#### 



# Meld: Exploring the feasibility of a framework-less framework

Kyle J. Knoepfel

11 May 2023

26<sup>th</sup> International Conference on Computing in High Energy & Nuclear Physics

#### What type of framework does DUNE need?



## What type of framework does DUNE need?

#### What DUNE has:

DUNE's current framework (art) originates from a collider-physics experiment, steeped in event-based concepts.

#### • But:

The "event" is not always a helpful concept for neutrino experiments.

• What DUNE needs...



#### What type of framework does DUNE need?

#### • What DUNE has:

DUNE's current framework (art) originates from a collider-physics experiment, steeped in event-based concepts.

#### • But:

The "event" is not always a helpful concept for neutrino experiments.

What DUNE needs...

**DUNE** Offline Computing

Conceptual Design Report

https://doi.org/10.48550/arXiv.2210.15665







Physics algorithms should be framework-agnostic.



Physics algorithms should be framework-agnostic.



Fine, assuming it's a requirement for those writing algorithms.



*Physics algorithms should be framework-agnostic.* 



The framework must be able to break apart events into smaller chunks for more granular processing, and then stitch those chunks back together into an event.



Physics algorithms should be framework-agnostic.



The framework must be able to break apart events into smaller chunks for more granular processing, and then stitch those chunks back together into an event.

•••

Okay, tricky but probably doable.



Physics algorithms should be framework-agnostic.



Fine, assuming it's a requirement for those writing algorithms.

The framework must be able to break apart events into smaller chunks for more granular processing, and then stitch those chunks back together into an event.

Okay, tricky but probably doable.

The framework should support "sliding event windows" to provide "edge effect" coverage for extended time readouts during supernovae events.



Physics algorithms should be framework-agnostic.



Fine, assuming it's a requirement for those writing algorithms.

The framework must be able to break apart events into smaller chunks for more granular processing, and then stitch those chunks back together into an event.



Okay, tricky but probably doable.

The framework should support "sliding event windows" to provide "edge effect" coverage for extended time readouts during supernovae events.



Cannot take advantage of statistical independence of events, memory issues, etc.



Physics algorithms should be framework-agnostic.



Fine, assuming it's a requirement for those writing algorithms.

The framework must be able to break apart events into smaller chunks for more granular processing, and then stitch those chunks back together into an event.



Okay, tricky but probably doable.

The framework should support "sliding event windows" to provide "edge effect" coverage for extended time readouts during supernovae events.



Cannot take advantage of statistical independence of events, memory issues, etc.

The framework should make minimal assumptions about the data model.



Physics algorithms should be framework-agnostic.



Fine, assuming it's a requirement for those writing algorithms.

The framework must be able to break apart events into smaller chunks for more granular processing, and then stitch those chunks back together into an event.



Okay, tricky but probably doable.

The framework should support "sliding event windows" to provide "edge effect" coverage for extended time readouts during supernovae events.



Cannot take advantage of statistical independence of events, memory issues, etc.

The framework should make minimal assumptions about the data model.



That sounds like a framework-less framework...



• How many of art's assumptions can be relaxed/removed to meet DUNE's needs?



- How many of art's assumptions can be relaxed/removed to meet DUNE's needs?
- Asking this question has resulted in a 2-year project called **Meld**, a laboratorydirected R&D project based at Fermilab.
- The goal is to explore options, not necessarily to provide software.



- How many of art's assumptions can be relaxed/removed to meet DUNE's needs?
- Asking this question has resulted in a 2-year project called Meld, a laboratorydirected R&D project based at Fermilab.
- The goal is to explore options, not necessarily to provide software.

Meld has been heavily influenced by:

Regular discussions with DUNE experts

Existing framework capabilities and limitations

Functional programming (e.g. Haskell)

Mathematics (set, graph, and category theory)



- How many of art's assumptions can be relaxed/removed to meet DUNE's needs?
- Asking this question has resulted in a 2-year project called Meld, a laboratorydirected R&D project based at Fermilab.
- The goal is to explore options, not necessarily to provide software.

 Meld has been heavily influenced by: Regular discussions with DUNE experts
 Existing framework capabilities and limitations
 Functional programming (e.g. Haskell)
 Mathematics (set, graph, and category theory)

#### **Prerequisites**

Support user-provided algorithms written in C++20 or newer

Design for concurrency

Favor community-provided software



#### The following discussion describes a logical organization of data.

It does not imply a specific in-memory representation of data.



# Looking at the data (set)











21 5/11/23 Kyle J. Knoepfel I Meld @ CHEP 2023





















# Looking at the data (product mappings)



























An operation that converts a sequence of elements (a)<sub>8</sub> to a sequence of elements (b)<sub>8</sub> of the same length using a function f:





An operation that converts a sequence of elements (a)<sub>8</sub> to a sequence of elements (b)<sub>8</sub> of the same length using a function f:

This is a map or transform.





An operation that converts a sequence of elements  $(a)_8$  to a sequence of elements  $(b)_8$  of the same length using a function f:

This is a map or transform.

• An operation that converts a sequence of elements  $(c)_8$  to a shorter sequence of elements  $(K)_4$  at a higher level of nesting, using a function  $g_0$ :





An operation that converts a sequence of elements (a)<sub>8</sub> to a sequence of elements (b)<sub>8</sub> of the same length using a function f:

This is a map or transform.

An operation that converts a sequence of elements (c)<sub>8</sub> to a shorter sequence of elements (K)<sub>4</sub> at a higher level of nesting, using a function g<sub>0</sub>:

This is a fold or reduction.





An operation that converts a sequence of elements (a)<sub>8</sub> to a sequence of elements (b)<sub>8</sub> of the same length using a function f:

This is a map or transform.

• An operation that converts a sequence of elements  $(c)_8$  to a shorter sequence of elements  $(K)_4$  at a higher level of nesting, using a function  $g_0$ :

This is a fold or reduction.

An operation that pairs element of two sequences
 (*J*)<sub>4</sub> and (*K*)<sub>4</sub> into one sequence (*J*, *K*)<sub>4</sub>:





An operation that converts a sequence of elements (a)<sub>8</sub> to a sequence of elements (b)<sub>8</sub> of the same length using a function f:

This is a map or transform.

An operation that converts a sequence of elements (*c*)<sub>8</sub> to a shorter sequence of elements (*K*)<sub>4</sub> at a higher level of nesting, using a function *g*<sub>0</sub>:

This is a fold or reduction.

An operation that pairs element of two sequences
 (*J*)<sub>4</sub> and (*K*)<sub>4</sub> into one sequence (*J*, *K*)<sub>4</sub>:

This is a zip.







# **Graph of data-product sequences**





## **Graph of data-product sequences**



View	Nodes	Edges	
Data-centric	Data products	Mappings	This work
Map-centric	Mappings	Data products	More common



# **Graph of data-product sequences**



View	Nodes	Edges	
Data-centric	Data products	Mappings	This work
Map-centric	Mappings	Data products	More common

#### The user specifications are the same with either view:

- Which data products to process
- The data set(s) that contain those products (event, etc.)
- Which higher-order function to use (transform, etc.)
- Which user-defined function to serve as the operation to the higher-order function.
- Allowed concurrency of each function.

The focus is just different.



#### How are data products and their mappings supported now?



With art, users do not transparently interact with data products. They instead:

- Implement functions based on datasets (e.g. event)
- "Open" the dataset to retrieve and insert products



## How are data products and their mappings supported now?



With art, users do not transparently interact with data products. They instead:

- Implement functions based on datasets (e.g. event)
- "Open" the dataset to retrieve and insert products

Some of this is historical and due to:

- The object-oriented nature of the framework.
- Technical limitations of C++ whenever the framework was designed.



## How are data products and their mappings supported now?



With art, users do not transparently interact with data products. They instead:

- Implement functions based on datasets (e.g. event)
- "Open" the dataset to retrieve and insert products

Some of this is historical and due to:

- The object-oriented nature of the framework.
- Technical limitations of C++ whenever the framework was designed.

Results in a lot of software mechanics...



• Create tracks from hits for each event.





• Create tracks from hits for each event.



Tracks make\_tracks(Hits const& hits) { ... }



• Create tracks from hits for each event.



Tracks make\_tracks(Hits const& hits) { ... }

```
namespace {
 Tracks make tracks(Hits const& hits) { ... }
}
namespace expt {
  class TrackMaker : public art::SharedProducer {
  public:
   TrackMaker(fhicl::ParameterSet const&) :
      consumes<Hits, art::InEvent>("GoodHits");
      produces<Tracks, art::InEvent>("GoodTracks");
      async<art::InEvent>();
   void produce(art::Event& e,
                 art::ProcessingFrame const&) override
      auto const& hits = e.getProduct<Hits>("GoodHits");
      auto tracks = make tracks(hits);
      e.put(std::make_unique<Tracks>(std::move(tracks)),
            "GoodTracks"):
 };
```

#include "art/Framework/Core/SharedProducer.h"
#include "art/Framework/Principal/Event.h"

DEFINE\_ART\_MODULE(expt::TrackMaker)



art

• Create tracks from hits for each event.



Tracks make\_tracks(Hits const& hits) { ... }



```
#include "art/Framework/Core/SharedProducer.h"
#include "art/Framework/Principal/Event.h"
```

```
namespace {
 Tracks make tracks(Hits const& hits) { ... }
namespace expt {
  class TrackMaker : public art::SharedProducer {
  public:
   TrackMaker(fhicl::ParameterSet const&) :
      consumes<Hits, art::InEvent>("GoodHits");
      produces<Tracks, art::InEvent>("GoodTracks");
      async<art::InEvent>();
   void produce(art::Event& e,
                 art::ProcessingFrame const&) override
      auto const& hits = e.getProduct<Hits>("GoodHits");
      auto tracks = make tracks(hits);
      e.put(std::make_unique<Tracks>(std::move(tracks)),
            "GoodTracks"):
 };
```

DEFINE\_ART\_MODULE(expt::TrackMaker)



art

• Create tracks from hits for each event.



Tracks make\_tracks(Hits const& hits) { ... }

# This is just a transform? 😁

#include "art/Framework/Core/SharedProducer.h"
#include "art/Framework/Principal/Event.h"





DEFINE\_ART\_MODULE(expt::TrackMaker)



• Create tracks from hits for each event.



Tracks make\_tracks(Hits const& hits) { ... }

This is just a transform? 

Nobody wants this.

#include "art/Framework/Core/SharedProducer.h"
#include "art/Framework/Principal/Event.h"



DEFINE\_ART\_MODULE(expt::TrackMaker)



art

• Create tracks from hits for each event.



Tracks make\_tracks(Hits const& hits) { ... }

# A better way...

```
#include "meld/module.hpp"
namespace {
  Tracks make_tracks(Hits const& hits) { ... }
}
DEFINE_MODULE(m, config) {
  m.with(make_tracks)
  .transform("GoodHits").in_each("Event")
  .to("GoodTracks")
```

.using\_concurrency(unlimited);

}



• Create tracks from hits for each event.



Tracks make\_tracks(Hits const& hits) { ... }

# A better way...

```
#include "meld/module.hpp"
namespace {
  Tracks make_tracks(Hits const& hits) { ... }
}
DEFINE_MODULE(m, config) {
  m.with(make_tracks)
   .transform("GoodHits").in_each("Event")
   .to("GoodTracks")
   .using_concurrency(unlimited);
```

}

• Minimal boilerplate.



Create tracks from hits for each event.



Tracks make\_tracks(Hits const& hits) { ... }

# A better way...

```
Meld
#include "meld/module.hpp"
namespace {
 Tracks make_tracks(Hits const& hits) { ... }
DEFINE_MODULE(m, config) {
  m.with(make tracks)
   .transform("GoodHits").in_each("Event"
   .to("GoodTracks")
   .using_concurrency(unlimited);
```

- Minimal boilerplate.
- Event is now a label.





• Create tracks from hits for each event.



Tracks make\_tracks(Hits const& hits) { ... }

## A better way...

```
#include "meld/module.hpp"
namespace {
  Tracks make_tracks(Hits const& hits) { ... }
}
DEFINE_MODULE(m, config) {
  m.with(make_tracks)
  .transform "GoodHits").in_each("Event")
  .to("GoodTracks")
  .using_concurrency(unlimited);
}
```

- Minimal boilerplate.
- Event is now a label.
- Higher-order function is now explicit.



# **Meld implementation**

- <u>https://github.com/knoepfel/meld</u> (not even alpha release)
- Implemented using oneTBB's flow graph

Supported construct	User function		
Transform (Map)	$f(\boldsymbol{a}) \rightarrow \boldsymbol{b}$		
Filter	$f(\boldsymbol{a}) \rightarrow \text{Boolean}$	Standard data-processing idioms	
Monitor	$f(\boldsymbol{a}) \rightarrow \text{Void}$		
Reduction (Fold)	$f_c(a) \to c$	For splitting and then combining events	
Splitter (Unfold)	$f_n(\boldsymbol{a}) \to (\boldsymbol{d})_n$		
Zip	_	For combining arguments to user functions	
Sliding window	_	To do: For sliding over adjacent events	



# Sample hierarchies tested by Meld



art-based hierarchy

Performance numbers are preliminary



# Sample hierarchies tested by Meld





# Sample hierarchies tested by Meld







*"Ways change, Stil."* —Paul from *Dune* by Frank Herbert

• Supporting DUNE's framework needs suggests rethinking framework concepts.





*"Ways change, Stil."* —Paul from *Dune* by Frank Herbert

- Supporting DUNE's framework needs suggests rethinking framework concepts.
- Meld seeks to address these needs by considering a framework job as a

(1) graph of data products connected by
 (2) user-provided operations of
 (3) higher-order functions.

- It is not a framework-less framework, but it *is* less framework coupling.
- Preliminary work indicates this is a productive avenue to pursue.





*"Ways change, Stil."* —Paul from *Dune* by Frank Herbert

- Supporting DUNE's framework needs suggests rethinking framework concepts.
- Meld seeks to address these needs by considering a framework job as a

(1) graph of data products connected by
 (2) user-provided operations of
 (3) higher-order functions.

- It is not a framework-less framework, but it *is* less framework coupling.
- Preliminary work indicates this is a productive avenue to pursue.

# Thank you for your time and attention.



#### **Backup slides**



# **Accessing provenance information**

```
#include "meld/module.hpp"
 namespace {
  Tracks make tracks(Hits const& hits) { ... }
-
  Tracks make_tracks(meld::handle<Hits> hits) { ... }
+
 }
 DEFINE MODULE(m, config) {
   m.with(make tracks)
    .transform("GoodHits").in each("Event")
    .to("GoodTracks")
    .using_concurrency(unlimited);
 }
```



# **Class example using lambda expression**

```
#include "meld/module.hpp"
```

```
DEFINE_MODULE(m, config)
{
    auto threshold = config.get<unsigned int>("threshold");
    m.with([threshold](Hits const& hits) { return hits.size() > threshold; })
    .filter("GoodHits").in_each("Event")
    .using_concurrency(unlimited);
}
```



#### **Class example registering two member functions**

```
#include "meld/module.hpp"
class Selector {
public:
  Selector(unsigned int n) : threshold{n} {}
  bool gt(Hits const& hits) const { return hits.size() > threshold; }
  bool le(Hits const& hits) const { return !gt(hits); }
private:
  unsigned int threshold;
};
DEFINE MODULE(m, config)
{
  auto threshold = config.get<unsigned int>("threshold");
  auto bound m = m.make<Selector>(threshold);
  bound m.with(&Selector::gt).filter("GoodHits").in each("Event");
  bound m.with(&Selector::le).filter("GoodHits").in each("Event");
}
```



#### **Reduction example**

```
class MyAccumulator : public art::EDProducer {
public:
  MyAccumulator(ParameterSet const&)
    produces<int, art::InSubRun>("sum");
  void produce(art::Event&) override
    ++counter_;
  void endSubRun(art::SubRun& sr) override
    sr.put(std::make_unique<int>(counter_), "sum");
    counter = 0;
private:
  int counter_ = 0;
};
DEFINE ART MODULE(MyAccumulator)
```







## **Higher-order functions**

• We are interested in the mappings of the form:

$$\left\{(\boldsymbol{a})_n \xrightarrow{f} (\boldsymbol{b})_m\right\} \in \mathcal{D}$$

- Each object *a* corresponds to a tuple of arguments passed to *f*.
- The signature of f and the value f(a), depends on the higher-order function.
- The above mapping happens within a domain  $\mathcal{D}$  (e.g. job, run, event).
- Each object a is an element of a subset of the domain  $\mathcal{D}$ .



# **Supported higher-order functions**

Meld term	CS term	Mathematical description		Domain
Transform	Мар	$(\boldsymbol{a})_n \xrightarrow{f} (\boldsymbol{b})_n$	where $f(a) \rightarrow b$	Same as $(a)_n$
Filter	Filter	$(a)_n \xrightarrow{f} (a)_m$ where $m \le n$	where $f(a) \rightarrow Boolean$	Same as $(a)_n$
Monitor	_	$(a)_n \xrightarrow{f} ()_0$	where $f(a) \rightarrow \text{Void}$	Same as $(a)_n$
Reduction	Fold	$(\boldsymbol{a})_n \xrightarrow{f_c} (\boldsymbol{c})_1$	where $f_c(a) \rightarrow c$	Above $(a)_n$
Splitter	Unfold	$(\boldsymbol{a})_1 \stackrel{f_n}{\to} (\boldsymbol{d})_m$	where $f_n(\boldsymbol{a}) \to (\boldsymbol{d})_n$	<b>Below</b> $(a)_n$
Zip	Zip	$((a)_n, (b)_n) \rightarrow (a, b)_n$		More nested domain

