Document Number: N4906
Date: 2022-02-14
Revises: None
Reply to: Michael Wong
Codeplay
fraggamuffin@gmail.com

Working Draft, Extensions to C++ for Transactional Memory Version 2

Note: this is an early draft. It's known to be incomplet and incorrekt, and it has lots of bad formatting.

Contents

Foreword [foreword]

ISO (the International Organization for Standardization) and IEC (the International Electrotechnical Commission) form the specialized system for worldwide standardization. National bodies that are members of ISO or IEC participate in the development of International Standards through technical committees established by the respective organization to deal with particular fields of technical activity. ISO and IEC technical committees collaborate in fields of mutual interest. Other international organizations, governmental and non-governmental, in liaison with ISO and IEC, also take part in the work.

The procedures used to develop this document and those intended for its further maintenance are described in the ISO/IEC Directives, Part 1. In particular the different approval criteria needed for the different types of document should be noted. This document was drafted in accordance with the editorial rules of the ISO/IEC Directives, Part 2 (see www.iso.org/directives).

Attention is drawn to the possibility that some of the elements of this document may be the subject of patent rights. ISO and IEC shall not be held responsible for identifying any or all such patent rights. Details of any patent rights identified during the development of the document will be in the Introduction and/or on the ISO list of patent declarations received (see www.iso.org/patents) or the IEC list of patent declarations received (see http://patents.iec.ch).

Any trade name used in this document is information given for the convenience of users and does not constitute an endorsement.

For an explanation on the voluntary nature of standards, the meaning of ISO specific terms and expressions related to conformity assessment, as well as information about ISO's adherence to the World Trade Organization (WTO) principles in the Technical Barriers to Trade (TBT), see www.iso.org/iso/foreword.html.

This document was prepared by Joint Technical Committee ISO/IEC JTC 1, *Information technology*, Subcommittee SC 22, *Programming languages, their environments and system software interfaces.*

Any feedback or questions on this document should be directed to the user's national standards body. A complete listing of these bodies can be found at www.iso.org/members.html.

1 Scope [scope]

1 This document describes requirements for implementations of an interface that computer programs written in the C++ programming language may use to invoke algorithms with concurrent execution. The algorithms described by this document are realizable across a broad class of computer architectures.

2 ISO/IEC 14882:2020 provide important context and specification for this document. This document is written as a set of changes against that specification. Instructions to modify or add paragraphs are written as explicit instructions. Modifications made directly to existing text from ISO/IEC 14882:2020 use underlining to represent added text and ~~strikethrough~~ to represent deleted text.

3 This document is non-normative. Some of the functionality described by this document may be considered for standardization in a future version of C++, but it is not currently part of any C++ standard. Some of the functionality in this document may never be standardized, and other functionality may be standardized in a substantially changed form.

4 The goal of this document is to build widespread existing practice for eventually adopting transactional memory in C++.

2 Normative references [refs]

1 The following referenced document is indispensable for the application of this document. For dated references, only the edition cited applies. For undated references, the latest edition of the referenced document (including any amendments) applies.

(1.1) — ISO/IEC 14882:2020, *Programming Languages — C++*

2 ISO/IEC 14882:2020 is herein called the C++ Standard. References to clauses within the C++ Standard are written as "C++20 §3.2".

3 Terms and definitions [defs]

1 No terms and definitions are listed in this document. ISO and IEC maintain terminological databases for use in standardization at the following addresses:

(1.1) — IEC Electropedia: available at https://www.electropedia.org/

(1.2) — ISO Online browsing platform: available at https://www.iso.org/obp

4 General [general]

4.1 Implementation compliance [general.compliance]

1 Conformance requirements for this document are those defined in C++20 §4.1, as applied to a merged document consisting of C++20 amended by this document.

[*Note 1*: Conformance is defined in terms of the behavior of programs. — *end note*]

4.2 Acknowledgments [general.ack]

This work is the result of a collaboration of researchers in industry and academia. We wish to thank the original authors of this document, Michael Wong, Jens Maurer, Hans Boehm, Michael Spear, Michael Scott, and Victor Luchangco. We also wish to thank people who made valuable contributions within and outside these groups, including all SG 5 members and the original TM group, and many others not named here who contributed to the discussion.

5 Lexical conventions [lex]

5.1 Identifiers [lex.name]

In C++20 §5.10, add `atomic` to the table of identifiers with special meaning (Table 4).

6 Basics [basic]

6.9 Program execution [basic.exec]

6.9.1 Sequential execution [intro.execution]

Change in C++20 §6.9.1 paragraph 5 as indicated:

5 A *full-expression* is

(5.1) — ...

(5.2) — an invocation of a destructor generated at the end of the lifetime of an object other than a temporary object (6.7.7) whose lifetime has not been extended, ~~or~~

(5.3) — the start and the end of an atomic block (8.8), or

(5.4) — an expression that is not a subexpression of another expression and that is not otherwise part of a full-expression.

6.9.2 Multi-threaded executions and data races [intro.multithread]

6.9.2.2 Data races [intro.races]

Change in C++20 §6.9.2.2 paragraph 6 as indicated:

6 ~~Certain~~ Atomic blocks as well as certain library calls may *synchronize with* other atomic blocks and library calls performed by another thread.

Add a new paragraph after C++20 §6.9.2.2 paragraph 20:

21 An atomic block that is not dynamically nested within another atomic block is termed a *transaction*.

[*Note 21*: Due to syntactic constraints, blocks cannot overlap unless one is nested within the other. — *end note*]

There is a global total order of execution for all transactions. If, in that total order, a transaction T1 is ordered before a transaction T2, then

(21.1) — no evaluation in T2 happens before any evaluation in T1 and

(21.2) — if T1 and T2 perform conflicting expression evaluations, then the end of T1 synchronizes with the start of T2.

[*Note 22*: If the evaluations in T1 and T2 do not conflict, they might be executed concurrently. — *end note*]

22 Two actions are *potentially concurrent* if ...

Change in C++20 §6.9.2.2 paragraph 21:

21 ...

[*Note 21*: It can be shown that programs that correctly use mutexes, atomic blocks, and `memory_order::seq_cst` operations to prevent all data races and use no other synchronization operations behave as if the operations executed by their constituent threads were simply interleaved, with each value computation of an object being taken from the last side effect on that object in that interleaving. This is normally referred to as "sequential consistency". ... — *end note*]

Add a new paragraph after C++20 §6.9.2.1 paragraph 21:

22 [*Note 22*: The following holds for a data-race-free program: If the start of an atomic block T is sequenced before an evaluation A, A is sequenced before the end of T, A strongly happens before some evaluation B, and B is not sequenced before the end of T, then the end of T strongly happens before B. If an evaluation C strongly happens before that evaluation A and C is not sequenced after the start of T, then C strongly happens before the start of T. These properties in turn imply that in any simple interleaved (sequentially consistent) execution, the operations of each atomic block appear to be contiguous in the interleaving. — *end note*]

6.9.2.3 Forward progress [intro.progress]

Change in C++20 §6.9.2.3 paragraph 1 as indicated:

1 ~~The implementation may assume that any thread will eventually do~~ An *inter-thread side effect* is one of the following:

(1.1) — ~~terminate,~~

(1.2) — a call to a library I/O function,

(1.3) — an access through a volatile glvalue, or

(1.4) — a synchronization operation or an atomic operation (Clause 31).

The implementation may assume that any thread will eventually terminate or evaluate an inter-thread side effect.

[*Note 1*: This is intended to allow compiler transformations such as removal of empty loops, even when termination cannot be proven. — *end note*]

8 Statements [stmt.stmt]

8.1 Preamble [stmt.pre]

1 Add a production to the grammar in C++20 §8.1 as indicated:

statement:
labeled-statement
attribute-specifier-seq$_{opt}$ *expression-statement*
attribute-specifier-seq$_{opt}$ *compound-statement*
attribute-specifier-seq$_{opt}$ *selection-statement*
attribute-specifier-seq$_{opt}$ *iteration-statement*
attribute-specifier-seq$_{opt}$ *jump-statement*
declaration-statement
attribute-specifier-seq$_{opt}$ *try-block*
atomic-statement

Add a new subclause before C++20 §8.8:

8.8 Atomic statement [stmt.tx]

atomic-statement:
`atomic do` *compound-statement*

1 An *atomic-statement* is also called an *atomic block*.

2 The start of the atomic block is immediately before the opening `{` of the *compound-statement*. The end of the atomic block is immediately after the closing `}` of the *compound-statement*.

[*Note 1*: Thus, variables with automatic storage duration declared in the *compound-statement* are destroyed prior to reaching the end of the atomic block; see 8.7. — *end note*]

3 A goto or switch statement shall not be used to transfer control into an atomic block.

4 If the execution of an atomic block evaluates an inter-thread side effect (6.9.2.3) or if an atomic block is exited via an exception, the behavior is undefined.

5 *Recommended practice*: In case an atomic block is exited via an exception, the program should be terminated without invoking a terminate handler (17.9.5) or destroying any objects with static or thread storage duration (6.9.3.4).

6 If the execution of an atomic block evaluates any of the following outside of a manifestly constant-evaluated context (), the behavior is implementation-defined:

(6.1) — an *asm-declaration* (9.10);

(6.2) — an invocation of a function other than one of the standard library functions specified in 16.4.6.17, unless the function is inline with a reachable definition;

(6.3) — a virtual function call (7.6.1.3);

(6.4) — a function call, unless overload resolution selects

(6.4.1) — a named function (12.4.2.2.2) or

(6.4.2) — a function call operator (12.4.2.2.3), but not a surrogate call function;

(6.5) — a `co_await` expression (7.6.2.4), a *yield-expression* (7.6.17), or a `co_return` statement (8.7.5);

(6.6) — dynamic initialization of a block-scope variable with static storage duration; or

(6.7) — dynamic initialization of a variable with thread storage duration.

[*Note 2*: The implementation can define that the behavior is undefined in some or all of the cases above. — *end note*]

[*Example 1*:

```
unsigned int f()
{
  static unsigned int i = 0;
```

```
    atomic do {
      ++i;
      return i;
    }
  }
```

Each invocation of `f` (even when called from several threads simultaneously) retrieves a unique value (ignoring wrap-around). — *end example*]

[*Note 3*: Atomic blocks are likely to perform best where they execute quickly and touch little data. — *end note*]

15 Preprocessor [cpp]

15.11 Predefined macro names [cpp.predefined]

Add a row to Table 19 in C++20 §15.11:

Table 19: Feature-test macros

Macro name	Value
`__cpp_transactional_memory`	`202110`

16 Library introduction [library]

16.4 Library-wide requirements [requirements]

Add a new subclause after C++20 §16.4.6.16:

16.4.6.17 Functions usable in an atomic block [atomic.use]

1 All library functions may be used in an atomic block (8.8), except

(1.1) — error category objects (19.5.3.5)

(1.2) — time zone database (19.5.3.5)

(1.3) — clocks (27.7)

(1.4) — `signal` (17.13.5) and `raise` (17.13.4)

(1.5) — `set_new_handler`, `set_terminate`, `get_new_handler`, `get_terminate` (16.4.5.7, 17.6.4, 17.9.2)

(1.6) — `system` (17.2.2)

(1.7) — startup and termination (17.5) except `abort`

(1.8) — `shared_ptr` (20.11.3) and `weak_ptr` (20.11.4)

(1.9) — `synchronized_pool_resource` (20.12.5)

(1.10) — program-wide `memory_resource` objects (20.12.4)

(1.11) — `setjmp` / `longjmp` (17.13.3)

(1.12) — parallel algorithms (25.3)

(1.13) — `random_device` (26.6.7)

(1.14) — `locale` construction (28.3.1.3)

(1.15) — input/output (Clause 29)

(1.16) — atomic operations (Clause 31)

(1.17) — thread support (Clause 32)