# An allocator is a handle to a heap

Lessons learned from std::pmr

#### **Outline**

- What are objects? What are values? [3–11]
- An Allocator is a handle to a MemoryResource [12–19]
- An Allocator is a "copy-only" type [20–23]
- An Allocator belongs to a rebindable family [24–30]
- An Allocator is more than a handle to a MemoryResource [31–45]
- Relating allocators to other parts of C++ [46–54]
- Examples and bonus slides [55–61]

Hey look! Slide numbers!

#### What is an object?

- An object, unlike a (pure) value, has an address.
- Address, pointer, name, unique identifier, handle all synonymous for our purposes.



Object a of type int

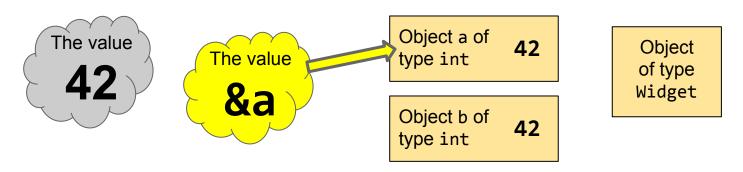
Object b of type int

42

Object of type Widget

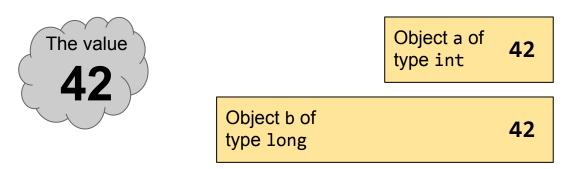
#### What is an object?

- An object, unlike a (pure) value, has an address.
- Address, pointer, name, unique identifier, handle all synonymous for our purposes.
- The name of an object *is itself* a value.



#### What is an object?

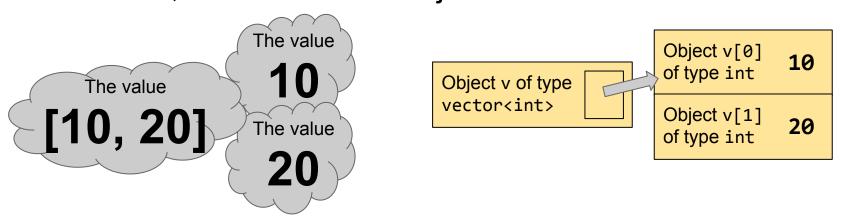
Where it gets confusing (for me at least): A C++ object is defined in part by its **in-memory representation**. And there is some sense in which some kinds of objects can "have" a value at any given moment.



#### What is a (sequence) container?

A container is a **value**, containing sub-values, which are called its elements.

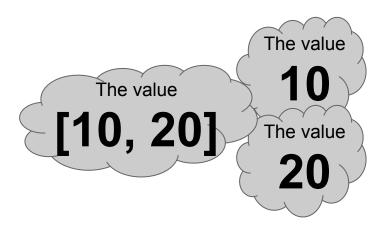
A container is an **object**, that holds and manages its elements, which are also objects.

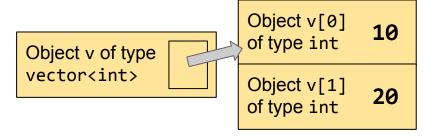


#### What is an allocator?

The classic C++ allocator model answers the questions implied by the object diagram on the right.

- Where does the memory for v[i] come from?
- What is the thing represented by in our diagram?





#### What is an allocator?

std::vector is parameterized on an allocator type A as well as on T.

- Where does the memory for v[i] come from?
  - It comes from A::allocate(n).
- What is the thing represented by in our diagram?
  - It's an object of type A::pointer.

The container holds an *instance* of the allocator type A within itself. Anything that can be funneled through that instance, is funneled.

What can we put inside the allocator instance?

#### What goes into an allocator?

The only allocator type in C++03 / 11 / 14 is std::allocator. std::allocator is stateless.

This led to people's trying to implement the wrong kind of stateful allocators.

```
template<class T>
struct Bad {
    alignas(16) char data[1024]; // "state" = big buffer to allocate from
    size_t used = 0;
    T *allocate(size_t n) {
        auto k = n*sizeof(T); used += k;
        return (T*)(data + used - k);
    }
}
```

#### std::pmr::polymorphic\_allocator

C++17 adds std::pmr::polymorphic\_allocator.

- Basically a pointer to a std::pmr::memory\_resource.
- All the shared state goes into the memory\_resource.

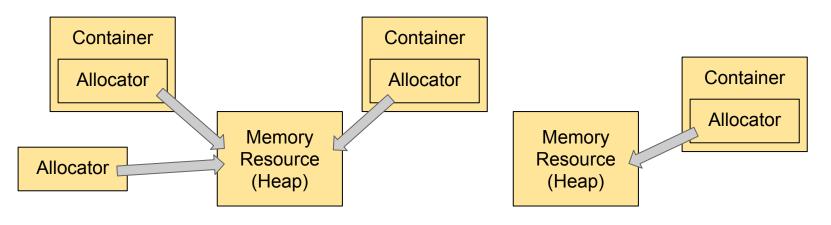
```
template<class T>
struct TrivialResource : std::pmr::memory_resource {
    alignas(16) char data[1024]; // "state" = big buffer to allocate from
    size_t used = 0;
    T *allocate(size_t n) {
        auto k = n*sizeof(T); used += k;
        return (T*)(data + used - k);
    }
};
TrivialResource<int, 10> mr;
std::vector<int, polymorphic_allocator<int>> fcvec(&mr);
```

### "Object-like" bad; "value-like" good

struct Bad	std::pmr::polymorphic_allocator
Object-like, contains mutable state	Value-like, contains only immutable state
"Source of memory" is stored directly in the object	"Source of memory" is shared among all copies of the allocator with the same <i>value</i>
allocate and deallocate are non-const	allocate and deallocate could be const
Moving/copying the allocator object is likely to cause bugs	Moving/copying the allocator is safe and even encouraged
There is no shared state	Need to think about lifetime of the shared state

#### Clarifies our thinking about allocators

- Old-style thinking: "an allocator object represents a source of memory" — WRONG!
- New-style thinking: "an allocator value represents a handle to a source of memory (plus some other, orthogonal pieces)."



#### But what about stateless allocators?

A stateless allocator [e.g. std::allocator<T>] represents a handle to a source of memory (plus some orthogonal pieces) where the source of memory is a global singleton [e.g. the new/delete heap].

- A datatype with k possible values needs only log<sub>2</sub> k bits.
- A pointer to a global singleton (with 1 possible value) needs log<sub>2</sub> 1 = 0 bits.

```
class memory resource {
public:
  void *allocate(size_t bytes, size_t align = alignof(max_align_t)) {
      return do_allocate(bytes, align);
  void deallocate(void *p, size_t bytes, size_t align = alignof(max_align_t)) {
      return do deallocate(p, bytes, align);
   bool is_equal(const memory_resource& rhs) const noexcept {
      return do is equal(rhs);
  virtual ~memory resource() = default;
private:
  virtual void *do allocate(size t bytes, size t align) = 0;
  virtual void do_deallocate(void *p, size_t bytes, size_t align) = 0;
  virtual bool do_is_equal(const memory_resource& rhs) const noexcept = 0;
};
bool operator == (const memory resource& a, const memory resource& b) noexcept {
   return (&a == &b) || a.is equal(b);
```

#### Standard new\_delete\_resource()

```
class singleton_new_delete_resource : public memory_resource {
  void *do allocate(size t bytes, size t align) override {
       return ::operator new(bytes, std::align val t(align));
  void do_deallocate(void *p, size_t bytes, size_t align) override {
       ::operator delete(p, bytes, std::align_val_t(align));
  bool do_is_equal(const memory_resource& rhs) const noexcept override {
       return (this == &rhs);
inline memory_resource *new_delete_resource() noexcept {
  static singleton new delete resource instance;
  return &instance;
```

```
template<class T> class polymorphic allocator {
                                                                                 This stateful allocator
   memory resource *m mr;
                                                                                        of size
public:
                                                                                 64 bits can "point to"
   using value type = T;
                                                                                   any of 2<sup>64</sup> different
   polymorphic allocator(memory resource *mr) : m mr(mr) {}
                                                                                  memory resources.
   template<class U>
   explicit polymorphic allocator(const polymorphic allocator<U>& rhs) noexcept
                                     { m mr = rhs.resource(); }
   polymorphic allocator()
                                     { m mr = get default resource(); }
   memory resource *resource() const { return m mr; }
   T *allocate(size t n) { return (T*)(m mr->allocate(n * sizeof(T), alignof(T))); }
   void deallocate(T *p, size t n) { m mr->deallocate((void*)(p), n * sizeof(T), alignof(T)); }
   polymorphic allocator select on container copy construction() const
                                     { return polymorphic allocator(); }
};
template<class A, class B>
bool operator == (const polymorphic allocator < A > & a, const polymorphic allocator < B > & b) noexcept
{ return *a.resource() == *b.resource(); }
```

```
static atomic refcounted ptr<memory resource> s table[256];
                                                                                This stateful allocator
                                                                                      of size
template<class T> class one byte allocator {
                                                                                8 bits can "point to"
  uint8 t m index = 0;
                                                                                any of 256 different
public:
  using value type = T;
                                                                                memory resources.
  one byte allocator() = delete;
  one byte allocator(memory_resource *mr) {
    /* find an empty slot in s table, insert mr, and inc ref it */
  one byte allocator(const one byte allocator% rhs) noexcept
                                    { m index = rhs.m index; s table[m index].inc ref(); }
  ~one byte allocator() { s table[m index].dec ref(); }
  template<class U>
  explicit one byte allocator(const one byte allocator<U>& rhs) noexcept
                                    { m index = rhs.m index; s table[m index].inc ref(); }
  memory resource *mr() const { return s table[m index].get(); }
  T *allocate(size t n) { return (T*)(mr()->allocate(n * sizeof(T), alignof(T))); }
  void deallocate(T *p, size t n) { mr()->deallocate((void*)(p), n * sizeof(T), alignof(T)); }
                                                                                                   17
};
```

```
template<class T> class zero byte allocator {
   /* no state */
public:
  using value type = T;
   zero byte allocator() = default;
  template<class U>
   explicit zero byte allocator(const zero byte allocator<U>& rhs) noexcept
  memory resource *mr() const { return std::pmr::new delete resource(); }
  T *allocate(size t n) { return (T*)(mr()->allocate(n * sizeof(T), alignof(T))); }
  void deallocate(T *p, size t n) { mr()->deallocate((void*)(p), n * sizeof(T), alignof(T)); }
};
```

This stateless allocator of size **0 bits** can "point to" any of 1 different memory resources.

This is essentially std::allocator!

#### Corollaries to the new way of thinking

- Allocator types should be copyable, just like pointers.
  - This was always true but now it's more obvious.
- Allocator types should be cheaply copyable, like pointers.
  - They need not be trivially copyable.
- Memory-resource types should generally be immobile.
  - A memory resource might allocate chunks out of a buffer stored inside itself as a data member.

## Allocators must "do as the pointers do" in one more subtle way...

```
static atomic refcounted ptr<memory resource> s table[256];
                                                                                 Did you notice that
                                                                                this allocator type is
template<class T> class one byte allocator {
  uint8 t m index = 0;
                                                                                  copyable but not
public:
                                                                                    (efficiently)
  using value type = T;
                                                                                    moveable?
  one byte allocator() = delete;
  one byte allocator(memory resource *mr) {
    /* find an empty slot in s table, insert mr, and inc ref it */
  one byte allocator(const one byte allocator% rhs) noexcept
                                    { m_index = rhs.m_index; s_table[m_index].inc_ref(); }
  ~one byte allocator() { s table[m index].dec ref(); }
  template<class U>
  explicit one byte allocator(const one byte allocator<U>& rhs) noexcept
                                    { m index = rhs.m index; s table[m index].inc ref(); }
  T *allocate(size t n) { return (T*)(mr()->allocate(n * sizeof(T), alignof(T))); }
  void deallocate(T *p, size t n) { mr()->deallocate((void*)(p), n * sizeof(T), alignof(T)); }
  memory resource *mr() const { return s table[m index].get(); }
};
```

#### Allocators must be "copy-only" types

This was LWG issue 2593.

Expression	Return type	Assertion/note/pre-/post-condition
X u(a); X u = a;		Shall not exit via an exception.  Postconditions: u == a.
<pre>X u(std::move(a)); X u = std::move(a);</pre>		Shall not exit via an exception.  Postconditions: The value of a is unchanged and is equal to u.

```
vector<int, A<int>> v1;
vector<int, A<int>> v2 = std::move(v1); // If we move-from A...
v1.clear(); v1.push_back(42); // ...this will allocate() using the moved-from A!
```

#### Allocators must be "copy-only" types

Given that we are not allowed to move allocators any more cheaply than we copy them, how worried do we have to be about the *cost* of a copy?

I wrote a little test to find out. https://wandbox.org/permlink/mHrj7Y55k3Gqu4Q5

Expression	Allocator copies+moves on libc++ (— if stateful)	Allocator copies+moves on libstdc++ (— if stateful)
list b(a);	2+2 — 2+2	1+1 — 1+1
<pre>list b = move(a);</pre>	1+2 — 1+2	0+1 — 0+1
vector b(a);	2+0 — 2+0	2+0 — 2+0
<pre>vector b = move(a);</pre>	1+0 — 1+0	0+1 — 0+1

## We see these extra copies/moves due to *rebinding*.

#### Allocators are "rebindable family" types

The type-to-allocate is baked into the allocator type.

- Alloc<int>{}.allocate(2) means "allocate 2 ints."
- This works great for std::vector and only for std::vector.
- std::list<T, Alloc<T>> wants to allocate \_\_node<T>, not T.
- std::map<K, V, Alloc<pair<const K, V>>> certainly doesn't want to allocate that messy type!

Every "allocator type" is really a whole family of related types: Alloc<int> and Alloc<double> and Alloc<void\*\*\*> and so on. An allocator value which is representable in *one* of the family's types must be representable in *all* of its types.

#### Allocators are "rebindable family" types

Rebinding is useful whenever your generic algorithm requires a "Foo of T," where *you* provide the T(s) but your *user* provides the (single) Foo.

```
// Rebinding, preferred by the STL
template<class T, class A_of_T>
class myContainer {
    using actualAllocator = allocator_traits<A_of_T>::rebind_alloc<U>;
};

// Template template parameters, unused in the STL
template<class T, template<class> class A>
class myContainer {
    using actualAllocator = A<U>;
};
```

#### Allocators are "rebindable family" types

Other "rebindable families" in C++:

- Allocator types
  - o allocator\_traits<Alloc<T>>::rebind\_alloc<U> == Alloc<U>
- Pointer types
  - o pointer\_traits<Ptr<T>>::rebind<U> == Ptr<U>
- Smart-pointer types
  - o decltype(reinterpret\_pointer\_cast<U>(Sptr<T>{})) == Sptr<U>
- Promise and future types
  - o decltype(Future<T>{}.then(
     [](auto&&... u) -> U { throw; })) == Future<U>

#### Each "rebindable family" has a prototype

or "representative" of the equivalence class —

- Pointer and smart-pointer families have a "void pointer" type
  - Ptr<void>, Sptr<void>
- Allocator families have\* a "proto-allocator" type
  - o Alloc<void>
- Promise and future types have a "future of void" type
  - o Future<void>

<sup>\* —</sup> Caveat. This "proto-allocator" concept is in the Networking and Executors TSes but not yet in the Standard. And somebody was recently agitating for the proto-allocator type to be spelled Alloc<std::byte> instead of Alloc<void>! But I hope that'll get cleared up soon.

#### If we were designing the STL today...

...we'd use the "proto-allocator" type in our interfaces to save on instantiations.

#### **Current STL:**

```
std::vector<int, Alloc<int>>
    std::list<int, Alloc<int>>
    std::map<int, int, Alloc<pair<const int, int>>>

Better, DRYer STL:
    std::vector<int, Alloc<void>>
    std::list<int, Alloc<void>>
    std::map<int, int, Alloc<void>>
```

#### We seem to be circling back to std::pmr

With the "proto-allocator" type, the container must rebind its allocator before any operation:

```
ProtoAlloc m_alloc = ...;
    using AllocT = allocator_traits<ProtoAlloc>::rebind alloc<T>;
    auto ptr = allocator traits<AllocT>::allocate(
        static cast<AllocT>(m_alloc), capacity
    );
So why don't we just standardize on this, instead?
    memory resource *m res = ...;
    auto ptr = m res->allocate(capacity * sizeof(T));
```

### Because an Allocator is not *merely* a pointer to a MemoryResource

#### Allocator > source of memory

ptr will be of type allocator\_traits<AllocT>::pointer

- probably T\*
- but could be something fancier, such as boost::interprocess::offset\_ptr<T>

It's completely up to the allocator to decide how its pointers are represented!

#### Allocator > source of memory

ptr will be of type allocator\_traits<AllocT>::pointer

- probably T\*
- but could be something fancier, such as boost::interprocess::offset\_ptr<T>

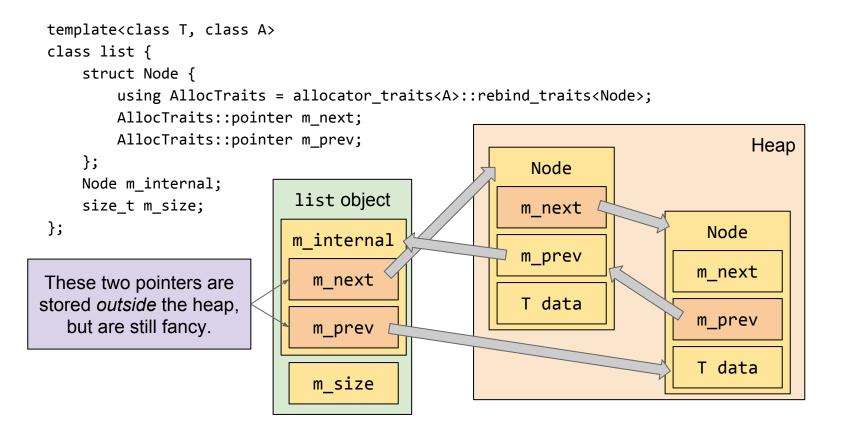
It's completely up to the allocator to decide how its pointers are represented!

But "its pointers" is awfully vague...

#### Container uses pointer for all allocations

```
template<class T, class A>
class list {
    struct Node {
       using AllocTraits = allocator_traits<A>::rebind_traits<Node>;
       AllocTraits::pointer m next;
       AllocTraits::pointer m prev;
                                                                               Heap
   };
                                                      Node
   Node m internal;
                           list object
    size_t m_size;
                                                     m_next
};
                                                                          Node
                           m internal
                                                     m_prev
                                                                         m_next
                             m_next
                                                     T data
                                                                         m_prev
                             m_prev
                                                                         T data
                             m_size
```

#### Container uses pointer for all allocations



#### Container uses pointer for all allocations

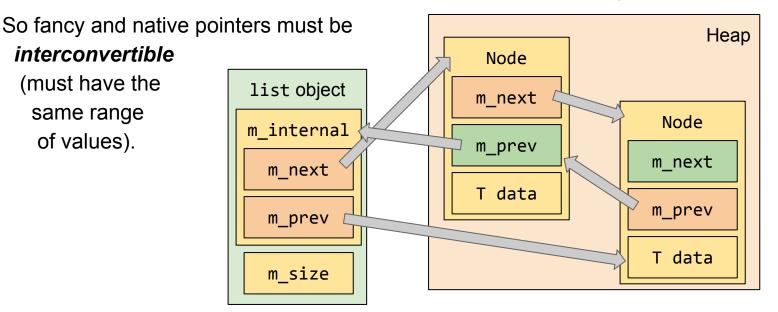
```
template<class T, class A>
 class list {
     struct Node {
         using AllocTraits = allocator_traits<A>::rebind_traits<Node>;
         AllocTraits::pointer m next;
         AllocTraits::pointer m prev;
                                                                                  Heap
    };
                                                        Node
     Node m internal;
                             list object
     size t m size;
                                                       m_next
                                                                             Node
                             m internal
                                                       m_prev
                                                                           m_next
                               m next
  These pointers are
                                                       T data
stored within the heap,
                                                                           m_prev
but point to objects living
                               m prev
   outside the heap.
                                                                            T data
                               m_size
```

## Fancy pointers' range = raw pointers' range

We must be able to convert fancy pointer m next into native reference \*m next, and then into native pointer this.

We must be able to convert native pointer &m internal into fancy pointer m prev.

interconvertible (must have the same range of values).



## So are fancy pointers just native pointers?

"Interconvertible" = same possible values. So are they the same type?

## So are fancy pointers just native pointers?

"Interconvertible" = same possible values. So are they the same type?

- No, because C++ type also involves object representation.
  - Boost offset ptr
  - Bob Steagall's "synthetic pointers"
  - C++ conflates valueish and objectish attributes: good idea or bad idea?

## So are fancy pointers just native pointers?

"Interconvertible" = same possible values. So are they the same type?

- No, because the fancy type might be augmented with extra data.
  - "Segmented" pointers carry metadata (e.g. slab number) for the deallocator
  - "Fat" pointers carry metadata (e.g. array bounds) used during dereferences
  - Vendor extensions support these cases only inconsistently / accidentally.
     P0773R0 suggests that they should be supported by the Standard.

Runtime source of memory (i.e., handle to a memory resource)

- Runtime source of memory (i.e., handle to a memory resource)
- Compile-time decider of the pointer type

- Runtime source of memory (i.e., handle to a memory resource)
- Compile-time decider of the pointer type
- Compile-time decider whether the source of memory should move with the container *value*, or stick to the original container *object* (POCCA, POCMA, POCS)
  - Bob calls this "lateral propagation"; I call it "stickiness"

- Runtime source of memory (i.e., handle to a memory resource)
- Compile-time decider of the pointer type
- Compile-time decider whether the source of memory should move with the container *value*, or stick to the original container *object* (POCCA, POCMA, POCS)
  - Bob calls this "lateral propagation"; I call it "stickiness"
- Runtime decider of how containers' sub-objects (elements) should be constructed (allocator\_traits::construct)
  - Bob calls this "vertical propagation"; I call it "scoped\_allocator\_adaptor is why we can't have nice things"
  - My other talk proposes traits::has\_trivial\_construct\_and\_destroy

## A handle to a heap plus orthogonal pieces

- Runtime source of memory (i.e., handle to a memory resource)
- Compile-time decider of the pointer type
- Compile-time decider whether the source of memory should move with the container value, or stick to the original container object (POCCA, POCMA, POCS)
  - Bob calls this "lateral propagation"; I call it "stickiness"
- Runtime decider of how containers' sub-objects (elements) should be constructed (allocator\_traits::construct)
  - Bob calls this "vertical propagation"; I call it "scoped\_allocator\_adaptor is why we can't have nice things"
  - My other talk proposes traits::has\_trivial\_construct\_and\_destroy

## We might separate some of these pieces

#### Imagine:

- nonsticky\_allocator\_adaptor (changing POCCA/POCMA without affecting source-of-memory)
- fancy\_allocator\_adaptor (changing pointer representation or fatness without affecting source-of-memory)
  - Boost.Interprocess has an allocator that deals in offset\_ptr, but it gets its memory from a segment\_manager. Getting offset\_ptrs from an arbitrary heap wouldn't be useful here.
  - Grafting fat pointers onto an arbitrary heap sounds potentially useful.

# How are analogous "handle" types handled in other areas of C++?

### Other "a Y is a handle to an X" in C++

An **allocator** is a cheaply copyable handle to a **memory resource** 

Plus some other bits, like pointer typedef and stickiness.

An **iterator** is a cheaply copyable handle to a container's contents

Plus some other bits, like iteration-direction (reverse\_iterator).

#### Other "a Y is a handle to an X" in C++

An **allocator** is a cheaply copyable handle to a **memory resource** 

Plus some other bits, like pointer typedef and stickiness.

An **iterator** is a cheaply copyable handle to a container's contents

Plus some other bits, like iteration-direction (reverse\_iterator).

Boost has iterator\_facade and iterator\_adaptor.

iterator\_facade means "you implement a complete set of primitive functions (and data members); then inherit from iterator\_facade to provide the standard zoo of operators."

iterator\_adaptor means "you override some but not all the primitive functions; then inherit from iterator\_adaptor<Base> to provide the missing parts."

### Other "a Y is a handle to an X" in C++

An **allocator** is a cheaply copyable handle to a **memory resource** 

Plus some other bits, like pointer typedef and stickiness.

An **iterator** is a cheaply copyable handle to a container's contents

Plus some other bits, like iteration-direction (reverse\_iterator).

P0443: An **executor** is a cheaply copyable handle to an **execution context** 

Plus some other bits, like bulkness.

## An executor is a handle (Executors TS)

P0443r5, also P0737r0

The context\_t property ... returns the **execution context** associated with the **executor**. An **execution context** is a program object that represents a specific collection of execution resources and the execution agents that exist within those resources. Execution agents are units of execution ...

P0443 proposes static\_thread\_pool as an example of an ExecutionContext.

### P0443 std::executor is like std::function

P0443 proposes std::executor as an example of an Executor.

std::executor is "polymorphic" in that it can hold (a shared copy of) any kind of Executor, period. This is like std::function or std::any.

OTOH, std::pmr::polymorphic\_allocator can hold merely (a non-owning pointer to) a concrete std::pmr::memory\_resource, which is just one model of the MemoryResource concept.

## "Truly polymorphic allocator"

```
template<class T>
class executor style allocator {
    shared ptr<memory resource> sptr;
public:
   using value type = T;
   template<class Alloc, class = enable if t<...>>
   executor style allocator(Alloc a) :
        sptr(std::make_shared<resource_adaptor<Alloc>>(std::move(a))) {}
    T *allocate(size t n) const {
       return (T*)sptr->allocate(n * sizeof(T), alignof(T));
   void deallocate(T *p, size_t n) const {
        sptr->deallocate((void*)p, n * sizeof(T), alignof(T));
```

## "Truly polymorphic allocator"

```
template<class T>
class executor style allocator {
    shared_ptr<memory_resource> sptr;
                                                                  Lib Fundamentals TS.
public:
                                                                    somehow omitted
   using value_type = T;
                                                                      from C++17
   template<class Alloc, class = enable_if_t<...>>
    executor style allocator(Alloc a) :
        sptr(std::make shared<resource adaptor<Alloc>>(std::move(a))) {}
    T *allocate(size t n) const {
        return (T*)sptr->allocate(n * sizeof(T), alignof(T));
    void deallocate(T *p, size_t n) const {
        sptr->deallocate((void*)p, n * sizeof(T), alignof(T));
```

## "Shutdown-safe allocator" (?)

```
template<class T>
class shutdown safe allocator {
   weak ptr<memory resource> wptr;
public:
   using value type = T;
    shutdown safe allocator(shared ptr<memory resource> mr) :
        wptr(mr) {}
   T *allocate(size t n) const {
        auto sptr = wptr->lock(); if (!sptr) throw bad alloc();
        return (T*)sptr->allocate(n * sizeof(T), alignof(T));
   void deallocate(T *p, size_t n) const {
        auto sptr = wptr->lock(); if (!sptr) return;
        sptr->deallocate((void*)p, n * sizeof(T), alignof(T));
```

## "Shutdown-quick allocator"

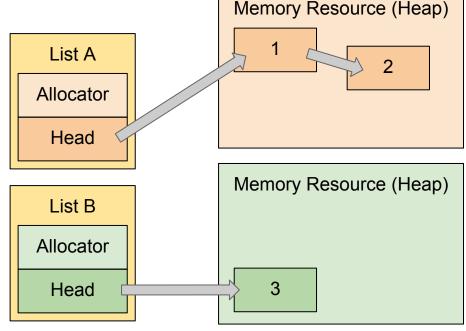
```
template<class T>
class shutdown quick allocator {
   memory resource *mr;
public:
   using value_type = T;
    shutdown_quick_allocator(memory_resource *mr) :
        mr(mr) {}
    T *allocate(size t n) const {
        return (T*)mr->allocate(n * sizeof(T), alignof(T));
    void deallocate(T *p, size_t n) const {
        if (ShuttingDownNow) return;
        mr->deallocate((void*)p, n * sizeof(T), alignof(T));
```



Propagating and/or swapping stateful allocators is still

```
broken in C++17; don't do it.
```

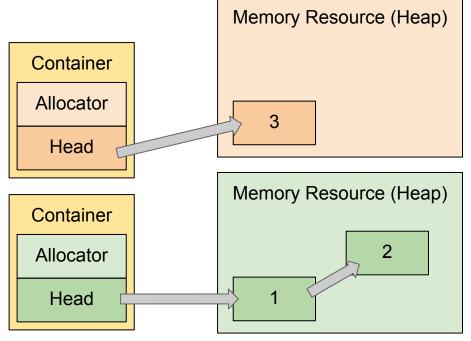
```
inline_buffer_resource<99> mr1;
inline_buffer_resource<99> mr2;
std::pmr::list<int> A{&mr1};
std::pmr::list<int> B{&mr2};
a.assign({1, 2});
b.assign({3});
swap(a, b);
```



Propagating and/or swapping stateful allocators is still

```
broken in C++17; don't do it.
```

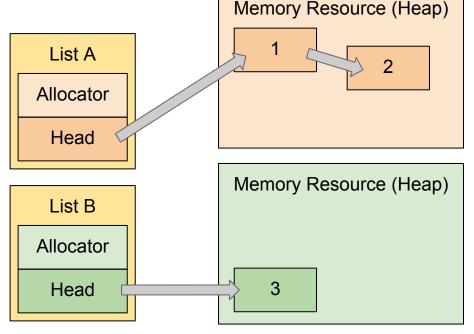
```
inline_buffer_resource<99> mr1;
inline_buffer_resource<99> mr2;
std::pmr::list<int> A{&mr1};
std::pmr::list<int> B{&mr2};
a.assign({1, 2});
b.assign({3});
swap(a, b);
```



Propagating and/or swapping stateful allocators is still

```
broken in C++17; don't do it.
```

```
inline_buffer_resource<99> mr1;
inline_buffer_resource<99> mr2;
std::pmr::list<int> A{&mr1};
std::pmr::list<int> B{&mr2};
a.assign({1, 2});
b.assign({3});
swap(a, b);
```



Propagating and/or swapping stateful allocators is still

broken in C++17; don't do it.

```
inline_buffer_resource<99> mr1;
inline_buffer_resource<99> mr2;
std::pmr::list<int> A{&mr1};
std::pmr::list<int> B{&mr2};
a.assign({1, 2});
b.assign({3});
swap(a, b);
```

