-  **-p, -pg, -qprofile**

Chapter 11. Optimizing your applications

The XL compilers enable development of high performance 32-bit and 64-bit applications by offering a comprehensive set of performance enhancing techniques that exploit the multilayered PowerPC® architecture. These performance advantages depend on good programming techniques, thorough testing and debugging, followed by optimization, and tuning.

Distinguishing between optimization and tuning

You can use optimization and tuning separately or in combination to increase the performance of your application. Understanding the difference between them is the first step in understanding how the different levels, settings, and techniques can increase performance.

Optimization

Optimization is a compiler driven process that searches for opportunities to restructure your source code and give your application better overall performance at run time, without significantly impacting development time. The XL compiler optimization suite, which you control using compiler options and directives, performs best on well-written source code that has already been through a thorough debugging and testing process. These optimization transformations can:

- Reduce the number of instructions your application executes to perform critical operations.
- Restructure your object code to make optimal use of the PowerPC architecture.
- Improve memory subsystem usage.
- Exploit the ability of the architecture to handle large amounts of shared memory parallelization.

Consider that although not all optimizations benefit all applications, even basic optimization techniques can result in a performance benefit. Consult the “Steps in the optimization process” on page 72 for an overview of the common sequence of steps you can use to increase the performance of your application.

Tuning

Where optimization applies increasingly aggressive transformations designed to improve the performance of any application in any supported environment, tuning offers you opportunities to adjust characteristics of your application to improve performance, or to target specific execution environments. Even at low optimization levels, tuning for your application and target architecture can have a positive impact on performance. With proper tuning the compiler can:

- Select more efficient machine instructions.
- Generate instruction sequences that are more relevant to your application.
- Write code that is more amenable to being optimized by the compiler.

For instructions, see “Tuning for your system architecture” on page 78.

Steps in the optimization process

As you begin the optimization process, consider that not all optimization techniques suit all applications. Trade-offs sometimes occur between an increase in compile time, a reduction in debugging capability, and the improvements that optimization can provide.

Learning about, and experimenting with different optimization techniques can help you strike the right balance for your XL compiler applications while achieving the best possible performance. Also, though it is unnecessary to hand-optimize your code, compiler-friendly programming can be extremely beneficial to the optimization process. Unusual constructs can obscure the characteristics of your application and make performance optimization difficult. Use the steps in this section as a guide for optimizing your application.

1. The Basic optimization step begins your optimization processes at levels 0 and 2.
2. The Advanced optimization step exposes your application to more intense optimizations at levels 3, 4 and 5.
3. The Using high-order loop analysis and transformations step can help you limit loop execution time.
4. The Using interprocedural analysis step can optimize your entire application at once.
5. The Using profile-directed feedback step focuses optimizations on specific characteristics of your application.
6. The Debugging optimized code step can help you identify issues and problems that can occur with optimized code.

Basic optimization

The XL compiler supports several levels of optimization, with each option level building on the levels below through increasingly aggressive transformations, and consequently using more machine resources.

Ensure that your application compiles and executes properly at low optimization levels before trying more aggressive optimizations. This topic discusses two optimizations levels, listed with complementary options in the *Basic optimizations* table. The table also includes a column for compiler options that can have a performance benefit at that optimization level for some applications.

Table 18. *Basic optimizations*

Optimization level	Additional options implied by default	Complementary options	Other options with possible benefits
-O0	None	-qarch	
-O2	-qmaxmem=8192	-qarch -qtune	-qmaxmem=-1 -qhot=level=0

Optimizing at level 0

Benefits at level 0

- Minimal performance improvement, with minimal impact on machine resources.
- Exposes some source code problems, helping in the debugging process.

Begin your optimization process at **-O0** which the compiler already specifies by default. In addition, for SMP programs, add the option **-qsmp=noopt**. This level performs basic analytical optimization by removing obviously redundant code, and can result in better compile time. It also ensures your code is algorithmically correct so you can move forward to more complex optimizations. **-O0** also includes some redundant instruction elimination and constant folding. The option **-qfloat=nofold** can be used to suppress folding floating-point operations. Optimizing at this level accurately preserves all debugging information and can expose problems in existing code, such as uninitialized variables and bad casting.

Additionally, specifying **-qarch** at this level targets your application for a particular machine and can significantly improve performance by ensuring your application takes advantage of all applicable architectural benefits.

For more information on tuning, see “Tuning for your system architecture” on page 78.

Optimizing at level 2

Benefits at level 2

- Eliminates redundant code
- Basic loop optimization
- Can structure code to take advantage of **-qarch** and **-qtune** settings

After successfully compiling, executing, and debugging your application using **-O0**, recompiling at **-O2** opens your application to a set of comprehensive low-level transformations that apply to subprogram or compilation unit scopes and can include some inlining. Optimizations at **-O2** are a relative balance between increasing performance while limiting the impact on compilation time and system resources. You can increase the memory available to some of the optimizations in the **-O2** portfolio by providing a larger value for the **-qmaxmem** option. Specifying **-qmaxmem=-1** allows the optimizer to use memory as needed without checking for limits but does not change the transformations the optimizer applies to your application at **-O2**.

In C, compile with **-qlibansi** unless your application defines functions with names identical to those of library functions. If you encounter problems with **-O2**, consider using **-qalias=noansi** rather than turning off optimization.

Also, ensure that pointers in your C code follow these type restrictions:

- Generic pointers can be `char*` or `void*`
- Mark all shared variables and pointers to shared variables `volatile`

Starting to tune at O2

Choosing the right hardware architecture target or family of targets becomes even more important at **-O2** and higher. By targeting the proper hardware, the optimizer can make the best use of the hardware facilities available. If you choose a family of hardware targets, the **-qtune** option can direct the compiler to emit code consistent with the architecture choice, but executes optimally on the chosen tuning hardware target. With this option, you can compile for a general set of targets but have the code run best on a particular target.

See the “Tuning for your system architecture” on page 78 section for details on the **-qarch** and **-qtune** options.

The **-O2** option can perform a number of additional optimizations, including:

- Common subexpression elimination: Eliminates redundant instructions.
- Constant propagation: Evaluates constant expressions at compile-time.
- Dead code elimination: Eliminates instructions that a particular control flow does not reach, or that generate an unused result.
- Dead store elimination: Eliminates unnecessary variable assignments.
- Graph coloring register allocation: Globally assigns user variables to registers.
- Value numbering: Simplifies algebraic expressions, by eliminating redundant computations.
- Instruction scheduling for the target machine.
- Loop unrolling and software pipelining.
- Moving invariant code out of loops.
- Simplifying control flow.
- Strength reduction and effective use of addressing modes.

Even with **-O2** optimizations, some useful information about your source code is made available to the debugger if you specify **-g**. Using a higher **-g** level increases the information provided to the debugger, but reduces the optimization that can be done. Conversely, higher optimization levels can transform code to an extent to which debugging information is no longer accurate. Use that information with discretion.

Advanced optimization

Higher optimization levels can have a tremendous impact on performance, but some trade-offs can occur in terms of code size, compile time, resource requirements, and numeric or algorithmic precision.

After applying “Basic optimization” on page 72 and successfully compiling and executing your application, you can apply more powerful optimization tools. The XL compiler optimization portfolio includes many options for directing advanced optimization, and the transformations your application undergoes are largely under your control. The discussion of each optimization level in Table 19 includes information on not only the performance benefits, and the possible trade-offs as well, but information on how you can help guide the optimizer to find the best solutions for your application.

Table 19. Advanced optimizations

Optimization Level	Additional options implied	Complementary options	Options with possible benefits
-O3	-qnostrict -qmaxmem=-1 -qhot=level=0	-qarch -qtune	-qpdf
-O4	-qnostrict -qmaxmem=-1 -qhot -qipa -qarch=auto -qtune=auto -qcache=auto	-qarch -qtune -qcache	-qpdf -qsmp=auto
-O5	All of -O4 -qipa=level=2	-qarch -qtune -qcache	-qpdf -qsmp=auto

When you compile programs with any of the following sets of options:

- **-qhot -qignerrno -qnostrict**
- **-qhot -O3**
- **-O4**
- **-O5**

the compiler automatically attempts to vectorize calls to system math functions by calling the equivalent vector functions in the Mathematical Acceleration Subsystem libraries (MASS), with the exceptions of functions `vdnint`, `vdint`, `vcosisin`, `vscoisin`, `vqdrct`, `vsqdrct`, `vrqdrct`, `vsrqdrct`, `vpopcnt4`, and `vpopcnt8`. If the compiler cannot vectorize, it automatically tries to call the equivalent MASS scalar functions. For automatic vectorization or scalarization, the compiler uses versions of the MASS functions contained in the system library `libxlopt.a`.

In addition to any of the preceding sets of options, when the **-qipa** option is in effect, if the compiler cannot vectorize, it tries to inline the MASS scalar functions before deciding to call them.

Optimizing at level 3

Benefits at level 3

- In-depth memory access analysis
- Better loop scheduling
- High-order loop analysis and transformations (**-qhot=level=0**)
- Inlining of small procedures within a compilation unit by default
- Eliminating implicit compile-time memory usage limits
- Widening, which merges adjacent load/stores and other operations
- Pointer aliasing improvements to enhance other optimizations

Specifying **-O3** initiates more intense low-level transformations that remove many of the limitations present at **-O2**. For instance, the optimizer no longer checks for memory limits, by defaulting to **-qmaxmem=-1**. Additionally, optimizations encompass larger program regions and attempt more in-depth analysis. While not all applications contain opportunities for the optimizer to provide a measurable increase in performance, most applications can benefit from this type of analysis.

Potential trade-offs at level 3

With the in-depth analysis of **-O3** comes a trade-off in terms of compilation time and memory resources. Also, since **-O3** implies **-qnostrict**, the optimizer can alter certain floating-point semantics in your application to gain execution speed. This typically involves precision trade-offs as follows:

- Reordering of floating-point computations.
- Reordering or elimination of possible exceptions, such as division by zero or overflow.
- Using alternative calculations that might give slightly less precise results or not handle infinities or NaNs in the same way.

You can still gain most of the **-O3** benefits while preserving precise floating-point semantics by specifying **-qstrict**. Compiling with **-qstrict** is necessary if you require the same absolute precision in floating-point computational accuracy as you get with **-O0**, **-O2**, or **-qnoopt** results. The option **-qstrict=ieee** also ensures

adherence to all IEEE semantics for floating-point operations. If your application is sensitive to floating-point exceptions or the order of evaluation for floating-point arithmetic, compiling with `-qstrict`, `-qstrict=exceptions`, or `-qstrict=order` helps to ensure accurate results. You should also consider the impact of the `-qstrict=precision` suboption group on floating-point computational accuracy. The precision suboption group includes the individual suboptions: `subnormals`, `operationprecision`, `association`, `reductionorder`, and `library` (described in the `-qstrict` option in the *XL C/C++ Compiler Reference*).

Without `-qstrict`, the difference in computation for any one source-level operation is very small in comparison to “Basic optimization” on page 72. Although a small difference can be compounded if the operation is in a loop structure where the difference becomes additive, most applications are not sensitive to the changes that can occur in floating-point semantics.

See “-O -qoptimize” in the *XL C/C++ Compiler Reference* for information on the `-O` level syntax.

An intermediate step: adding `-qhot` suboptions at level 3

At `-O3`, the optimization includes minimal `-qhot` loop transformations at `level=0` to increase performance. You can further increase your performance benefit by increasing the level and therefore the aggressiveness of `-qhot`. Try specifying `-qhot` without any suboptions, or `-qhot=level=1`.

For more information on `-qhot`, see “Using high-order loop analysis and transformations” on page 80.

Conversely, if the application does not use loops processing arrays (which `-qhot` improves), you can improve compile speed with minimal performance loss by using `-qnohot` after `-O3`.

Optimizing at level 4

Benefits at level 4

- Propagation of global and argument values between compilation units
- Inlining code from one compilation unit to another
- Reorganization or elimination of global data structures
- An increase in the precision of aliasing analysis

Optimizing at `-O4` builds on `-O3` by triggering `-qipa=level=1` which performs interprocedural analysis (IPA), optimizing your entire application as a unit. This option is particularly pertinent to applications that contain a large number of frequently used routines.

To make full use of IPA optimizations, you must specify `-O4` on the compilation and link steps of your application build as interprocedural analysis occurs in stages at both compile and link time.

Potential trade-offs at level 4

In addition to the trade-offs already mentioned for `-O3`, specifying `-qipa` can significantly increase compilation time, especially at the link step.

The IPA process

1. At compile time optimizations occur on a file-by-file basis, as well as preparation for the link stage. IPA writes analysis information directly into the object files the compiler produces.
2. At the link stage, IPA reads the information from the object files and analyzes the entire application.
3. This analysis guides the optimizer on how to rewrite and restructure your application and apply appropriate **-O3** level optimizations.

The “Using interprocedural analysis” on page 83 section contains more information on IPA including details on IPA suboptions.

Beyond **-qipa**, **-O4** enables other optimization options:

- **-qhot**
Enables more aggressive HOT transformations to optimize loop constructs and array language.
- **-qarch=auto** and **-qtune=auto**
Optimizes your application to execute on a hardware architecture identical to your build machine. If the architecture of your build machine is incompatible with your application's execution environment, you must specify a different **-qarch** suboption after the **-O4** option. This overrides **-qarch=auto**.
- **-qcache=auto**
Optimizes your cache configuration for execution on specific hardware architecture. The **auto** suboption assumes that the cache configuration of your build machine is identical to the configuration of your execution architecture. Specifying a cache configuration can increase program performance, particularly loop operations by blocking them to process only the amount of data that can fit into the data cache.
If you want to execute your application on a different machine, specify correct cache values.

Optimizing at level 5

Benefits at level 5

- Most aggressive optimizations available
- Makes full use of loop optimizations and IPA

As the highest optimization level, **-O5** includes all **-O4** optimizations and deepens whole program analysis by increasing the **-qipa** level to 2. Compiling with **-O5** also increases how aggressively the optimizer pursues aliasing improvements. Additionally, if your application contains a mix of C/C++ and Fortran code that you compile using the XL compilers, you can increase performance by compiling and linking your code with the **-O5** option.

Potential trade-offs at level 5

Compiling at **-O5** requires more compile time and machine resources than any other optimization levels, particularly if you include **-O5** on the IPA link step. Compile at **-O5** as the final phase in your optimization process after successfully compiling and executing your application at **-O4**.

Tuning for your system architecture

You can instruct the compiler to generate code for optimal execution on a given microprocessor or architecture family. By selecting appropriate target machine options, you can optimize to suit the broadest possible selection of target processors, a range of processors within a given family of processor architectures, or a specific processor.


The following table lists the optimization options that affect individual aspects of the target machine. Using a predefined optimization level sets default values for these individual options.

Table 20. Target machine options

Option	Behavior
-q32	Generates code for a 32-bit (4 byte integer / 4 byte long / 4 byte pointer) addressing model (32-bit execution mode). This is the default setting.
-q64	Generates code for a 64-bit (4 byte integer / 8 byte long / 8 byte pointer) addressing model (64-bit execution mode).
-qarch	Selects a family of processor architectures for which instruction code should be generated. This option restricts the instruction set generated to a subset of that for the PowerPC architecture. Using <code>-O4</code> or <code>-O5</code> sets the default to <code>-qarch=auto</code> . See “Getting the most out of target machine options” on page 79 below for more information on this option.
-qtune	Biases optimization toward execution on a given microprocessor, without implying anything about the instruction set architecture to use as a target. See “Getting the most out of target machine options” on page 79 below for more information on this option.
-qcache	Defines a specific cache or memory geometry. The defaults are determined through the setting of <code>-qtune</code> . See “Getting the most out of target machine options” on page 79 below for more information on this option.


For a complete listing of valid hardware-related suboptions and combinations of suboptions, see *Acceptable -qarch/-qtune combinations* in the **-qtune** section of the *XL C/C++ Compiler Reference* and see *Specifying Compiler Options for Architecture-Specific, 32- or 64-bit Compilation* in the *XL C/C++ Compiler Reference*.


Related information in the XL C/C++ Compiler Reference

 [-q32, -q64](#)

 [-qarch](#)

 [-qipa](#)

 [-qtune](#)

 [-qcache](#)

 [Specifying compiler options for architecture-specific, 32-bit or 64-bit compilation](#)

Getting the most out of target machine options

Using `-qarch` options

You use the `-qarch` compiler option to generate instructions that are optimized for a specific machine architecture. For example, if you want to generate an object code that contains instructions optimized for POWER7[®], you use `-qarch=pwr7`. If your application runs on the same machine on which you are compiling it, you can use the `-qarch=auto` option, which automatically detects the specific architecture of the compiling machine, and generates code to take advantage of instructions available only on that machine (or on a system that supports the equivalent processor architecture). Otherwise, use the `-qarch` option to specify the smallest possible family of the machines that can run your code reasonably well.

If you want to run your application on a system architecture that provides specific feature supports, you must specify a corresponding `-qarch` suboption to generate the object code for your system architecture. For example, if you want to deploy your application on a POWER6[®] or POWER7 machine and to fully exploit VMX vector processing and large-page support, you must specify `-qarch=pwr6` for POWER6 and `-qarch=pwr7` for POWER7 on your compiling machine. Specifying `-qarch=auto` or `-qarch` does not give you the support you want. However, if you deploy your application on both POWER6 and POWER7, you must make sure that `-qarch` is set to the lowest common architecture. This way your application will only contain instructions that are common to all processors the application is deployed on. In this example, the lowest common architecture is POWER6 so you must use `-qarch=pwr6`. For details about `-qarch` and its suboptions, see `-qarch` in the *XL C/C++ Compiler Reference*. For details about the corresponding system architectures each `-qarch` suboption supports, see the Features support in processor architectures table in `-qarch`.

Using `-qtune` options

You use the `-qtune` compiler option to control the scheduling of instructions that are optimized for your machine architecture. If you specify a particular architecture with `-qarch`, `-qtune` automatically selects the suboption that generates instruction sequences with the best performance for that architecture. If you specify a *group* of architectures with `-qarch`, compiling with `-qtune=auto` generates code that runs on all of the architectures in the specified group, but the instruction sequences are those with the best performance on the architecture of the compiling machine.

Try to specify with `-qtune` the particular architecture that the compiler should target for best performance but still allow execution of the produced object file on all architectures specified in the `-qarch` option. For information on the valid combinations of `-qarch` and `-qtune`, see *Acceptable `-qarch`/`-qtune` combinations* in the `-qtune` section of the *XL C/C++ Compiler Reference*.

If you need to create a single binary file that runs on a range of PowerPC hardware, consider using the `-qtune=balanced` option. With this option in effect, optimization decisions made by the compiler are not targeted to a specific version of hardware. Instead, tuning decisions try to include features that are generally helpful across a broad range of hardware and avoid those optimizations that might be harmful on some hardware.

Note: You must verify the performance of code compiled with the `-qtune=balanced` option before distributing it.

The main difference between using **-qtune=balanced** and **-qtune=auto** is that, with **-qtune=auto** and a specified **-qarch** suboption, the compiler generates instructions that are optimized for that specified versions of hardware architecture and might not perform well on others. For example, if you want to use **-qtune=auto** to generate optimized instructions that are deployable on a POWER7 machine, you use **-qarch=pwr7 -qtune=auto**. To generate instructions that perform reasonably well across a range of Power hardware, use **-qtune=balanced** instead. For details, see **-qtune** in the *XL C/C++ Compiler Reference*.

Using -qcache options

Before using the **-qcache** option, use the **-qlistopt** option to generate a listing of the current settings and verify if they are satisfactory. If you decide to specify your own **-qcache** suboptions, use **-qhot** or **-qsmp** along with it. For the full set of suboptions, option syntax, and guidelines for use, see **-qcache** in the *XL C/C++ Compiler Reference*.

Related information in the *XL C/C++ Compiler Reference*



-qhot



-qsmp



-qcache



-qlistopt



-qarch



-qtune

Using high-order loop analysis and transformations

High-order transformations are optimizations that specifically improve the performance of loops through techniques such as interchange, fusion, and unrolling.

The goals of these loop optimizations include:






- Reducing the costs of memory access through the effective use of caches and translation look-aside buffers.
- Overlapping computation and memory access through effective utilization of the data prefetching capabilities provided by the hardware.
- Improving the utilization of microprocessor resources through reordering and balancing the usage of instructions with complementary resource requirements.
- Generating vector instructions.
- Generating calls to vector math library functions.

To enable high-order loop analysis and transformations, you use the **-qhot** option, which implies an optimization level of **-O2**. The following table lists the suboptions available for **-qhot**.

Table 21. *-qhot* suboptions

Suboption	Behavior
level=0	Instructs the compiler to perform a subset of high-order transformations that enhance performance by improving data locality. This suboption implies -qhot=novector and -qhot=noarraypad . This level is automatically enabled if you compile with -O3 .
level=1	This is the default suboption if you specify -qhot with no suboptions. This level is also automatically enabled if you compile with -O4 or -O5 . This is equivalent to specifying -qhot=vector .
level=2	When used with -qsmp , instructs the compiler to perform the transformations of -qhot=level=1 plus some additional transformation on nested loops. The resulting loop analysis and transformations can lead to more cache reuse and loop parallelization.
vector	When specified with -qnostrict and -qignerrno , or -O3 or a higher optimization level, instructs the compiler to transform some loops to use the optimized versions of various math functions contained in the MASS libraries, rather than use the system versions. The optimized versions make different trade-offs with respect to accuracy and exception-handling versus performance. This suboption is enabled by default if you specify -qhot with no suboptions. Also, specifying -qhot=vector with -O3 implies -qhot=level=1 .
arraypad	Instructs the compiler to pad any arrays where it infers there might be a benefit and to pad by whatever amount it chooses.

Related information in the XL C/C++ Compiler Reference

-  **-qhot**
-  **-qstrict**
-  **-qignerrno**
-  **-qarch**
-  **-qsimd**

Getting the most out of **-qhot**

Here are some suggestions for using **-qhot**:

- Try using **-qhot** along with **-O3** for all of your code. It is designed to have a neutral effect when no opportunities for transformation exist. However, it might increase compile time and have little benefit if the program has no loop processing vectors or arrays.
- If the runtime performance of your code can significantly benefit from automatic inlining and memory locality optimizations, try using **-O4** with **-qhot=level=0** or **-qhot=novector**.
- If you encounter unacceptably long compile time (this can happen with complex loop nests), try **-qhot=level=0** or **-qnohot**.
- If your code size is unacceptably large, try using **-qcompact** along with **-qhot**.
- You can compile some source files with the **-qhot** option and some files without the **-qhot** option, allowing the compiler to improve only the parts of your code that need optimization.
- Use **-qreport** along with **-qsimd=auto** to generate a loop transformation listing. The listing file identifies how loops are transformed in a section marked LOOP TRANSFORMATION SECTION. Use the listing information as feedback about how the


loops in your program are being transformed. Based on this information, you may want to adjust your code so that the compiler can transform loops more effectively. For example, you can use this section of the listing to identify non-stride-one references that may prevent loop vectorization.

- Use **-qreport** along with **-qhot** or any optimization option that implies **-qhot** to generate information about nested loops in the LOOP TRANSFORMATION SECTION of the listing file. In addition, when you use **-qprefetch=assistthread** to generate prefetching assist threads, a message Assist thread for data prefetching was generated is also displayed in this section of the report. To generate a list of aggressive loop transformations and parallelizations performed on loop nests in the LOOP TRANSFORMATION SECTION of the listing file, use **-qhot=level=2** and **-qsmp** together with **-qreport**.
- If you specify **-qassert=refalign**, you assert to the compiler that all pointers inside the compilation unit only point to data that is naturally aligned with respect to the length of the pointer types. With this assertion, the compiler might generate more efficient code. This assertion is particularly useful when you target a SIMD architecture with **-qhot=level=0** or **-qhot=level=1** with the **-qsimd=auto** option.

Related information in the XL C/C++ Compiler Reference

 **-qcompact**

 **-qhot**

 **-qsimd**

 **-qprefetch**

 **-qstrict**

Using shared-memory parallelism (SMP)

Many IBM pSeries[®] machines are capable of shared-memory parallel processing. You can compile with **-qsmp** to generate the threaded code needed to exploit this capability. The option implies an optimization level of at least **-O2**.

The following table lists the most commonly used suboptions. Descriptions and syntax of all the suboptions are provided in **-qsmp** in the *XL C/C++ Compiler Reference*. An overview of automatic parallelization, as well as of IBM SMP and OpenMP directives is provided in Chapter 15, “Parallelizing your programs,” on page 135.


Table 22. Commonly used **-qsmp** suboptions

suboption	Behavior
auto	Instructs the compiler to automatically generate parallel code where possible without user assistance. Any SMP programming constructs in the source code, including IBM SMP and OpenMP directives, are also recognized. This is the default setting if you do not specify any -qsmp suboptions, and it also implies the opt suboption.
omp	Instructs the compiler to enforce strict conformance to the OpenMP API for specifying explicit parallelism. Only language constructs that conform to the OpenMP standard are recognized. Note that -qsmp=omp is currently incompatible with -qsmp=auto .
opt	Instructs the compiler to optimize as well as parallelize. The optimization is equivalent to -O2 -qhot in the absence of other optimization options.

Table 22. Commonly used `-qsmp` suboptions (continued)

suboption	Behavior
<code>noopt</code>	All optimization is turned off. During development, it can be useful to turn off optimization to facilitate debugging.
<i>fine_tuning</i>	Other values for the suboption provide control over thread scheduling, nested parallelism, locking, etc.

Related information in the XL C/C++ Compiler Reference

 `-O, -qoptimize`

 `-qsmp`

 `-qhot`

Getting the most out of `-qsmp`


Here are some suggestions for using the `-qsmp` option:

- Before using `-qsmp` with automatic parallelization, test your programs using optimization and `-qhot` in a single-threaded manner.
- If you are compiling an OpenMP program and do not want automatic parallelization, use `-qsmp=omp:noauto`.
- Always use the reentrant compiler invocations (the `_r` invocations) when using `-qsmp`.
- By default, the runtime environment uses all available processors. Do not set the `XLSMPOPTS=PARTHDS` or `OMP_NUM_THREADS` environment variables unless you want to use fewer than the number of available processors. You might want to set the number of executing threads to a small number or to 1 to ease debugging.
- If you are using a dedicated machine or node, consider setting the `SPINS` and `YIELDS` environment variables (suboptions of the `XLSMPOPTS` environment variable) to 0. Doing so prevents the operating system from intervening in the scheduling of threads across synchronization boundaries such as barriers.
- When debugging an OpenMP program, try using `-qsmp=noopt` (without `-O`) to make the debugging information produced by the compiler more precise.

Related information in the XL C/C++ Compiler Reference

 `-qsmp`

 `-qhot`

 Invoking the compiler

 `XLSMPOPTS`

 Environment variables for parallel processing

Using interprocedural analysis

Interprocedural analysis (IPA) enables the compiler to optimize across different files (whole-program analysis), and can result in significant performance improvements.

You can specify interprocedural analysis on the compilation step only or on both compilation and link steps in whole program mode. Whole program mode expands the scope of optimization to an entire program unit, which can be an

executable or shared object. As IPA can significantly increase compile time, you should limit using IPA to the final performance tuning stage of development.

You can generate relinkable objects while preserving IPA information by specifying **-r -qipa=relink**. This creates a nonexecutable package that contains all object files. By using this suboption, you can postpone linking until the very last stage.

If you want to use your own archive files while generating the nonexecutable package, you can use the **ar** tool and set the **XL_AR** environment variable to point to the **ar** tool. For details, refer to the **-qipa** section of the *XL C/C++ Compiler Reference*.

You enable IPA by specifying the **-qipa** option. The most commonly used suboptions and their effects are described in the following table. The full set of suboptions and syntax is described in the **-qipa** section of the *XL C/C++ Compiler Reference*.

The steps to use IPA are:

1. Do preliminary performance analysis and tuning before compiling with the **-qipa** option, because the IPA analysis uses a two-pass mechanism that increases compile and link time. You can reduce some compilation and link overhead by using the **-qipa=noobject** option.
2. Specify the **-qipa** option on both the compilation and the link steps of the entire application, or as much of it as possible. Use suboptions to indicate assumptions to be made about parts of the program *not* compiled with **-qipa**.

Table 23. Commonly used **-qipa** suboptions

Suboption	Behavior
level=0	<p>Program partitioning and simple interprocedural optimization, which consists of:</p> <ul style="list-style-type: none"> • Automatic recognition of standard libraries. • Localization of statically bound variables and procedures. • Partitioning and layout of procedures according to their calling relationships. (Procedures that call each other frequently are located closer together in memory.) • Expansion of scope for some optimizations, notably register allocation.
level=1	<p>Inlining and global data mapping. Specifically:</p> <ul style="list-style-type: none"> • Procedure inlining. • Partitioning and layout of static data according to reference affinity. (Data that is frequently referenced together will be located closer together in memory.) <p>This is the default level if you do not specify any suboptions with the -qipa option.</p>

Table 23. Commonly used **-qipa** suboptions (continued)

Suboption	Behavior
level=2	Global alias analysis, specialization, interprocedural data flow: <ul style="list-style-type: none"> • Whole-program alias analysis. This level includes the disambiguation of pointer dereferences and indirect function calls, and the refinement of information about the side effects of a function call. • Intensive intraprocedural optimizations. This can take the form of value numbering, code propagation and simplification, moving code into conditions or out of loops, and elimination of redundancy. • Interprocedural constant propagation, dead code elimination, pointer analysis, code motion across functions, and interprocedural strength reduction. • Procedure specialization (cloning). • Whole program data reorganization.
inline= <i>suboptions</i>	Provides precise control over function inlining.
<i>fine_tuning</i>	Other values for -qipa provide the ability to specify the behavior of library code, tune program partitioning, read commands from a file, etc.
relink	Creates a nonexecutable package that contains all of your object files while preserving IPA information.



Related information in the *XL C/C++ Compiler Reference*



-qipa

Getting the most from **-qipa**

It is not necessary to compile everything with **-qipa**, but try to apply it to as much of your program as possible. Here are some suggestions:

- Specify the **-qipa** option on both the compile and link steps of the entire application. Although you can also use **-qipa** with libraries, shared objects, and executable files, be sure to use **-qipa** to compile the main and exported functions.
- When compiling and linking separately, use **-qipa=noobject** on the compile step for faster compilation.
- When specifying optimization options in a makefile, remember to use the compiler driver (**xlc**) to link, and to include all compiler options on the link step.
- As IPA can generate significantly larger object files than traditional compilations, ensure that there is enough space in the `/tmp` directory (at least 200 MB). You can use the `TMPDIR` environment variable to specify a directory with sufficient free space.
- Try varying the **level** suboption if link time is too long. Compiling with **-qipa=level=0** can still be very beneficial for little additional link time.
- Use **-qipa=list=long** to generate a report of functions that were previously inlined. If too few or too many functions are inlined, consider using **-qinline** or **-qnoinline**.  To control the inlining of specific functions, use **-qinline+function_name** or **-qinline-function_name**. 
- To generate data reorganization information in the listing file, specify the optimization level **-qipa=level=2** or **-O5** together with **-qreport**. During the IPA link pass, the data reorganization messages for program variable data will be produced to the data reorganization section of the listing file with the label `DATA`


REORGANIZATION SECTION. Reorganizations include array splitting, array transposing, memory allocation merging, array interleaving, and array coalescing.

- Use **-r -qipa=relink** to create a nonexecutable package that contains all of your object files while preserving IPA information. If you want to use your archive files while generating the package, you can use the **ar** tool and set the **XL_AR** environment variable to point to the **ar** tool. For details, refer to the section of the *XL C/C++ Compiler Reference*.

Note: While IPA's interprocedural optimizations can significantly improve performance of a program, they can also cause incorrect but previously functioning programs to fail. Here are examples of programming practices that can work by accident without aggressive optimization but are exposed with IPA:

- Relying on the allocation order or location of automatic variables, such as taking the address of an automatic variable and then later comparing it with the address of another local variable to determine the growth direction of a stack. The C language does not guarantee where an automatic variable is allocated, or its position relative to other automatic variables. Do not compile such a function with IPA.
- Accessing a pointer that is either invalid or beyond an array's bounds. Because IPA can reorganize global data structures, a wayward pointer which might have previously modified unused memory might now conflict with user-allocated storage.

Related information in the *XL C/C++ Compiler Reference*

 -qinline

 -qlist

 -qipa

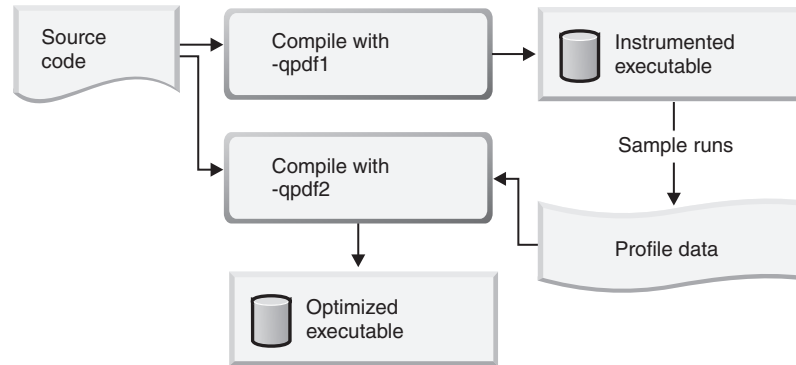
Using profile-directed feedback

You can use profile-directed feedback (PDF) to tune the performance of your application for a typical usage scenario. The compiler optimizes the application based on an analysis of how often branches are taken and blocks of code are run.

Use PDF process after other debugging and tuning is finished, as one of the last steps before putting the application into production. Other optimizations such as the **-qipa** option and optimization levels **-O4** and **-O5** can also benefit when using with PDF process.

The following diagram illustrates the PDF process:

Figure 2. Profile-directed feedback



To use the PDF process to optimize your application, follow these steps:

1. Compile some or all of the source files in a program with the **-qpdf1** option. You must specify at least the **-O2** optimization level.

Notes:

- A PDF map file is generated at this step. It is used for the **showpdf** utility to display part of the profiling information in text or XML format. For details, see “Viewing profiling information with showpdf” on page 89. If you do not need to view the profiling information, specify the **-qnoshowpdf** option at this step so that the PDF map file is not generated. For details of **-qnoshowpdf**, see **-qshowpdf** in the *XL C/C++ Compiler Reference*.
 - Although you can specify PDF optimization (**-qpdf**) as early in the optimization level as **-O2**, PDF optimization is recommended at **-O4** and higher.
 - You do not have to compile all of the codes of the programs with the **-qpdf1** option. In a large application, you can concentrate on those areas of the code that can benefit most from the optimization.
2. Run the resulting application with a typical data set. When the application exits, profile information is written to a PDF file. You can run the resulting application multiple times with different data sets. The profiling information is accumulated to provide a count of how often branches are taken and blocks of code are run, based on the input data used. By default, the PDF file is named **._pdf**, and it is placed in the current working directory or the directory specified by the **PDFDIR** environment variable. If the **PDFDIR** environment variable is set but the specified directory does not exist, the compiler issues a warning message. To override the defaults, use the **-qpdf1=pdfname** or **-qpdf1=exename** option.

If you recompile your program by using either of the **-qpdf1=level=0** or **-qpdf1=level=1** option, single-pass profiling is supported. The compiler removes the existing PDF file before generating a new application.

If you recompile your program by using **-qpdf1=level=2** option, multiple-pass profiling is supported. You can repeat compiling your program and running the resulting application, then new PDF files are generated up to five times.

Notes:

- When compiling your program with the **-qpdf1** or **-qpdf2** option, by default, the **-qipa** option is also invoked with **level=0**.
- To avoid wasting compile and run time, make sure that the **PDFDIR** environment variable is set to an absolute path. Otherwise, you might run the application from a wrong directory, and the compiler cannot locate the

profiling information files. When it happens, the program might not be optimized correctly or might be stopped by a segmentation fault. A segmentation fault might also happen if you change the value of the PDFDIR environment variable and run the application before finishing the PDF process.

- Avoid using atypical data that can distort the analysis to infrequently executed code paths.
3. If you have several PDF files, use the **mergepdf** utility to combine these PDF files into one PDF file. For example, if you produce three PDF files that represent usage patterns that occur 53%, 32%, and 15% of the time respectively, you can use this command:

```
mergepdf -r 53 path1 -r 32 path2 -r 15 path3
```

Notes:

- Avoid mixing the PDF files created by different version levels of the XL C/C++ compiler.
 - You cannot edit PDF files that are generated by the resulting application. Otherwise, the performance or function of the generated executable application might be affected.
4. Recompile your program using the same compiler options as before, but change **-qpdf1** to **-qpdf2**. In this second compilation, the accumulated profiling information is used to fine-tune the optimizations. The resulting program contains no profiling overhead and runs at full speed.

Notes:

- You are highly recommended to use the same optimization level at all compilation steps for a particular program. Otherwise, the PDF process cannot optimize your program correctly and might even slow it down. All compiler settings that affect optimization must be the same, including any supplied by configuration files.
 - You can modify your source code and use the **-qpdf1** and **-qpdf2** options to compile your program. Old profiling information can still be preserved and used during the second stage of the PDF process. The compiler issues a list of warnings but the compilation does not stop. An information message is also issued with a number in the range of 0 - 100 to indicate how outdated the old profiling information is.
 - When using the **-qreport** option with the **-qpdf2** option, you can get additional information in your listing file to help you tune your program. This information is written to the PDF Report section.
5. If you want to erase the PDF information, use the **cleanpdf** or **resetpdf** utility.

Instead of step 4, you can use the **-qpdf2** option to link the object files created during the **-qpdf1** phase without recompiling your program during the **-qpdf2** phase. This alternative approach can save considerable time and help tune large applications for optimization.

Examples

The following example demonstrates that you can concentrate on compiling those codes that can benefit most from the optimization, instead of compiling all the code of applications with the **-qpdf1** option:

```
#Set the PDFDIR variable
export PDFDIR=$HOME/project_dir
```



```

#Compile most of the files with -qpdf1
xlc -qpdf1 -O3 -c file1.c file2.c file3.c

#This file does not need optimization
xlc -c file4.c

#Non-PDF object files such as file4.o can be linked
xlc -qpdf1 -O3 file1.o file2.o file3.o file4.o

#Run several times with different input data
./a.out < polar_orbit.data
./a.out < elliptical_orbit.data
./a.out < geosynchronous_orbit.data

#No need to recompile the source of non-PDF object files
#(file4.c).
xlc -qpdf2 -O3 file1.c file2.c file3.c

#Link all the object files into the final application
xlc -qpdf2 -O3 file1.o file2.o file3.o file4.o

```

The following example bypasses recompiling the source with the **-qpdf2** option:

```

#Compile source with -qpdf1
xlc -c -qpdf1 -O3 file1.c file2.c




#Link object files
xlc -qpdf1 -O3 file1.o file2.o

#Run with one set of input data
./a.out < sample.data

#Link the mix of pdf1 and pdf2 objects
xlc -qpdf2 -O3 file1.o file2.o

```

Related information in the *XL C/C++ Compiler Reference*

-  -qpdf1, -qpdf2
-  -O, -qoptimize
-  Runtime environment variables

Viewing profiling information with showpdf

With the **showpdf** utility, you can view the following types of profiling information that is gathered from your application:

- Block-counter profiling
- Call-counter profiling
- Value profiling
- Cache-miss profiling, if you specified the **-qpdf1=level=2** option during the **-qpdf1** phase.

You can view the first two types of profiling information in either text or XML format. However, you can view value profiling and cache-miss profiling information only in XML format.

Syntax

```

>> showpdf [pdfdir] [-f pdfname] [-m pdfmapdir] [--xml]

```

Parameters

pdfdir

is the directory that contains the profile-directed feedback (PDF) file. If the PDFDIR environment variable is not changed after the **-qpdf1** phase, the PDF map file is also contained in this directory. If this parameter is not specified, the compiler uses the value of the PDFDIR environment variable as the name of the directory.

pdfname

is the name of the PDF file. If this parameter is not specified, the compiler uses `._pdf` as the name of the PDF file.

pdfmapdir

is the directory that contains the PDF map file. If this parameter is not specified, the compiler uses the value of the PDFDIR environment variable as the name of the directory.

-xml

determines the display format of the PDF information. If this parameter is specified, the PDF information is displayed in XML format; otherwise it is displayed in text format. Because value profiling and cache-miss profiling information can be displayed only in XML format, the PDF report in XML format contains more information than the report in text format.

Usage

A PDF map file that contains static information is generated during the **-qpdf1** phase, and a PDF file is generated during the execution of the resulting application. The **showpdf** utility needs both the PDF and PDF map files to display PDF information in either text or XML format.

If the **-qpdf1=level=2** option is specified during the **-qpdf1** phase, several PDF and PDF map files might be generated. Then if you want to view the profiling information, you need to run the **showpdf** utility for each pair of PDF and PDF map files.

By default, the PDF file is named `._pdf`, and the PDF map file is named `._pdf_map`. If the PDFDIR environment variable is set, the compiler places the PDF and PDF map files in the directory specified by PDFDIR. Otherwise, if the PDFDIR environment variable is not set, the compiler places these files in the current working directory. If the PDFDIR environment variable is set but the specified directory does not exist, the compiler issues a warning message. To override the defaults, use the **-qpdf1=pdfname** option to specify the paths and names for the PDF and PDF map files. For example, if you specify the **-qpdf1=pdfname=/home/joe/func** option, the resulting PDF file is named `func`, and the PDF map file is named `func_map`. Both of the files are placed in the `/home/joe` directory.

If the PDFDIR environment variable is changed between the **-qpdf1** phase and the execution of the resulting application, the PDF and PDF map files are generated in separate directories. In this case, you must specify the directories for both of these files to the **showpdf** utility.

Notes:

- PDF and PDF map files must be generated from the same compilation instance. Otherwise, the compiler issues an error.

- PDF and PDF map files must be generated during the same profiling process. It means that you cannot mix and match PDF and PDF map files that are generated from different profiling processes.
- You must use the same version and PTF level of the compiler to generate the PDF file and the PDF map file.
- The **showpdf** utility accepts only PDF files that are in binary format.

The following example shows how to use the **showpdf** utility to view the profiling information for a Hello World application:

The source for the program file `hello.c` is as follows:

```
#include <stdio.h>
void HelloWorld()
{
    printf("Hello World");
}
main()
{
    HelloWorld();
    return 0;
}
```

1. Compile the source file.
`xlc -qpdf1 -o hello.c`
2. Run the resulting executable program **a.out** using a typical data set or several typical data sets.
3. If you want to view the profiling information for the executable file in text format, run the **showpdf** utility without any parameters.

```
showpdf
```

The result is as follows:

```
HelloWorld(67):  1 (hello.c)
```

```
Call Counters:
```

```
4 | 1  printf(69)
```

```
Call coverage = 100% ( 1/1 )
```

```
Block Counters:
```

```
2-4 | 1
```

```
5 |
5 | 1
```

```
Block coverage = 100% ( 2/2 )
```

```
-----
main(68):  1 (hello.c)
```

```
Call Counters:
```

```
8 | 1  HelloWorld(67)
```

```
Call coverage = 100% ( 1/1 )
```

```
Block Counters:
```

```
6-9 | 1
```

```
10 |
```

```
Block coverage = 100% ( 1/1 )
```

```
Total Call coverage = 100% ( 2/2 )
```

```
Total Block coverage = 100% ( 3/3 )
```

If you want to view the profiling information in XML format, run the `showpdf` utility with the `-xml` parameter.

```
showpdf -xml
```

The result is as follows:

```
<?xml version="1.0" encoding="UTF-8" ?>
- <XLTransformationReport xmlns="http://www.ibm.com/2010/04/CompilerTransformation" version="1.0">
- <CompilationStep name="showpdf">
- <ProgramHierarchy>
- <FileList>
- <File id="1" name="hello.c">
- <RegionList>
- <Region id="67" name="HelloWorld" startLineNumber="2" />
- <Region id="68" name="main" startLineNumber="6" />
- </RegionList>
- </File>
- </FileList>
- </ProgramHierarchy>
- <TransformationHierarchy />
- <ProfilingReports>
- <BlockCounterList>
- <BlockCounter regionId="67" execCount="1" coveredBlock="2" totalBlock="2">
- <BlockList>
- <Block index="3" execCount="1" startLineNumber="2" endLineNumber="4" />
- <Block index="2" execCount="0" startLineNumber="5" endLineNumber="5" />
- <Block index="4" execCount="1" startLineNumber="5" endLineNumber="5" />
- </BlockList>
- </BlockCounter>
- <BlockCounter regionId="68" execCount="1" coveredBlock="1" totalBlock="1">
- <BlockList>
- <Block index="3" execCount="1" startLineNumber="6" endLineNumber="9" />
- <Block index="2" execCount="0" startLineNumber="10" endLineNumber="10" />
- </BlockList>
- </BlockCounter>
- </BlockCounterList>
- <CallCounterList>
- <CallCounter regionId="67" execCount="1" coveredCall="0" totalCall="0">
- <CallList>
- <Call name="printf" execCount="1" lineNumber="4" />
- </CallList>
- </CallCounter>
- <CallCounter regionId="68" execCount="1" coveredCall="0" totalCall="0">
- <CallList>
- <Call name="HelloWorld" execCount="1" lineNumber="8" />
- </CallList>
- </CallCounter>
- </CallCounterList>
- <ValueProfileList />
- <CacheMissList />
- </ProfilingReports>
- </CompilationStep>
- </XLTransformationReport>
```

Related information in the *XL C/C++ Compiler Reference*



`-qpdf1, -qpdf2`



`-qshowpdf`

Object level profile-directed feedback

About this task

In addition to optimizing entire executables, profile-directed feedback (PDF) can also be applied to specific objects. This can be an advantage in applications where patches or updates are distributed as object files or libraries rather than as executables. Also, specific areas of functionality in your application can be optimized without you needing to go through the process of relinking the entire application. In large applications, you can save the time and trouble that otherwise need to be spent relinking the application.

The process for using object level PDF is essentially the same as the standard PDF process but with a small change to the `-qpdf2` step. For object level PDF, compile

your program using the **-qpdf1** option, execute the resulting application with representative data, compile the program again with the **-qpdf2** option, but now also use the **-qnoipa** option so that the linking step is skipped.

The steps below outline this process:

1. Compile your program using the **-qpdf1** option. For example:

```
xlc -c -O3 -qpdf1 file1.c file2.c file3.c
```

In this example, we are using the option **-O3** to indicate that we want a moderate level of optimization.

2. Link the object files to get an instrumented executable:

```
xlc -O3 -qpdf1 file1.o file2.o file3.o
```

Note: you must use the same optimization options. In this example, the optimization option **-O3**.

3. Run the instrumented executable with sample data that is representative of the data you want to optimize for.

```
a.out < sample_data
```

4. Compile the program again using the **-qpdf2** option. Specify the **-qnoipa** option so that the linking step is skipped and PDF optimization is applied to the object files rather than to the entire executable.

Note: you must use the same optimization options as in the previous steps. In this example, the optimization option **-O3**.

```
xlc -c -O3 -qpdf2 -qnoipa file1.c file2.c file3.c
```

The resulting output of this step are object files optimized for the sample data processed by the original instrumented executable. In this example, the optimized object files would be `file1.o`, `file2.o`, and `file3.o`. These can be linked using the system loader `ld` or by omitting the `-c` option in the **-qpdf2** step.

Notes:

- If you want to specify a file name for the profile that is created, use the **pdfname** suboption in both the **-qpdf1** and **-qpdf2** steps. For example:

```
xlc -O3 -qpdf1=pdfname=myprofile file1.c file2.c file3.c
```

Without the **pdfname** suboption, by default the file name is `._pdf`; the location of the file is the current working directory or whatever directory you have set using the `PDFDIR` environment variable. If the `PDFDIR` environment variable is set but the specified directory does not exist, the compiler issues a warning message.

- Because the **-qnoipa** option needs to be specified in the **-qpdf2** step so that linking of your object files is skipped, you cannot use interprocedural analysis (IPA) optimizations and object level PDF at the same time.

For details, see the `-qpdf1`, `-qpdf2` section in the XL C/C++ Compiler Reference.

Using compiler reports to diagnose optimization opportunities

You can use the **-qlistfmt** option to generate a compiler report in XML or HTML format that indicates some of the details of how your program was optimized. You can also use the **genhtml** tool to convert an existing XML report to HTML format. This information can be used to understand your application code and to tune your code for better performance.

The compiler report in XML format can be viewed in a browser that supports XSLT. If you compile with the stylesheet suboption, **-qlistfmt=xml=all:stylesheet=xstyle.xsl**, the report contains a link to a stylesheet that renders the XML readable and provides you with opportunities to improve the optimization of your code. You can also create tools to parse this information.

Inline reports

If compiled with **-qinline** and one of **-qlistfmt=xml=inlines**, **-qlistfmt=html=inlines**, **-qlistfmt=xml** or **-qlistfmt=html**, the compiler report that is generated includes a list of inline attempts during the compilation. The report also specifies the type of attempt and its outcome.

For each function that the compiler has attempted to inline, there is an indication of whether the inline was successful. The report might contain any number of explanations for a named function that has not been successfully inlined. Some examples of these explanations are:

- **FunctionTooBig** - The function is too big to be inlined.
- **RecursiveCall** - The function is not inlined because it is recursive.
- **ProhibitedByUser** - Inlining was not performed because of a user specified pragma or directive.
- **CallerIsNoopt** - No inlining was performed because the caller was compiled without optimization.
- **WeakAndNotExplicitlyInline** - The calling function is weak and not marked as inline.

For a complete list of the possible explanations, see the **Inline optimization types** section of the XML schema file called **XMLContent.html** that is in the `/usr/vacpp/listings/` directory.

Loop transformations

If compiled with **-qhot** and one of **-qlistfmt=xml=transforms**, **-qlistfmt=html=transforms**, **-qlistfmt=xml** or **-qlistfmt=html**, the compiler report that is generated includes a list of the transformations performed on all loops in the file during the compilation. It also lists reasons why some transformations were not performed.

- Reasons why a loop cannot be automatically parallelized
- Reasons why a loop cannot be unrolled
- Reasons why SIMD vectorization failed

For a complete list of the possible transformation problems, see the **Loop transformation types** section of the XML schema file called **XMLContent.html** that is in the `/usr/vacpp/listings/` directory.

Data reorganizations

If compiled with **-qhot** and one of **-qlistfmt=xml=data**, **-qlistfmt=html=data**, **-qlistfmt=xml** or **-qlistfmt=html**, the compiler report that is generated includes a list of data reorganizations performed on the program during compilation. Here are some examples of data reorganizations:

- Array splitting
- Array coalescing

- Array interleaving
- Array transposition
- Memory merge

For each of these reorganizations, the report contains details about the name of the data, file names, line numbers, and the region names.

Profile-directed feedback reports

If compiled with `-qpdf` and one of `-qlistfmt=xml=pdf`, `-qlistfmt=html=pdf`, `-qlistfmt=xml` or `-qlistfmt=html`, the compiler report that is generated includes the following information:

- Loop iteration counts
- Block and call counts
- Cache misses (if compiled with `-qpdf1=level=2`)
- Relevance of profiling data
- Missing profiling data
- Outdated profiling data

Parsing compiler reports with development tools

Software development tools can be created to parse the compiler reports produced in XML or HTML format. These tools can help direct you to opportunities to improve the performance of your application.

The compiler includes an XML schema that you can use to create a tool to parse the compiler reports and display aspects of your code that may represent performance improvement opportunities. The schema, `xllisting.xsd`, is located in the `/usr/vacpp/listings/` directory. There is also a version of the file designed for you to read in your browser. It is called `XMLContent.html`.

This schema presents the information from the report in a tree structure.

Other optimization options


Options are available to control particular aspects of optimization. They are often enabled as a group or given default values when you enable a more general optimization option or level.

For more information on these options, see the heading for each option in the *XL C/C++ Compiler Reference*.








Table 24. Selected compiler options for optimizing performance

Option	Description
<code>-qignerrno</code>	Allows the compiler to assume that <code>errno</code> is not modified by library function calls, so that such calls can be optimized. Also allows optimization of square root operations, by generating inline code rather than calling a library function. (For processors that support <code>sqrt</code> .)
<code>-qsmallstack</code>	Instructs the compiler to compact stack storage. Doing so might increase heap usage, which might increase execution time. However, it might be necessary for the program to run or to be optimally multithreaded.
<code>-qinline</code>	Controls inlining.

Table 24. Selected compiler options for optimizing performance (continued)

Option	Description
-qunroll	Independently controls loop unrolling. -qunroll is implicitly activated under -O3 .
-qinlglue	Instructs the compiler to inline the "glue code" generated by the linker and used to make a call to an external function or a call made through a function pointer.
-qtbtable	Controls the generation of traceback table information.
 -qnoeh	Informs the compiler that no C++ exceptions will be thrown and that cleanup code can be omitted. If your program does not throw any C++ exceptions, use this option to compact your program by removing exception-handling code.
-qnounwind	Informs the compiler that the stack will not be unwound while any routine in this compilation is active. This option can improve optimization of non-volatile register saves and restores. In C++, the -qnounwind option implies the -qnoeh option. It should not be used if the program uses <code>setjmp/longjmp</code> or any other form of exception handling.
-qstrict	Disables all transformations that change program semantics. In general, compiling a program correctly with -qstrict and any levels of optimization produces the same results as without optimization. For details about -qstrict and all of its suboptions, see <code>-qstrict</code> in the <i>XL C/C++ Compiler Reference</i> .
-qnostrict	Allows the compiler to reorder floating-point calculations and potentially excepting instructions. A potentially excepting instruction is one that might raise an interrupt due to erroneous execution (for example, floating-point overflow, a memory access violation). -qnostrict is used by default for optimization levels -O3 and higher.
-qlargepage	Supports large 16M pages in addition to the default 4K pages, to allow hardware prefetching to be done more efficiently. Informs the compiler that heap and static data will be allocated from large pages at execution time.
-qprefetch	Inserts prefetch instructions in compiled code to improve code performance. In situations where you are working with applications that generate a high cache-miss rate, you can use its suboption assistthread to generate prefetching assist threads (for example, <code>-qprefetch=assistthread</code>). -qnoprefetch is the default option. For details, see <code>-qprefetch</code> in the <i>XL C/C++ Compiler Reference</i> .

Related information in the *XL C/C++ Compiler Reference*

-  `-qignerrno`
-  `-qsmallstack`
-  `-qinline`
-  `-qunroll / #pragma unroll`
-  `-qinlglue`
-  `-qtbtable`
-  `-qeh` (C++ only)

 -qunwind

 -qstrict

 -qlargepage

 -qprefetch

Chapter 12. Debugging optimized code

Debugging optimized programs presents special usability problems. Optimization can change the sequence of operations, add or remove code, change variable data locations, and perform other transformations that make it difficult to associate the generated code with the original source statements.

For example:

Data location issues

With an optimized program, it is not always certain where the most current value for a variable is located. For example, a value in memory may not be current if the most current value is being stored in a register. Most debuggers are incapable of following the removal of stores to a variable, and to the debugger it appears as though that variable is never updated, or possibly even set. This contrasts with no optimization where all values are flushed back to memory and debugging can be more effective and usable.

Instruction scheduling issues

With an optimized program, the compiler may reorder instructions. That is, instructions may not be executed in the order the programmer would expect based on the sequence of lines in their original source code. Also, the sequence of instructions may not be contiguous. As the user steps through their program with a debugger, it may appear as if they are returning to a previously executed line in their code (interleaving of instructions).

Consolidating variable values

Optimizations can result in the removal and consolidation of variables. For example, if a program has two expressions that assign the same value to two different variables, the compiler may substitute a single variable. This can inhibit debug usability because a variable that a programmer is expecting to see is no longer available in the optimized program.

There are a couple of different approaches you can take to improve debug capabilities while also optimizing your program:

Debug non-optimized code first

Debug a non-optimized version of your program first, then recompile it with your desired optimization options. See “Debugging in the presence of optimization” on page 100 for some compiler options that are useful in this approach.

Use -g level

Use the **-g** level suboption to control the amount of debugging information made available. Increasing it improves debug capability, but prevents some optimizations.

Use -qoptdebug

When compiling with **-O3** optimization or higher, use the compiler option **-qoptdebug** to generate a pseudocode file that more accurately maps to how instructions and variable values will operate in an optimized program. With this option, when you load your program into a debugger,

you will be debugging the pseudocode for the optimized program. See “Using `-qoptdebug` to help debug optimized programs” on page 101 for more information.

Understanding different results in optimized programs

Here are some reasons why an optimized program might produce different results from one that has not undergone the optimization process:

- Optimized code can fail if a program contains code that is not valid. The optimization process relies on your application conforming to language standards.
- If a program that works without optimization fails when you optimize, check the cross-reference listing and the execution flow of the program for variables that are used before they are initialized. Compile with the `-qinitauto=hex_value` option to try to produce the incorrect results consistently. For example, using `-qinitauto=FF` gives variables an initial value of "negative not a number" (-NAN). Any operations on these variables will also result in NAN values. Other bit patterns (*hex_value*) may yield different results and provide further clues as to what is going on. Programs with uninitialized variables can appear to work properly when compiled without optimization, because of the default assumptions the compiler makes, but can fail when you optimize. Similarly, a program can appear to execute correctly after optimization, but fails at lower optimization levels or when run in a different environment.
- A variation on uninitialized storage. Referring to an automatic-storage variable by its address after the owning function has gone out of scope leads to a reference to a memory location that can be overwritten as other auto variables come into scope as new functions are called.

Use with caution debugging techniques that rely on examining values in storage. The compiler might have deleted or moved a common expression evaluation. It might have assigned some variables to registers, so that they do not appear in storage at all.

Debugging in the presence of optimization

Debug and compile your program with your desired optimization options. Test the optimized program before placing it into production. If the optimized code does not produce the expected results, you can attempt to isolate the specific optimization problems in a debugging session.

The following list presents options that provide specialized information, which can be helpful during the development of optimized code:

- qlist** Instructs the compiler to emit an object listing. The object listing includes hex and pseudo-assembly representations of the generated instructions, traceback tables, and text constants.
- qreport** Instructs the compiler to produce a report of the loop transformations it performed and how the program was parallelized. For **-qreport** to generate a listing, the options **-qhot** or **-qsmp** should also be specified.
- qipa=list** Instructs the compiler to emit an object listing that provides information for IPA optimization.

-qcheck

Generates code that performs certain types of runtime checking.

-qsmp=noopt

If you are debugging SMP code, **-qsmp=noopt** ensures that the compiler performs only the minimum transformations necessary to parallelize your code and preserves maximum debug capability.

-qoptdebug

When used with high levels of optimization, produces files containing optimized pseudocode that can be read by a debugger.

-qkeepparam

Ensures that procedure parameters are stored on the stack even during optimization. This can negatively impact execution performance. The **-qkeepparam** option then provides access to the values of incoming parameters to tools, such as debuggers, simply by preserving those values on the stack.

-qinitauto

Instructs the compiler to emit code that initializes all automatic variables to a given value.

-qextchk

Generates additional symbolic information to allow the linker to do cross-file type checking of external variables and functions. This option requires the linker **-btypchk** option to be active.

-g

Generates debugging information for use by a symbolic debugger. You can use different **-g** levels to debug optimized code by viewing or possibly modifying accessible variables at selected source locations in the debugger.

In addition, you can also use the **snapshot** pragma to ensure that certain variables are visible to the debugger at points in your application.

Using **-qoptdebug** to help debug optimized programs

The purpose of the **-qoptdebug** compiler option is to aid the debugging of optimized programs. It does this by creating pseudocode that maps more closely to the instructions and values of an optimized program than the original source code. When a program compiled with this option is loaded into a debugger, you will be debugging the pseudocode rather than your original source. By making optimizations explicit in pseudocode, you can gain a better understanding of how your program is really behaving under optimization. Files containing the pseudocode for your program are generated with the file suffix `.optdbg`. Only line debugging is supported for this feature.

Compile your program as in the following example:

```
xlc myprogram.c -O3 -qhot -g -qoptdebug
```

In this example, your source file is compiled to `a.out`. The pseudocode for the optimized program is written to a file called `myprogram.optdbg` which can be referred to while debugging your program.

Notes:

- The **-g** or the **-qlinedebug** option must also be specified in order for the compiled executable to be debuggable. However, if neither of these options are specified, the pseudocode file `<output_file>.optdbg` containing the optimized pseudocode is still generated.
- The **-qoptdebug** option only has an effect when one or more of the optimization options **-qhot**, **-qsmp**, **-qpdf**, or **-qipa** are specified, or when the optimization levels that imply these options are specified; that is, the optimization levels **-O3**, **-O4**, and **-O5**. The example shows the optimization options **-qhot** and **-O3**.

Debugging the optimized program

From the following examples, you can see how the compiler might apply optimizations to a simple program and how debugging it can differ from debugging your original source.

Example 1: Represents the original non-optimized code for a simple program. It presents a couple of optimization opportunities to the compiler. For example, the variables `z` and `d` are both assigned by the equivalent expressions `x + y`. Therefore, these two variables can be consolidated in the optimized source. Also, the loop can be unrolled. In the optimized source, you can see iterations of the loop listed explicitly.

Example 2: Represents a listing of the optimized source as shown in the debugger. Note the unrolled loop and the consolidation of values assigned by the `x + y` expression.

Example 3: Shows an example of stepping through the optimized source using the debugger. Note, there is no longer a correspondence between the line numbers for these statements in the optimized source as compared to the line numbers in the original source.

Example 1: Original code

```
#include "stdio.h"

void foo(int x, int y, char* w)
{
    char* s = w+1;
    char* t = w+1;
    int z = x + y;
    int d = x + y;
    int a = printf("TEST\n");

    for (int i = 0; i < 4; i++)
        printf("%d %d %d %s %s\n", a, z, d, s, t);
}

int main()
{
    char d[] = "DEBUG";
    foo(3, 4, d);
    return 0;
}
```

Example 2: dbx debugger listing

```
(dbx) list
   1          3 | void foo(long x, long y, char * w)
   2          4 | {
   3          9 |     a = printf("TEST/n");
   4         12 |     printf("%d %d %d %s %s/n",a,x + y,x + y,
```

```

5          ((char *)w + 1),((char *)w + 1));
          printf("%d %d %d %s %s/n",a,x + y,x + y,
          ((char *)w + 1),((char *)w + 1));
6          printf("%d %d %d %s %s/n",a,x + y,x + y,
          ((char *)w + 1),((char *)w + 1));
7          printf("%d %d %d %s %s/n",a,x + y,x + y,
          ((char *)w + 1),((char *)w + 1));
8      13 |   return;
9          } /* function */
10
11
12      15 |   long main()
13      16 |   {
14      17 |       d$init$0 = "DEBUG";
15      18 |       @PARM.x0 = 3;
16          @PARM.y1 = 4;
17          @PARM.w2 = &d;
18      9 |       a = printf("TEST/n");
19      12 |       printf("%d %d %d %s %s/n",a,@PARM.x0 + @PARM.y1,
          @PARM.x0 + @PARM.y1,((char *)@PARM.w2 + 1),
          ((char *)@PARM.w2 + 1));
20          printf("%d %d %d %s %s/n",a,@PARM.x0 + @PARM.y1,
          @PARM.x0 + @PARM.y1,((char *)@PARM.w2 + 1),
          ((char *)@PARM.w2 + 1));
21          printf("%d %d %d %s %s/n",a,@PARM.x0 + @PARM.y1,
          @PARM.x0 + @PARM.y1,((char *)@PARM.w2 + 1),
          ((char *)@PARM.w2 + 1));
22          printf("%d %d %d %s %s/n",a,@PARM.x0 + @PARM.y1,
          @PARM.x0 + @PARM.y1,((char *)@PARM.w2 + 1),
          ((char *)@PARM.w2 + 1));
23      19 |       rstr = 0;
24          return rstr;
25      20 |   } /* function */

```

Example 3: Stepping through optimized source

```

(dbx) stop at 18
[1] stop at "myprogram.o.optdbg":18
(dbx) run
[1] stopped in main at line 18 in file "myprogram.o.optdbg"
    18      9 |   a = printf("TEST/n");
(dbx) cont
TEST
5 7 7 EBUG EBUG
5 7 7 EBUG EBUG
5 7 7 EBUG EBUG
5 7 7 EBUG EBUG

execution completed

```

Chapter 13. Coding your application to improve performance

Chapter 11, “Optimizing your applications,” on page 71 discusses the various compiler options that the XL C/C++ compiler provides for optimizing your code with minimal coding effort. If you want to take your application a step further, to complement and take the most advantage of compiler optimizations, the following sections discuss C and C++ programming techniques that can improve performance of your code:

- “Finding faster input/output techniques”
- “Reducing function-call overhead”
- “Using delegating constructors (C++0x)” on page 107
- “Using template explicit instantiation declarations (C++0x)” on page 107
- “Managing memory efficiently” on page 108
- “Optimizing variables” on page 108
- “Manipulating strings efficiently” on page 109
- “Optimizing expressions and program logic” on page 110
- “Optimizing operations in 64-bit mode” on page 110

Finding faster input/output techniques







There are a number of ways to improve your program's performance of input and output:


- If your file I/O accesses do not exhibit locality (that is truly random access such as in a database), implement your own buffering or caching mechanism on the low-level I/O functions.
- If you do your own I/O buffering, make the buffer a multiple of 4K, which is the size of a page.
- Use buffered I/O to handle text files.
- If you know you have to process an entire file, determine the size of the data to be read in, allocate a single buffer to read it to, read the whole file into that buffer at once using read, and then process the data in the buffer. This reduces disk I/O, provided the file is not so big that excessive swapping will occur. Consider using the `mmap` function to access the file.

Reducing function-call overhead

When you write a function or call a library function, consider the following guidelines:

- Call a function directly, rather than using function pointers.
- Use `const` arguments in inlined functions whenever possible. Functions with constant arguments provide more opportunities for optimization.
- Use the `#pragma expected_value` preprocessor directive so that the compiler can optimize for common values used with a function.
- Use the `#pragma isolated_call` preprocessor directive to list functions that have no side effects and do not depend on side effects.
- Use the `restrict` keyword for pointers that can never point to the same memory.

- Use `#pragma disjoint` within functions for pointers or reference parameters that can never point to the same memory.
- Declare a nonmember function as static whenever possible. This can speed up calls to the function and increase the likelihood that the function will be inlined.
-  Usually, you should not declare all your virtual functions inline. If all virtual functions in a class are inline, the virtual function table and all the virtual function bodies will be replicated in each compilation unit that uses the class.
-  When declaring functions, use the `const` specifier whenever possible.
-  Fully prototype all functions. A full prototype gives the compiler and optimizer complete information about the types of the parameters. As a result, promotions from unwidened types to widened types are not required, and parameters can be passed in appropriate registers.
-  Avoid using unprototyped variable argument functions.
- Design functions so that they have few parameters and the most frequently used parameters are in the leftmost positions in the function prototype.
- Avoid passing by value large structures or unions as function parameters or returning a large structure or a union. Passing such aggregates requires the compiler to copy and store many values. This is worse in C++ programs in which class objects are passed by value because a constructor and destructor are called when the function is called. Instead, pass or return a pointer to the structure or union, or pass it by reference.
- Pass non-aggregate types such as `int` and `short` or small aggregates by value rather than passing by reference, whenever possible.
- If your function exits by returning the value of another function with the same parameters that were passed to your function, put the parameters in the same order in the function prototypes. The compiler can then branch directly to the other function.
- Use the built-in functions, which include string manipulation, floating-point, and trigonometric functions, instead of coding your own. Intrinsic functions require less overhead and are faster than a function call, and often allow the compiler to perform better optimization.
 -  Many functions from the C++ standard libraries are mapped to optimized built-in functions by the compiler.
 -  Many functions from `string.h` and `math.h` are mapped to optimized built-in functions by the compiler.
- Selectively mark your functions for inlining, using the `inline` keyword. An inlined function requires less overhead and is generally faster than a function call. The best candidates for inlining are small functions that are called frequently from a few places, or functions called with one or more compile-time constant parameters, especially those that affect `if`, `switch` or `for` statements. You might also want to put these functions into header files, which allows automatic inlining across file boundaries even at low optimization levels. Be sure to inline all functions that only load or store a value, or use simple operators such as comparison or arithmetic operators. Large functions and functions that are called rarely are generally not good candidates for inlining. Neither are medium size functions that are called from many places.
- Avoid breaking your program into too many small functions. If you must use small functions, seriously consider using the `-qipa` compiler option, which can automatically inline such functions, and uses other techniques for optimizing calls between functions.

-  C++ Avoid virtual functions and virtual inheritance unless required for class extensibility. These language features are costly in object space and function invocation performance.

Related information in the *XL C/C++ Compiler Reference*

 #pragma expected_value

 -qisolated_call / #pragma isolated_call

 #pragma disjoint

 -qipa

Using delegating constructors (C++0x)

Note: C++0x is a new version of the C++ programming language standard. IBM continues to develop and implement the features of the new standard. The implementation of the language level is based on IBM's interpretation of the standard. Until IBM's implementation of all the features of the C++0x standard is complete, including the support of a new C++ standard library, the implementation may change from release to release. IBM makes no attempt to maintain compatibility, in source, binary, or listings and other compiler interfaces, with earlier releases of IBM's implementation of the new features of the C++0x standard and therefore they should not be relied on as a stable programming interface.

Use the delegating constructors feature to concentrate common initializations in one constructor. This helps reduce the code size and make program more readable and maintainable.

This technique is described in “Using delegating constructors (C++0x)” on page 41.

Using template explicit instantiation declarations (C++0x)






Note: C++0x is a new version of the C++ programming language standard. IBM continues to develop and implement the features of the new standard. The implementation of the language level is based on IBM's interpretation of the standard. Until IBM's implementation of all the features of the C++0x standard is complete, including the support of a new C++ standard library, the implementation may change from release to release. IBM makes no attempt to maintain compatibility, in source, binary, or listings and other compiler interfaces, with earlier releases of IBM's implementation of the new features of the C++0x standard and therefore they should not be relied on as a stable programming interface.

Use the explicit instantiation declarations feature to suppress the implicit instantiation of a template specialization or its members. This helps reduce the collective size of the object files and shorten compile time.

This technique is described in “Using explicit instantiation declarations (C++0x)” on page 49.

Managing memory efficiently

Because C++ objects are often allocated from the heap and have limited scope, memory use affects performance more in C++ programs than it does in C programs. For that reason, consider the following guidelines when you develop C++ applications:

- In a structure, declare the largest aligned members first. Members of similar alignment should be grouped together where possible.
- In a structure, place variables near each other if they are frequently used together.
-  C++ Ensure that objects that are no longer needed are freed or otherwise made available for reuse. One way to do this is to use an *object manager*. Each time you create an instance of an object, pass the pointer to that object to the object manager. The object manager maintains a list of these pointers. To access an object, you can call an object manager member function to return the information to you. The object manager can then manage memory usage and object reuse.
- Storage pools are a good way of keeping track of used memory (and reclaiming it) without having to resort to an object manager or reference counting.
- For XL C/C++ for AIX, V11.1 compiler with the September PTF (11.1.0.03) and later, consider using the `<ssstring>` header file that is supplied by IBM for programs that create large numbers of small strings. The header file uses the Small Buffer Optimization (SBO) technique that can reduce the number of dynamic memory allocations at program execution time so runtime overhead is reduced and runtime performance is improved. The public interface of the header file is identical to the `<string>` header file in the standard C++ library. For more information about using the header file, see Small String Optimized (SSO) string class and `<string>`.
-  C++ Avoid copying large, complicated objects.
-  C++ Avoid performing a *deep copy* if a *shallow copy* is all you require. For an object that contains pointers to other objects, a shallow copy copies only the pointers and not the objects to which they point. The result is two objects that point to the same contained object. A deep copy, however, copies the pointers and the objects they point to, as well as any pointers or objects contained within that object, and so on. A deep copy must be performed in multithreaded environments, because it reduces sharing and synchronization.
-  C++ Use virtual methods only when absolutely necessary.
-  C++ Use the "Resource Acquisition is Initialization" (RAII) pattern.
- Use `boost::shared_ptr` and `boost::weak_ptr`.

Optimizing variables

Consider the following guidelines:

- Use local variables, preferably automatic variables, as much as possible. The compiler must make several worst-case assumptions about global variables. For example, if a function uses external variables and also calls external functions, the compiler assumes that every call to an external function could use and change the value of every external variable. If you know that a global variable is not read or affected by any function call, and this variable is read several times with function calls interspersed, copy the global variable to a local variable and then use this local variable.

- If you must use global variables, use static variables with file scope rather than external variables whenever possible. In a file with several related functions and static variables, the optimizer can gather and use more information about how the variables are affected.
- If you must use external variables, group external data into structures or arrays whenever it makes sense to do so. All elements of an external structure use the same base address. Do not group variables whose addresses are taken with variables whose addresses are not taken.
- The `#pragma isolated_call` preprocessor directive can improve the runtime performance of optimized code by allowing the compiler to make less pessimistic assumptions about the storage of external and static variables. Isolated call functions with constant or loop-invariant parameters can be moved out of loops, and multiple calls with the same parameters can be replaced with a single call.
- Avoid taking the address of a variable. If you use a local variable as a temporary variable and must take its address, avoid reusing the temporary variable for a different purpose. Taking the address of a local variable can inhibit optimizations that would otherwise be done on calculations involving that variable.
- Use constants instead of variables where possible. The optimizer is able to do a better job reducing runtime calculations by doing them at compile time instead. For instance, if a loop body has a constant number of iterations, use constants in the loop condition to improve optimization (`for (i=0; i<4; i++)` can be better optimized than `for (i=0; i<x; i++)`).
- Use register-sized integers (long data type) for scalars to avoid sign extension instructions after each change in 64-bit mode. For large arrays of integers, consider using one- or two-byte integers or bit fields.
- Use the smallest floating-point precision appropriate to your computation. Use the long double data type only when high precision is required.

Related information in the *XL C/C++ Compiler Reference*



`-qisolated_call / #pragma isolated_call`

Manipulating strings efficiently

The handling of string operations can affect the performance of your program.

- When you store strings into allocated storage, align the start of the string on an 8-byte boundary.
- Keep track of the length of your strings. If you know the length of a string, you can use `mem` functions instead of `str` functions. For example, `memcpy` is faster than `strcpy` because it does not have to search for the end of the string.
- If you are certain that the source and target do not overlap, use `memcpy` instead of `memmove`. This is because `memcpy` copies directly from the source to the destination, while `memmove` might copy the source to a temporary location in memory before copying to the destination (depending on the length of the string).
- When manipulating strings using `mem` functions, faster code can be generated if the `count` parameter is a constant rather than a variable. This is especially true for small count values.
- Make string literals read-only, whenever possible. This improves certain optimization techniques and reduces memory usage if there are multiple uses of

the same string. You can explicitly set strings to read-only by using **#pragma strings (readonly)** in your source files or **-qro** (this is enabled by default) to avoid changing your source files.

Related information in the XL C/C++ Compiler Reference

 `-qro / #pragma strings`

Optimizing expressions and program logic

Consider the following guidelines:

- If components of an expression are used in other expressions and they include function calls or there are function calls between the uses, assign the duplicated values to a local variable.
- Avoid forcing the compiler to convert numbers between integer and floating-point internal representations. For example:

```
float array[10];
float x = 1.0;
int i;
for (i = 0; i < 9; i++) {      /* No conversions needed */
    array[i] = array[i]*x;
    x = x + 1.0;
}
for (i = 0; i < 9; i++) {      /* Multiple conversions needed */
    array[i] = array[i]*i;
}
```


When you must use mixed-mode arithmetic, code the integer and floating-point arithmetic in separate computations whenever possible.

- Do not use global variables as loop indices or bounds.
- Avoid goto statements that jump into the middle of loops. Such statements inhibit certain optimizations.
- Improve the predictability of your code by making the fall-through path more probable. Code such as:

```
if (error) {handle error} else {real code}
```

should be written as:

```
if (!error) {real code} else {error}
```

- If one or two cases of a switch statement are typically executed much more frequently than other cases, break out those cases by handling them separately before the switch statement. If possible, replace the switch statement by checking whether the value is in range to be obtained from an array.
-  Use try blocks for exception handling only when necessary because they can inhibit optimization.
- Keep array index expressions as simple as possible.

Optimizing operations in 64-bit mode

The ability to handle larger amounts of data directly in physical memory rather than relying on disk I/O is perhaps the most significant performance benefit of 64-bit machines. However, some applications compiled in 32-bit mode perform better than when they are recompiled in 64-bit mode. Some reasons for this include:

- 64-bit programs are larger. The increase in program size places greater demands on physical memory.
- 64-bit long division is more time-consuming than 32-bit integer division.

- 64-bit programs that use 32-bit signed integers as array indexes or loop counts might require additional instructions to perform sign extension each time the array is referenced or the loop count is incremented.

Some ways to compensate for the performance liabilities of 64-bit programs include:

- Avoid performing mixed 32- and 64-bit operations. For example, adding a 32-bit data type to a 64-bit data type requires that the 32-bit type be sign-extended to clear or set the upper 32 bits of the register. This slows the computation.
- Use long types instead of signed, unsigned, and plain int types for variables which will be frequently accessed, such as loop counters and array indexes. Doing so frees the compiler from having to truncate or sign-extend array references, parameters during function calls, and function results during returns.

Tracing functions in your code

You can instruct the compiler to insert calls to user-defined tracing functions to aid in debugging or timing the execution of other functions.

Using tracing functions in your program requires the following steps:

1. Writing tracing functions.
2. Specifying which functions to trace with the **-qfunctrace** option.

Using the **-qfunctrace** option causes the compiler to insert calls to these tracing functions at key points in the function body; however you are responsible for defining these tracing functions. The following list describes at which points the tracing functions are called:

- The compiler inserts calls to the tracing function at the entry point of a function. The line number passed to the routine is the line number of the first executable statement in the instrumented function.
- The compiler inserts calls to the tracing function at the exit point of a function. The line number that is passed to the function is the line number of the statement causing the exit in the instrumented function.
- The catch tracing function is called at the beginning of the C++ catch block when the exception occurs.

You can use the **-qnofunctrace** compiler option or the `#pragma nofunctrace` pragma to disable function tracing.

How to write tracing functions

To trace functions in your code, define the following tracing functions:

- `__func_trace_enter` is the entry point tracing function.
- `__func_trace_exit` is the exit point tracing function.
- `__func_trace_catch` is the catch tracing function.

The prototypes of these functions are as follows:

- `void __func_trace_enter(const char *const function_name, const char *const file_name, int line_number, void **const user_data);`
- `void __func_trace_exit(const char *const function_name, const char *const file_name, int line_number, void **const user_data);`
- `void __func_trace_catch(const char *const function_name, const char *const file_name, int line_number, void **const user_data);`

In the preceding tracing functions, the descriptions for their variables are as follows:

- `function_name` is the name of the function you want to trace.
- `file_name` is the name of the file.
- `line_number` is the line number at entry or exit point of the function. This is a 4-byte number.
- `user_data` is the address of a static pointer variable. The static pointer variable is generated by the compiler and initialized to NULL; in addition, because the pointer variable is static, its address is the same for all instrumentation calls inside the same function.

Notes:

- The exit function is not called if the function has an abnormal exit. The abnormal exit can be caused by C++ exception throws, raise the signal, or calls `exit`.
- The `-qfunctrace` option does not support `setjmp` and `longjmp`. For example, a call to `longjmp()` that leaves `function1` and returns from `setjmp()` in `function2` will have a missing call to `__func_trace_exit` in `function1` and a missing a call to `__func_trace_enter` in `function2`.
- The catch function is called at the point where the C++ exception is caught by user code.
- To define tracing functions in C++ programs, use the `extern "C"` linkage directive before your function definition.
- The function calls are only inserted into the function definition, and if a function is inlined, no tracing is done within the inlined code.
- If you develop a multithreaded program, make sure the tracing functions have the proper synchronization. Calls to the tracing functions are not thread-safe.
- If you specify a function that does not exist with the option, the function is ignored.

Rules

The following rules apply when you trace functions in your code:

- When optimization is enabled, line numbers might not be accurate.
- The tracing function must not call instrumented function; otherwise an infinite loop might occur.
- If you instruct the compiler to trace recursive functions, make sure that your tracing functions can handle recursion.
- Inlined functions are not instrumented.
- Tracing functions are not instrumented.
- Compiler-generated functions are not instrumented, except for the outlined functions generated by optimization such as OpenMP. In those cases, the name of the outlined function contains the name of the original user function as prefix.
- Tracing functions might be called during static initialization. You must be careful that anything used in the tracing functions are initialized before the first possible call to the tracing function.

Examples

The following C example shows how you can trace functions in your code using function prototypes. Assume you want to trace the entry and exit points of `function1` and `function2`, as well as how much time it takes the compiler to trace them in the following code:

```
#include <stdio.h>
#include <stdlib.h>
#include <unistd.h>
#include <time.h>
#ifdef __cplusplus
extern "C"
#endif
void __func_trace_enter(const char *function_name, const char *file_name,
                      int line_number, void** const user_data){
    if((*user_data)==NULL)
        (*user_data)=(time_t *)malloc(sizeof(time_t));
    (*(time_t *)*user_data)=time(NULL);
    printf("begin function: name=%s file=%s line=%d\n",function_name,file_name,
          line_number);
}
#ifdef __cplusplus
extern "C"
#endif
void __func_trace_exit(const char *function_name, const char*file_name,
                     int line_number, void** const user_data){
    printf("end function: name=%s file=%s line=%d. It took %g seconds\n",
          function_name,file_name,line_number, difftime(time(NULL),
        *(time_t *)*user_data));
}
void function2(void){
    sleep(3);
}
void function1(void){
    sleep(5);
    function2();
}
int main(){
    function1();
}
```

Run `xlc t.c -qfunctrace+function1:function2` to output function trace results:

```
begin function: name=function1 file=t.c line=27
begin function: name=function2 file=t.c line=24
end function: name=function2 file=t.c line=25. It took 3 seconds
end function: name=function1 file=t.c line=29. It took 8 seconds
```

As you see from the preceding example, the `user_data` parameter is defined to use the system time as basis for time calculation. The following steps explain how `user_data` is defined to achieve this goal:

1. The function reserves a memory area for storing the value of `user_data`.
2. The system time is used as the value for `user_data`.
3. In the `__func_trace_exit` function, the `difftime` function uses `user_data` to calculate time differences. The result is displayed in the form of `It took %g seconds` in the output.

The following C++ example shows how tracing functions are called. The following example traces class `myStack`, function `foo`, and disables tracing for `int main()` using `#pragma nofunctrace`:

```

#include <iostream>
#include <vector>
#include <stdexcept>
#include <stdio.h>
#include <stdlib.h>
#include <unistd.h>
#include <time.h>
extern "C"
void __func_trace_enter(const char *function_name, const char *file_name,
                       int line_number, void** const user_data){
    if((*user_data)==NULL)
        (*user_data)=(time_t *)malloc(sizeof(time_t));
    (*(time_t *)*user_data)=time(NULL);
    printf("enter function: name=%s file=%s line=%d\n",function_name,file_name,
          line_number);
}
extern "C"
void __func_trace_exit(const char *function_name, const char*file_name,
                      int line_number, void** const user_data){
    printf("exit function: name=%s file=%s line=%d. It took %g seconds\n",
          function_name, file_name, line_number, difftime(time(NULL),
          *(time_t *)*user_data));
}
extern "C"
void __func_trace_catch(const char *function_name, const char*file_name,
                       int line_number, void** const user_data){
    printf("catch function: name=%s file=%s line=%d. It took %g seconds\n",
          function_name, file_name,line_number, difftime(time(NULL),
          *(time_t *)*user_data));
}

template <typename T> class myStack{
private:
    std::vector<T> elements;
public:
    void push(T const&);
    void pop();
};

template <typename T>
void myStack<T>::push(T const& value){
    sleep(3);
    std::cout<< "\tpush(" << value << ")" <<std::endl;
    elements.push_back(value);
}
template <typename T>
void myStack<T>::pop(){
    sleep(5);
    std::cout<< "\tpop()" <<std::endl;
    if(elements.empty()){
        throw std::out_of_range("myStack is empty");
    }
    elements.pop_back();
}
void foo(){
    myStack<int> intValues;
    myStack<float> floatValues;
    myStack<double> doubleValues;
    intValues.push(4);
    floatValues.push(5.5f);
    try{
        intValues.pop();
        floatValues.pop();
        doubleValues.pop(); // cause exception
    } catch(std::exception const& e){
        std::cout<< "\tException: "<<e.what()<<std::endl;
    }
}

```

```

    std::cout<<"\tdone"<<std::endl;
}
#pragma nofunctrace(main)
int main(){
    foo();
}

```

Run `x1C t.cpp -qfunctrace+myStack:foo` or `x1C t.cpp -qfunctrace` to output function trace results:

```

enter function: name=foo__Fv file=t.cpp line=53
enter function: name=push__7myStackXTi_FRCi file=t.cpp line=39
    push(4)
exit function: name=push__7myStackXTi_FRCi file=t.cpp line=42. It took 3 seconds
enter function: name=push__7myStackXTf_FRCf file=t.cpp line=39
    push(5.5)
exit function: name=push__7myStackXTf_FRCf file=t.cpp line=42. It took 3 seconds
enter function: name=pop__7myStackXTi_Fv file=t.cpp line=45
    pop()
exit function: name=pop__7myStackXTi_Fv file=t.cpp line=51. It took 5 seconds
enter function: name=pop__7myStackXTf_Fv file=t.cpp line=45
    pop()
exit function: name=pop__7myStackXTf_Fv file=t.cpp line=51. It took 5 seconds
enter function: name=pop__7myStackXTd_Fv file=t.cpp line=45
    pop()
catch function: name=foo__Fv file=t.cpp line=62. It took 21 seconds
    Exception: myStack is empty
    done
exit function: name=foo__Fv file=t.cpp line=66. It took 21 seconds

```

Related information

- For details about the `-qfunctrace` compiler option, see `-qfunctrace` in the *XL C/C++ Compiler Reference*.
- See `#pragma nofunctrace` in the *XL C/C++ Compiler Reference* for details about the `#pragma nofunctrace`.

Using rvalue references (C++0x)

Note: C++0x is a new version of the C++ programming language standard. IBM continues to develop and implement the features of the new standard. The implementation of the language level is based on IBM's interpretation of the standard. Until IBM's implementation of all the features of the C++0x standard is complete, including the support of a new C++ standard library, the implementation may change from release to release. IBM makes no attempt to maintain compatibility, in source, binary, or listings and other compiler interfaces, with earlier releases of IBM's implementation of the new features of the C++0x standard and therefore they should not be relied on as a stable programming interface.

In C++0x, you can overload functions based on the value categories of arguments and similarly have lvalueness detected by template argument deduction. You can also have an rvalue bound to an rvalue reference and modify the rvalue through the reference. This enables a programming technique with which you can reuse the resources of expiring objects and therefore improve the performance of your libraries, especially if you use generic code with class types, for example, template data structures. Additionally, the value category can be considered when writing a forwarding function.

Move semantics

When you want to optimize the use of temporary values, you can use a move operation in what is known as destructive copying. Consider the following string concatenation and assignment:

```
std::string a, b, c;  
c = a + b;
```

In this program, the compiler first stores the result of `a + b` in an internal temporary variable, that is, an rvalue.

The signature of a normal copy assignment operator is as follows:

```
string& operator = (const string&)
```

With this copy assignment operator, the assignment consists of the following steps:

1. Copy the temporary variable into `c` using a deep-copy operation.
2. Discard the temporary variable.

Deep copying the temporary variable into `c` is not efficient because the temporary variable is discarded at the next step.

To avoid the needless duplication of the temporary variable, you can implement an assignment operator that moves the variable instead of copying the variable. That is, the argument of the operator is modified by the operation. A move operation is faster because it is done through pointer manipulation, but it requires a reference through which the source variable can be manipulated. However, `a + b` is a temporary value, which is not easily differentiated from a const-qualified value in C++ before C++0x for the purposes of overload resolution.

With rvalue references, you can create a move assignment operator as follows:

```
string& operator= (string&&)
```

With this move assignment operator, the memory allocated for the underlying C-style string in the result of `a + b` is assigned to `c`. Therefore, it is not necessary to allocate new memory to hold the underlying string in `c` and to copy the contents to the new memory.

The following code can be an implementation of the `string` move assignment operator:

```
string& string::operator=(string&& str)  
{  
    // The named rvalue reference str acts like an lvalue  
    std::swap(_capacity, str._capacity);  
    std::swap(_length, str._length);  
  
    // char* _str points to a character array and is a  
    // member variable of the string class  
    std::swap(_str, str._str);  
    return *this;  
}
```

However, in this implementation, the memory originally held by the string being assigned to is not freed until `str` is destroyed. The following implementation that uses a local variable is more memory efficient:

```
string& string::operator=(string&& parm_str)  
{  
    // The named rvalue reference parm_str acts like an lvalue
```

```

string sink_str;
std::swap(sink_str, parm_str);
std::swap(*this, sink_str);
return *this;
}

```

In a similar manner, the following program is a possible implementation of a string concatenation operator:

```

string operator+(string&& a, const string& b)
{
    return std::move(a+=b);
}

```

Note: The `std::move` function only casts the result of `a+=b` to an rvalue reference, without moving anything. The return value is constructed using a move constructor because the expression `std::move(a+=b)` is an rvalue. The relationship between a move constructor and a copy constructor is analogous to the relationship between a move assignment operator and a copy assignment operator.

Perfect forwarding

The `std::forward` function is a helper template, much like `std::move`. It returns a reference to its function argument, with the resulting value category determined by the template type argument. In an instantiation of a forwarding function template, the value category of an argument is encoded as part of the deduced type for the related template type parameter. The deduced type is passed to the `std::forward` function.

The wrapper function in the following example is a forwarding function template that forwards to the `do_work` function. Use `std::forward` in forwarding functions on the calls to the target functions. The following example also uses the `decltype` and trailing return type features to produce a forwarding function that forwards to one of the `do_work` functions. Calling the wrapper function with any argument results in a call to a `do_work` function if a suitable overload function exists. Extra temporaries are not created and overload resolution on the forwarding call resolves to the same overload as it would if the `do_work` function were called directly.

```




struct s1 *do_work(const int&);           // #1
struct s2 *do_work(const double&);      // #2
struct s3 *do_work(int&&);              // #3
struct s4 *do_work(double&&);           // #4
template <typename T> auto wrapper(T && a)->
    decltype(do_work(std::forward<T>(*static_cast<typename std::remove_reference<T>
        ::type*>(0))))
{
    return do_work(std::forward<T>(a));
}
template <typename T> void tPtr(T *t);
int main()
{
    int x;
    double y;
    tPtr<s1>(wrapper(x)); // calls #1
    tPtr<s2>(wrapper(y)); // calls #2
    tPtr<s3>(wrapper(0)); // calls #3
    tPtr<s4>(wrapper(1.0)); // calls #4
}

```

Related information in the XL C/C++ Compiler Reference

 **-qlanglvl**

Related information in the XL C/C++ Language Reference

-  Reference collapsing(C++0x)
-  The decltype(expression) type specifier (C++0x)
-  Trailing return type (C++0x)

Chapter 14. Using the high performance libraries

IBM XL C/C++ for AIX, V12.1 is shipped with a set of libraries for high-performance mathematical computing:

- The Mathematical Acceleration Subsystem (MASS) is a set of libraries of tuned mathematical intrinsic functions that provide improved performance over the corresponding standard system math library functions. MASS is described in “Using the Mathematical Acceleration Subsystem libraries (MASS).”
- The Basic Linear Algebra Subprograms (BLAS) are a set of routines which provide matrix/vector multiplication functions tuned for PowerPC architectures. The BLAS functions are described in “Using the Basic Linear Algebra Subprograms – BLAS” on page 132.

Using the Mathematical Acceleration Subsystem libraries (MASS)

XL C/C++ is shipped with a set of Mathematical Acceleration Subsystem (MASS) libraries for high-performance mathematical computing.

The MASS libraries consist of a library of scalar C/C++ functions described in “Using the scalar library” on page 120, a set of vector libraries tuned for specific architectures described in “Using the vector libraries” on page 122, and a SIMD library tuned for POWER7 described in “Using the SIMD library for POWER7” on page 128. The functions contained in both scalar and vector libraries are automatically called at certain levels of optimization, but you can also call them explicitly in your programs. Note that the accuracy and exception handling might not be identical in MASS functions and system library functions.

The MASS functions must run with the default rounding mode and floating-point exception trapping settings.

When you compile programs with any of the following sets of options:


- **-qhot -qignerrno -qnostrict**
- **-qhot -O3**
- **-O4**
- **-O5**

the compiler automatically attempts to vectorize calls to system math functions by calling the equivalent MASS vector functions (with the exceptions of functions `vdnint`, `vdint`, `vcosin`, `vscosin`, `vqdr`, `vsqdr`, `vrqdr`, `vsrqdr`, `vpopcnt4`, `vpopcnt8`, `vexp2`, `vexp2m1`, `vexp2`, `vexp2m1`, `vlog2`, `vlog21p`, `vslog2`, and `vslog21p`). If it cannot vectorize, it automatically tries to call the equivalent MASS scalar functions. For automatic vectorization or scalarization, the compiler uses versions of the MASS functions contained in the XLOPT library `libxlopt.a`.

In addition to any of the preceding sets of options, when the **-qipa** option is in effect, if the compiler cannot vectorize, it tries to inline the MASS scalar functions before deciding to call them.

“Compiling and linking a program with MASS” on page 131 describes how to compile and link a program that uses the MASS libraries, and how to selectively use the MASS scalar library functions in conjunction with the regular system libraries.

Related external information

 Mathematical Acceleration Subsystem website, available at <http://www.ibm.com/software/awdtools/mass/>

Using the scalar library

The MASS scalar library `libmass.a` contains an accelerated set of frequently used math intrinsic functions that provide improved performance over the corresponding standard system library functions. The MASS scalar functions are used when explicitly linking `libmass.a`.

If you want to explicitly call the MASS scalar functions, you can take the following steps:

1. Provide the prototypes for the functions (except `anint`, `cosisin`, `dnint`, and `sincos`), by including `math.h` in your source files.
2. Provide the prototypes for `anint`, `cosisin`, `dnint`, and `sincos`, by including `mass.h` in your source files.
3. Link the MASS scalar library `libmass.a` with your application. For instructions, see “Compiling and linking a program with MASS” on page 131.

The MASS scalar functions accept double-precision parameters and return a double-precision result, or accept single-precision parameters and return a single-precision result, except `sincos` which gives 2 double-precision results. They are summarized in Table 25.

Table 25. MASS scalar functions

Double-precision function	Single-precision function	Description	Double-precision function prototype	Single-precision function prototype
<code>acos</code>	<code>acosf</code>	Returns the arccosine of x	<code>double acos (double x);</code>	<code>float acosf (float x);</code>
<code>acosh</code>	<code>acoshf</code>	Returns the hyperbolic arccosine of x	<code>double acosh (double x);</code>	<code>float acoshf (float x);</code>
	<code>anint</code>	Returns the rounded integer value of x		<code>float anint (float x);</code>
<code>asin</code>	<code>asinf</code>	Returns the arcsine of x	<code>double asin (double x);</code>	<code>float asinf (float x);</code>
<code>asinh</code>	<code>asinhf</code>	Returns the hyperbolic arcsine of x	<code>double asinh (double x);</code>	<code>float asinhf (float x);</code>
<code>atan2</code>	<code>atan2f</code>	Returns the arctangent of x/y	<code>double atan2 (double x, double y);</code>	<code>float atan2f (float x, float y);</code>
<code>atan</code>	<code>atanf</code>	Returns the arctangent of x	<code>double atan (double x);</code>	<code>float atanf (float x);</code>
<code>atanh</code>	<code>atanhf</code>	Returns the hyperbolic arctangent of x	<code>double atanh (double x);</code>	<code>float atanhf (float x);</code>
<code>cbrt</code>	<code>cbrtf</code>	Returns the cube root of x	<code>double cbrt (double x);</code>	<code>float cbrtf (float x);</code>
<code>copysign</code>	<code>copysignf</code>	Returns x with the sign of y	<code>double copysign (double x, double y);</code>	<code>float copysignf (float x);</code>
<code>cos</code>	<code>cosf</code>	Returns the cosine of x	<code>double cos (double x);</code>	<code>float cosf (float x);</code>
<code>cosh</code>	<code>coshf</code>	Returns the hyperbolic cosine of x	<code>double cosh (double x);</code>	<code>float coshf (float x);</code>


Table 25. MASS scalar functions (continued)

Double-precision function	Single-precision function	Description	Double-precision function prototype	Single-precision function prototype
cosisin		Returns a complex number with the real part the cosine of x and the imaginary part the sine of x .	double_Complex cosisin (double);	
dnint		Returns the nearest integer to x (as a double)	double dnint (double x);	
erf	erff	Returns the error function of x	double erf (double x);	float erff (float x);
erfc	erfcf	Returns the complementary error function of x	double erfc (double x);	float erfcf (float x);
exp	expf	Returns the exponential function of x	double exp (double x);	float expf (float x);
expm1	expm1f	Returns (the exponential function of x) - 1	double expm1 (double x);	float expm1f (float x);
hypot	hypotf	Returns the square root of $x^2 + y^2$	double hypot (double x , double y);	float hypotf (float x , float y);
lgamma	lgammaf	Returns the natural logarithm of the absolute value of the Gamma function of x	double lgamma (double x);	float lgammaf (float x);
log	logf	Returns the natural logarithm of x	double log (double x);	float logf (float x);
log10	log10f	Returns the base 10 logarithm of x	double log10 (double x);	float log10f (float x);
log1p	log1pf	Returns the natural logarithm of ($x + 1$)	double log1p (double x);	float log1pf (float x);
pow	powf	Returns x raised to the power y	double pow (double x , double y);	float powf (float x , float y);
rsqrt		Returns the reciprocal of the square root of x	double rsqrt (double x);	
sin	sinf	Returns the sine of x	double sin (double x);	float sinf (float x);
sincos		Sets $*s$ to the sine of x and $*c$ to the cosine of x	void sincos (double x , double* s , double* c);	
sinh	sinhf	Returns the hyperbolic sine of x	double sinh (double x);	float sinhf (float x);
sqrt		Returns the square root of x	double sqrt (double x);	
tan	tanf	Returns the tangent of x	double tan (double x);	float tanf (float x);
tanh	tanhf	Returns the hyperbolic tangent of x	double tanh (double x);	float tanhf (float x);

Notes:

- The trigonometric functions (`sin`, `cos`, `tan`) return NaN (Not-a-Number) for large arguments (where the absolute value is greater than $2^{50}\pi$).
- In some cases, the MASS functions are not as accurate as the `libm.a` library, and they might handle edge cases differently (`sqrt(Inf)`, for example).
- See the *Mathematical Acceleration Subsystem website* for accuracy comparisons with `libm.a`.

Related external information

 Mathematical Acceleration Subsystem website, available at <http://www.ibm.com/software/awdtools/mass/>

Using the vector libraries

If you want to explicitly call any of the MASS vector functions, you can do so by including `massv.h` in your source files and linking your application with the appropriate vector library. (Information about linking is provided in “Compiling and linking a program with MASS” on page 131.)

libmassv.a

The generic vector library that runs on any POWER® processor. Unless your application requires this portability, use the appropriate architecture-specific library below for maximum performance.

libmassvp3.a

Contains some functions that have been tuned for the POWER3 architecture. The remaining functions are identical to those in `libmassv.a`.

libmassvp4.a

Contains some functions that have been tuned for the POWER4 architecture. The remaining functions are identical to those in `libmassv.a`. If you are using a PPC970 machine, this library is the recommended choice.

libmassvp5.a

Contains some functions that have been tuned for the POWER5 architecture. The remaining functions are identical to those in `libmassv.a`.

libmassvp6.a

Contains some functions that have been tuned for the POWER6 architecture. The remaining functions are identical to those in `libmassv.a`.

libmassvp7.a

Contains functions that have been tuned for the POWER7 architecture.

All libraries can be used in either 32-bit or 64-bit mode.

The single-precision and double-precision floating-point functions contained in the vector libraries are summarized in Table 26 on page 123. The integer functions contained in the vector libraries are summarized in Table 27 on page 126. Note that in C and C++ applications, only call by reference is supported, even for scalar arguments.

With the exception of a few functions (described in the following paragraph), all of the floating-point functions in the vector libraries accept three parameters:

- A double-precision (for double-precision functions) or single-precision (for single-precision functions) vector output parameter
- A double-precision (for double-precision functions) or single-precision (for single-precision functions) vector input parameter
- An integer vector-length parameter.

The functions are of the form

function_name (*y,x,n*)

where *y* is the target vector, *x* is the source vector, and *n* is the vector length. The parameters *y* and *x* are assumed to be double-precision for functions with the prefix *v*, and single-precision for functions with the prefix *vs*. As an example, the following code:

```
#include <massv.h>

double x[500], y[500];
int n;
n = 500;
...
vexp (y, x, &n);
```

outputs a vector *y* of length 500 whose elements are $\exp(x[i])$, where $i=0,\dots,499$.

The functions *vdiv*, *vsincos*, *vpow*, and *vatan2* (and their single-precision versions, *vsdiv*, *vssincos*, *vspow*, and *vsatan2*) take four arguments. The functions *vdiv*, *vpow*, and *vatan2* take the arguments (*z,x,y,n*). The function *vdiv* outputs a vector *z* whose elements are $x[i]/y[i]$, where $i=0,\dots,*n-1$. The function *vpow* outputs a vector *z* whose elements are $x[i]^{y[i]}$, where $i=0,\dots,*n-1$. The function *vatan2* outputs a vector *z* whose elements are $\text{atan}(x[i]/y[i])$, where $i=0,\dots,*n-1$. The function *vsincos* takes the arguments (*y,z,x,n*), and outputs two vectors, *y* and *z*, whose elements are $\sin(x[i])$ and $\cos(x[i])$, respectively.

In *vcosisin*(*y,x,n*) and *vscosisin*(*y,x,n*), *x* is a vector of *n* elements and the function outputs a vector *y* of *n* `__Complex` elements of the form $(\cos(x[i]),\sin(x[i]))$. If `-D__nocomplex` is used (see note in Table 26), the output vector holds $y[0][i] = \cos(x[i])$ and $y[1][i] = \sin(x[i])$, where $i=0,\dots,*n-1$.

Table 26. MASS floating-point vector functions

Double-precision function	Single-precision function	Description	Double-precision function prototype	Single-precision function prototype
<i>vacos</i>	<i>vsacos</i>	Sets $y[i]$ to the arc cosine of $x[i]$, for $i=0,\dots,*n-1$	<code>void vacos (double y[], double x[], int *n);</code>	<code>void vsacos (float y[], float x[], int *n);</code>
<i>vacosh</i>	<i>vsacosh</i>	Sets $y[i]$ to the hyperbolic arc cosine of $x[i]$, for $i=0,\dots,*n-1$	<code>void vacosh (double y[], double x[], int *n);</code>	<code>void vsacosh (float y[], float x[], int *n);</code>
<i>vasin</i>	<i>vsasin</i>	Sets $y[i]$ to the arc sine of $x[i]$, for $i=0,\dots,*n-1$	<code>void vasin (double y[], double x[], int *n);</code>	<code>void vsasin (float y[], float x[], int *n);</code>
<i>vasinh</i>	<i>vsasinh</i>	Sets $y[i]$ to the hyperbolic arc sine of $x[i]$, for $i=0,\dots,*n-1$	<code>void vasinh (double y[], double x[], int *n);</code>	<code>void vsasinh (float y[], float x[], int *n);</code>
<i>vatan2</i>	<i>vsatan2</i>	Sets $z[i]$ to the arc tangent of $x[i]/y[i]$, for $i=0,\dots,*n-1$	<code>void vatan2 (double z[], double x[], double y[], int *n);</code>	<code>void vsatan2 (float z[], float x[], float y[], int *n);</code>
<i>vatanh</i>	<i>vsatanh</i>	Sets $y[i]$ to the hyperbolic arc tangent of $x[i]$, for $i=0,\dots,*n-1$	<code>void vatanh (double y[], double x[], int *n);</code>	<code>void vsatanh (float y[], float x[], int *n);</code>
<i>vcbrt</i>	<i>vscbrt</i>	Sets $y[i]$ to the cube root of $x[i]$, for $i=0,\dots,*n-1$	<code>void vcbrt (double y[], double x[], int *n);</code>	<code>void vscbrt (float y[], float x[], int *n);</code>

Table 26. MASS floating-point vector functions (continued)

Double-precision function	Single-precision function	Description	Double-precision function prototype	Single-precision function prototype
vcos	vscos	Sets $y[i]$ to the cosine of $x[i]$, for $i=0,..,*n-1$	void vcos (double y[], double x[], int *n);	void vscos (float y[], float x[], int *n);
vcosh	vscosh	Sets $y[i]$ to the hyperbolic cosine of $x[i]$, for $i=0,..,*n-1$	void vcosh (double y[], double x[], int *n);	void vscosh (float y[], float x[], int *n);
vcosisin ¹	vscosisin ¹	Sets the real part of $y[i]$ to the cosine of $x[i]$ and the imaginary part of $y[i]$ to the sine of $x[i]$, for $i=0,..,*n-1$	void vcosisin (double _Complex y[], double x[], int *n);	void vscosisin (float _Complex y[], float x[], int *n);
vdint		Sets $y[i]$ to the integer truncation of $x[i]$, for $i=0,..,*n-1$	void vdint (double y[], double x[], int *n);	
vdiv	vsdiv	Sets $z[i]$ to $x[i]/y[i]$, for $i=0,..,*n-1$	void vdiv (double z[], double x[], double y[], int *n);	void vsdiv (float z[], float x[], float y[], int *n);
vdnint		Sets $y[i]$ to the nearest integer to $x[i]$, for $i=0,..,*n-1$	void vdnint (double y[], double x[], int *n);	
verf	vserf	Sets $y[i]$ to the error function of $x[i]$, for $i=0,..,*n-1$	void verf (double y[], double x[], int *n)	void vserf (float y[], float x[], int *n)
verfc	vserfc	Sets $y[i]$ to the complimentary error function of $x[i]$, for $i=0,..,*n-1$	void verfc (double y[], double x[], int *n)	void vserfc (float y[], float x[], int *n)
vexp	vsexp	Sets $y[i]$ to the exponential function of $x[i]$, for $i=0,..,*n-1$	void vexp (double y[], double x[], int *n);	void vsexp (float y[], float x[], int *n);
vexp2	vsexp2	Sets $y[i]$ to 2 raised to the power of $x[i]$, for $i=1,..,*n-1$	void vexp2 (double y[], double x[], int *n);	void vsexp2 (float y[], float x[], int *n);
vexpm1	vsexpm1	Sets $y[i]$ to (the exponential function of $x[i]$)-1, for $i=0,..,*n-1$	void vexpm1 (double y[], double x[], int *n);	void vsexpm1 (float y[], float x[], int *n);
vexp2m1	vsexp2m1	Sets $y[i]$ to (2 raised to the power of $x[i]$) - 1, for $i=1,..,*n-1$	void vexp2m1 (double y[], double x[], int *n);	void vsexp2m1 (float y[], float x[], int *n);
vhypot	vshypot	Sets $z[i]$ to the square root of the sum of the squares of $x[i]$ and $y[i]$, for $i=0,..,*n-1$	void vhypot (double z[], double x[], double y[], int *n)	void vshypot (float z[], float x[], float y[], int *n)
vlog	vslog	Sets $y[i]$ to the natural logarithm of $x[i]$, for $i=0,..,*n-1$	void vlog (double y[], double x[], int *n);	void vslog (float y[], float x[], int *n);
vlog2	vslog2	Sets $y[i]$ to the base-2 logarithm of $x[i]$, for $i=1,..,*n-1$	void vlog2 (double y[], double x[], int *n);	void vslog2 (float y[], float x[], int *n);

Table 26. MASS floating-point vector functions (continued)

Double-precision function	Single-precision function	Description	Double-precision function prototype	Single-precision function prototype
vlog10	vslog10	Sets $y[i]$ to the base-10 logarithm of $x[i]$, for $i=0,\dots,*n-1$	<code>void vlog10 (double y[], double x[], int *n);</code>	<code>void vslog10 (float y[], float x[], int *n);</code>
vlog1p	vslog1p	Sets $y[i]$ to the natural logarithm of $(x[i]+1)$, for $i=0,\dots,*n-1$	<code>void vlog1p (double y[], double x[], int *n);</code>	<code>void vslog1p (float y[], float x[], int *n);</code>
vlog21p	vslog21p	Sets $y[i]$ to the base-2 logarithm of $(x[i]+1)$, for $i=1,\dots,*n-1$	<code>void vlog21p (double y[], double x[], int *n);</code>	<code>void vslog21p (float y[], float x[], int *n);</code>
vpow	vspow	Sets $z[i]$ to $x[i]$ raised to the power $y[i]$, for $i=0,\dots,*n-1$	<code>void vpow (double z[], double x[], double y[], int *n);</code>	<code>void vspow (float z[], float x[], float y[], int *n);</code>
vqdrft	vsqdrft	Sets $y[i]$ to the fourth root of $x[i]$, for $i=0,\dots,*n-1$	<code>void vqdrft (double y[], double x[], int *n);</code>	<code>void vsqdrft (float y[], float x[], int *n);</code>
vrcbrt	vsrbrt	Sets $y[i]$ to the reciprocal of the cube root of $x[i]$, for $i=0,\dots,*n-1$	<code>void vrcbrt (double y[], double x[], int *n);</code>	<code>void vsrbrt (float y[], float x[], int *n);</code>
vrec	vsrec	Sets $y[i]$ to the reciprocal of $x[i]$, for $i=0,\dots,*n-1$	<code>void vrec (double y[], double x[], int *n);</code>	<code>void vsrec (float y[], float x[], int *n);</code>
vrqdrft	vsrqdrft	Sets $y[i]$ to the reciprocal of the fourth root of $x[i]$, for $i=0,\dots,*n-1$	<code>void vrqdrft (double y[], double x[], int *n);</code>	<code>void vsrqdrft (float y[], float x[], int *n);</code>
vrsqrt	vsrsqrt	Sets $y[i]$ to the reciprocal of the square root of $x[i]$, for $i=0,\dots,*n-1$	<code>void vrsqrt (double y[], double x[], int *n);</code>	<code>void vsrsqrt (float y[], float x[], int *n);</code>
vsin	vssin	Sets $y[i]$ to the sine of $x[i]$, for $i=0,\dots,*n-1$	<code>void vsin (double y[], double x[], int *n);</code>	<code>void vssin (float y[], float x[], int *n);</code>
vsincos	vssincos	Sets $y[i]$ to the sine of $x[i]$ and $z[i]$ to the cosine of $x[i]$, for $i=0,\dots,*n-1$	<code>void vsincos (double y[], double z[], double x[], int *n);</code>	<code>void vssincos (float y[], float z[], float x[], int *n);</code>
vsinh	vssinh	Sets $y[i]$ to the hyperbolic sine of $x[i]$, for $i=0,\dots,*n-1$	<code>void vsinh (double y[], double x[], int *n);</code>	<code>void vssinh (float y[], float x[], int *n);</code>
vsqrt	vssqrt	Sets $y[i]$ to the square root of $x[i]$, for $i=0,\dots,*n-1$	<code>void vsqrt (double y[], double x[], int *n);</code>	<code>void vssqrt (float y[], float x[], int *n);</code>
vtan	vstan	Sets $y[i]$ to the tangent of $x[i]$, for $i=0,\dots,*n-1$	<code>void vtan (double y[], double x[], int *n);</code>	<code>void vstan (float y[], float x[], int *n);</code>
vtanh	vstanh	Sets $y[i]$ to the hyperbolic tangent of $x[i]$, for $i=0,\dots,*n-1$	<code>void vtanh (double y[], double x[], int *n);</code>	<code>void vstanh (float y[], float x[], int *n);</code>

Note:

- By default, these functions use the `__Complex` data type, which is only available for AIX 5.2 and later, and does not compile on older versions of the operating system. To get an alternate prototype for these functions, compile with `-D__nocomplex`. This defines the functions as: `void vcosisin (double y[][2], double *x, int *n);` and `void vscosisin(float y[][2], float *x, int *n);`

Integer functions are of the form *function_name* (*x*[], **n*), where *x*[] is a vector of 4-byte (for *vpopcnt4*) or 8-byte (for *vpopcnt8*) numeric objects (integral or floating-point), and **n* is the vector length.

Table 27. MASS integer vector library functions

Function	Description	Prototype
<i>vpopcnt4</i>	Returns the total number of 1 bits in the concatenation of the binary representation of <i>x</i> [<i>i</i>], for <i>i</i> =0,..., <i>*n</i> -1 , where <i>x</i> is a vector of 32-bit objects.	unsigned int <i>vpopcnt4</i> (void * <i>x</i> , int * <i>n</i>)
<i>vpopcnt8</i>	Returns the total number of 1 bits in the concatenation of the binary representation of <i>x</i> [<i>i</i>], for <i>i</i> =0,..., <i>*n</i> -1 , where <i>x</i> is a vector of 64-bit objects.	unsigned int <i>vpopcnt8</i> (void * <i>x</i> , int * <i>n</i>)

Overlap of input and output vectors

In most applications, the MASS vector functions are called with disjoint input and output vectors; that is, the two vectors do not overlap in memory. Another common usage scenario is to call them with the same vector for both input and output parameters (for example, *vsin* (*y*, *y*, &*n*)). For other kinds of overlap, be sure to observe the following restrictions, to ensure correct operation of your application:

- For calls to vector functions that take one input and one output vector (for example, *vsin* (*y*, *x*, &*n*)):

The vectors *x*[0:*n*-1] and *y*[0:*n*-1] must be either disjoint or identical, or the address of *x*[0] must be greater than the address of *y*[0]. That is, if *x* and *y* are not the same vector, the address of *y*[0] must not fall within the range of addresses spanned by *x*[0:*n*-1], or unexpected results may be obtained.
- For calls to vector functions that take two input vectors (for example, *vatan2* (*y*, *x1*, *x2*, &*n*)):

The previous restriction applies to both pairs of vectors *y*,*x1* and *y*,*x2*. That is, if *y* is not the same vector as *x1*, the address of *y*[0] must not fall within the range of addresses spanned by *x1*[0:*n*-1]; if *y* is not the same vector as *x2*, the address of *y*[0] must not fall within the range of addresses spanned by *x2*[0:*n*-1].
- For calls to vector functions that take two output vectors (for example, *vsincos* (*x*, *y1*, *y2*, &*n*)):

The above restriction applies to both pairs of vectors *y1*,*x* and *y2*,*x*. That is, if *y1* and *x* are not the same vector, the address of *y1*[0] must not fall within the range of addresses spanned by *x*[0:*n*-1]; if *y2* and *x* are not the same vector, the address of *y2*[0] must not fall within the range of addresses spanned by *x*[0:*n*-1]. Also, the vectors *y1*[0:*n*-1] and *y2*[0:*n*-1] must be disjoint.

Alignment of input and output vectors

To get the best performance from the vector library, align the input and output vectors on 8-byte boundaries.

Consistency of MASS vector functions

The accuracy of the vector functions is comparable to that of the corresponding scalar functions in *libmass.a*, though results might not be bitwise-identical.

In the interest of speed, the MASS libraries make certain trade-offs. One of these involves the consistency of certain MASS vector functions. For certain functions, it is possible that the result computed for a particular input value varies slightly (usually only in the least significant bit) depending on its position in the vector, the vector length, and nearby elements of the input vector. Also, the results produced by the different MASS libraries are not necessarily bit-wise identical.

All the functions in `libmassvp7.a` are consistent.

The following functions are consistent in all versions of the library in which they appear.

double-precision functions

`vacos`, `vacosh`, `vasin`, `vasinh`, `vatan2`, `vatanh`, `vcbrt`, `vcos`, `vcosh`, `vcosisin`, `vdint`, `vdnint`, `vexp2`, `vexpm1`, `vexp2m1`, `vlog`, `vlog2`, `vlog10`, `vlog1p`, `vlog21p`, `vpow`, `vqdrct`, `vrcbrt`, `vrqdrct`, `vsin`, `vsincos`, `vsinh`, `vatan`, `vatanh`

single-precision functions

`vsacos`, `vsacosh`, `vsasin`, `vsasinh`, `vsatan2`, `vsatanh`, `vsbrt`, `vsacos`, `vsacosh`, `vsosisin`, `vsexp`, `vsexp2`, `vsexp1`, `vsexp2m1`, `vslog`, `vslog2`, `vslog10`, `vslog1p`, `vslog21p`, `vspow`, `vsqdrct`, `vsrbrt`, `vsrqdrct`, `vssin`, `vssincos`, `vssinh`, `vssqrt`, `vstan`, `vstanh`

The following functions are consistent in `libmassvp3.a`, `libmassvp4.a`, `libmassvp5.a`, and `libmassvp6.a`:

`vsqrt` and `vrsqrt`.

The following functions are consistent in `libmassvp4.a`, `libmassvp5.a`, and `libmassvp6.a`:

`vrec`, `vsrec`, `vdiv`, `vsdiv`, and `vexp`.

The following function is consistent in `libmassv.a`, `libmassvp5.a`, and `libmassvp6.a`:


`vsrsqrt`.


Older, inconsistent versions of some of these functions are available on the *Mathematical Acceleration Subsystem for AIX website*. If consistency is not required, there may be a performance advantage to using the older versions. For more information on consistency and avoiding inconsistency with the vector libraries, as well as performance and accuracy data, see the *Mathematical Acceleration Subsystem website*.

Related information in the XL C/C++ Compiler Reference



Related external information

 [Mathematical Acceleration Subsystem for AIX website, available at `http://www.ibm.com/software/awdtools/mass/aix`](http://www.ibm.com/software/awdtools/mass/aix)

 [Mathematical Acceleration Subsystem website, available at `http://www.ibm.com/software/awdtools/mass/`](http://www.ibm.com/software/awdtools/mass/)

Using the SIMD library for POWER7

The MASS SIMD library `libmass_simdp7.a` contains a set of frequently used math intrinsic functions that provide improved performance over the corresponding standard system library functions. If you want to use the MASS SIMD functions, you can do so as follows:

1. Provide the prototypes for the functions by including `mass_simdp7.h` in your source files.
2. Link the MASS SIMD library `libmass_simdp7.a` with your application. For instructions, see “Compiling and linking a program with MASS” on page 131.

The single/double-precision MASS SIMD functions accept single/double-precision arguments and return single/double-precision results. They are summarized in Table 28.

Table 28. MASS SIMD functions

Double-precision function	Single-precision function	Description	Double-precision function prototype	Single-precision function prototype
<code>acosd2</code>	<code>acosf4</code>	Computes the arc cosine of each element of <code>vx</code> .	<code>vector double acosd2 (vector double vx);</code>	<code>vector float acosf4 (vector float vx);</code>
<code>acoshd2</code>	<code>acoshf4</code>	Computes the arc hyperbolic cosine of each element of <code>vx</code> .	<code>vector double acoshd2 (vector double vx);</code>	<code>vector float acoshf4 (vector float vx);</code>
<code>asind2</code>	<code>asinf4</code>	Computes the arc sine of each element of <code>vx</code> .	<code>vector double asind2 (vector double vx);</code>	<code>vector float asinf4 (vector float vx);</code>
<code>asinhd2</code>	<code>asinhf4</code>	Computes the arc hyperbolic sine of each element of <code>vx</code> .	<code>vector double asinhd2 (vector double vx);</code>	<code>vector float asinhf4 (vector float vx);</code>
<code>atand2</code>	<code>atanf4</code>	Computes the arc tangent of each element of <code>vx</code> .	<code>vector double atand2 (vector double vx);</code>	<code>vector float atanf4 (vector float vx);</code>
<code>atan2d2</code>	<code>atan2f4</code>	Computes the arc tangent of each element of <code>vy/vx</code> .	<code>vector double atan2d2 (vector double vx, vector double vy);</code>	<code>vector float atan2f4 (vector float vx, vector float vy);</code>
<code>atanhd2</code>	<code>atanhf4</code>	Computes the arc hyperbolic tangent of each element of <code>vx</code> .	<code>vector double atanhd2 (vector double vx);</code>	<code>vector float atanhf4 (vector float vx);</code>
<code>cbrtd2</code>	<code>cbrtf4</code>	Computes the cube root of each element of <code>vx</code> .	<code>vector double cbrtd2 (vector double vx);</code>	<code>vector float cbrtf4 (vector float vx);</code>
<code>cosd2</code>	<code>cosf4</code>	Computes the cosine of each element of <code>vx</code> .	<code>vector double cosd2 (vector double vx);</code>	<code>vector float cosf4 (vector float vx);</code>
<code>coshd2</code>	<code>coshf4</code>	Computes the hyperbolic cosine of each element of <code>vx</code> .	<code>vector double coshd2 (vector double vx);</code>	<code>vector float coshf4 (vector float vx);</code>

Table 28. MASS SIMD functions (continued)

Double-precision function	Single-precision function	Description	Double-precision function prototype	Single-precision function prototype
cosisind2	cosisinf4	<p>Computes the cosine and sine of each element of x, and stores the results in y and z as follows:</p> <p><code>cosisind2 (x,y,z)</code> sets y and z to $\{\cos(x_1), \sin(x_1)\}$ and $\{\cos(x_2), \sin(x_2)\}$ where $x=\{x_1,x_2\}$.</p> <p><code>cosisinf4 (x,y,z)</code> sets y and z to $\{\cos(x_1), \sin(x_1), \cos(x_2), \sin(x_2)\}$ and $\{\cos(x_3), \sin(x_3), \cos(x_4), \sin(x_4)\}$ where $x=\{x_1,x_2,x_3,x_4\}$.</p>	void cosisind2 (vector double x , vector double $*y$, vector double $*z$)	void cosisinf4 (vector float x , vector float $*y$, vector float $*z$)
divd2	divf4	Computes the quotient vx/vy .	vector double divd2 (vector double vx , vector double vy);	vector float divf4 (vector float vx , vector float vy);
erfcd2	erfcf4	Computes the complementary error function of each element of vx .	vector double erfcd2 (vector double vx);	vector float erfcf4 (vector float vx);
erfd2	erff4	Computes the error function of each element of vx .	vector double erfd2 (vector double vx);	vector float erff4 (vector float vx);
expd2	expf4	Computes the exponential function of each element of vx .	vector double expd2 (vector double vx);	vector float expf4 (vector float vx);
exp2d2	exp2f4	Computes 2 raised to the power of each element of vx .	vector double exp2d2 (vector double vx);	vector float exp2f4 (vector float vx);
expm1d2	expm1f4	Computes (the exponential function of each element of vx) - 1.	vector double expm1d2 (vector double vx);	vector float expm1f4 (vector float vx);
exp2m1d2	exp2m1f4	Computes (2 raised to the power of each element of vx) -1.	vector double exp2m1d2 (vector double vx);	vector float exp2m1f4 (vector float vx);
hypotd2	hypotf4	For each element of vx and the corresponding element of vy , computes $\sqrt{x*x+y*y}$.	vector double hypotd2 (vector double vx , vector double vy);	vector float hypotf4 (vector float vx , vector float vy);
lgammad2	lgammaf4	Computes the natural logarithm of the absolute value of the Gamma function of each element of vx .	vector double lgammad2 (vector double vx);	vector float lgammaf4 (vector float vx);

Table 28. MASS SIMD functions (continued)

Double-precision function	Single-precision function	Description	Double-precision function prototype	Single-precision function prototype
logd2	logf4	Computes the natural logarithm of each element of vx.	vector double logd2 (vector double vx);	vector float logf4 (vector float vx);
log2d2	log2f4	Computes the base-2 logarithm of each element of vx.	vector double log2d2 (vector double vx);	vector float log2f4 (vector float vx);
log10d2	log10f4	Computes the base-10 logarithm of each element of vx.	vector double log10d2 (vector double vx);	vector float log10f4 (vector float vx);
log1pd2	log1pf4	Computes the natural logarithm of each element of (vx + 1).	vector double log1pd2 (vector double vx);	vector float log1pf4 (vector float vx);
log21pd2	log21pf4	Computes the base-2 logarithm of each element of (vx + 1).	vector double log21pd2 (vector double vx);	vector float log21pf4 (vector float vx);
powd2	powf4	Computes each element of vx raised to the power of the corresponding element of vy.	vector double powd2 (vector double vx, vector double vy);	vector float powf4 (vector float vx, vector float vy);
qdrtd2	qdrtf4	Computes the quad root of each element of vx.	vector double qdrtd2 (vector double vx);	vector float qdrtf4 (vector float vx);
rcbrtd2	rcbrtf4	Computes the reciprocal of the cube root of each element of vx.	vector double rcbrtd2 (vector double vx);	vector float rcbrtf4 (vector float vx);
recipd2	recipf4	Computes the reciprocal of each element of vx.	vector double recipd2 (vector double vx);	vector float recipf4 (vector float vx);
rqrtd2	rqrdf4	Computes the reciprocal of the quad root of each element of vx.	vector double rqrtd2 (vector double vx);	vector float rqrdf4 (vector float vx);
rsqrtd2	rsqrdf4	Computes the reciprocal of the square root of each element of vx.	vector double rsqrtd2 (vector double vx);	vector float rsqrdf4 (vector float vx);
sincosd2	sincosf4	Computes the sine and cosine of each element of vx.	void sincosd2 (vector double vx, vector double *vs, vector double *vc);	void sincosf4 (vector float vx, vector float *vs, vector float *vc);
sind2	sinf4	Computes the sine of each element of vx.	vector double sind2 (vector double vx);	vector float sinf4 (vector float vx);
sinhd2	sinhf4	Computes the hyperbolic sine of each element of vx.	vector double sinhd2 (vector double vx);	vector float sinhf4 (vector float vx);
sqrtd2	sqrdf4	Computes the square root of each element of vx.	vector double sqrtd2 (vector double vx);	vector float sqrdf4 (vector float vx);

Table 28. MASS SIMD functions (continued)

Double-precision function	Single-precision function	Description	Double-precision function prototype	Single-precision function prototype
tand2	tanf4	Computes the tangent of each element of vx.	vector double tand2 (vector double vx);	vector float tanf4 (vector float vx);
tanh2	tanhf4	Computes the hyperbolic tangent of each element of vx.	vector double tanhd2 (vector double vx);	vector float tanhf4 (vector float vx);

Compiling and linking a program with MASS

To compile an application that calls the functions in the scalar, SIMD, or vector MASS libraries, specify **mass**, **mass_simdp7**, and/or one of **massv**, **massvp3**, **massvp4**, **massvp5**, **massvp6**, **massvp7** on the **-l** linker option respectively.

For example, if the MASS libraries are installed in the default directory, you can specify one of the following:

Link with scalar library **libmass.a** and vector library **libmassvp7.a**

```
xlc -qarch=pwr7 prog.c -o prog -lmass -lmassvp7
```

Link with SIMD library **libmass_simdp7.a**

```
xlc -qarch=pwr7 prog.c -o prog -lmass_simdp7
```

Using **libmass.a** with the math system library

If you want to use the **libmass.a** scalar library for some functions and the normal math library **libm.a** for other functions, follow this procedure to compile and link your program:

1. Create an export list (this can be a flat text file) containing the names of the desired functions. For example, to select only the fast tangent function from **libmass.a** for use with the C program **sample.c**, create a file called **fasttan.exp** with the following line:

```
tan
```

2. Create a shared object from the export list with the **ld** command, linking with the **libmass.a** library. For example:

```
ld -bexport:fasttan.exp -o fasttan.o -bnoentry -lmass -bmodtype:SRE
```

3. Archive the shared object into a library with the **ar** command. For example:

```
ar -q libfasttan.a fasttan.o
```

4. Create the final executable using **xlc**, specifying the object file containing the MASS functions *before* the standard math library, **libm.a**. This links only the functions specified in the object file (in this example, the **tan** function) and the remainder of the math functions from the standard math library. For example:

```
xlc sample.c -o sample -Ldir_containing_libfasttan -lfasttan -lm
```

Notes:

- The MASS **sincos** function is automatically linked if you export MASS **cosisin**;
- The MASS **cos** function is automatically linked if you export MASS **sin**;
- The MASS **atan2** is automatically linked if you export MASS **atan**.

Related external information

- **ar** and **ld** in the *AIX Commands Reference, Volumes 1 - 6*

Using the Basic Linear Algebra Subprograms – BLAS

Four Basic Linear Algebra Subprograms (BLAS) functions are shipped with the XL C/C++ compiler in the `libxlopt` library. The functions consist of the following:

- `sgemv` (single-precision) and `dgemv` (double-precision), which compute the matrix-vector product for a general matrix or its transpose
- `sgermm` (single-precision) and `dgermm` (double-precision), which perform combined matrix multiplication and addition for general matrices or their transposes

Because the BLAS routines are written in Fortran, all parameters are passed to them by reference, and all arrays are stored in column-major order.

Note: Some error-handling code has been removed from the BLAS functions in `libxlopt`, and no error messages are emitted for calls to these functions.

“BLAS function syntax” describes the prototypes and parameters for the XL C/C++ BLAS functions. The interfaces for these functions are similar to those of the equivalent BLAS functions shipped in IBM's Engineering and Scientific Subroutine Library (ESSL); for more information and examples of usage of these functions, see *Engineering and Scientific Subroutine Library Guide and Reference*, available at the Engineering and Scientific Subroutine Library (ESSL) and Parallel ESSL web page.

“Linking the `libxlopt` library” on page 134 describes how to link to the XL C/C++ `libxlopt` library if you are also using a third-party BLAS library.

BLAS function syntax

The prototypes for the `sgemv` and `dgemv` functions are as follows:

```
void sgemv(const char *trans, int *m, int *n, float *alpha,
          void *a, int *lda, void *x, int *incx,
          float *beta, void *y, int *incy);
void dgemv(const char *trans, int *m, int *n, double *alpha,
          void *a, int *lda, void *x, int *incx,
          double *beta, void *y, int *incy);
```

The parameters are as follows:

trans

is a single character indicating the form of the input matrix *a*, where:

- 'N' or 'n' indicates that *a* is to be used in the computation
- 'T' or 't' indicates that the transpose of *a* is to be used in the computation

m represents:

- the number of rows in input matrix *a*
- the length of vector *y*, if 'N' or 'n' is used for the *trans* parameter
- the length of vector *x*, if 'T' or 't' is used for the *trans* parameter

The number of rows must be greater than or equal to zero, and less than the leading dimension of the matrix *a* (specified in *lda*)

n represents:

- the number of columns in input matrix *a*
- the length of vector *x*, if 'N' or 'n' is used for the *trans* parameter
- the length of vector *y*, if 'T' or 't' is used for the *trans* parameter

The number of columns must be greater than or equal to zero.

alpha

is the scaling constant for matrix *a*

a is the input matrix of float (for sgemv) or double (for dgemv) values

lda

is the leading dimension of the array specified by *a*. The leading dimension must be greater than zero. The leading dimension must be greater than or equal to 1 and greater than or equal to the value specified in *m*.

x is the input vector of float (for sgemv) or double (for dgemv) values.

incx

is the stride for vector *x*. It can have any value.

beta

is the scaling constant for vector *y*

y is the output vector of float (for sgemv) or double (for dgemv) values.

incy

is the stride for vector *y*. It must not be zero.

Note: Vector *y* must have no common elements with matrix *a* or vector *x*; otherwise, the results are unpredictable.

The prototypes for the sgemm and dgemm functions are as follows:

```
void sgemm(const char *transa, const char *transb,
           int *l, int *n, int *m, float *alpha,
           const void *a, int *lda, void *b, int *ldb,
           float *beta, void *c, int *ldc);
void dgemm(const char *transa, const char *transb,
           int *l, int *n, int *m, double *alpha,
           const void *a, int *lda, void *b, int *ldb,
           double *beta, void *c, int *ldc);
```

The parameters are as follows:

transa

is a single character indicating the form of the input matrix *a*, where:

- 'N' or 'n' indicates that *a* is to be used in the computation
- 'T' or 't' indicates that the transpose of *a* is to be used in the computation

transb

is a single character indicating the form of the input matrix *b*, where:

- 'N' or 'n' indicates that *b* is to be used in the computation
- 'T' or 't' indicates that the transpose of *b* is to be used in the computation

l represents the number of rows in output matrix *c*. The number of rows must be greater than or equal to zero, and less than the leading dimension of *c*.

n represents the number of columns in output matrix *c*. The number of columns must be greater than or equal to zero.

m represents:

- the number of columns in matrix *a*, if 'N' or 'n' is used for the *transa* parameter
- the number of rows in matrix *a*, if 'T' or 't' is used for the *transa* parameter

and:

- the number of rows in matrix b , if 'N' or 'n' is used for the *transb* parameter
- the number of columns in matrix b , if 'T' or 't' is used for the *transb* parameter

m must be greater than or equal to zero.

alpha

is the scaling constant for matrix a

a is the input matrix a of float (for sgemm) or double (for dgemm) values

lda

is the leading dimension of the array specified by a . The leading dimension must be greater than zero. If *transa* is specified as 'N' or 'n', the leading dimension must be greater than or equal to 1. If *transa* is specified as 'T' or 't', the leading dimension must be greater than or equal to the value specified in m .

b is the input matrix b of float (for sgemm) or double (for dgemm) values.

ldb

is the leading dimension of the array specified by b . The leading dimension must be greater than zero. If *transb* is specified as 'N' or 'n', the leading dimension must be greater than or equal to the value specified in m . If *transa* is specified as 'T' or 't', the leading dimension must be greater than or equal to the value specified in n .

beta

is the scaling constant for matrix c

c is the output matrix c of float (for sgemm) or double (for dgemm) values.

ldc

is the leading dimension of the array specified by c . The leading dimension must be greater than zero. If *transb* is specified as 'N' or 'n', the leading dimension must be greater than or equal to 0 and greater than or equal to the value specified in l .

Note: Matrix c must have no common elements with matrices a or b ; otherwise, the results are unpredictable.

Linking the libxlopt library


By default, the libxlopt library is linked with any application you compile with the XL C/C++ compiler. However, if you are using a third-party BLAS library, but want to use the BLAS routines shipped with libxlopt, you must specify the libxlopt library before any other BLAS library on the command line at link time. For example, if your other BLAS library is called libblas.a, you would compile your code with the following command:

```
xlc app.c -lxlopt -lblas
```

The compiler will call the sgemv, dgemv, sgemm, and dgemm functions from the libxlopt library, and all other BLAS functions in the libblas.a library.

Chapter 15. Parallelizing your programs

The compiler offers you the following methods of implementing shared memory program parallelization:

- Automatic parallelization of countable program loops, which are defined in “Countable loops” on page 136. An overview of the compiler's automatic parallelization capabilities is provided in “Enabling automatic parallelization” on page 137.
-  Explicit parallelization of countable loops using IBM SMP directives. An overview of the IBM SMP directives is provided in “Using IBM SMP directives (C only)” on page 137.
- Explicit parallelization of C and C++ program code using pragma directives compliant to the OpenMP Application Program Interface specification. An overview of the OpenMP directives is provided in “Using OpenMP directives” on page 138.

All methods of program parallelization are enabled when the **-qsmp** compiler option is in effect without the **omp** suboption. You can enable strict OpenMP compliance with the **-qsmp=omp** compiler option, but doing so will disable automatic parallelization.

Note: The **-qsmp** option must only be used together with thread-safe compiler invocation modes (those that contain the **_r** suffix).

Parallel regions of program code are executed by multiple threads, possibly running on multiple processors. The number of threads created is determined by environment variables and calls to library functions. Work is distributed among available threads according to scheduling algorithms specified by the environment variables. For any of the methods of parallelization, you can use the **XLSMPOPTS** environment variable and its suboptions to control thread scheduling; for more information on this environment variable, see *XLSMPOPTS* in the *XL C/C++ Compiler Reference*. If you are using OpenMP constructs, you can use the OpenMP environment variables to control thread scheduling; for information on OpenMP environment variables, see *OpenMP environment variables for parallel processing* in the *XL C/C++ Compiler Reference*. For more information on both IBM SMP and OpenMP built-in functions, see *Built-in functions for parallel processing* in the *XL C/C++ Compiler Reference*.

For a complete discussion on how threads are created and utilized, refer to the *OpenMP Application Program Interface Specification*, available at <http://www.openmp.org>.

Related information:

“Using shared-memory parallelism (SMP)” on page 82

Related information in the *XL C/C++ Compiler Reference*

 **XLSMPOPTS**

 **OpenMP environment variables for parallel processing**

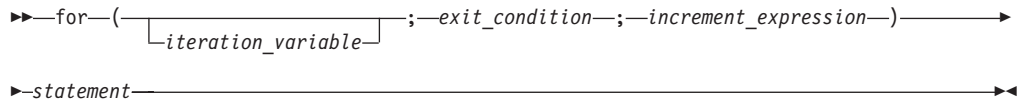
 **Built-in functions for parallel processing**

Related external information

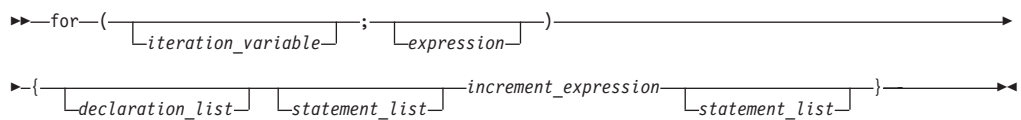
Countable loops

Loops are considered to be countable if they take any of the following forms:

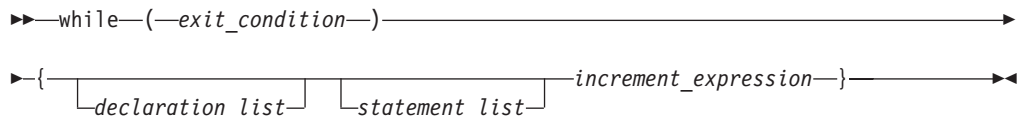
Countable for loop syntax with single statement



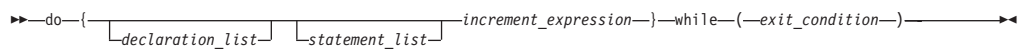
Countable for loop syntax with statement block



Countable while loop syntax



Countable do while loop syntax



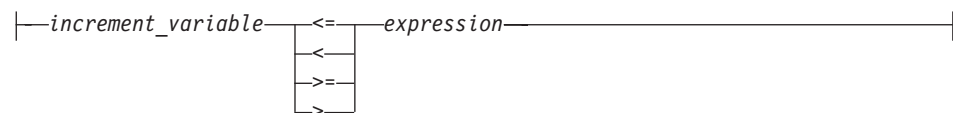
The following definitions apply to these syntax diagrams:

iteration_variable

is a signed integer that has either automatic or register storage class, does not have its address taken, and is not modified anywhere in the loop except in the *increment_expression*.

exit_condition

takes the following form:



where *expression* is a loop-invariant signed integer expression. *expression* cannot reference external or static variables, pointers or pointer expressions, function calls, or variables that have their address taken.

increment_expression

takes any of the following forms:

- ++*iteration_variable*
- --*iteration_variable*

- *iteration_variable*++
- *iteration_variable*--
- *iteration_variable* += *increment*
- *iteration_variable* -= *increment*
- *iteration_variable* = *iteration_variable* + *increment*
- *iteration_variable* = *increment* + *iteration_variable*
- *iteration_variable* = *iteration_variable* - *increment*

where *increment* is a loop-invariant signed integer expression. The value of the expression is known at run time and is not 0. *increment* cannot reference external or static variables, pointers or pointer expressions, function calls, or variables that have their address taken.

Enabling automatic parallelization

The compiler can automatically locate and parallelize all countable loops where possible in your program code. A loop is considered to be countable if it has any of the forms shown in “Countable loops” on page 136, and:

- There is no branching into or out of the loop.
- The *increment_expression* is not within a critical section.

In general, a countable loop is automatically parallelized only if all of the following conditions are met:

- The order in which loop iterations start or end does not affect the results of the program.
- The loop does not contain I/O operations.
- Floating point reductions inside the loop are not affected by round-off error, unless the **-qnostrict** option is in effect.
- The **-qnostrict_induction** compiler option is in effect.
- The **-qsmp=auto** compiler option is in effect.
- The compiler is invoked with a thread-safe compiler mode.

Using IBM SMP directives (C only)

Note: The pragma directives **#pragma ibm critical**, **#pragma ibm parallel_loop**, and **#pragma ibm schedule** have been deprecated and might be removed in a future release. You can use the corresponding OpenMP directives or clauses to obtain the same behavior.

For detailed information about how to replace the deprecated pragma directives with corresponding OpenMP directives, refer to “Deprecated directives” in the *XL C/C++ Compiler Reference*.

IBM SMP directives exploit shared memory parallelism through the parallelization of countable loops. A loop is considered to be countable if it has any of the forms described in “Countable loops” on page 136. The XL C compiler provides pragma directives that you can use to improve on automatic parallelization performed by the compiler. Pragmas fall into two general categories:

1. Pragmas that give you explicit control over parallelization. Use these pragmas to force or suppress parallelization of a loop (**#pragma ibm parallel_loop** and

#pragma ibm sequential_loop), apply specific parallelization algorithms to a loop (**#pragma ibm schedule**), and synchronize access to shared variables using critical sections (**#pragma ibm critical**).

2. Pragmas that let you give the compiler information on the characteristics of a specific countable loop (**#pragma ibm independent_calls**, **#pragma ibm independent_loop**, **#pragma ibm iterations**, **#pragma ibm permutation**). The compiler uses this information to perform more efficient automatic parallelization of the loop.

IBM SMP directive syntax

►—#pragma ibm—*pragma_name_and_args*—*countable_loop*—◄

Pragma directives must appear immediately before the countable loop to which they apply. More than one parallel processing pragma directive can be applied to a countable loop. For example:

```
#pragma ibm independent_loop
#pragma ibm independent_calls
#pragma ibm schedule(static,5)
countable_loop
```

Some pragma directives are mutually exclusive of each other, such as, for example, the **parallel_loop** and **sequential_loop** directives. If mutually exclusive pragmas are specified for the same loop, the pragma last specified applies to the loop.

Other pragmas, if specified repeatedly for a given loop, have an additive effect. For example:

```
#pragma ibm permutation (a,b)
#pragma ibm permutation (c)
```

is equivalent to:

```
#pragma ibm permutation
(a,b,c)
```

For a pragma-by-pragma description of the IBM SMP directives, refer to *Pragma directives for parallel processing* in the *XL C/C++ Compiler Reference*.

Related information in the XL C/C++ Compiler Reference



Pragma directives for parallel processing

Using OpenMP directives

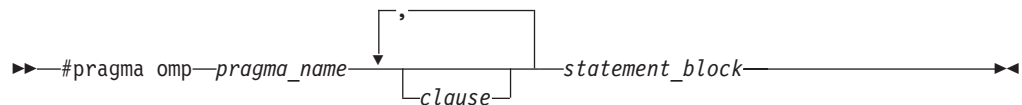
OpenMP directives exploit shared memory parallelism by defining various types of parallel regions. Parallel regions can include both iterative and non-iterative segments of program code.

Pragmas fall into these general categories:

1. Pragmas that let you define parallel regions in which work is done by threads in parallel (**#pragma omp parallel**). Most of the OpenMP directives either statically or dynamically bind to an enclosing parallel region.
2. Pragmas that let you define how work is distributed or shared across the threads in a parallel region (**#pragma omp section**, **#pragma omp for**, **#pragma omp single**, **#pragma omp task**).

3. Pragma that let you control synchronization among threads (`#pragma omp atomic`, `#pragma omp master`, `#pragma omp barrier`, `#pragma omp critical`, `#pragma omp flush`, `#pragma omp ordered`).
4. Pragma that let you define the scope of data visibility across threads (`#pragma omp threadprivate`).
5. Pragma for task synchronization (`#pragma omp taskwait`, `#pragma omp barrier`)

OpenMP directive syntax



Pragmas can be controlled by clauses. For example, a `num_threads` clause can be used to control a parallel region pragma.

Pragma directives generally appear immediately before the section of code to which they apply. For example, the following example defines a parallel region in which iterations of a for loop can run in parallel:




```
#pragma omp parallel
{
  #pragma omp for
  for (i=0; i<n; i++)
    ...
}
```

This example defines a parallel region in which two or more non-iterative sections of program code can run in parallel:

```
#pragma omp parallel
{
  #pragma omp sections
  {
    #pragma omp section
    structured_block_1
    ...
    #pragma omp section
    structured_block_2
    ...
  }
}
```

For a pragma-by-pragma description of the OpenMP directives, refer to *Pragma directives for parallel processing* in the *XL C/C++ Compiler Reference*.

Related information in the XL C/C++ Compiler Reference

-  Pragma directives for parallel processing
-  OpenMP built-in functions
-  OpenMP environment variables for parallel processing

Shared and private variables in a parallel environment

Variables can have either shared or private context in a parallel environment. Variables in shared context are visible to all threads running in associated parallel loops. Variables in private context are hidden from other threads. Each thread has its own private copy of the variable, and modifications made by a thread to its copy are not visible to other threads.

The default context of a variable is determined by the following rules:

- Variables with `static` storage duration are shared.
- Dynamically allocated objects are shared.
- Variables with automatic storage duration are private.
- Variables in heap allocated memory are shared. There can be only one shared heap.
- All variables defined outside a parallel construct become shared when the parallel loop is encountered.
- Loop iteration variables are private within their loops. The value of the iteration variable after the loop is the same as if the loop were run sequentially.
- Memory allocated within a parallel loop by the `alloca` function persists only for the duration of one iteration of that loop, and is private for each thread.

The following code segments show examples of these default rules:

```
int E1;                                /* shared static */

void main (argc,...) {                 /* argc is shared */
    int i;                             /* shared automatic */

    void *p = malloc(...);            /* memory allocated by malloc */
                                        /* is accessible by all threads */
                                        /* and cannot be privatized */

#pragma omp parallel firstprivate (p)
{
    int b;                             /* private automatic */
    static int s;                       /* shared static */

    #pragma omp for
    for (i =0;...) {
        b = 1;                          /* b is still private here ! */
        foo (i);                        /* i is private here because it */
                                        /* is an iteration variable */
    }

#pragma omp parallel
{
    b = 1;                               /* b is shared here because it */
                                        /* is another parallel region */
}
}

int E2;                                /*shared static */

void foo (int x) {                    /* x is private for the parallel */
                                        /* region it was called from */

    int c;                              /* the same */
    ... }
}
```

Some OpenMP clauses let you specify visibility context for selected data variables. A brief summary of data scope attribute clauses are listed below:

Data scope attribute clause	Description
private	The private clause declares the variables in the list to be private to each thread in a team.
firstprivate	The firstprivate clause provides a superset of the functionality provided by the private clause.
lastprivate	The lastprivate clause provides a superset of the functionality provided by the private clause.
shared	The shared clause shares variables that appear in the list among all the threads in a team. All threads within a team access the same storage area for shared variables.
reduction	The reduction clause performs a reduction on the scalar variables that appear in the list, with a specified operator.
default	The default clause allows the user to affect the data scope attributes of variables.

For more information, see the OpenMP directive descriptions in "Pragma directives for parallel processing" in the *XL C/C++ Compiler Reference*. You can also refer to the *OpenMP Application Program Interface Language Specification*, which is available at <http://www.openmp.org>.

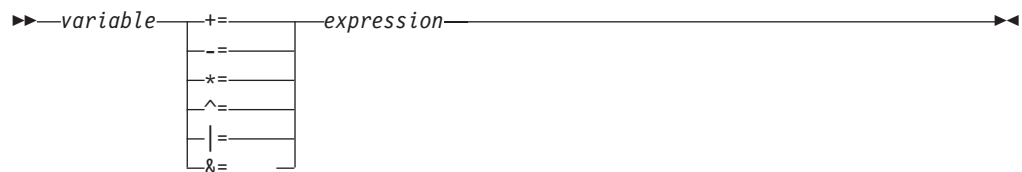
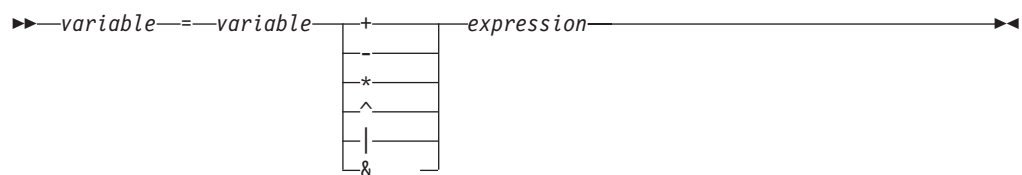
Related information in the *XL C/C++ Compiler Reference*



Pragma directives for parallel processing

Reduction operations in parallelized loops

The compiler can recognize and properly handle most reduction operations in a loop during both automatic and explicit parallelization. In particular, it can handle reduction statements that have either of the following forms:



where:

variable

is an identifier designating an automatic or register variable that does not have

its address taken and is not referenced anywhere else in the loop, including all loops that are nested. For example, in the following code, only *S* in the nested loop is recognized as a reduction:

```
int i,j, S=0;
for (i= 0 ;i < N; i++) {
    S = S+ i;
    for (j=0;j< M; j++) {
        S = S + j;
    }
}
```

expression

is any valid expression.

Recognized reductions are listed by the **-qinfo=reduction** option. When using IBM directives, use critical sections to synchronize access to all reduction variables not recognized by the compiler. OpenMP directives provide you with mechanisms to specify reduction variables explicitly.

Chapter 16. Selecting the standard allocation method to suit performance (C++)

You can select the allocation method used by the standard allocator of the C++ Standard Library to suit the performance needs of the application. For instance, in a non-threaded application you might want to use a pooled allocation strategy for standard allocators. This tends to be faster than the classic allocator for allocating small objects. However, there might be some instances when the classic allocator is preferable, as in the case where an application provides its own fast versions of the operator `new()` and operator `delete()` operators. In this case, the pooled allocators might not increase the performance and will incur a greater memory overhead.

To use the pooled `std::allocator`, do one of the following steps:

- On the command line, specify the option:
`-D__IBM_ENABLE_POOLED_ALLOCATORS__=1`
- In the source code, use the following line:

```
#define __IBM_ENABLE_POOLED_ALLOCATORS__ 1
```

To use the "classic", non-pooled `std::allocator`, do one of the following steps:

- On the command line, specify the option:
`-D__IBM_ENABLE_CLASSIC_ALLOCATORS__=1`
- In the source code, use the following line:

```
#define __IBM_ENABLE_CLASSIC_ALLOCATORS__ 1
```

In all applications, threaded and non-threaded, if no macro is specified, the "classic", non-pooled allocators are used by default; they have no effect on performance. In applications that link to `libpthread.a`, the "classic", non-pooled allocators are always used, and the macro `__IBM_ENABLE_POOLED_ALLOCATORS__` has no effect. Pooled allocators are not thread-safe and are not available in threaded programs.

Note: All compilation units in a program, including those located in a static or shared library, must be compiled with the same allocation strategy macro. If a program is composed of a mix of object files where some objects have been compiled with `__IBM_ENABLE_POOLED_ALLOCATORS__`, while other objects have been compiled with `__IBM_ENABLE_CLASSIC_ALLOCATORS__` (this includes cases where neither macro was defined, and so the default "classic" setting was selected), unexpected behavior, including program crashes, can occur.

Related information in the *Standard C++ Library Reference*

 `<memory>allocator`

Chapter 17. Ensuring thread safety (C++)

If you are building multithreaded C++ applications, there are some thread-safety issues which you need to consider when using objects defined in the C++ Standard Template Library or in the stream classes.

Ensuring thread safety of template objects

The following headers in the Standard Template Library are reentrant:

- **algorithm**
- **deque**
- **functional**
- **iterator**
- **list**
- **map**
- **memory**
- **numeric**
- **queue**
- **set**
- **stack**
- **string**
- **unordered_map**
- **unordered_set**
- **utility**
- **valarray**
- **vector**

The XL C/C++ compiler supports reentrancy to the extent that you can safely read a single object from multiple threads simultaneously. This level of reentrancy is intrinsic. No locks or other globally allocated resources are used.

However, the headers are not reentrant in these cases:

- A single container object is written by multiple threads simultaneously.
- A single container object is written in one thread, while being read in one or more other threads.

If multiple threads write to a single container, or a single thread writes to a single container while other threads are reading from that container, it is your responsibility to serialize access to this container. If multiple threads read from a single container, and no processes write to the container, no serialization is necessary.

Ensuring thread safety of stream objects

All classes declared in the **iostream** standard library are reentrant, and use a single lock to ensure thread-safety while preventing deadlock from occurring. However, on multiprocessor machines, there is a chance, although rare, that livelock can occur when two different threads attempt to concurrently access a shared stream object, or when a stream object holds a lock while waiting for input (for example, from the keyboard). If you want to avoid the possibility of livelock, you can disable locking in input stream objects, output stream objects, or both, by using the following macros at compile time:

`__NOLOCK_ON_INPUT`

Disables input locking.

`__NOLOCK_ON_OUTPUT`

Disables output locking.

To use one or both of these macros, prefix the macro name with the **-D** option on the compilation command line. For example:

```
xlc_r -D__NOLOCK_ON_INPUT -D__NOLOCK_ON_OUTPUT a.C
```

However, if you disable locking on input or output objects, you must provide the appropriate locking mechanisms in your source code if stream objects are shared between threads. If you do not, the behavior is undefined, with the possibility of data corruption or application crash.

Note: If you use OpenMP directives or the **-qsmp** option to automatically parallelize code which shares input/output stream objects, in conjunction with the lock-disabling macros, you run the same risks as with code that implements Pthreads or other multithreading constructs, and you will need to synchronize the threads accordingly.

Related information in the *XL C/C++ Compiler Reference*



Chapter 18. Memory debug library functions

This appendix contains reference information about the XL C/C++ compiler memory debug library functions, which are extensions of the standard C memory management functions. The appendix is divided into two sections:

- “Memory allocation debug functions” describes the debug versions of the standard library functions for allocating heap memory.
- “String handling debug functions” on page 155 describes the debug versions of the standard library functions for manipulating strings.

To use these debug versions, you can do either of the following operations:

- In your source code, prefix any of the default or user-defined-heap memory management functions with `_debug_`.
- If you do not want to make changes to the source code, compile with the `-qheapdebug` option. This option maps all calls to memory management functions to their debug version counterparts. To prevent a call from being mapped, parenthesize the function name.

All of the examples provided in this appendix assume compilation with the `-qheapdebug` option.

Related information in the XL C/C++ Compiler Reference

 `-qheapdebug`

Memory allocation debug functions

This section describes the debug versions of standard and user-created heap memory allocation functions. All of these functions automatically make a call to `_heap_check` or `_uheap_check` to check the validity of the heap. You can then use the `_dump_allocated` or `_dump_allocated_delta` functions to print the information returned by the heap-checking functions.

Related information:

“Functions for debugging memory heaps” on page 38

`_debug_calloc` — Allocate and initialize memory

Format

```
#include <stdlib.h> /* also in <malloc.h> */
void *_debug_calloc(size_t num, size_t size, const char *file, size_t line);
```

Purpose

This is the debug version of `calloc`. Like `calloc`, it allocates memory from the default heap for an array of `num` elements, each of length `size` bytes. It then initializes all bits of each element to 0. In addition, `_debug_calloc` makes an implicit call to `_heap_check`, and stores the name of the file `file` and the line number `line` where the storage is allocated.

Return values

Returns a pointer to the reserved space. If not enough memory is available, or if `num` or `size` is 0, returns NULL.

Examples

This example reserves storage of 100 bytes. It then attempts to write to storage that was not allocated. When `_debug_calloc` is called again, `_heap_check` detects the error, generates several messages, and stops the program.

```
/* _debug_calloc.c */
#include <stdlib.h>
#include <stdio.h>
#include <string.h>

int main(void)
{
    char *ptr1, *ptr2;

    if (NULL == (ptr1 = (char*)calloc(1, 100))) {
        puts("Could not allocate memory block.");
        exit(EXIT_FAILURE);
    }
    memset(ptr1, 'a', 105);          /* overwrites storage that was not allocated */
    ptr2 = (char*)calloc(2, 20);     /* this call to calloc invokes _heap_check */
    puts("_debug_calloc did not detect that a memory block was overwritten.");
    return 0;
}
```

The output is similar to:

```
End of allocated object 0x00073890 was overwritten at 0x000738f4.
The first eight bytes of the memory block (in hex) are: 6161616161616161.
This memory block was (re)allocated at line number 9 in _debug_calloc.c.
Heap state was valid at line 9 of _debug_calloc.c.
Memory error detected at line 14 of _debug_calloc.c.
```

`_debug_free` — Free allocated memory

Format

```
#include <stdlib.h> /* also in <malloc.h> */
void _debug_free(void *ptr, const char *file, size_t line);
```

Purpose

This is the debug version of `free`. Like `free`, it frees the block of memory pointed to by `ptr`. `_debug_free` also sets each block of freed memory to `0xFB`, so you can easily locate instances where your program uses the data in freed memory. In addition, `_debug_free` makes an implicit call to the `_heap_check` function, and stores the file name `file` and the line number `line` where the memory is freed.

Because `_debug_free` always checks the type of heap from which the memory was allocated, you can use this function to free memory blocks allocated by the regular, heap-specific, or debug versions of the memory management functions. However, if the memory was not allocated by the memory management functions, or was previously freed, `_debug_free` generates an error message and the program ends.

Return values

There is no return value.

Examples

This example reserves two blocks, one of 10 bytes and the other of 20 bytes. It then frees the first block and attempts to overwrite the freed storage. When `_debug_free` is called a second time, `_heap_check` detects the error, prints out several messages, and stops the program.

```
/* _debug_free.c */
#include <stdlib.h>
#include <stdio.h>
#include <string.h>
int main(void)
{
    char *ptr1, *ptr2;
    if (NULL == (ptr1 = (char*)malloc(10)) || NULL == (ptr2 = (char*)malloc(20))) {
        puts("Could not allocate memory block.");
        exit(EXIT_FAILURE);
    }
    free(ptr1);
    memset(ptr1, 'a', 5);      /* overwrites storage that has been freed */
    free(ptr2);               /* this call to free invokes _heap_check */
    puts("_debug_free did not detect that a freed memory block was overwritten.");
    return 0;
}
```

The output is similar to:

```
Free heap was overwritten at 0x00073890.
Heap state was valid at line 12 of _debug_free.c.
Memory error detected at line 14 of _debug_free.c.
```

`_debug_heapmin` — Free unused memory in the default heap Format

```
#include <stdlib.h> /* also in <malloc.h> */
int _debug_heapmin(const char *file, size_t line);
```

Purpose

This is the debug version of `_heapmin`. Like `_heapmin`, it returns all unused memory from the default runtime heap to the operating system. In addition, `_debug_heapmin` makes an implicit call to `_heap_check`, and stores the file name *file* and the line number *line* where the memory is returned.

Return values

If successful, returns 0; otherwise, returns -1.

Examples

This example allocates 10000 bytes of storage, changes the storage size to 10 bytes, and then uses `_debug_heapmin` to return the unused memory to the operating system. The program then attempts to overwrite memory that was not allocated. When `_debug_heapmin` is called again, `_heap_check` detects the error, generates several messages, and stops the program.

```
/* _debug_heapmin.c */
#include <stdlib.h>
#include <stdio.h>

int main(void)
{
```

```

char *ptr;

/* Allocate a large object from the system */
if (NULL == (ptr = (char*)malloc(100000))) {
    puts("Could not allocate memory block.");
    exit(EXIT_FAILURE);
}
ptr = (char*)realloc(ptr, 10);
_heapmin();          /* No allocation problems to detect      */

*(ptr - 1) = 'a';    /* Overwrite memory that was not allocated      */
_heapmin();          /* This call to _heapmin invokes _heap_check    */

puts("_debug_heapmin did not detect that a non-allocated memory block"
     "was overwritten.");
return 0;
}

```

Possible output is:

```

Header information of object 0x000738b0 was overwritten at 0x000738ac.
The first eight bytes of the memory block (in hex) are: AAAAAAAAAAAAAAAA.
This memory block was (re)allocated at line number 13 in _debug_heapmin.c.
Heap state was valid at line 14 of _debug_heapmin.c.
Memory error detected at line 17 of _debug_heapmin.c.

```

_debug_malloc — Allocate memory

Format

```

#include <stdlib.h> /* also in <malloc.h> */
void *_debug_malloc(size_t size, const char *file, size_t line);

```

Purpose

This is the debug version of malloc. Like malloc, it reserves a block of storage of *size* bytes from the default heap. `_debug_malloc` also sets all the memory it allocates to 0xAA, so you can easily locate instances where your program uses the data in the memory without initializing it first. In addition, `_debug_malloc` makes an implicit call to `_heap_check`, and stores the file name *file* and the line number *line* where the storage is allocated.

Return values

Returns a pointer to the reserved space. If not enough memory is available or if *size* is 0, returns NULL.

Examples

This example allocates 100 bytes of storage. It then attempts to write to storage that was not allocated. When `_debug_malloc` is called again, `_heap_check` detects the error, generates several messages, and stops the program.

```

/* _debug_malloc.c */
#include <stdlib.h>
#include <stdio.h>

int main(void)
{
    char *ptr1, *ptr2;

    if (NULL == (ptr1 = (char*)malloc(100))) {
        puts("Could not allocate memory block.");
        exit(EXIT_FAILURE);
    }
}

```

```

    }
    *(ptr1 - 1) = 'a';          /* overwrites storage that was not allocated */
    ptr2 = (char*)malloc(10); /* this call to malloc invokes _heap_check */
    puts("_debug_malloc did not detect that a memory block was overwritten.");
    return 0;
}

```

Possible output is:

```

Header information of object 0x00073890 was overwritten at 0x0007388c.
The first eight bytes of the memory block (in hex) are: AAAAAAAAAAAAAAAA.
This memory block was (re)allocated at line number 8 in _debug_malloc.c.
Heap state was valid at line 8 of _debug_malloc.c.
Memory error detected at line 13 of _debug_malloc.c.

```

`_debug_ucalloc` — Reserve and initialize memory from a user-created heap

Format

```

#include <umalloc.h>
void *_debug_ucalloc(Heap_t heap, size_t num, size_t size, const char *file,
                    size_t line);

```

Purpose

This is the debug version of `_ucalloc`. Like `_ucalloc`, it allocates memory from the *heap* you specify for an array of *num* elements, each of length *size* bytes. It then initializes all bits of each element to 0. In addition, `_debug_ucalloc` makes an implicit call to `_uheap_check`, and stores the name of the file *file* and the line number *line* where the storage is allocated.

If the *heap* does not have enough memory for the request, `_debug_ucalloc` calls the heap-expanding function that you specify when you create the heap with `_ucreate`.

Note: Passing `_debug_ucalloc` a heap that is not valid results in undefined behavior.

Return values

Returns a pointer to the reserved space. If *size* or *num* was specified as zero, or if your heap-expanding function cannot provide enough memory, returns NULL.

Examples

This example creates a user heap and allocates memory from it with `_debug_ucalloc`. It then attempts to write to memory that was not allocated. When `_debug_free` is called, `_uheap_check` detects the error, generates several messages, and stops the program.

```

/* _debug_ucalloc.c */
#include <stdlib.h>
#include <stdio.h>
#include <umalloc.h>
#include <string.h>
int main(void)
{
    Heap_t myheap;
    char *ptr;

    /* Use default heap as user heap */
    myheap = _udefault(NULL);

```

```

    if (NULL == (ptr = (char*)_umalloc(myheap, 100, 1))) {
        puts("Cannot allocate memory from user heap.");
        exit(EXIT_FAILURE);
    }
    memset(ptr, 'x', 105); /* Overwrites storage that was not allocated */
    free(ptr);
    return 0;
}

```

The output is similar to :

```

End of allocated object 0x00073890 was overwritten at 0x000738f4.
The first eight bytes of the memory block (in hex) are: 7878787878787878.
This memory block was (re)allocated at line number 14 in _debug_umalloc.c.
Heap state was valid at line 14 of _debug_umalloc.c.
Memory error detected at line 19 of _debug_umalloc.c.

```

`_debug_uheapmin` — Free unused memory in a user-created heap

Format

```

#include <umalloc.h>
int _debug_uheapmin(Heap_t heap, const char *file, size_t line);

```

Purpose

This is the debug version of `_uheapmin`. Like `_uheapmin`, it returns all unused memory blocks from the specified *heap* to the operating system.

To return the memory, `_debug_uheapmin` calls the heap-shrinking function you supply when you create the heap with `_ucreate`. If you do not supply a heap-shrinking function, `_debug_uheapmin` has no effect and returns 0.

In addition, `_debug_uheapmin` makes an implicit call to `_uheap_check` to validate the heap.

Return values

If successful, returns 0. A nonzero return value indicates failure. If the heap specified is not valid, generates an error message with the file name and line number in which the call to `_debug_uheapmin` was made.

Examples

This example creates a heap and allocates memory from it, then uses `_debug_heapmin` to release the memory.

```

/* _debug_uheapmin.c */
#include <stdlib.h>
#include <stdio.h>
#include <string.h>
#include <umalloc.h>

int main(void)
{
    Heap_t myheap;
    char *ptr;

    /* Use default heap as user heap */
    myheap = _udefault(NULL);

    /* Allocate a large object */
    if (NULL == (ptr = (char*)_umalloc(myheap, 60000))) {
        puts("Cannot allocate memory from user heap.\n");
    }
}

```



```

        exit(EXIT_FAILURE);
    }
    memset(ptr, 'x', 60000);
    free(ptr);

    /* _debug_uheapmin will attempt to return the freed object to the system */
    if (0 != _uheapmin(myheap)) {
        puts("_debug_uheapmin returns failed.\n");
        exit(EXIT_FAILURE);
    }
    return 0;
}

```

`_debug_umalloc` — Reserve memory blocks from a user-created heap

Format

```

#include <umalloc.h>
void *_debug_umalloc(Heap_t heap, size_t size, const char *file, size_t line);

```

Purpose

This is the debug version of `_umalloc`. Like `_umalloc`, it reserves storage space from the *heap* you specify for a block of *size* bytes. `_debug_umalloc` also sets all the memory it allocates to `0xAA`, so you can easily locate instances where your program uses the data in the memory without initializing it first.

In addition, `_debug_umalloc` makes an implicit call to `_uheap_check`, and stores the name of the file *file* and the line number *line* where the storage is allocated.

If the *heap* does not have enough memory for the request, `_debug_umalloc` calls the heap-expanding function that you specify when you create the heap with `_ucreate`.

Note: Passing `_debug_umalloc` a heap that is not valid results in undefined behavior.

Return values

Returns a pointer to the reserved space. If *size* was specified as zero, or your heap-expanding function cannot provide enough memory, returns `NULL`.

Examples

This example creates a heap `myheap` and uses `_debug_umalloc` to allocate 100 bytes from it. It then attempts to overwrite storage that was not allocated. The call to `_debug_free` invokes `_uheap_check`, which detects the error, generates messages, and ends the program.

```

/* _debug_umalloc.c */
#include <stdlib.h>
#include <stdio.h>
#include <umalloc.h>
#include <string.h>

int main(void)
{
    Heap_t myheap;
    char *ptr;

    /* Use default heap as user heap */
    myheap = _udefault(NULL);
}

```

```

    if (NULL == (ptr = (char*)_umalloc(myheap, 100))) {
        puts("Cannot allocate memory from user heap.\n");
        exit(EXIT_FAILURE);
    }
    memset(ptr, 'x', 105); /* Overwrites storage that was not allocated */
    free(ptr);
    return 0;
}

```

The output is similar to :

```

End of allocated object 0x00073890 was overwritten at 0x000738f4.
The first eight bytes of the memory block (in hex) are: 7878787878787878.
This memory block was (re)allocated at line number 14 in _debug_umalloc.c.
Heap state was valid at line 14 of _debug_umalloc.c.
Memory error detected at line 19 of _debug_umalloc.c.

```

`_debug_realloc` — Reallocate memory block

Format

```

#include <stdlib.h> /* also in <malloc.h> */
void *_debug_realloc(void *ptr, size_t size, const char *file, size_t line);

```

Purpose

This is the debug version of `realloc`. Like `realloc`, it reallocates the block of memory pointed to by `ptr` to a new `size`, specified in bytes. It also sets any new memory it allocates to `0xAA`, so you can easily locate instances where your program tries to use the data in that memory without initializing it first. In addition, `_debug_realloc` makes an implicit call to `_heap_check`, and stores the file name `file` and the line number `line` where the storage is reallocated.

If `ptr` is `NULL`, `_debug_realloc` behaves like `_debug_malloc` (or `malloc`) and allocates the block of memory.

Because `_debug_realloc` always checks to determine the heap from which the memory was allocated, you can use `_debug_realloc` to reallocate memory blocks allocated by the regular or debug versions of the memory management functions. However, if the memory was not allocated by the memory management functions, or was previously freed, `_debug_realloc` generates an error message and the program ends.

Return values

Returns a pointer to the reallocated memory block. The `ptr` argument is not the same as the return value; `_debug_realloc` always changes the memory location to help you locate references to the memory that were not freed before the memory was reallocated.

If `size` is 0, returns `NULL`. If not enough memory is available to expand the block to the given size, the original block is unchanged and `NULL` is returned.

Examples

This example uses `_debug_realloc` to allocate 100 bytes of storage. It then attempts to write to storage that was not allocated. When `_debug_realloc` is called again, `_heap_check` detects the error, generates several messages, and stops the program.

```

/* _debug_realloc.c */
#include <stdlib.h>
#include <stdio.h>
#include <string.h>

int main(void)
{
    char *ptr;

    if (NULL == (ptr = (char*)realloc(NULL, 100))) {
        puts("Could not allocate memory block.");
        exit(EXIT_FAILURE);
    }
    memset(ptr, 'a', 105); /* overwrites storage that was not allocated */
    ptr = (char*)realloc(ptr, 200); /* realloc invokes _heap_check */
    puts("_debug_realloc did not detect that a memory block was overwritten." );
    return 0;
}

```

The output is similar to:

```

End of allocated object 0x00073890 was overwritten at 0x000738f4.
The first eight bytes of the memory block (in hex) are: 6161616161616161.
This memory block was (re)allocated at line number 8 in _debug_realloc.c.
Heap state was valid at line 8 of _debug_realloc.c.
Memory error detected at line 13 of _debug_realloc.c.

```

String handling debug functions

This section describes the debug versions of the string manipulation and memory functions of the standard C string handling library. Note that these functions check only the current default heap; they do not check all heaps in applications that use multiple user-created heaps.

`_debug_memcpy` — Copy bytes

Format

```

#include <string.h>
void *_debug_memcpy(void *dest, const void *src, size_t count, const char *file,
                    size_t line);

```

Purpose

This is the debug version of `memcpy`. Like `memcpy`, it copies *count* bytes of *src* to *dest*, where the behavior is undefined if copying takes place between objects that overlap.

`_debug_memcpy` validates the heap after copying the bytes to the target location, and performs this check only when the target is within a heap. `_debug_memcpy` makes an implicit call to `_heap_check`. If `_debug_memcpy` detects a corrupted heap when it makes a call to `_heap_check`, `_debug_memcpy` reports the file name *file* and line number *line* in a message.

Return values

Returns a pointer to *dest*.

Examples

This example contains a programming error. On the call to `memcpy` used to initialize the target location, the count is more than the size of the target object, and the `memcpy` operation copies bytes past the end of the allocated object.

```
/* _debug_memcpy.c */
#include <stdlib.h>
#include <string.h>
#include <stdio.h>

#define MAX_LEN 10

int main(void)
{
    char *source, *target;

    target = (char*)malloc(MAX_LEN);
    memcpy(target, "This is the target string", 11);

    printf("Target is \"%s\"\n", target);
    return 0;
}
```

The output is similar to:

```
End of allocated object 0x00073c80 was overwritten at 0x00073c8a.
The first eight bytes of the memory block (in hex) are: 5468697320697320.
This memory block was (re)allocated at line number 11 in _debug_memcpy.c.
Heap state was valid at line 11 of _debug_memcpy.c.
Memory error detected at line 12 of _debug_memcpy.c.
```

`_debug_memset` — Set bytes to value

Format

```
#include <string.h>
void *_debug_memset(void *dest, int c, size_t count, const char *file, size_t line);
```

Purpose

This is the debug version of `memset`. Like `memset`, it sets the first *count* bytes of *dest* to the value *c*. The value of *c* is converted to an unsigned character.

`_debug_memset` validates the heap after setting the bytes, and performs this check only when the target is within a heap. `_debug_memset` makes an implicit call to `_heap_check`. If `_debug_memset` detects a corrupted heap when it makes a call to `_heap_check`, `_debug_memset` reports the file name *file* and line number *line* in a message.

Return values

Returns a pointer to *dest*.

Examples

This example contains a programming error. The invocation of `memset` that puts 'B' in the buffer specifies the wrong count, and stores bytes past the end of the buffer.

```
/* _debug_memset.c */
#include <stdlib.h>
#include <string.h>
#include <stdio.h>
```

```

#define BUF_SIZE    20

int main(void)
{
    char *buffer, *buffer2;
    char *string;

    buffer = (char*)calloc(1, BUF_SIZE+1);    /* +1 for null-terminator */

    string = (char*)memset(buffer, 'A', 10);
    printf("\nBuffer contents: %s\n", string);
    memset(buffer+10, 'B', 20);

    return 0;
}

```

The output is similar to:

```

Buffer contents: AAAAAAAAAA
End of allocated object 0x00073c80 was overwritten at 0x00073c95.
The first eight bytes of the memory block (in hex) are: 4141414141414141.
This memory block was (re)allocated at line number 12 in _debug_memset.c.
Heap state was valid at line 14 of _debug_memset.c.
Memory error detected at line 16 of _debug_memset.c.

```

_debug_strcat — Concatenate strings

Format

```

#include <string.h>
char *_debug_strcat(char *string1, const char *string2, const char *file,
                    size_t file);

```

Purpose

This is the debug version of `strcat`. Like `strcat`, it concatenates `string2` to `string1` and ends the resulting string with the null character.

`_debug_strcat` validates the heap after concatenating the strings, and performs this check only when the target is within a heap. `_debug_strcat` makes an implicit call to `_heap_check`. If `_debug_strcat` detects a corrupted heap when it makes a call to `_heap_check`, `_debug_strcat` reports the file name `file` and line number `file` in a message.

Return values

Returns a pointer to the concatenated string `string1`.

Examples

This example contains a programming error. The `buffer1` object is not large enough to store the result after the string " program" is concatenated.

```

/* _debug_strcat.hc */
#include <stdlib.h>
#include <stdio.h>
#include <string.h>

#define SIZE    10

int main(void)
{

```

```

char *buffer1;
char *ptr;

buffer1 = (char*)malloc(SIZE);
strcpy(buffer1, "computer");

ptr = strcat(buffer1, " program");
printf("buffer1 = %s\n", buffer1);
return 0;
}

```

The output is similar to:

```

End of allocated object 0x00073c80 was overwritten at 0x00073c8a.
The first eight bytes of the memory block (in hex) are: 636F6D7075746572.
This memory block was (re)allocated at line number 12 in _debug_strcat.c.
Heap state was valid at line 13 of _debug_strcat.c.
Memory error detected at line 15 of _debug_strcat.c.

```

`_debug_strcpy` — Copy strings

Format

```

#include <string.h>
char *_debug_strcpy(char *string1, const char *string2, const char *file,
                   size_t line);

```

Purpose

This is the debug version of `strcpy`. Like `strcpy`, it copies *string2*, including the ending null character, to the location specified by *string1*.

`_debug_strcpy` validates the heap after copying the string to the target location, and performs this check only when the target is within a heap. `_debug_strcpy` makes an implicit call to `_heap_check`. If `_debug_strcpy` detects a corrupted heap when it makes a call to `_heap_check`, `_debug_strcpy` reports the file name *file* and line number *line* in a message.

Return values

Returns a pointer to the copied string *string1*.

Examples

This example contains a programming error. The source string is too long for the destination buffer, and the `strcpy` operation damages the heap.

```

/* _debug_strcpy.c */
#include <stdlib.h>
#include <stdio.h>
#include <string.h>

#define SIZE 10

int main(void)
{
    char *source = "1234567890123456789";
    char *destination;
    char *return_string;

    destination = (char*)malloc(SIZE);
    strcpy(destination, "abcdefg"),

```

```

printf("destination is originally = '%s'\n", destination);
return_string = strcpy(destination, source);
printf("After strcpy, destination becomes '%s'\n\n", destination);
return 0;
}

```

The output is similar to:

```

destination is originally = 'abcdefg'
End of allocated object 0x00073c80 was overwritten at 0x00073c8a.
The first eight bytes of the memory block (in hex) are: 3132333435363738.
This memory block was (re)allocated at line number 13 in _debug_strcpy.c.
Heap state was valid at line 14 of _debug_strcpy.c.
Memory error detected at line 17 of _debug_strcpy.c.

```

`_debug_strncat` — Concatenate strings

Format

```

#include <string.h>
char *_debug_strncat(char *string1, const char *string2, size_t count,
                    const char *file, size_t line);

```

Purpose

This is the debug version of `strncat`. Like `strncat`, it appends the first `count` characters of `string2` to `string1` and ends the resulting string with a null character (`\0`). If `count` is greater than the length of `string2`, the length of `string2` is used in place of `count`.

`_debug_strncat` validates the heap after appending the characters, and performs this check only when the target is within a heap. `_debug_strncat` makes an implicit call to `_heap_check`. If `_debug_strncat` detects a corrupted heap when it makes a call to `_heap_check`, `_debug_strncat` reports the file name `file` and line number `line` in a message.

Return values

Returns a pointer to the joined string `string1`.

Examples

This example contains a programming error. The `buffer1` object is not large enough to store the result after eight characters from the string " programming" are concatenated.

```

/* _debug_strncat.c */
#include <stdlib.h>
#include <stdio.h>
#include <string.h>
#define SIZE          10
int main(void)
{
    char *buffer1;
    char *ptr;

    buffer1 = (char*)malloc(SIZE);
    strcpy(buffer1, "computer");

    /* Call strncat with buffer1 and " programming" */
    ptr = strncat(buffer1, " programming", 8);
    printf("strncat: buffer1 = \"%s\"\n", buffer1);
    return 0;
}

```

The output is similar to:

```
End of allocated object 0x00073c80 was overwritten at 0x00073c8a.
The first eight bytes of the memory block (in hex) are: 636F6D7075746572.
This memory block was (re)allocated at line number 12 in _debug_strncat.c.
Heap state was valid at line 13 of _debug_strncat.c.
Memory error detected at line 17 of _debug_strncat.c.
```

`_debug_strncpy` — Copy strings

Format

```
#include <string.h>
char *_debug_strncpy(char *string1, const char *string2, size_t count,
                    const char *file, size_t line);
```

Purpose

This is the debug version of `strncpy`. Like `strncpy`, it copies *count* characters of *string2* to *string1*. If *count* is less than or equal to the length of *string2*, a null character (`\0`) is not appended to the copied string. If *count* is greater than the length of *string2*, the *string1* result is padded with null characters (`\0`) up to length *count*.

`_debug_strncpy` validates the heap after copying the strings to the target location, and performs this check only when the target is within a heap. `_debug_strncpy` makes an implicit call to `_heap_check`. If `_debug_strncpy` detects a corrupted heap when it makes a call to `_heap_check`, `_debug_strncpy` reports the file name *file* and line number *line* in a message.

Return values

Returns a pointer to *string1*.

Examples

This example contains a programming error. The source string is too long for the destination buffer, and the `strncpy` operation damages the heap.

```
/* _debug_strncpy */
#include <stdlib.h>
#include <stdio.h>
#include <string.h>
#define SIZE 10
int main(void)
{
    char *source = "1234567890123456789";
    char *destination;
    char *return_string;
    int index = 15;

    destination = (char*)malloc(SIZE);
    strcpy(destination, "abcdefg"),

    printf("destination is originally = '%s'\n", destination);
    return_string = strncpy(destination, source, index);
    printf("After strncpy, destination becomes '%s'\n\n", destination);
    return 0;
}
```

The output is similar to:

```
destination is originally = 'abcdefg'
End of allocated object 0x00073c80 was overwritten at 0x00073c8a.
The first eight bytes of the memory block (in hex) are: 3132333435363738.
```


This memory block was (re)allocated at line number 14 in `_debug_strncpy.c`.
Heap state was valid at line 15 of `_debug_strncpy.c`.
Memory error detected at line 18 of `_debug_strncpy.c`.

`_debug_strnset` — Set characters in a string

Format

```
#include <string.h>
char *_debug_strnset(char *string, int c, size_t n, const char *file, size_t line);
```

Purpose

This is the debug version of `strnset`. Like `strnset`, it sets, at most, the first n characters of `string` to c (converted to a char), where if n is greater than the length of `string`, the length of `string` is used in place of n .

`_debug_strnset` validates the heap after setting the bytes, and performs this check only when the target is within a heap. `_debug_strnset` makes an implicit call to `_heap_check`. If `_debug_strnset` detects a corrupted heap when it makes a call to `_heap_check`, `_debug_strnset` reports the file name `file` and line number `line` in a message.

Return values

Returns a pointer to the altered `string`. There is no error return value.

Examples

This example contains two programming errors. The string, `str`, was created without a null-terminator to mark the end of the string, and without the terminator `strnset` with a count of 10 stores bytes past the end of the allocated object.

```
/* _debug_strnset */
#include <stdlib.h>
#include <stdio.h>
#include <string.h>
int main(void)
{
    char *str;
    str = (char*)malloc(10);
    printf("This is the string after strnset: %s\n", str);
    return 0;
}
```

The output is similar to:

```
End of allocated object 0x00073c80 was overwritten at 0x00073c8a.
The first eight bytes of the memory block (in hex) are: 7878787878797979.
This memory block was (re)allocated at line number 9 in _debug_strnset.c.
Heap state was valid at line 11 of _debug_strnset.c.
```

`_debug_strset` — Set characters in a string

Format

```
#include <string.h>
char *_debug_strset(char *string, size_t c, const char *file, size_t line);
```

Purpose

This is the debug version of `strset`. Like `strset`, it sets all characters of *string*, except the ending null character (`\0`), to *c* (converted to a char).

`_debug_strset` validates the heap after setting all characters of *string*, and performs this check only when the target is within a heap. `_debug_strset` makes an implicit call to `_heap_check`. If `_debug_strset` detects a corrupted heap when it makes a call to `_heap_check`, `_debug_strset` reports the file name *file* and line number *line* in a message.

Return values

Returns a pointer to the altered string. There is no error return value.

Examples

This example contains a programming error. The string, *str*, was created without a null-terminator, and `strset` propagates the letter 'k' until it finds what it thinks is the null-terminator.

```
/* file: _debug_strset.c */
#include <stdlib.h>
#include <stdio.h>
#include <string.h>
int main(void)
{
    char *str;
    str = (char*)malloc(10);
    strnset(str, 'x', 5);
    strset(str+5, 'k');
    printf("This is the string after strset: %s\n", str);
    return 0;
}
```

The output is similar to:

```
End of allocated object 0x00073c80 was overwritten at 0x00073c8a.
The first eight bytes of the memory block (in hex) are: 78787878786B6B6B.
This memory block was (re)allocated at line number 9 in _debug_strset.c.
Heap state was valid at line 11 of _debug_strset.c.
Memory error detected at line 12 of _debug_strset.c.
```

Notices

This information was developed for products and services offered in the U.S.A. IBM may not offer the products, services, or features discussed in this document in other countries. Consult your local IBM representative for information on the products and services currently available in your area. Any reference to an IBM product, program, or service is not intended to state or imply that only that IBM product, program, or service may be used. Any functionally equivalent product, program, or service that does not infringe any IBM intellectual property right may be used instead. However, it is the user's responsibility to evaluate and verify the operation of any non-IBM product, program, or service.

IBM may have patents or pending patent applications covering subject matter described in this document. The furnishing of this document does not give you any license to these patents. You can send license inquiries, in writing, to:

IBM Director of Licensing
IBM Corporation
North Castle Drive
Armonk, NY 10504-1785
U.S.A.

For license inquiries regarding double-byte (DBCS) information, contact the IBM Intellectual Property Department in your country or send inquiries, in writing, to:

IBM World Trade Asia Corporation
Licensing
2-31 Roppongi 3-chome, Minato-ku
Tokyo 106, Japan

The following paragraph does not apply to the United Kingdom or any other country where such provisions are inconsistent with local law:

INTERNATIONAL BUSINESS MACHINES CORPORATION PROVIDES THIS PUBLICATION "AS IS" WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESS OR IMPLIED, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF NON-INFRINGEMENT, MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. Some states do not allow disclaimer of express or implied warranties in certain transactions, therefore, this statement may not apply to you.

This information could include technical inaccuracies or typographical errors. Changes are periodically made to the information herein; these changes will be incorporated in new editions of the publication. IBM may make improvements and/or changes in the product(s) and/or the program(s) described in this publication at any time without notice.

Any references in this information to non-IBM websites are provided for convenience only and do not in any manner serve as an endorsement of those websites. The materials at those websites are not part of the materials for this IBM product and use of those websites is at your own risk.

IBM may use or distribute any of the information you supply in any way it believes appropriate without incurring any obligation to you.

Licensees of this program who wish to have information about it for the purpose of enabling: (i) the exchange of information between independently created programs and other programs (including this one) and (ii) the mutual use of the information which has been exchanged, should contact:

Lab Director
IBM Canada Ltd. Laboratory
8200 Warden Avenue
Markham, Ontario L6G 1C7
Canada

Such information may be available, subject to appropriate terms and conditions, including in some cases, payment of a fee.

The licensed program described in this document and all licensed material available for it are provided by IBM under terms of the IBM Customer Agreement, IBM International Program License Agreement or any equivalent agreement between us.

Any performance data contained herein was determined in a controlled environment. Therefore, the results obtained in other operating environments may vary significantly. Some measurements may have been made on development-level systems and there is no guarantee that these measurements will be the same on generally available systems. Furthermore, some measurements may have been estimated through extrapolation. Actual results may vary. Users of this document should verify the applicable data for their specific environment.

Information concerning non-IBM products was obtained from the suppliers of those products, their published announcements or other publicly available sources. IBM has not tested those products and cannot confirm the accuracy of performance, compatibility or any other claims related to non-IBM products. Questions on the capabilities of non-IBM products should be addressed to the suppliers of those products.

All statements regarding IBM's future direction or intent are subject to change or withdrawal without notice, and represent goals and objectives only.

This information contains examples of data and reports used in daily business operations. To illustrate them as completely as possible, the examples include the names of individuals, companies, brands, and products. All of these names are fictitious and any similarity to the names and addresses used by an actual business enterprise is entirely coincidental.

COPYRIGHT LICENSE:

This information contains sample application programs in source language, which illustrates programming techniques on various operating platforms. You may copy, modify, and distribute these sample programs in any form without payment to IBM, for the purposes of developing, using, marketing or distributing application programs conforming to the application programming interface for the operating platform for which the sample programs are written. These examples have not been thoroughly tested under all conditions. IBM, therefore, cannot guarantee or imply reliability, serviceability, or function of these programs. You may copy, modify, and distribute these sample programs in any form without payment to IBM for the purposes of developing, using, marketing, or distributing application programs conforming to IBM's application programming interfaces.

Each copy or any portion of these sample programs or any derivative work, must include a copyright notice as follows:

© (your company name) (year). Portions of this code are derived from IBM Corp. Sample Programs. © Copyright IBM Corp. 1998, 2012. All rights reserved.

Trademarks and service marks

IBM, the IBM logo, and [ibm.com](http://www.ibm.com) are trademarks or registered trademarks of International Business Machines Corp., registered in many jurisdictions worldwide. Other product and service names might be trademarks of IBM or other companies. A current list of IBM trademarks is available on the web at “Copyright and trademark information” at <http://www.ibm.com/legal/copytrade.shtml>.

Adobe, the Adobe logo, PostScript, and the PostScript logo are either registered trademarks or trademarks of Adobe Systems Incorporated in the United States, other countries, or both.

Linux is a registered trademark of Linus Torvalds in the United States, other countries, or both.

Microsoft and Windows are trademarks of Microsoft Corporation in the United States, other countries, or both.

Cell Broadband Engine is a trademark of Sony Computer Entertainment, Inc. in the United States, other countries, or both and is used under license therefrom.

UNIX is a registered trademark of The Open Group in the United States and other countries.

Other company, product, or service names may be trademarks or service marks of others.

Index

Special characters

- `__align` specifier 15
- `-O0` 72
- `-O2` 73
- `-O3` 75
 - trade-offs 75
- `-O4` 76
 - trade-offs 76
- `-O5` 77
 - trade-offs 77
- `-q32` 1, 78
- `-q64` 1
- `-qalign` 9
- `-qarch` 78, 79
- `-qcache` 76, 78, 79
- `-qfloat` 20, 23
 - IEEE conformance 20
 - multiply-add operations 19
- `-qflttrap` 23
- `-qfunctrace` 111
- `-qheapdebug` 36, 147
- `-qhot` 80
- `-qipa` 76, 78, 83
 - IPA process 77
- `-qlistfmt` compiler option 94
- `-qlongdouble`
 - 32-bit and 64-bit precision 19
 - corresponding Fortran types 5
- `-qmkshrobj` 51
- `-qnofunctrace` 111
- `-qpdf` 86
- `-qpriority` 55
- `-qsmp` 82, 135, 137
- `-qstrict` 20, 75
- `-qtempinc` 43
- `-qtemplatercompile` 49
- `-qtemplaterregistry` 43
- `-qtune` 78, 79
- `-qwarn64` 1
- `-W` 57
- `-y` 20
- `#pragma nofunctrace` 111

Numerics

- 64-bit mode 4
 - alignment 4
 - bit-shifting 3
 - data types 1
 - Fortran 4
 - long constants 2
 - long types 2
 - optimization 110
 - pointers 3

A

- advanced optimization 74
- aggregate
 - alignment 4, 9, 11

aggregate (*continued*)

- Fortran 6
- aligned attribute 15
- alignment 4, 9
 - bit fields 14
 - modes 9
 - modifiers 15
- architecture
 - optimization 78
- arrays, Fortran 6
- attribute
 - aligned 15
 - packed 15

B

- basic example, described xi
- basic optimization 72
- bit field 14
 - alignment 14
- bit-shifting 3
- BLAS library 132

C

- C++0x
 - delegating constructors 41, 107
 - explicit instantiation declarations 43, 49, 107
 - rvalue references 115
 - target constructors 41
 - variadic templates 43
- c++filt utility 65
- cloning, function 78, 83
- constants
 - folding 20
 - long types 2
 - rounding 20

D

- data types
 - 32-bit and 64-bit modes 1
 - 64-bit mode 1
 - Fortran 4, 5
 - long 2
 - size and alignment 9
- debugging 99
 - heap memory 25, 147
 - string handling functions 155
- dynamic library 51
- dynamic loading 59
- dynamic memory allocation 25, 147

E

- environment variable
 - HD_FILL 39
 - HD_STACK 40

environment variable (*continued*)

- OBJECT_MODE 1
- errors, floating-point 23
- exceptions, floating-point 23
- export list 51

F

- floating-point
 - exceptions 23
 - folding 20
 - IEEE conformance 20
 - range and precision 19
 - rounding 20
- folding, floating-point 20
- Fortran
 - 64-bit mode 4
 - aggregates 6
 - arrays 6
 - data types 4, 5
 - function calls 7
 - function pointers 7
 - identifiers 5
- function calls
 - Fortran 7
 - optimizing 105
- function cloning 78, 83
- function pointers, Fortran 7

H

- hardware optimization 78
- HD_FILL environment variable 39
- HD_STACK environment variable 40
- heap memory 25, 147

I

- IBM SMP 141
- IBM SMP directives 137
- IEEE conformance 20
- initialization order of C++ static objects 54
- input/output
 - floating-point rounding 22
 - optimizing 105
 - thread safety 145
- instantiating templates 43
- interlanguage calls 7
- interprocedural analysis (IPA) 83

L

- libmass library 120
- libmassv library 122
- library
 - BLAS 132
 - MASS 119
 - scalar 120

library (*continued*)
 shared (dynamic) 51
 static 51
 vector 122
linear algebra functions 132
long constants, 64-bit mode 2
long data type, 64-bit mode 2
loop optimization 80, 135

M

mangled names 65
MASS libraries 119
 scalar functions 120
 vector functions 122
matrix multiplication functions 132
memory
 allocation 25, 147
 debugging 25, 147
 management 108
 user heaps 25, 147
mergepdf 86
move 115
multithreading 82, 135, 145

N

name mangling 65

O

OBJECT_MODE environment variable 1
OpenMP 82, 140, 141
OpenMP directives 138
optimization 105
 -O0 72
 -O2 73
 -O3 75
 -O4 76
 -O5 77
 64-bit mode 110
 across program units 83
 advanced 74
 architecture 78
 basic 72
 debugging 99
 hardware 78
 loop 80
 loops 135
 math functions 119
optimization and tuning
 optimizing 71
 tuning 71
optimization trade-offs
 -O3 75
 -O4 76
 -O5 77
optimization, diagnostics 94, 95
optimizing
 applications 71
option
 -qheapdebug 36, 147

P

packed attribute 15
parallelization 82, 135
 automatic 137
 IBM SMP directives 137
 OpenMP directives 138
perfect forwarding 115
performance tuning 105
pointers
 64-bit mode 3
 Fortran 7
pragma
 align 9
 ibm 137
 implementation 44
 omp 138
 pack 15
 priority 55
pragma nofunctrace 111
precision, floating-point numbers 19
priority of static objects 54
profile-directed feedback (PDF) 86
profiling 86

R

range, floating-point numbers 19
reentrancy 145
rounding, floating-point 20

S

scalar MASS library 120
shared (dynamic) library 51, 59
shared memory parallelism (SMP) 82,
 135, 137, 138, 140, 141
showpdf 86
Standard Template Library 145
static library 51
static objects, C++ 54
strings
 debug functions 155
 optimizing 109
structure alignment 11
 64-bit mode 4

T

template instantiation 43
thread safety 145
 stream objects 145
 template objects 145
tracing
 functions 111
tuning for performance 78

V

vector MASS library 122

X

xlopt library 132
XML report schema 95



Product Number: 5765-J02; 5725-C72

Printed in USA

SC14-7332-00

