std::generator: Synchronous Coroutine Generator for Ranges

Document #: P2502R2 Date: 2022-06-03

Project: Programming Language C++

Audience: LEWG

Reply-to: Casey Carter < Casey@Carter.net>

Abstract

We propose a standard library type std::generator which implements a coroutine generator that models std::ranges::input_range.

Acknowledgements

I'd like to thank Lewis Baker and Corentin Jabot, whose P2168 [3] I looted shamelessly for this proposal. This paper is presented as a new revision of their unpublished D2168R4 for continuity of design.

Revisions

P2502R2

- Rebase onto N4910.
- Change directions to the Editor from blue "Drafting Note" to magenta "Editor's note" in square brackets. Hopefully this makes them distinctive so as not to be confused with the actual wording changes.
- Changes to wording for elements_of:
 - Don't present as a subclause that holds a single subclause; flatten.
 - Don't reuse the same input range in the example, which could result in undefined behavior with some input ranges. Instead add a bool template parameter that selects between the two different uses with if constexpr.
 - Annotate member range with [[no_unique_address]] because a range type *could* be empty even if they typically are not.
- Editorial clarifications in [coroutine.generator.overview]:
 - Change "it was returned from" to "from which it was returned."

- Strike the sentence that redundantly describes how co_yield works.
- Change "element of the generator" to "element of the sequence."
- Rename the iota generator in the example to ints to avoid confusion with std::iota, std::ranges::iota_view, and std::ranges::views::iota.
- <generator> no longer explicitly includes <coroutine> and <ranges>.
- Rename generator's first template parameter from R to Ref to avoid confusion with range parameters which are usually R.
- generator publicly derives from ranges::view_interface instead of specializing ranges:: enable_view.
- Unify all "Mandates" conditions on generator's template parameters into a single bulleted list.
- Express the type requirement on generator's Allocator template parameter as a precondition.
- Allocators must use real pointers; both generator's template parameter Allocator and any allocator used in an operator new.
- Replace all occurrences (all instances?) of "instance" in the wording (primarily "generator instance" becomes "generator object").
- generator moves don't invalidate iterators.
- Simplify the specification of generator's assignment operator: it now takes a generator argument by-value and swaps coroutine handles between *this and the argument. (Usage should be equivalent to the prior formulation.)
- "Initial suspend point" has no hyphens, and is defined in [dcl.fct.def.coroutine].
- We don't need to emphasize that the stack is empty in begin.
- Strike a couple of friend sometype declarations.
- Strike exposition-only constructors.
- Coroutine handles do not "denote" coroutines, they "refer to" coroutines ([expr.await]/3.5 is the first of many usages).
- get_return_object returns generator(...) instead of generator{...}.
- Clarify behavior of promise_type::final_suspend, especially when the stack of associated coroutine handles becomes empty.
- Rename template parameter Alloc2 of promise's yield_value(elements_of) member function template to Alloc.
- Massively simplify and un-template the yield_value overload that takes const remove_reference_t<yielded>&. (We don't need to artificially template to ensure it's distinct from

- yield_value(yielded), the added constraints make it distinct even when yielded is a const lvalue reference.)
- In the description of the <code>yield_value</code> overload that takes <code>elements_of<generator<...,</code> remove the metasyntactic variable <code>r</code> that refers to "an Ivalue denoting the <code>generator</code> referenced by <code>g.range</code>". <code>g.range</code> is already such an Ivalue, just use it directly. Use the language "stored exception" instead of "captured exception" to avoid any possible confusion with lambda captures. Strike the "Variables with automatic storage duration" note as extraneous.
- In the description of the yield_value overload that takes elements_of<Rng, similarly use r.range directly instead of defining rng as an identical lvalue. In the "effects equivalent to" code, define a local name for range_reference_t<Rng> and use that instead of auto and decltype(e).
- Specify that *yield-expression* s that call one of the promise::yield_value overloads have type void.
- Corrected spelling of __STDCPP_DEFAULT_NEW_ALIGNMENT__.
- Remove extraneous static from all declarations of class-member operator new and operator delete; class-member operator new and operator delete are static members regardless.
- Merge the specifications of the operator new overloads into a single specification.
- Fix precondition on iterator::operator*() and iterator::operator++().
- Remove precondition from *iterator*::operator++(int) which is extraneous thanks to "Effects: equivalent to".
- Make operator== a hidden friend instead of a member for consistency with the rest of the library. Also use "Effects: equivalent to" for operator== to avoid explicit preconditions.
- Add "top of stack coroutine is suspended" preconditions to iterator::operator* (explicitly) and iterator::operator++ (implicit via effects-equivalent-to).
- Since all "parameter preview" arguments to operator new are Ivalues, operator new overloads now take the allocator, implicit object, and "ignored" parameters by const&.
- Make the "stack of associated coroutines" an exposition-only unique_ptr<stack<coroutine
 _handle<>>> member of generator.
- Don't use "Preconditions" for a template parameter type requirement.
- (Not requested by LWG) Change implementation of generic yield_value per Lewis's suggestion to encourage HALO. (Yes, this overload is overspecified.)
- Add missing *Returns:* element to iterator::operator=.
- Fix "Returns: return *this;" for iterator::operator++().
- Define the nested classes within namespace std.

· Update implementation.

P2502R1

- The type generator::yielded is now publicly-visible and not exposition-only per 2021-12-13 LEWG discussion.
- Feature-test macro insertion is now ordered properly with the other wording.
- Coroutines for generator<T&&> can now yield Ivalues; the generated element is an xvalue denoting a copy of the Ivalue. This saves folks the trouble of writing co_yield auto(1value) instead of co_yield 1value.
- Reorder generator's template parameters (again, for the last last time). Coroutine Task
 Force discussions convinced all participants (including the author) that it was too big
 a break with the rest of the Standard Library to have the Allocator not be the final
 template parameter.
- Simplify elements_of into a vanilla two-element struct.
- Update implementations accordingly.
- Per 2022-01-025 LEWG direction, generator<T> now behaves like generator<T&&> instead of like generator<const T&> per the design in P2529R0 [9].

P2502R0

- Reorder generator's template parameters. This allows the reference type to be more
 easily defaulted to a true reference, while still respecting requests for differing value
 and reference types. This preserves the previous design's ease-of-use, while providing
 full generality.
- Remove concerns about the $\mathcal{O}(1)$ destruction requirement for view, which has been relaxed by P2415R2 "What is a view?" [7].

D2168R4

Wording improvements

P2168R3

Wording improvements

P2168R2

- Some wording fixes
- Improve the section on allocator support
- Updated implementation

P2168R1

- Add benchmarks results and discussion about performance
- Introduce elements_of to avoid ambiguities when a generator is convertible to the reference type of the parent generator.
- Add allocator support
- Symmetric transfer works with generators of different value / allocator types
- Remove iterator::operator->
- Put generator in a new <generator> header.
- Add an other example to motivate the Value template parameter

Example

```
std::generator<int> fib() {
    auto a = 0, b = 1;
    while (true) {
        co_yield std::exchange(a, std::exchange(b, a + b));
    }
}
int answer_to_the_universe() {
    auto rng = fib() | std::views::drop(6) | std::views::take(3);
    return std::ranges::fold_left(std::move(range), 0, std::plus{});
}
```

Motivation

C++ 20 had very minimalist library support for coroutines. Synchronous generators are an important use case for coroutines, one that cannot be supported without the machinery presented in this paper. Writing an efficient recursive generator is non-trivial; the standard should provide one.

Design

While the proposed std::generator interface is fairly straightforward, a few decisions are worth pointing out.

input_view

std::generator is a move-only view which models input_range and has move-only iterators. This is because the coroutine state is a unique resource (even if the coroutine *handle* is copyable).

Header

Multiple options are available as to where to put the generator class.

- <coroutine>, but <coroutine> is a low level header, and generator depends on bits of <type_traits> and <iterator>.
- < <ranges>
- A new <generator>

This paper uses a new <generator> header since P2168R3 did so, and LEWG has provided no guidance to do otherwise. We do note on our MSVC STL branch implementation that #include<ranges> includes 52.6k lines of code, and #include<generator> 53.3k lines. (Note that <generator> is specified to include both <ranges> and <coroutines>.) Defining generator in <ranges> together with a #include<coroutine> would penalize people who want <ranges> but not generator by about 610 LoC.

Reference type

generator has 3 template parameters: generator<R, V = void, Allocator = void>

From R and V, we derive types:

```
using Value = conditional_t<is_void_v<V>, remove_cvref_t<R>, V>;}
using Reference = conditional_t<is_void_v<V>, R&&, R>;
using Yielded = conditional_t<is_reference_v<Reference>, Reference, const Reference&>;
```

- Value is a cv-unqualified object type that specifies the value type of the generator's range and iterators,
- Reference specifies the reference type (not necessarily a core language reference type) of the generator's range and iterators, and
- Allocator is the type of allocator used for the coroutine state, which can be void to typeerase any allocator specified as a coroutine argument, defaulting to allocator<byte> when none is specified.
- Yielded (necessarily a reference type) is the type of the parameter to the primary overload
 of yield_value in the generator's associated promise type.

generator<meow>

Our expectation is that 98% of use cases will need to specify only one parameter. The resulting generator:

- has a value type of remove_cvref_t<meow>
- has a reference type of meow, if it is a reference type, or meow&& otherwise,
- expects co_yield to appear in the body of the generator with operands that are convertible to the reference type, and

• can use any allocator (via type-erasure) defaulting to allocator

byte>.

This avoids the performance pitfall from earlier revisions of the proposal that used the first argument type directly as reference type; users who naively chose generator<std::string> got an iterator that produces independent copies of the yielded value on every dereference, when they may have been satisfied by yielding a reference to the same constant value.

As a nod to ease of use, and to more readily reflect the semantics of prvalue-returning functions, we allow yielding lvalues from generators with an rvalue reference type when the lvalue is copyable. The result is that the yielded lvalue is copied into temporary storage, and an iterator for the generator will return an xvalue denoting that copy when dereferenced.

generator<meow, woof>

For the rare user who needs generator to step outside the box and use a proxy reference type, or who needs to generate a range whose iterators yield prvalues for whatever reason, we have two-argument generator. If woof is void, this is generator<meow>. Otherwise, the resulting generator:

- has a value type of woof,
- has a reference type of meow,
- expects co_yield to appear in the body of the generator with operands that are convertible to meow, if it is reference type, and otherwise const meow&, and
- can use any allocator (via type-erasure) if woof is void, or otherwise can use any allocator convertible to woof.

Your iterators can yield a prvalue, but it must be a prvalue of copy_constructible type so a copy of the operand of a single co_yield can be returned multiple times from repeated dereferences of the same iterator value.

generator<meow, woof, quack>

For use cases that want to specify an allocator type statically so they need not constantly pass pairs of allocator_arg, my_allocator arguments to every coroutine, we have three-argument generator. The resulting generator can use any allocator convertible to quack, defaulting to a default-constructed woof if it is default_initializable. The value, reference, and yielded types are as described for generator<meow> if woof is void, and as described for generator<meow, woof> otherwise.

Our expectation is that libraries that wish to declare many functions with the same statically-specified allocator type will define a template alias like

```
template <class R, class V = void>
using my_generator = generator<R, V, my_allocator>;
```

to ease declarations. In practice, this should mean <code>generator<R, void, my_allocator></code> (with an explicit second argument of <code>void</code>) appears quite rarely in real code.

Allocator = void

In the 2022-01-13 Coroutine Task Force meeting, Lewis Baker brought up that LEWG feedback on some revision of P2168 had expressed dissatisfaction with the use of void to indicate "type-erase any allocator." We discussed some alternatives:

- Devise a new type for this specific purpose. We felt that single-use tag types have low return-on-investment; every such tag type consumes a few more bytes of memory in nearly every compilation, and a few more neurons of memory in every programmer.
- Reuse an existing type for this purpose. void has the advantage that it can never be confused for a conforming allocator type, even if it someday does become a regular type. void requires no predeclaration, and no memory is is expended to store its definition.

monostate was another suggestion that was universally reviled; that type already has a very clear purpose in the Standard Library that we didn't feel inclined to muddle.

No one suggested another type that we found to be an improvement.

Our conclusion was that we didn't find the use of void for this purpose confusing, we were happy to keep it given that more-recent LEWG discussion had resulted in no further feedback on this topic.

Obsolete discussion about reference specification

[*Note:* Before P2502R0, generator's first parameter Type denoted the reference type of the range / iterators, and the value type was defaulted to remove_cvref_t<Type>. The following sections of design discussion from that era are preserved here. — end note]

In earlier versions of this paper, the reference type was exactly the first template parameter. This had the advantage of being simple. But it was a terrible performance trap:

Consider the behavior of the following code assuming the reference type is exactly the first template argument:

```
std::generator<std::string> f() {
    std::string hello = "hello";
    co_yield hello;    // 0 or 1 copy depending on implementation
    co_yield "Hello";    // 1 copy (conversion from const char* to std::string)
}

for (auto&& str : f()) {} // 1 copy (*it returns std::string)
```

Of course the solution, which we advocated for, is for the user to manually specify an explicit reference type:

```
generator<const std::string&> f() {
    std::string hello = "hello";
    co_yield hello; // 0 or 1 copy depending on implementation
    co_yield "Hello"; // 1 copy (conversion from const char* to std::string)
}
```

```
for (auto&& str : f()) {} // 0 copy
```

This works, can be explained, and is even logical. You get what you asked for. It is nonetheless surprising for non-experts that using the simple generator<string> would create 2 copies per co_yield.

To hope users would not routinely forget to use a reference type when using std::generator calls for a heaping barrel of optimism.

We later proposed that for a generator<T>, its reference type be conditional_t<is_reference_-v<T>, T, const T&>.

First parameter	reference type	default value	can yield mutable lvalue ref?
int	const int&	int	No
const int&	const int&	int	No
int&	int&	int	Yes
int&&	int&&	int	No
const int&&	const int&&	int	No

Attempts have been made to characterize the exact relations between reference, value, storage, and co_yield exception types and categories. Ultimately, a simpler mental model is to characterize what expressions can be yielded for a given reference type and how many copies are made for each scenario.

First parameter	co_yield const T&	co_yield T&	co_yield T&&	co_yield U&&
Т	0	0	0	1
const T&	0	0	0	1
T&	Ill-formed	0	Ill-formed	Ill-formed
T&&	Ill-formed	Ill-formed	0	1
const T&&	Ill-formed	Ill-formed	0	1

In this table, we see that only co_yield that requires conversion incurs copy, which is expected. Coroutines guarantee that the yielded expression exceeds the lifetime of the co_yield expression, so generator can usefully store a pointer to the object denoted by a yielded xvalue.

co_yield expressions involving conversion can store the yielded value in an awaiter. The type of the stored expression is the reference type with its reference qualifiers stripped, but that is an implementation detail that is not observable and is therefore of limited interest. Of course, that type needs to be constructible from yielded values.

Besides the T case, this behaves very much like returning from a function that is intended.

Move-only and immovable types

LEWG was interested in how this works with generator of move-only and immovable types.

First parameter	co_yield const T&	co_yield T&	co_yield T&&	
move_only	0	0	0	
const move_only&	0	0	0	
move_only&	Ill-formed	0	Ill-formed	
move_only&&	Ill-formed	Ill-formed	0	
const move_only&&	Ill-formed	Ill-formed	0	
immovable	0	0	0	
const immovable&	0	0	0	
immovable&	Ill-formed	0	Ill-formed	
immovable&&	Ill-formed	Ill-formed	0	
const immovable&&	Ill-formed	Ill-formed	0	

As that table shows, these types work exactly like other types. However, to be able to move from a move only reference type, the coroutine has to explicitly state so:

```
auto f = []() -> std::generator<move_only> { co_yield move_only{}; }();
for (auto&& x : f) {
    move_only mo = std::move(x); // ill-formed, decltype(x) is const move_only&}

auto f = []() -> std::generator<move_only&&> { co_yield move_only{}; }();
for (auto&& x : f) {
    move_only mo = x; // ok
}

auto f = []() -> std::generator<move_only&> { move_only m; co_yield m; }();
for (auto&& x : f) {
    move_only mo = std::move(x); // dicey but okay
}
```

Potential downsides

```
auto f = []() -> std::generator<MyType> {
    MyType t;
    co_yield std::move(t);
}();
```

In the example above std::move doesn't move. Arguably more than usual. Indeed the code expands to something similar to:

```
auto&& __temp = std::move(t);
yield_value(_temp); // <=> promise.value = std::addressof(__temp); // no move
```

Of course, a move would not have occurred for a std::generator<const MyType&> either as these things are identical. It might be suprising? The only way to avoid that is to create temporary value for rvalue reference, which would force a move to actually occurs, at the cost

of performance.

Alternatives considered

Mandating a reference as the first parameter We could make generator<int> ill-formed and force people to specify a reference type like generator<const int&>. We do not think this is very user-friendly, given that we can provide a reasonable default.

We rejected this option.

Using T& as the default There are two issues with mutable references:

- They are mutable (They allow mutating the coroutine frame), which would be an *interest-ing* default.
- They are very restrictive as to the set of co_yield expression allowed with them.

We rejected this option.

Using T&& as the default This avoids a copy when doing auto object = *it (where it is a std::generator::iterator), but it is easy to misuse, consider:

```
auto f = [] -> std::generator<std::string> { co_yield "footgun"; }();
for (auto&& x : f | std::views::filter([](std::string s) { return s.size() > 0; })) {
    std::cout << x << '\n'; // outputs a single empty line
}</pre>
```

We rejected this option. [*Note:* Nevertheless, we came back to this option after much discussion of pros and cons in both LEWG and several additional meetings if a Coroutine Task Force formed by LEWG Chair specifically to build consensus. — *end note*]

Doing something clever for move-only types We considered returning T% for move_only types so that they can be moved from by default. We realized this was too clever and inconsistent. Notably, adding a copy constructor to T would change the meaning of the code.

We rejected this option.

Doing something clever for reference types By default generator<reference_wrapper<T>> could yield reference_wrapper<T> has that is already a "reference-like" type. However, no other view does that, "reference-like" is fuzzily defined, and this would probably cause more trouble than it's worth.

We rejected this option.

Keeping the D2168R4 design Returning values has the potential to severely impact performance, is inconsistent with other views, and is not necessary. It also did not work with move-only types.

The change, along with an implementation strategy described in the "How to store the yielded value in the promise type?" guarantees that no copy needs to be made if the reference and yielded types are the same (with qualifiers stripped).

We think this new approach keeps the simplicity of the original design, improves performance, and works with more types.

Thank you LEWG, and in particular Mathias, for highlighting these concerns!

Separately specifyable Value Type

This proposal supports specifying both the "yielded" type, which is the iterator's reference type (not required to be a reference) and its corresponding value type. This allow ranges to handle proxy types and wrapped reference, like this implementation of zip:

```
namespace ranges = std::ranges;

template<ranges::input_range Rng1, ranges::input_range Rng2>
std::generator<
    std::tuple<ranges::range_reference_t<Rng1>, ranges::range_reference_t<Rng2>>
    std::tuple<ranges::range_value_t<Rng1>, ranges::range_value_t<Rng2>>>
zip(Rng1 r1, Rng2 r2) {
    auto it1 = ranges::begin(r1);
    auto it2 = ranges::begin(r2);
    auto end1 = ranges::end(r1);
    auto end2 = ranges::end(r2);
    for (; it1 != end1 && it2 != end2; ++it1, ++it2) {
        co_yield {*it1, *it2};
    }
}
```

In this second example, using string as value type ensures that calling code can take the necessary steps to make sure iterating over a generator would not invalidate any of the yielded values.

```
// Yielding string literals : always fine
std::generator<std::string_view, std::string_view> string_views() {
    co_yield "foo";
    co_yield "bar";
}

std::generator<std::string_view, std::string> strings() {
    co_yield "start";
    std::string s;
    for (auto sv : string_views()) {
        s = sv;
        s.push_back('!');
        co_yield s;
    }
    co_yield "end";
}
```

```
// conversion to a vector of strings
// If the value_type was string_view, it would convert to a vector of string_view,
// which would lead to undefined behavior operating on elements of v that were
// invalidated while iterating through the generator.
auto v = std::ranges::to<vector>(strings()); // (P1206R3 [4])
```

How to store the yielded value in the promise type?

There are multiple implementation strategies possible to store the value in the generator. An early revision of this paper always stored a copy of the yielded value, leading to an extra copy. Later revisions supported storing the yielded value in an awaitable object returned from the promise's yield_value function.

However, the object denoted by a glvalue yield expression is guaranteed to live until the coroutine resumes. We can take advantage of that fact by storing only a pointer to the denoted object in the promise, if the result of dereferencing that pointer is convertible to the generator's reference type. We guarantee this is the case by providing a yield_value whose parameter type is always a reference type (conditional_t<is_reference_v<Reference>, Reference, const Reference&>). This forces any conversions to happen inside the coroutine itself, yielding a temporary glvalue that can later be dereferenced to an lvalue which is trivially static_casted to Reference in the iterator's operator*.

A drawback of this solution is that the yielded value is only destroyed at the end of the full expression in which co_yield appears, so given

```
(co_yield x, co_yield y); // x is destroyed after y is yielded.
```

We think this is a reasonable tradeoff given that this approach minimizes the number of copies must be made of the yielded value. We force the coroutine to materialize the element to be yielded, but after doing so can cleanly pass a reference to that element through the coroutine and iterator machinery and directly to consuming code.

Recursive generator

A "recursive generator" is a coroutine that supports the ability to directly co_yield a generator of the same type as a way of emitting the elements of that generator as elements of the current generator.

Example: A generator can co_yield other generators of the same type

```
std::generator<const std::string&> delete_rows(std::string table, std::vector<int> ids) {
    for (int id : ids) {
        co_yield std::format("DELETE FROM {0} WHERE id = {1};", table, id);
    }
}
std::generator<const std::string&> all_queries() {
    co_yield std::ranges::elements_of(delete_rows("user", {4, 7, 9 10}));
```

```
co_yield std::ranges::elements_of(delete_rows("order", {11, 19}));
}
```

Example: A generator can also be used recursively

```
using namespace std;

struct Tree {
    Tree* left;
    Tree* right;
    int value;
};

generator<int> visit(Tree& tree) {
    if (tree.left) co_yield ranges::elements_of(visit(*tree.left));
    co_yield tree.value;
    if (tree.right) co_yield ranges::elements_of(visit(*tree.right));
}
```

In addition to being more concise, the ability to directly yield a nested generator has some performance benefits compared to iterating over the contents of the nested generator and manually yielding each of its elements.

Yielding a nested generator allows the consumer of the top-level coroutine to directly resume the current leaf generator when incrementing the iterator, whereas a solution that has each generator manually iterating over elements of the child generator requires O(depth) coroutine resumptions/suspensions per element of the sequence.

Example: Non-recursive form incurs O(depth) resumptions/suspensions per element and is more cumbersome to write:

```
using namespace std;
generator<int> slow_visit(Tree& tree) {
   if (tree.left) {
      for (int x : ranges::elements_of(visit(*tree.left)))
            co_yield x;
   }
   co_yield tree.value;
   if (tree.right) {
      for (int x : ranges::elements_of(visit(*tree.right)))
            co_yield x;
   }
}
```

Exceptions that propagate out of the body of nested generator coroutines are rethrown into the parent coroutine from the co_yield expression rather than propagating out of the top-level iterator::operator++(). This follows the mental model that co_yield someGenerator is semantically equivalent to manually iterating over the elements and yielding each element.

For example: nested_ints() is semantically equivalent to manual_ints()

```
std::generator<int> might_throw() {
    co_yield 0;
    throw some_error{};
}
std::generator<int> nested_ints() {
    try {
        co_yield std::ranges::elements_of(might_throw());
    } catch (const some_error&) {}
   co_yield 1;
}
// nested_ints() is semantically equivalent to the following:
std::generator<int> manual_ints() {
    try {
        for (int x : might_throw()) {
            co_yield x;
        }
    } catch (const some_error&) {}
   co_yield 1;
}
void consumer() {
   for (int x : nested_ints()) {
        std::cout << x << " "; // outputs 0 1
    }
   for (int x : manual_ints()) {
        std::cout << x << " "; // also outputs 0 1</pre>
}
```

std::ranges::elements_of

ranges::elements_of is a utility function that prevents ambiguity when a nested generator type is convertible to the value type of the present generator

```
generator<int> f()
{
    co_yield 42;
}
generator<any> g()
{
    co_yield f(); // should we yield 42 or generator<int> ?
}
```

To avoid this issue, we propose that:

• co_yield <expression> yields the value directly, and

• co_yield elements_of(<expression>) yields successive elements from the nested generator.

For convenience, we further propose that $co_yield\ elements_of(x)$ be extended to support yielding the values of arbitrary ranges beyond generators, ie

```
std::generator<int> f()
{
    std::vector<int> v = /*... */;
    co_yield std::ranges::elements_of(v);
}
```

Symmetric transfer

The recursive form can be implemented efficiently with symmetric transfer. Earlier works in [CppCoro] implemented this feature in a distinct recursive_generator type.

However, it appears that a single type is reasonably efficient thanks to HALO optimizations and symmetric transfer. The memory cost of that feature is two extra pointers per generator¹. It is difficult to evaluate the runtime cost of our design given the current coroutine support in compilers. However our tests show no noticeable difference between a generator and a recursive_generator which is called non-recursively. It is worth noting that the proposed design makes sure that HALO [8] optimizations are possible.

While we think a single generator type is sufficient and offers a better API, there are three options:

- A single generator type supporting recursive calls (this proposal).
- A separate type recursive_generator that can yield values from either a recursive_generator or a generator. That may offer very negligible performance benefits, same
 memory usage.
- A separate recursive_generator type which can only yield values from other recursive_generators.

That third option would make the following ill-formed:

```
generator<int> f();
recursive_generator<int> g() {
    co_yield f(); // incompatible types
}
```

Instead you would need to write:

```
recursive_generator<int> g() {
   for (int x : f()) co_yield x;
}
```

¹The two pointers in our implementation have non-overlapping active times; we believe the pair can be optimized into a single pointer's space with some bit hacking to store a discriminator in the unused lower bits.

Such a limitation can make it difficult to decide at the time of writing a generator coroutine whether or not you should return a generator or recursive_generator as you may not know at the time whether or not this particular generator will be used within recursive_generator or not.

If you choose the <code>generator</code> return-type and then later someone wants to yield its elements from a <code>recursive_generator</code> then you either need to manually yield its elements one-by-one or use a helper function that adapts the <code>generator</code> into a <code>recursive_generator</code>. Both of these options can add runtime cost compared to the case where the <code>generator</code> was originally written to return a <code>recursive_generator</code>, as it requires two coroutine resumptions per element instead of a single coroutine resumption.

Because of these limitations, we are not recommending this approach.

Symmetric transfer is possible for different generator types as long as the reference type is the same, aka, different value type or allocator type does not preclude symmetric transfer (see the section on allocators).

Allocator support

In line with the design exploration done in section 2 of P1681R0 [6], std::generator supports both stateless and stateful allocators and strives to minimize the interface verbosity for stateless allocators by templating both the generator itself and the promise_type's new operator on the allocator type. Details for this interface are found in P1681R0 [6].

coroutine_parameter_preview_t such as discussed in section 3 of P1681R0 [6] has not been explored in this paper.

```
std::generator<int> stateless_example() {
    co_yield 42;
}

template <class Allocator>
std::generator<int> allocator_example(std::allocator_arg_t, Allocator alloc) {
    co_yield 42;
}

my_allocator<std::byte> alloc;
input_range auto rng = allocator_example(std::allocator_arg, alloc);
```

The proposed interface requires that, if an allocator is provided, it is the second argument to the coroutine function, immediately preceded by an argument of type std::allocator_arg_t. This approach is necessary to distinguish the allocator desired to allocate the coroutine state from allocators whose purpose is to be used in the body of the coroutine function. The required argument order might be a limitation if any other argument is required to be the first. However, we cannot think of any scenario where that would be the case.

We think it is important that all standard and user coroutine types can accommodate similar interfaces for allocator support. In fact, the implementation for that allocator support can be

shared amongst generator, lazy, and other standard types.

By default std::generator type erases the allocator type, and uses std::allocator unless an allocator is provided to the coroutine function. Then:

Type erased allocator(default)

```
template <class Allocator>
std::generator<int> f(std::allocator_arg_t, Allocator alloc) {}
f(std::allocator_arg, my_alloc{});
```

Returns a generator of type std::generator<int, void, void> where the final void denotes that the allocator is type erased. The allocator is stored in the same allocation as the coroutine state if it is stateful or not default constructible; a pointer is always stored so that the deallocate method of the type erased allocator can be called.

No allocator

```
std::generator<int> f() {}
f();
```

Again, returns a generator of type std::generator<int, void, void> where the final void denotes that the allocator is type erased. A pointer is stored so that the deallocate method of the type-erased allocator can be called, but the default allocator (std::allocator) need not be stored since it is stateless.

Explicit stateless allocator

```
std::generator<int, void, std::stateless_allocator<int>> f() {}
f();
```

No extra storage is used for the allocator because it is stateless.

Explicit stateful allocator

```
std::generator<int, void, some_stateful_allocator<int>>
    f(std::allocator_arg_t, some_stateful_allocator<int>> alloc) {}
f(std::allocator_arg, some_allocator); // must be convertible to some_stateful_allocator
```

The allocator is copied in the coroutine state.

Can we postpone adding allocator support?

A case can be made that allocator support could be added to std::generator later. However, because the proposed design has the allocator as a template parameter, adding allocator after std::generator ships would represent an ABI break. We recommend that we add allocator support as proposed in this paper now and make sure that the design remains consistent as work on std::lazy is made in this cycle. However, it would be possible to extend support for different mechanisms (such as presented in section 3 of P1681R0 [6] later.

Interaction of symmetric transfer and allocator support

The allocator must necessarily be part of a coroutine's promise type since implementations query the promise for allocation functions. Nonetheless, it would seem silly for a generator to be unable to nest another generator with identical element type but differing allocator. For that matter, even differing value types shouldn't be problematic: the only interface between the generator and the coroutine it wraps that differs depending on the type arguments to generator is yield_value. Ideally, generators would be able to recurse into other generators whose yield_value has the same parameter type even if all three template arguments to generator differ.

Our implementation uses a base class to implement the non-allocation behaviors for generator's promise type so that generators with different allocator types can yield each other. Doing so, however, requires that we partially erase the type of a coroutine_handle so we can resume it later knowing only that its promise type derives from a particular base.

There are at least two ways to implement this partial type erasure:

- Storing a pointer in the common base to a component with full type knowledge, which can then resume the targeted coroutine,
- Relax the preconditions on some of the coroutine_handle functions to allow conversion from coroutine_handle<void> to coroutine_handle<T> when the source's corresponding address() value was obtained from a coroutine_handle referring to a coroutine whose promise object is pointer-interconvertible with an object of type T.

Our current plan is to standardize the intent to allow yielding nested generators with different allocator and value types, leaving the details of the implementation unspecified, and to later separately propose the changes to coroutine_handle that enable that implementation to be maximally efficient.

Implementation and experience

generator has been provided as part of cppcoro and folly. However, cppcoro offers a separate recursive_generator type, which is different than the proposed design.

Folly uses a single generator type, which can be recursive but doesn't implement symmetric transfer. Despite that, Folly users found the use of Folly:::Generator to be a lot more efficient than the eager algorithm they replaced with it.

ranges-v3 also implements a generator type, which is never recursive and predates the work on move-only views and iterators [1], [2] which forces this implementation to ref-count the coroutine handler.

Our implementation [Implementation] consists of a single type that takes advantage of symmetric transfer to implement recursion - it notably works well with three different major standard libraries.

Performance & benchmarks

[*Note:* These benchmark results are fairly dated now - roughly a year old - and should be taken with a grain of salt. — *end note*]

Because implementations are still being perfected, and because performance is extremely dependant on whether HALO optimization (see P0981R0 [8]) occurs, it is difficult at this time to make definitive statements about the performance of the proposed design.

At the time of the writing of this paper, Clang is able to inline non-nested coroutines whether the implementation supports nested coroutines or not, while GCC never performs HALO optimization.

When the coroutine is not inlined, support for recursion does not noticeably impact performance. And, when the coroutine yields another generator, the performance of the recursive version is noticeably faster than yielding each element of the range. This is especially noticeable with deep recursion.

	Clang	Clang ST ¹	GCC	GCC ST ¹	MSVC	MSVC ST ¹
Single value	(1) 0.235	(2) 2.36	12.4	13.4	61.9	63.7
Single value, noinline (3)	13.5	13.7	14.1	15.2	63.8	64.4
Deep nesting	43670266.0	(4) 427955.0	58801348	338736	224052033	4760914

¹ Symmetric transfer.

The values are expressed in nanoseconds. However, please note that the comparison of the same result across compiler is not meaningful, notably because the MSVC results were obtained on different hardware. That being said, we observe:

- Only Clang can perform constant folding of values yielded by simple coroutine (1)
- When the generator supports symmetric transfer, clang is not able to fully inline the generator construction, but HALO is still performed (2).
- When HALO is not performed, the relative performance of both approaches is similar (3).
- Supporting recursion is greatly beneficial to nested/recursive algorithms (4).

The code for these benchmarks, as well as more detailed results, can be found on Github.

Wording

[Editor's note: Wording is relative to Working Draft N4910 [5]. Insert in lexicographical order in [version.syn] (updating YYYYXXL to the date of merge):]

[Editor's note: Modify [ranges.general] as follows:]

� General [ranges.general]

This Clause describes components for dealing with ranges of elements.

The following subclauses describe range and view requirements, and components for range primitives and range generators as summarized in Table [tab:range.summary].

[Editor's note: Add a new row at the end of [tab:range.summary] "Range generators" with header <generator> referring to the new subclause [coroutine.generator] added below.]

[Editor's note: Add the declaration of ranges::elements_of to the <ranges> synopsis:]

Header < ranges > synopsis

[ranges.syn]

```
namespace std::ranges {
[...]

template<input_or_output_iterator I, sentinel_for<I> S, subrange_kind K>
inline constexpr bool enable_borrowed_range<subrange<I, S, K>> = true;

// [range.dangling], dangling iterator handling
struct dangling;

// [elementsof.overview], class template elements_of
template<range R, class Allocator = allocator<byte>>
struct elements_of;

template<range R>
using borrowed_iterator_t = conditional_t<borrowed_range<R>, iterator_t<R>, dangling>;
[...]
}
```

[Editor's note: Insert the following new subclause immediately after [range.dangling]:]

class template elements_of

[ranges.elementsof]

Specializations of elements_of encapsulate a range and act as a tag in overload sets to disambiguate when a range should be treated as a sequence rather than a single value.

[Example:

```
template <bool YieldElements>
        std::generator<any> f(std::ranges::input_range auto&& rng) {
            if constexpr (YieldElements)
                co_yield std::ranges::elements_of(rng); // yield each element of rng
            else
               co_yield rng; // yield rng as a single value
        }
— end example]
namespace std::ranges {
  template<range R, class Allocator = allocator<byte>>
  struct elements_of {
   [[no_unique_address]] R range;
    [[no_unique_address]] Allocator allocator = Allocator();
 };
 template<class R, class Allocator = allocator<byte>>
 elements_of(R&&, Allocator = Allocator()) -> elements_of<R&&, Allocator>;
}
```

[Editor's note: Add the following subclause to the end of [ranges]:]

Range Generators

[coroutine.generator]



[coroutine.generator.overview]

Class template generator presents a view of the elements yielded by the evaluation of a coroutine.

A generator generates a sequence of elements by repeatedly resuming the coroutine from which it was returned. Elements of the sequence are produced by the coroutine each time a co_yield statement is evaluated. When the co_yield statement is of the form co_yield elements_of(rng), each element of the range rng is successively produced as an element of the sequence.

[Example:

```
std::generator<int> ints(int start = 0) {
        while (true)
            co_yield start++;
}

void f() {
    for (auto i : ints() | std::views::take(3))
        std::cout << i << ' '; // prints 0 1 2
}</pre>
```

— end example]

Header < generator > synopsis

[generator.syn]

```
namespace std {
   // [coroutine.generator.class], class template generator
   template<class Ref, class V = void, class Allocator = void>
   class generator;
}
```

Class template generator

[coroutine.generator.class]

```
namespace std {
  template<class Ref, class V = void, class Allocator = void>
  class generator : public ranges::view_interface<generator<Ref, V, Allocator>> {
 private:
   using value =
                      // exposition only
     conditional_t<is_void_v<V>, remove_cvref_t<Ref>, V>;
   using reference = // exposition only
     conditional_t<is_void_v<V>, Ref&&, Ref>;
   class iterator; // exposition only
  public:
    using yielded =
     conditional_t<is_reference_v<reference>, reference, const reference&>;
    class promise_type;
    generator(const generator&) = delete;
    generator(generator&& other) noexcept;
    ~generator();
    generator& operator=(generator other) noexcept;
    iterator begin();
    default_sentinel_t end() const noexcept;
 private:
   coroutine_handlecoroutine_ = nullptr; // exposition only
   unique_ptr<stack<coroutine_handle<>>> active_; // exposition only
 };
}
```

Mandates:

- If Allocator is not void, allocator_traits<Allocator>::pointer is a pointer type.
- *value* is a cv-unqualified object type.
- reference is either a reference type, or a cv-unqualified object type that models copy_constructible.
- Let RRef denote remove_reference_t<*reference*>&& if *reference* is a reference type, or *reference* otherwise. Each of:

- common_reference_with<reference&&, value&>,
- common_reference_with<reference&&, RRef&&>, and
- common_reference_with<RRef&&, const value&>

is modeled. [*Note:* These requirements ensure the exposition-only *iterator* type can model indirectly_readable and thus input_iterator. — *end note*]

If Allocator is not void, it shall meet the Cpp17Allocator requirements.

Specializations of generator model view and input_range.

The behavior of a program that adds a specialization for generator is undefined.

Members

[generator.members]

generator(generator&& other) noexcept;

Effects: Initializes coroutine_ with exchange(other.coroutine_, {}) and active_ with exchange(other.active_, nullptr).

[Note: Iterators previously obtained from other are not invalidated; they become iterators into *this. — end note]

~generator();

Effects: Equivalent to:

```
if (coroutine_) {
  coroutine_.destroy();
}
```

[*Note:* Ownership of recursively yielded generators is held in awaitable objects in the coroutine frame of the yielding generator, so destroying the root generator effectively destroys the entire stack of yielded generators. — *end note*]

generator& operator=(generator other) noexcept;

Effects: Equivalent to:

```
swap(coroutine_, other.coroutine_);
swap(active_, other.active_);
```

Returns: *this.

[Note: Iterators previously obtained from other are not invalidated; they become iterators into *this. — end note]

```
iterator begin();
```

Preconditions: coroutine_ refers to a coroutine suspended at its initial suspend point ([dcl.fct.def.coroutine]).

Effects: Pushes *coroutine_* into **active_*, then evaluates *coroutine_*.resume().

Returns: An *iterator* object whose member *coroutine*_ refers to the same coroutine as does *coroutine*_.

[*Note:* A program that calls begin more than once on the same generator has undefined behavior. — *end note*]

default_sentinel_t end() const noexcept;

Returns: default_sentinel.

class generator::promise_type

[coroutine.generator.promise]

```
namespace std {
  template<class Ref, class V, class Allocator>
  class generator<Ref, V, Allocator>::promise_type {
  public:
    generator get_return_object() noexcept;
    suspend_always initial_suspend() const noexcept { return {}; }
    auto final_suspend() noexcept;
    suspend_always yield_value(yielded val) noexcept;
    auto yield_value(const remove_reference_t<yielded>& lval)
      requires is_rvalue_reference_v<yielded> &&
        constructible_from<remove_cvref_t<yielded>, const remove_reference_t<yielded>&>;
    template<class R2, class V2, class Alloc2, class Unused>
      requires same_as<typename generator<T2, V2, Alloc2>::yielded, yielded>
        auto yield_value(ranges::elements_of<generator<T2, V2, Alloc2>&&, Unused> g) noexcept;
    template<ranges::input_range Rng, class Alloc>
      requires convertible_to<ranges::range_reference_t<Rng>, yielded>
        auto yield_value(ranges::elements_of<Rng, Alloc> r) noexcept;
    void await_transform() = delete;
    void return_void() const noexcept {}
   void unhandled_exception();
    void* operator new(size_t size)
      requires same_as<Allocator, void> || default_initializable<Allocator>;
    template<class Alloc, class... Args>
      requires same_as<Allocator, void> || convertible_to<const Alloc&, Allocator>
        void* operator new(size_t size, allocator_arg_t, const Alloc& alloc, const Args&...);
    template<class This, class Alloc, class... Args>
      requires same_as<Allocator, void> || convertible_to<const Alloc&, Allocator>
```

Returns: A generator object whose member *coroutine_* is coroutine_handlecoroutine_handlecoroutine_handlecoroutine_handlecoroutine_handlecoroutine_handlecoroutine_handlecoroutine_handlecoroutine_handlecoroutine_handlecoroutine_handlecoroutine_handlecoroutine_handlecoroutine_handlecoroutine_handlecoroutine_handlecoroutine_handlecoroutine_handlecoroutine_handlecoroutine_handlecoroutine_handlecoroutine_handlecoroutine_handlecoroutine_handlecoroutine_handlecoroutine_handlecoroutine_handlecoroutine_handlecoroutine_handlecoroutine_handlecoroutine_handlecoroutine_handlecoroutine_handlecoroutine_handlecoroutine_handlecoroutine_handlecoroutine_handlecoroutine_handlecoroutine_handlecoroutine_handlecoroutine_handlecoroutine_handlecoroutine_handlecoroutine_handlecoroutine_handle

```
auto final_suspend() noexcept;
```

Preconditions: A handle referring to the coroutine whose promise object is *this is at the top of *active_ of some generator object x. This function is called by that coroutine upon reaching its final suspend point ([dcl.fct.def.coroutine]).

Returns: An awaitable object of unspecified type whose member functions arrange for the calling coroutine to be suspended, pop the coroutine handle from the top of *x.active_, and resume execution of the coroutine referred to by x.active_->top() if *x.active_ is not empty. If it is empty, control flow returns to the current coroutine caller or resumer ([dcl.fct.def.coroutine]).

Preconditions: A handle referring to the coroutine whose promise object is *this is at the top of * $active_{-}$ of some generator object x.

Returns: An awaitable object of an unspecified type ([expr.await]) that stores an object of type remove_cvref_t<yielded> direct-non-list-initialized with lval, whose member functions arrange for value_ to point to that stored object and then suspend the coroutine.

Throws: Any exception thrown by the initialization of the stored object.

Remarks: A *yield-expression* that calls this function has type void ([expr.yield]).

```
template<class T2, class V2, class Alloc2, class Unused>
  requires same_as<typename generator<T2, V2, Alloc2>::yielded, yielded>
  auto yield_value(ranges::elements_of<generator<T2, V2, Alloc2>&&, Unused> g) noexcept;
```

Preconditions: A handle referring to the coroutine whose promise object is *this is at the top of *active_ of some generator object x. The coroutine referred to by g.range .coroutine_ is suspended at its initial suspend point.

Returns: An awaitable object of an unspecified type ([expr.await]) which takes ownership of the generator <code>g.range</code>, whose member <code>await_ready</code> returns false, whose member <code>await_suspend</code> pushes <code>g.range.coroutine_into*x.active_and</code> resumes execution of the coroutine referred to by <code>g.range.coroutine_</code>, and whose member <code>await_resume</code> evaluates rethrow_exception(<code>except_</code>) if bool(<code>except_</code>) is true. If bool(<code>except_</code>) is false, the <code>await_resume</code> member has no effects.

Remarks: A *yield-expression* that calls this function has type void ([expr.yield]).

Remarks: A *yield-expression* that calls this function has type void ([expr.yield]).

void unhandled_exception();

Preconditions: A handle referring to the coroutine whose promise object is *this is at the top of * $active_{-}$ of some generator object x.

Effects: If the handle referring to the coroutine whose promise object is *this is the sole element of *x.active_, equivalent to: throw; Otherwise, assigns current_exception() to except_.

Let A be

- Allocator, if it is not void,
- Alloc for the overloads with a template parameter Alloc, or
- allocator<void> otherwise.

Let B be allocator_traits<A>::template rebind_alloc<U> where U is an unspecified type whose size and alignment are both __STDCPP_DEFAULT_NEW_ALIGNMENT__.

Mandates: allocator_traits::pointer is a pointer type.

Effects: Initializes an allocator b of type B with A(alloc), for the overloads with a function parameter alloc, or with A() otherwise. Uses b to allocate storage for the smallest array of U sufficient to provide storage for a coroutine state of size size, and unspecified additional state necessary to ensure that operator delete can later deallocate this memory block with an allocator equal to b.

Returns: A pointer to the allocated storage.

```
void operator delete(void* pointer, size_t size) noexcept;
```

Preconditions: pointer was returned from an invocation of one of the above overloads of operator new with a size argument equal to size.

Effects: Deallocates the storage pointed to by pointer using an allocator equivalent to that used to allocate it.

Class template generator::iterator

[coroutine.generator.iterator]

```
namespace std {
  template<class Ref, class V, class Allocator>
  class generator<Ref, V, Allocator>::iterator {
  public:
     using value_type = value;
     using difference_type = ptrdiff_t;

     iterator(iterator&& other) noexcept;

     iterator& operator=(iterator&& other) noexcept;

     reference operator*() const noexcept(is_nothrow_copy_constructible_v<reference>);

     iterator& operator++();
     void operator++(int);

     friend bool operator==(iterator i, default_sentinel_t);

private:
     coroutine_handle<promise_type> coroutine_; // exposition only
  };
}
```

```
iterator(iterator&& other) noexcept;
     Effects: Initializes coroutine_ with exchange(other.coroutine_, {}).
iterator& operator=(iterator&& other) noexcept;
     Effects: Equivalent to coroutine_ = exchange(other.coroutine_, {}).
     Returns: *this.
reference operator*() const noexcept(is_nothrow_copy_constructible_v<reference>);
     Preconditions: For some generator object x, coroutine_ is in *x.active_ and x.active_-
     >top() refers to a suspended coroutine with promise object p.
     Effects: Equivalent to: return static_cast<reference>(*p.value_);
iterator& operator++();
     Preconditions: For some generator object x, coroutine_ is in *x.active_.
     Effects: Equivalent to x.active_->top().resume().
     Returns: *this.
void operator++(int);
     Effects: Equivalent to ++*this.
friend bool operator==(iterator i, default_sentinel_t);
     Effects: Equivalent to: return i.coroutine_.done();
```

References

- [1] Casey Carter. P1456R1: Move-only views. https://wg21.link/p1456r1, 11 2019.
- [2] Corentin Jabot. P1207R0: Movability of single-pass iterators. https://wg21.link/p1207r0, 8 2018.
- [3] Corentin Jabot and Lewis Baker. P2168R3: generator: A synchronous coroutine generator compatible with ranges. https://wg21.link/p2168r3, 4 2021.
- [4] Corentin Jabot, Eric Niebler, and Casey Carter. P1206R3: ranges::to: A function to convert any range to a container. https://wg21.link/p1206r3, 11 2020.
- [5] Thomas Köppe. N4910: Working draft, standard for programming language c++. https://wg21.link/n4910, 3 2022.
- [6] Gor Nishanov. P1681R0: Revisiting allocator model for coroutine lazy/task/generator. https://wg21.link/p1681r0, 6 2019.
- [7] Barry Revzin and Tim Song. P2415R2: What is a view? https://wg21.link/p2415r2, 10 2021.
- [8] Richard Smith and Gor Nishanov. P0981R0: Halo: coroutine heap allocation elision optimization: the joint response. https://wg21.link/p0981r0, 3 2018.
- [9] Mathias Stearn. P2529R0: generator should have T&& reference_type. https://wg21.link/p2529r0, 1 2022.
- [CppCoro] Lewis Baker. CppCoro: A library of C++ coroutine abstractions for the coroutines TS. https://github.com/lewissbaker/cppcoro
- [Folly] Facebook. Folly: An open-source C++ library developed and used at Facebook. https://github.com/facebook/folly
- [range] Eric Niebler. range-v3 Range library for C++14/17/20. https://github.com/ericniebler/range-v3
- [Implementation] Casey Carter, Lewis Baker, Corentin Jabot. std::generator implementation. https://godbolt.org/z/5hcaPcfvP