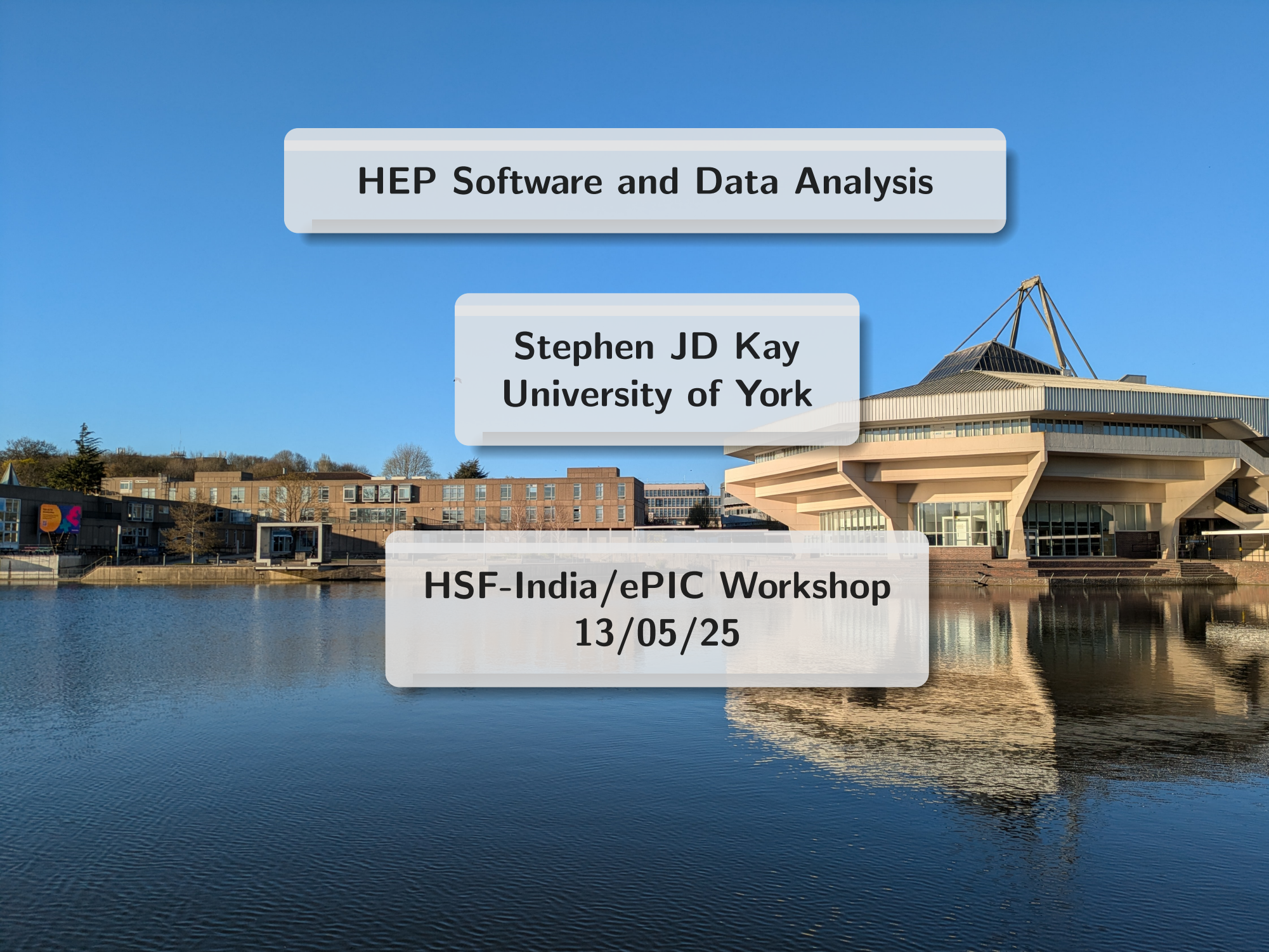


HEP Software and Data Analysis

Stephen JD Kay
University of York

HSF-India/ePIC Workshop
13/05/25

HEP Software

- We utilise software in **H**igh **E**nergy **P**hysics for a variety of reasons in a range of contexts
- Primarily, we will be using it in the context of data analysis

HEP Software

- We utilise software in **H**igh **E**nergy **P**hysics for a variety of reasons in a range of contexts
- Primarily, we will be using it in the context of data analysis
- We will look at where this software comes in, what form it takes, and how we can utilise it
- Before that though, I want to step back a bit

HEP Software

- We utilise software in **H**igh **E**nergy **P**hysics for a variety of reasons in a range of contexts
- Primarily, we will be using it in the context of data analysis
- We will look at where this software comes in, what form it takes, and how we can utilise it
- Before that though, I want to step back a bit
- Before we look at our software, we need to think about what it is trying to do
- We are using it as a tool to study HEP

HEP Software

- We utilise software in **H**igh **E**nergy **P**hysics for a variety of reasons in a range of contexts
- Primarily, we will be using it in the context of data analysis
- We will look at where this software comes in, what form it takes, and how we can utilise it
- Before that though, I want to step back a bit
- Before we look at our software, we need to think about what it is trying to do
- We are using it as a tool to study HEP
- So what is High Energy Physics about?
- What information will our software be used on, **what hardware does it come from?**

Disclaimer

Disclaimer

I am be covering lots of topics here in a very brief manner

Disclaimer

I am be covering lots of topics here in a very brief manner

Follow the references/links in many slides if you want to know more

Disclaimer

I am be covering lots of topics here in a very brief manner

Follow the references/links in many slides if you want to know more

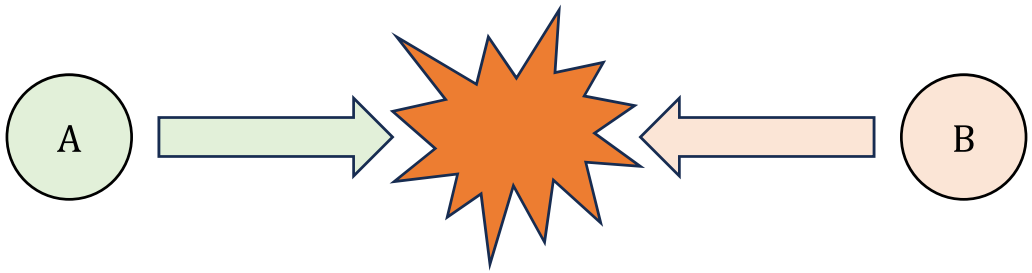
... or, just ask!

A HEP Experiment in a Nutshell

- Boiling it down, High Energy Physics can be described easily

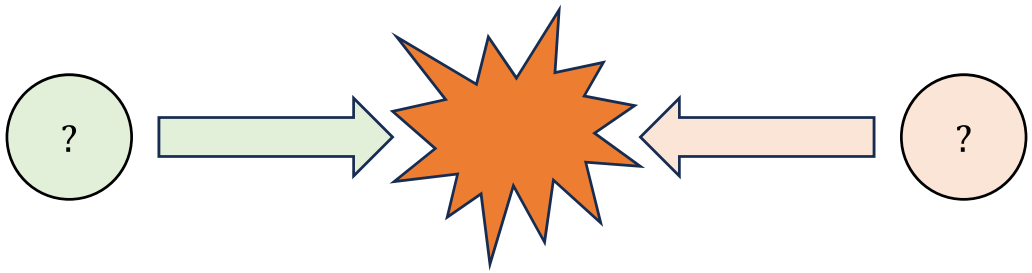
A HEP Experiment in a Nutshell

- Boiling it down, High Energy Physics can be described easily
  - Smash things together and see what happens



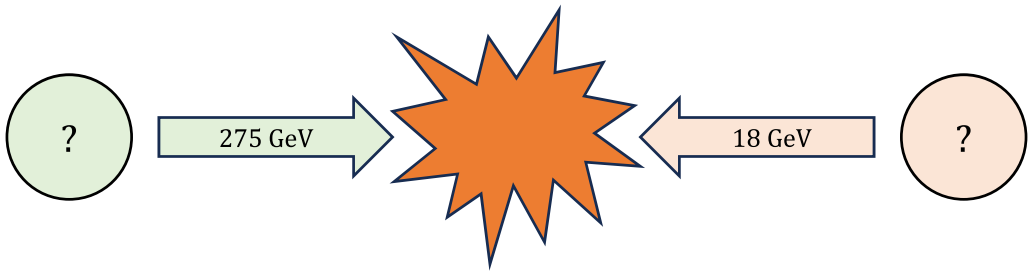
A HEP Experiment in a Nutshell

- Boiling it down, High Energy Physics can be described easily
  - Smash things together and see what happens
- There's nuance to that of course
  - What are we smashing together?



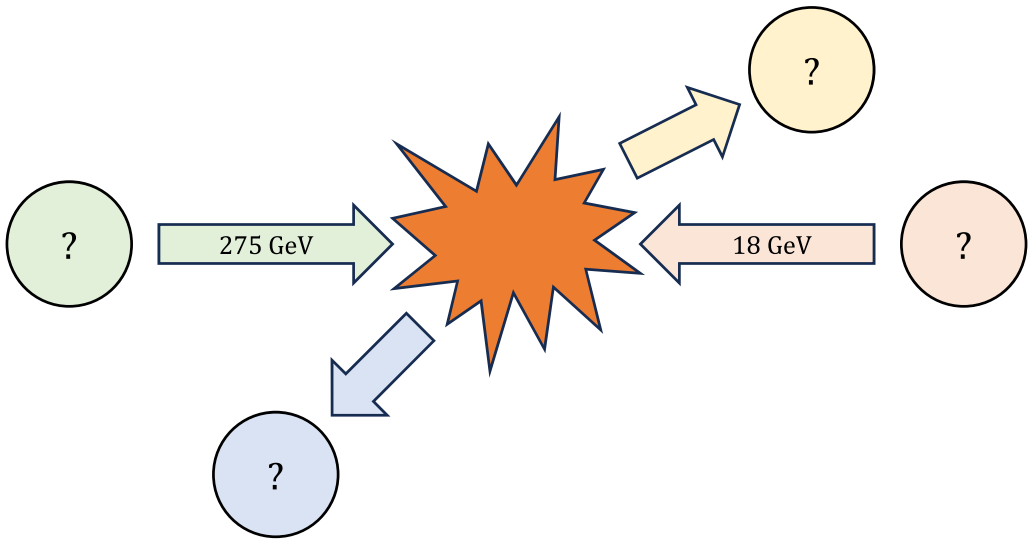
A HEP Experiment in a Nutshell

- Boiling it down, High Energy Physics can be described easily
  - Smash things together and see what happens
- There's nuance to that of course
  - What are we smashing together?
  - How energetically are we smashing them together?



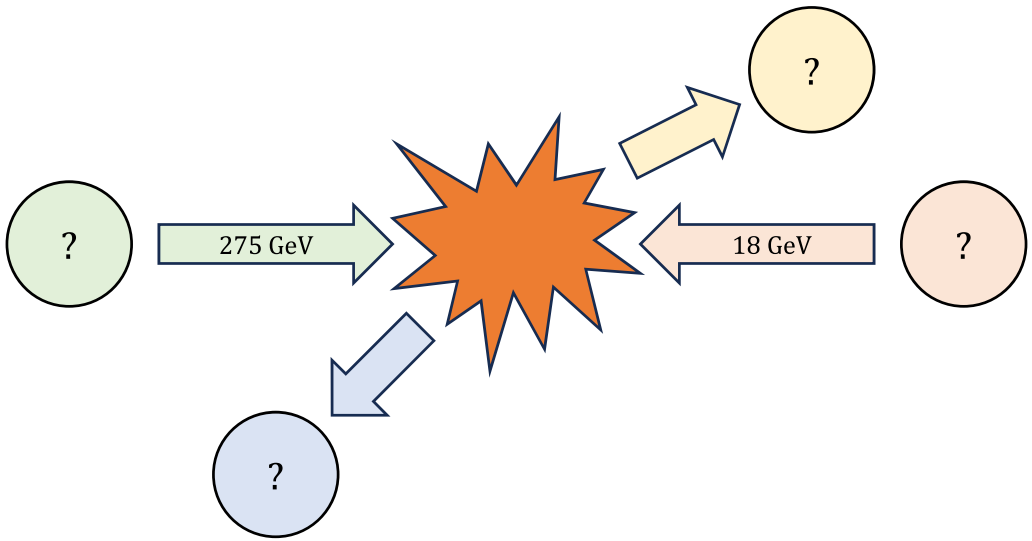
A HEP Experiment in a Nutshell

- Boiling it down, High Energy Physics can be described easily
  - Smash things together and see what happens
- There's nuance to that of course
  - What are we smashing together?
  - How energetically are we smashing them together?
  - What is being produced in our collision?



A HEP Experiment in a Nutshell

- Boiling it down, High Energy Physics can be described easily
  - Smash things together and see what happens
- There's nuance to that of course
  - What are we smashing together?
  - How energetically are we smashing them together?
  - What is being produced in our collision?
  - How **often** are specific things produced?



What is an event?

- We generally talk about our distinct beam on beam/target collisions as discrete "events"
  - One collision $\rightarrow$ One event
- Luckily, this means we have a nice logical way to structure software - event by event

What is an event?

- We generally talk about our distinct beam on beam/target collisions as discrete "events"
  - One collision $\rightarrow$ One event
- Luckily, this means we have a nice logical way to structure software - event by event
- **In reality, we might have overlapping events**
- Our task then would be to identify the discrete, individual events we're interested in

What is an event?

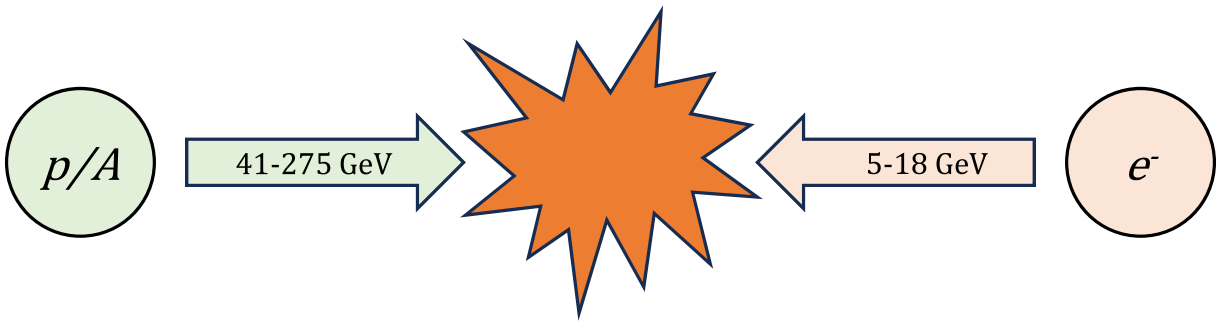
- We generally talk about our distinct beam on beam/target collisions as discrete "events"
  - One collision $\rightarrow$ One event
- Luckily, this means we have a nice logical way to structure software - event by event
- **In reality, we might have overlapping events**
- Our task then would be to identify the discrete, individual events we're interested in
- Typically, we can group events into categories **based upon the types of particles we detect in them**
- At the **E**lectron-**I**on **C**ollider, EIC, there are a few broad categories of events we wish to observe

Example Events: DIS

- **D**eep **I**nelastic **S**cattering - **DIS**
- A golden process to probe nucleons and nuclei

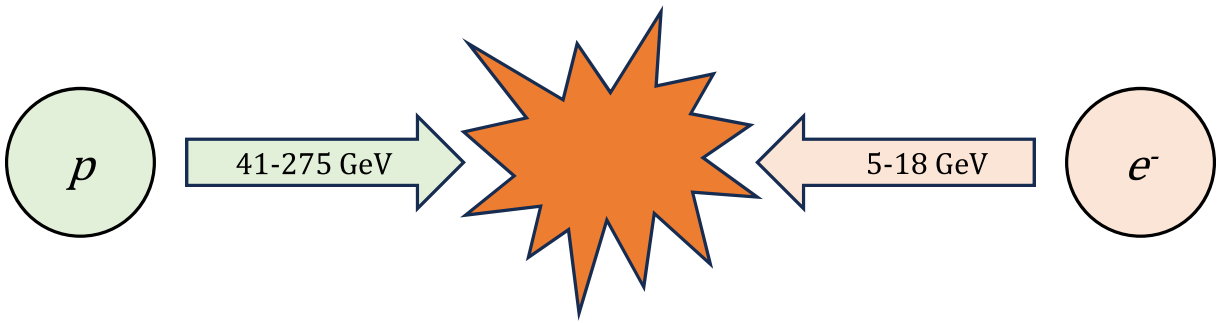
Example Events: DIS

- **D**eep **I**nelastic **S**cattering - **DIS**
- A golden process to probe nucleons and nuclei



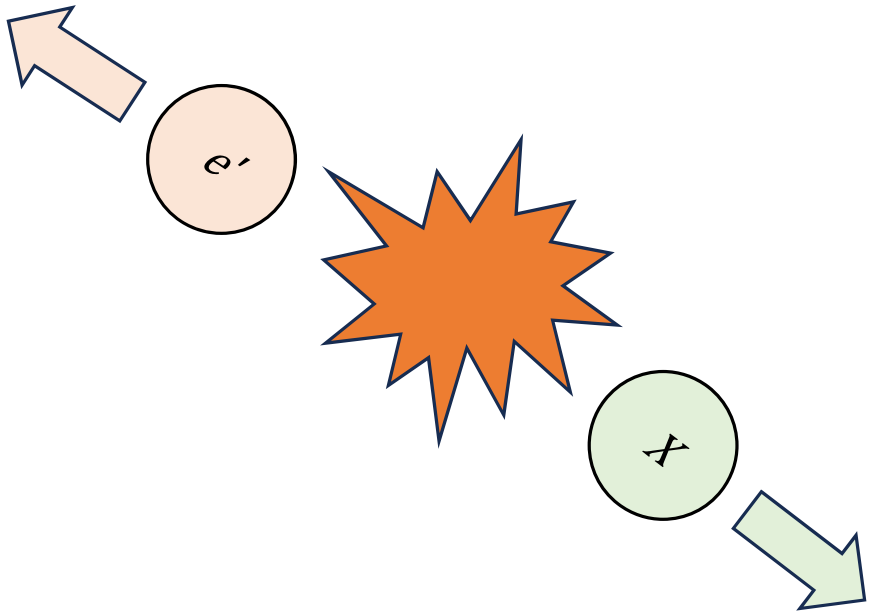
Example Events: DIS

- **D**eep **I**nelastic **S**cattering - **DIS**
- A golden process to probe nucleons and nuclei



Example Events: DIS

- **D**eep **I**nelastic **S**cattering - **DIS**
- A golden process to probe nucleons and nuclei
- In DIS, we have a scattered electron, e'
- ... and a shower of hadronic "stuff", X

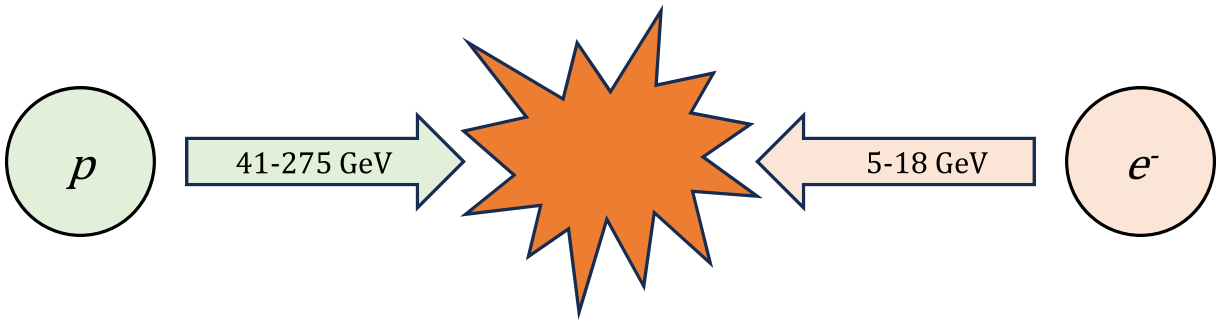


Example Events: SIDIS

- **S**emi-**I**nclusive **D**eep **I**nelastic **S**cattering - **SIDIS**
- Similar, to DIS, except we "tag" a specific hadron

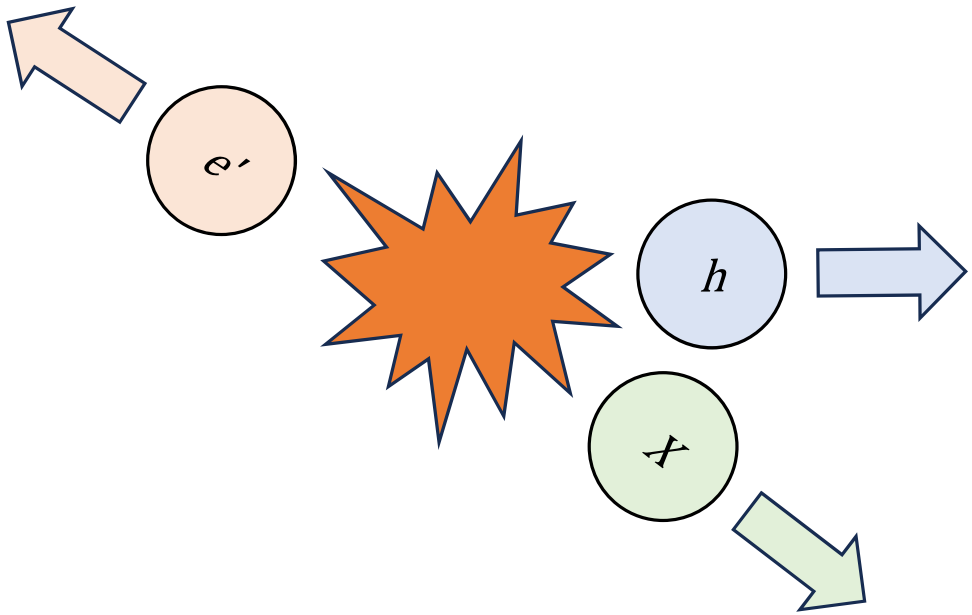
Example Events: SIDIS

- **S**emi-**I**nclusive **D**eep **I**nelastic **S**cattering - **SIDIS**
- Similar, to DIS, except we "tag" a specific hadron



Example Events: SIDIS

- **S**emi-**I**nclusive **D**eep **I**nelastic **S**cattering - **SIDIS**
- Similar, to DIS, except we "tag" a specific hadron
- Again we have, e' and X
- But we also tag a specific hadron, h

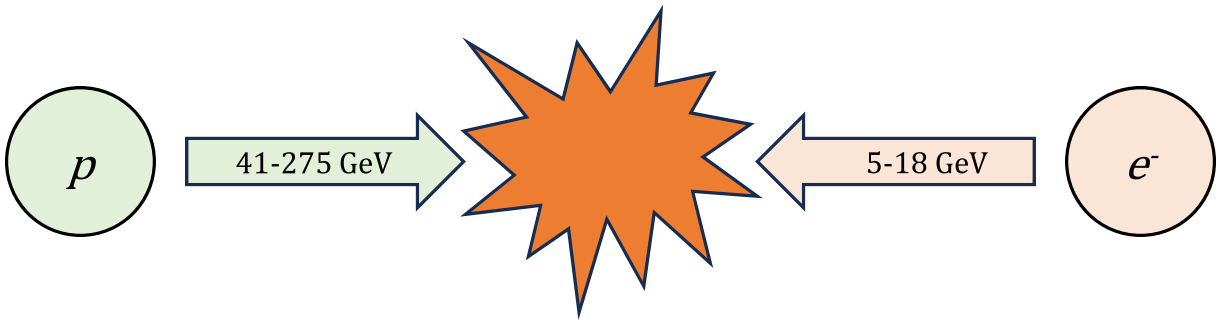


Example Events: Exclusive Reactions

- Exclusive Reactions - We detect all of the reaction products

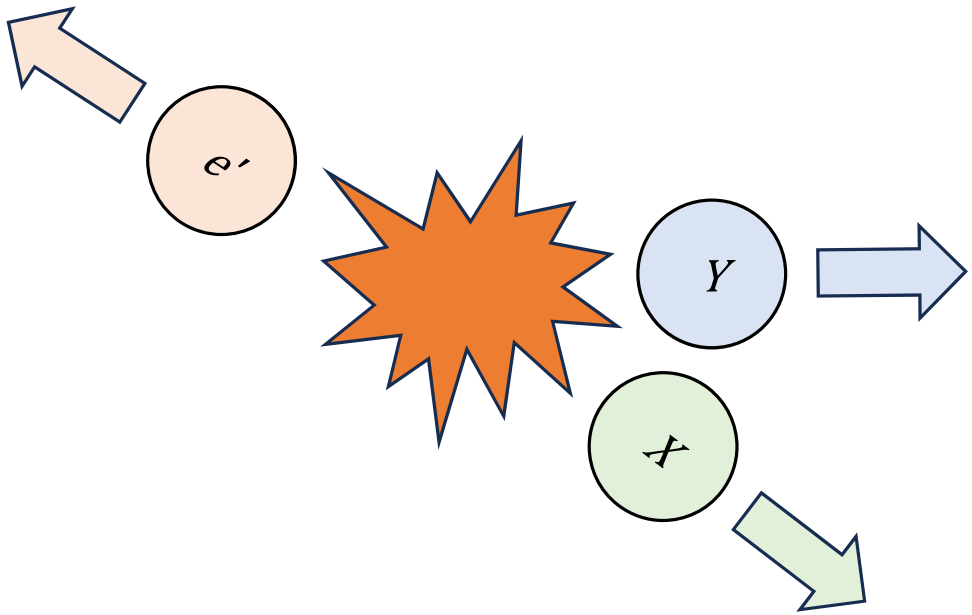
Example Events: Exclusive Reactions

- Exclusive Reactions - We detect all of the reaction products



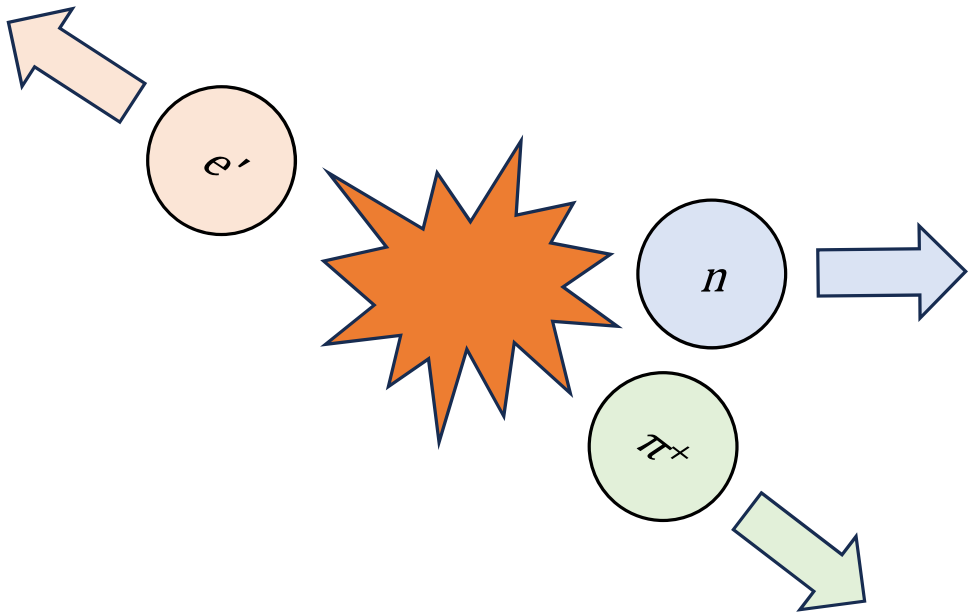
Example Events: Exclusive Reactions

- Exclusive Reactions - We detect all of the reaction products
- Many, many different reactions
- To generalise, we have e' and two other products, X and Y



Example Events: Exclusive Reactions

- Exclusive Reactions - We detect all of the reaction products
- Many, many different reactions
- To generalise, we have e' and two other products, X and Y
- As a specific example, **D**eep **E**xclusive **M**eson **P**roduction



Seeing Our Event

- When an event happens in our system, we need to detect and interpret it somehow
- To do so, we need an interplay of hardware and software

Seeing Our Event

- When an event happens in our system, we need to detect and interpret it somehow
- To do so, we need an interplay of hardware and software
- Detectors - Our particles need to interact with something and leave a signal we can detect

Seeing Our Event

- When an event happens in our system, we need to detect and interpret it somehow
- To do so, we need an interplay of hardware and software
- Detectors - Our particles need to interact with something and leave a signal we can detect
- Data Acquisition and readout - We need to interpret the signals from our detectors in some way and pass that on to systems that can process it further

Seeing Our Event

- When an event happens in our system, we need to detect and interpret it somehow
- To do so, we need an interplay of hardware and software
- Detectors - Our particles need to interact with something and leave a signal we can detect
- Data Acquisition and readout - We need to interpret the signals from our detectors in some way and pass that on to systems that can process it further
- Analysis - We need to process readout from our detectors and use it to interpret the event
  - What particles were produced?
  - How much energy does each particle have?
  - ...

Detectors: What are they "Detecting"?

- Generally, three broad categories of detectors

Detectors: What are they "Detecting"?

- Generally, three broad categories of detectors
  - Calorimetry
  - Tracking
  - **P**article **ID**entification, PID

Detectors: What are they "Detecting"?

- Generally, three broad categories of detectors
  - Calorimetry
  - Tracking
  - **P**article **ID**entification, PID
- Calorimetry - Determination of incident particle energy
  - Electromagnetic Calorimeters - Charged particles, photons
  - Hadronic Calorimeters - Hadrons

Detectors: What are they "Detecting"?

- Generally, three broad categories of detectors
  - Calorimetry
  - Tracking
  - **P**article **ID**entification, PID
- Calorimetry - Determination of incident particle energy
  - Electromagnetic Calorimeters - Charged particles, photons
  - Hadronic Calorimeters - Hadrons
- Tracking - Charged particles, vertex and momentum determination
  - Magnetic fields might be utilised

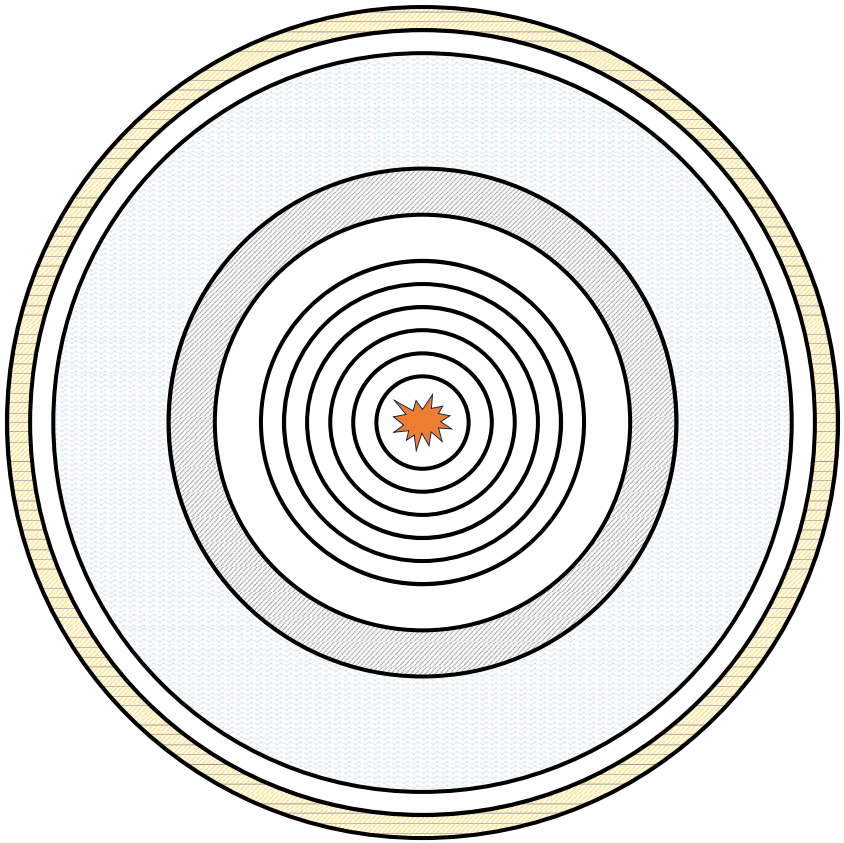
Detectors: What are they "Detecting"?

- Generally, three broad categories of detectors
  - Calorimetry
  - Tracking
  - **P**article **ID**entification, PID
- Calorimetry - Determination of incident particle energy
  - Electromagnetic Calorimeters - Charged particles, photons
  - Hadronic Calorimeters - Hadrons
- Tracking - Charged particles, vertex and momentum determination
  - Magnetic fields might be utilised
- PID - Determination of particle species
  - Many particles may look similar in other systems - π^+/K^+
  - Various different systems, threshold Cherenkov detectors, Ring Imaging Cherenkov detectors, Time of Flight detectors...

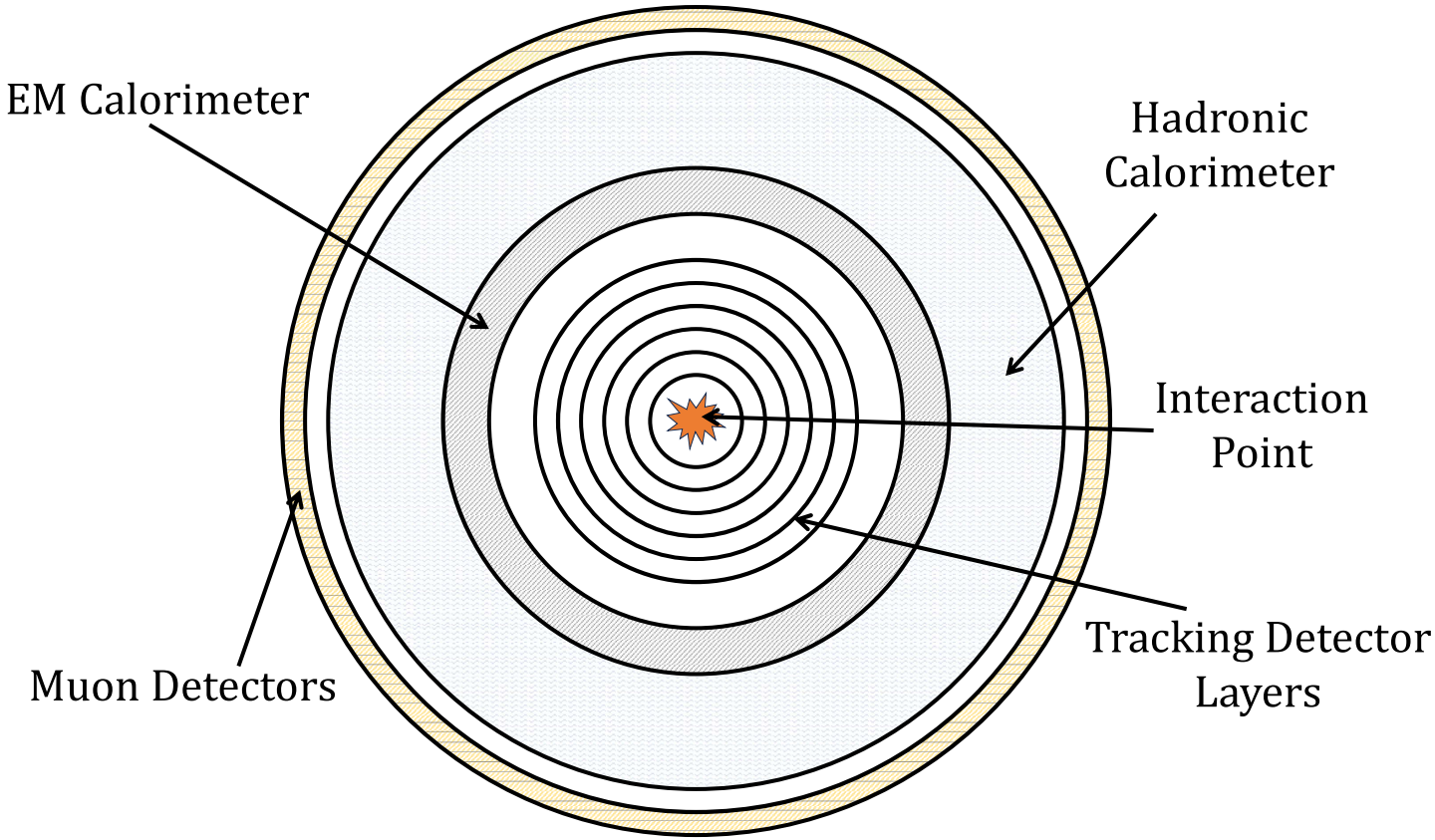
Detectors: What are they "Detecting"?

- Generally, three broad categories of detectors
  - Calorimetry
  - Tracking
  - **P**article **ID**entification, PID
- Calorimetry - Determination of incident particle energy
  - Electromagnetic Calorimeters - Charged particles, photons
  - Hadronic Calorimeters - Hadrons
- Tracking - Charged particles, vertex and momentum determination
  - Magnetic fields might be utilised
- PID - Determination of particle species
  - Many particles may look similar in other systems - π^+/K^+
  - Various different systems, threshold Cherenkov detectors, Ring Imaging Cherenkov detectors, Time of Flight detectors...
- Systems for other purposes too, polarimetry, timing etc

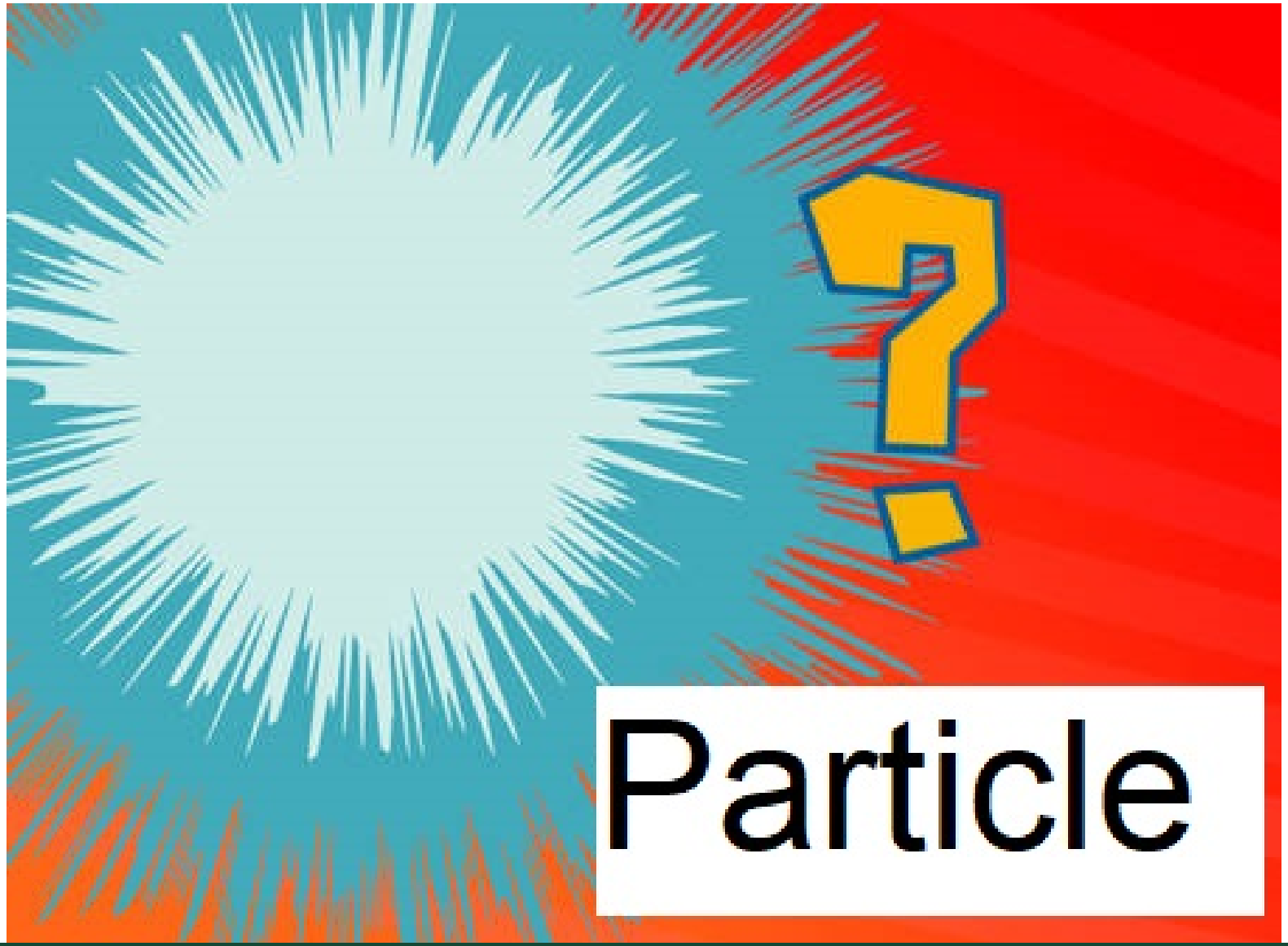
Detectors: Simplistic Model



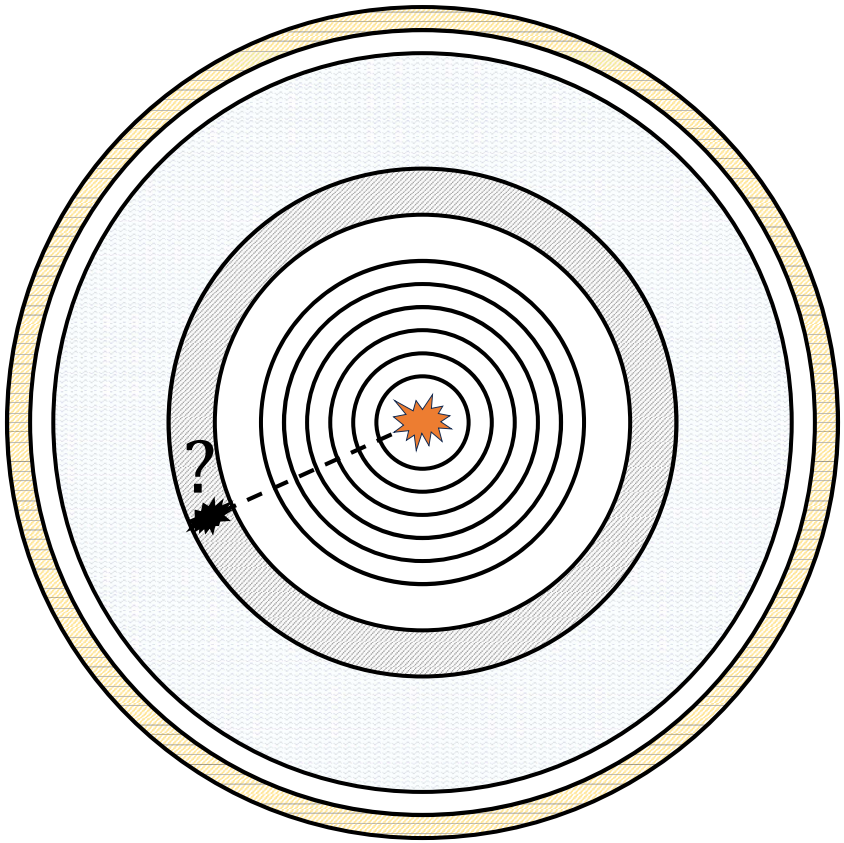
Detectors: Simplistic Model



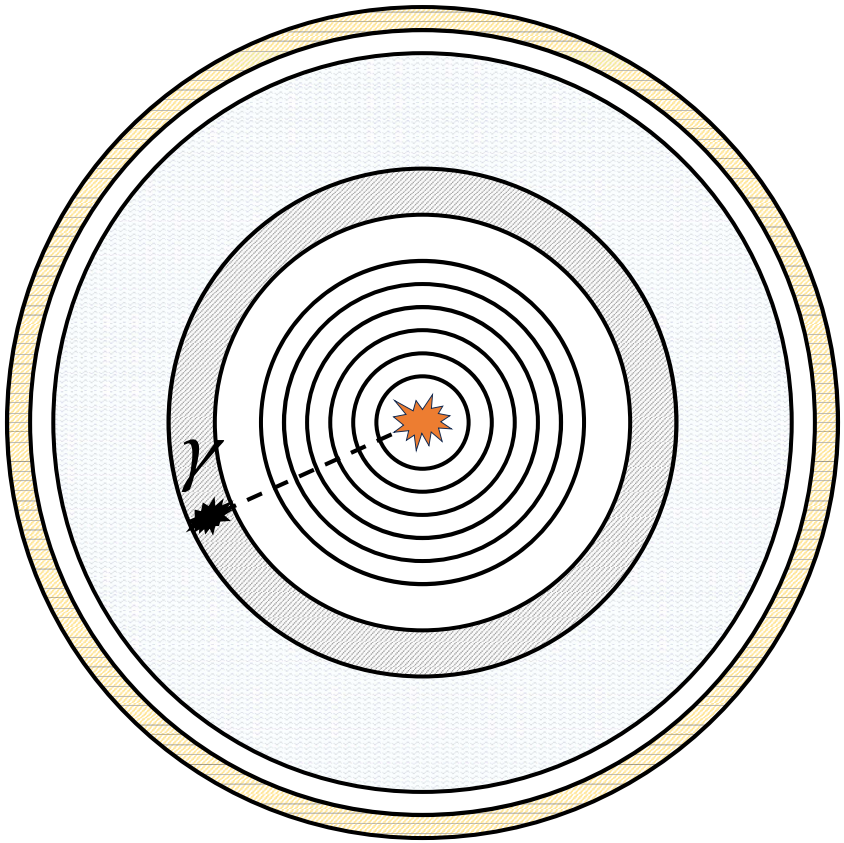
Who's that ~~Pokemon~~ Particle?



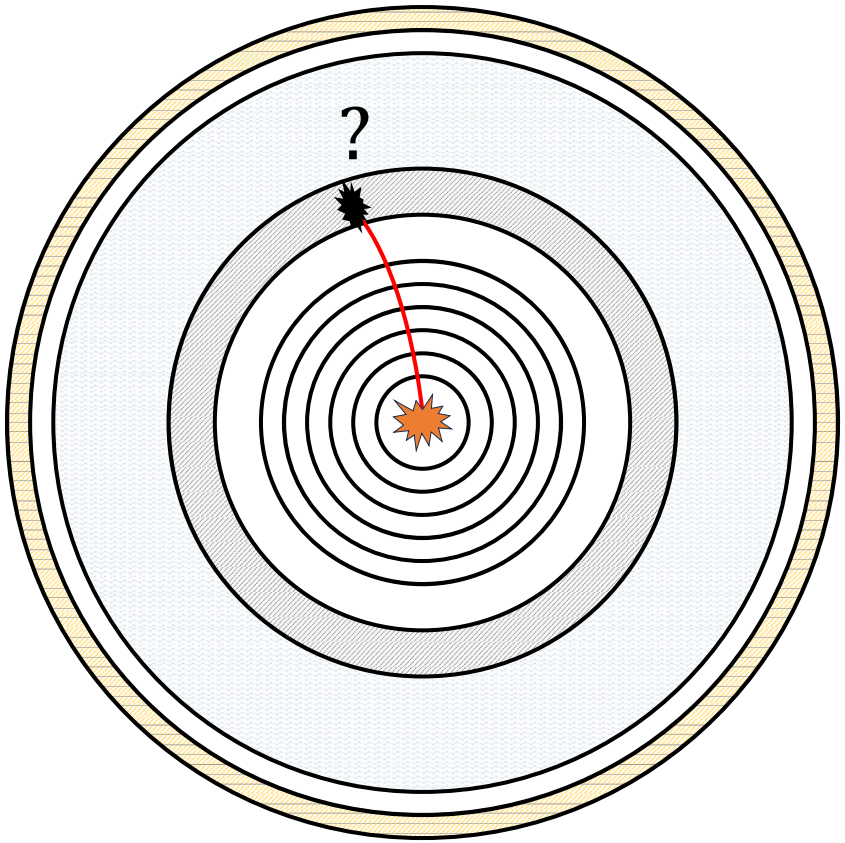
Detectors: Quick Quiz



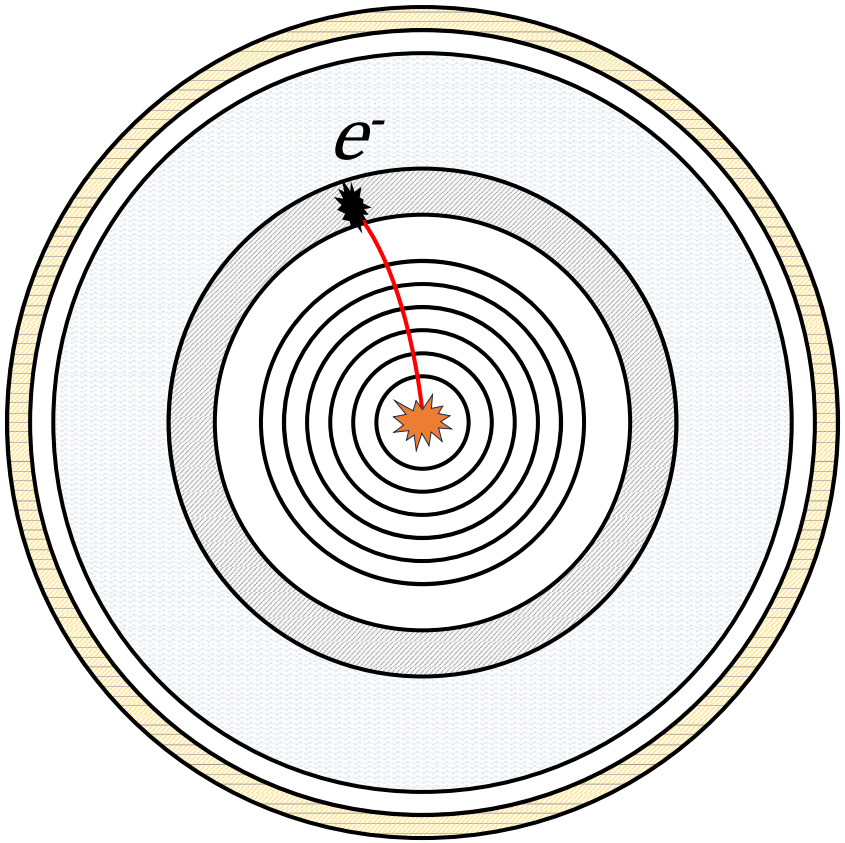
Detectors: Quick Quiz



Detectors: Quick Quiz

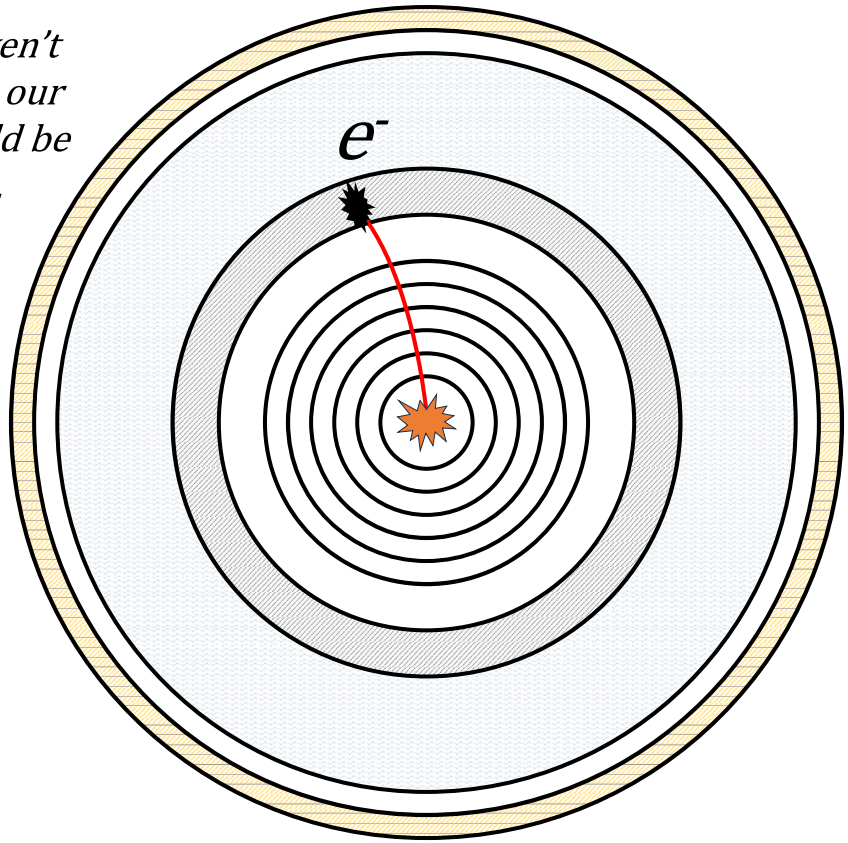


Detectors: Quick Quiz

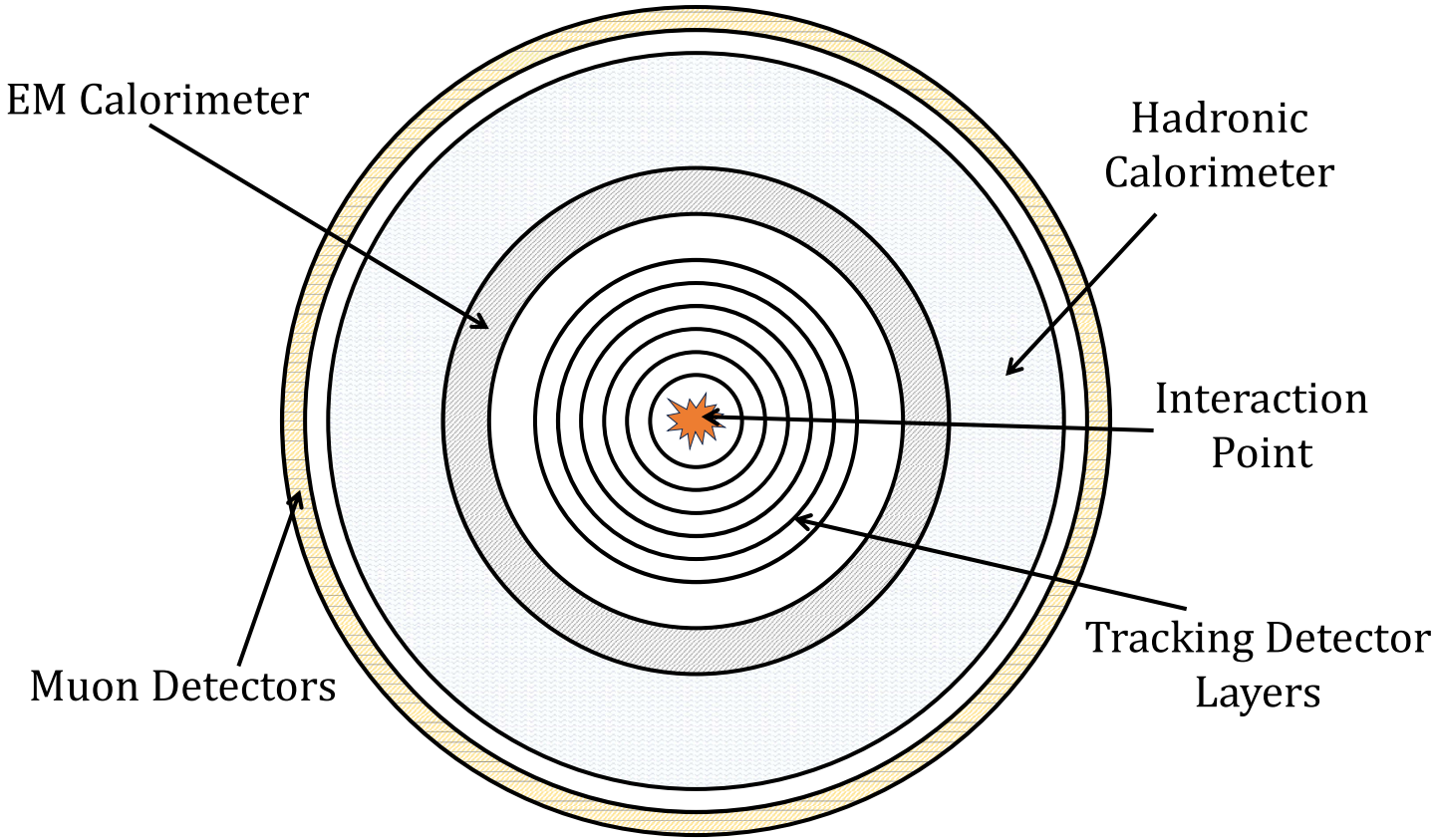


Detectors: Quick Quiz

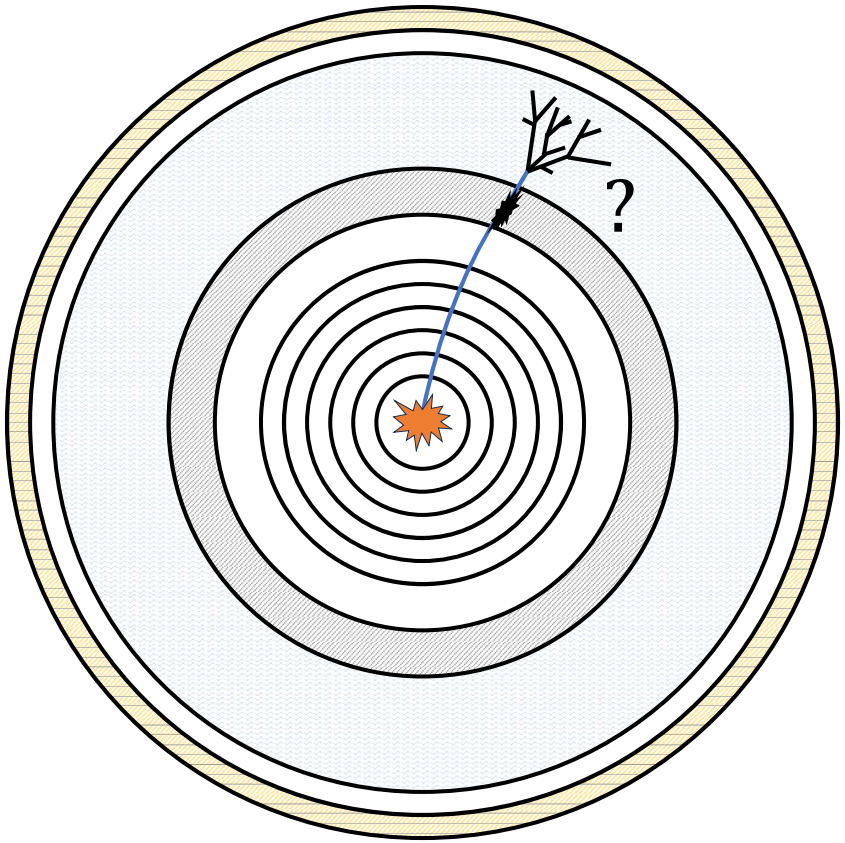
OR e^+, *we haven't really defined our field, so it could be a positron.*



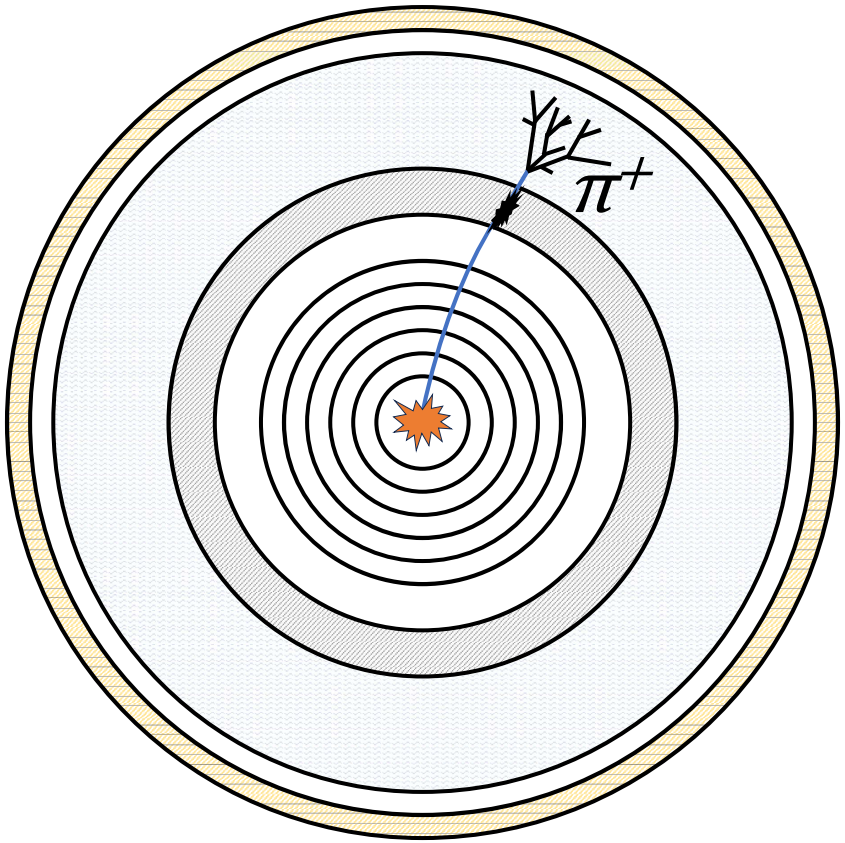
Detectors: Quick Quiz



Detectors: Quick Quiz

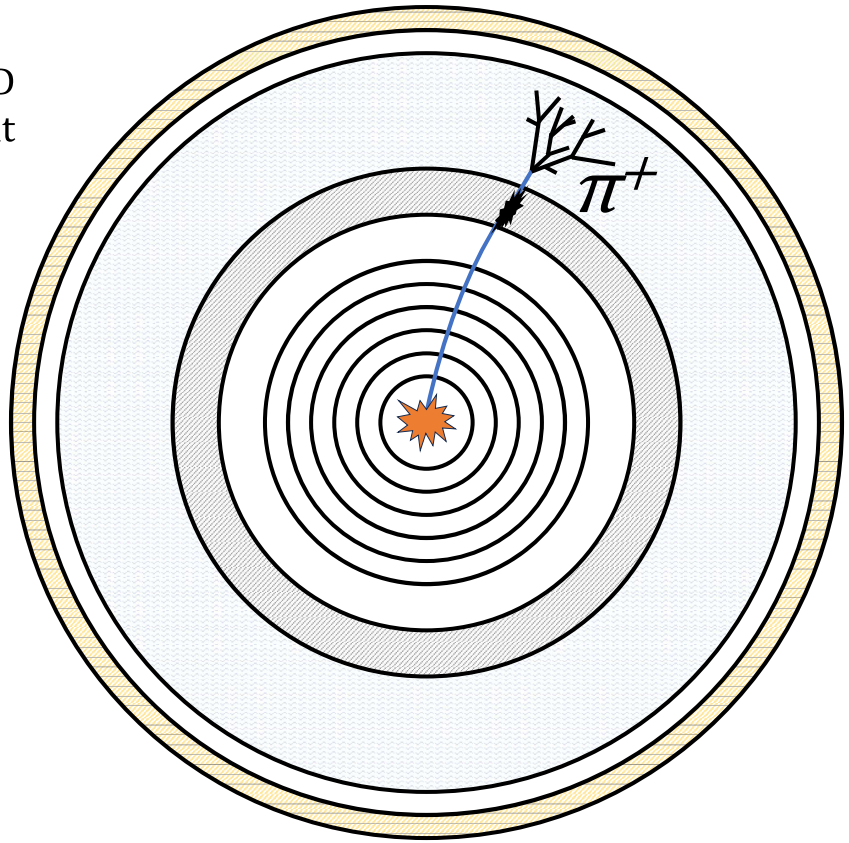


Detectors: Quick Quiz

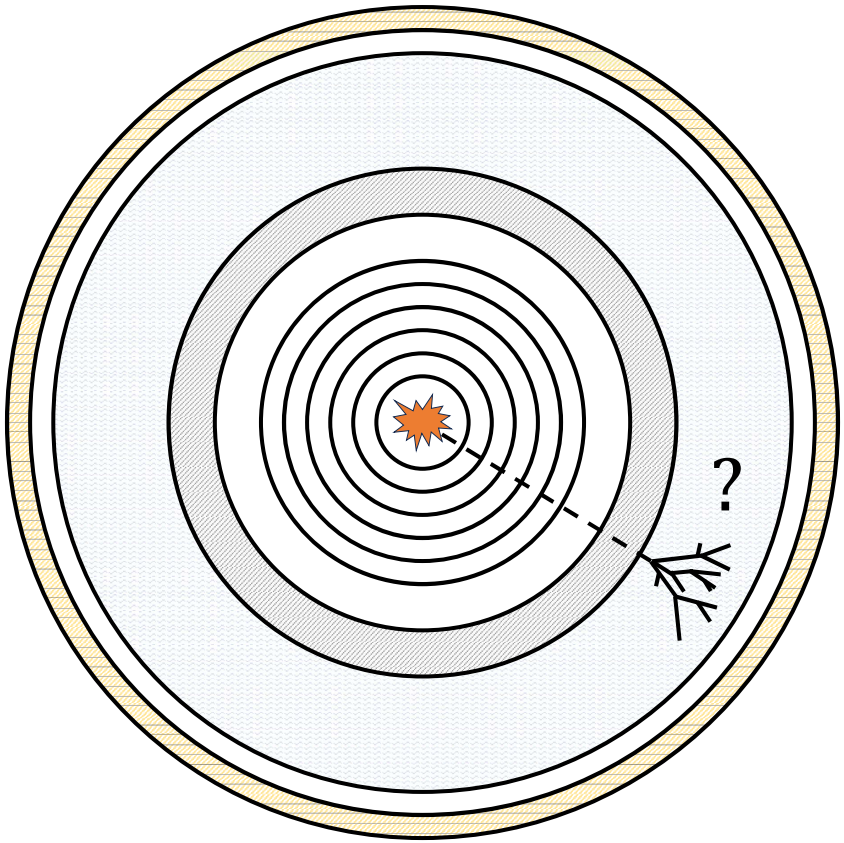


Detectors: Quick Quiz

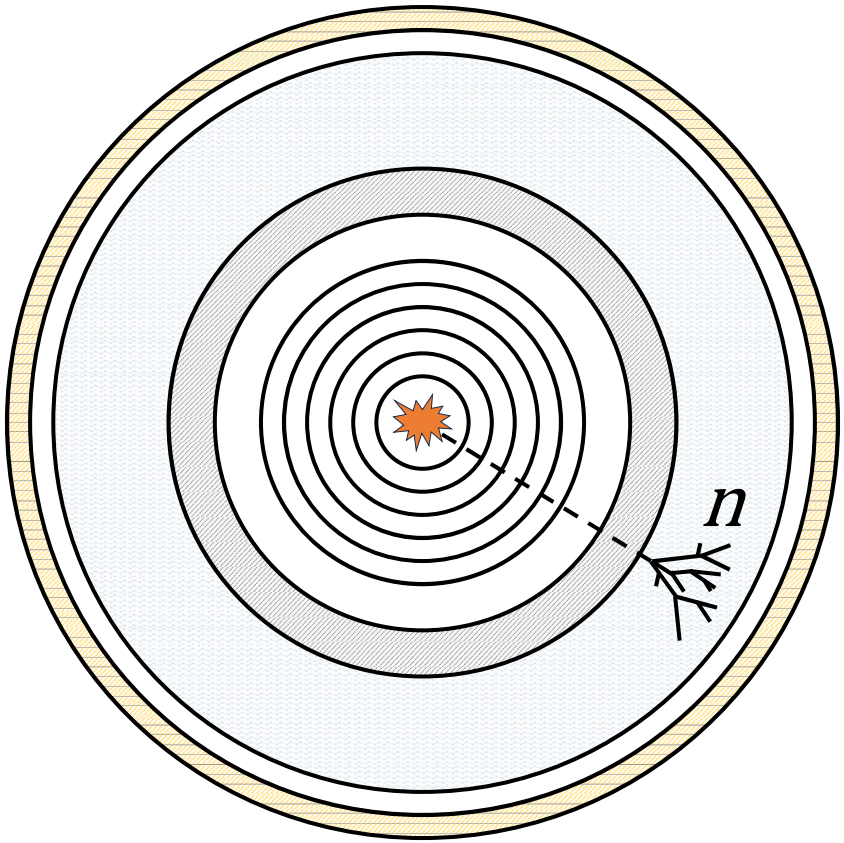
Or some other +ve hadron. Without PID info, we can't be that definitive here.



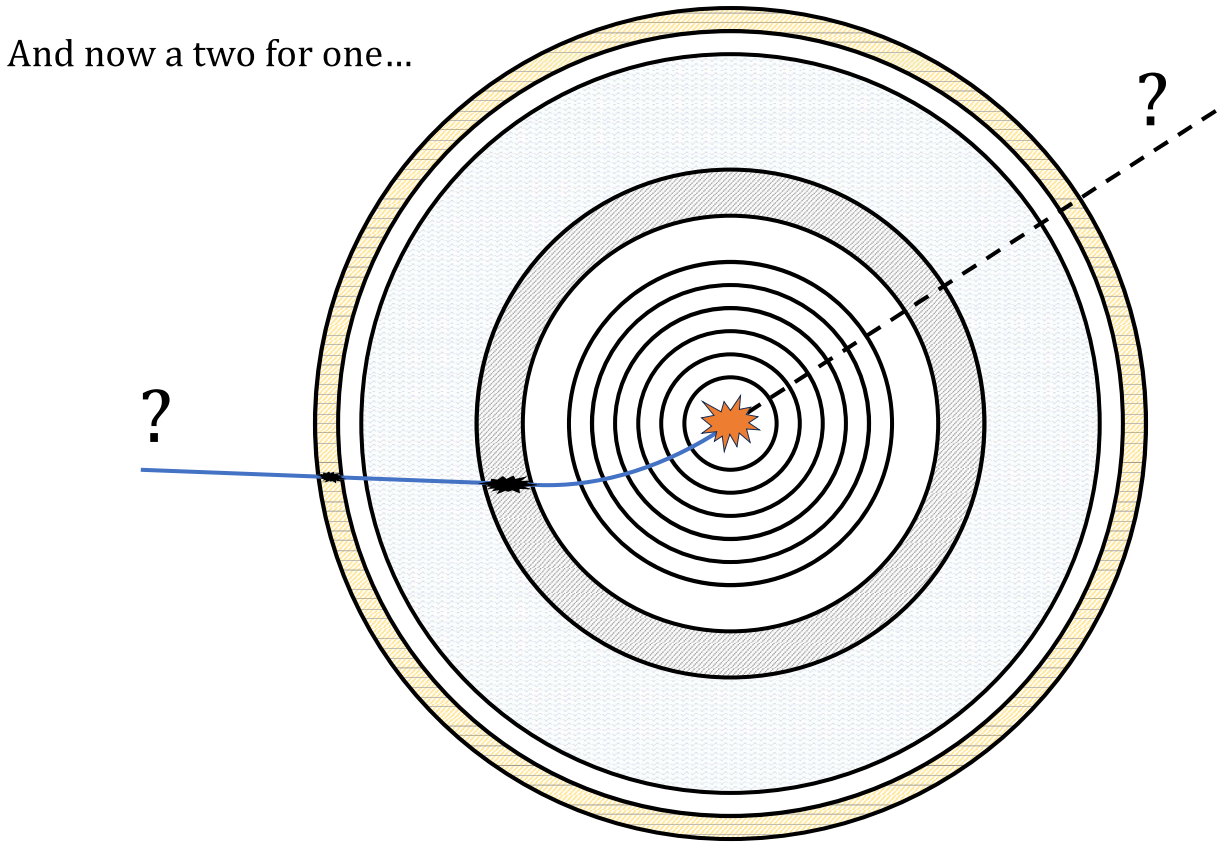
Detectors: Quick Quiz



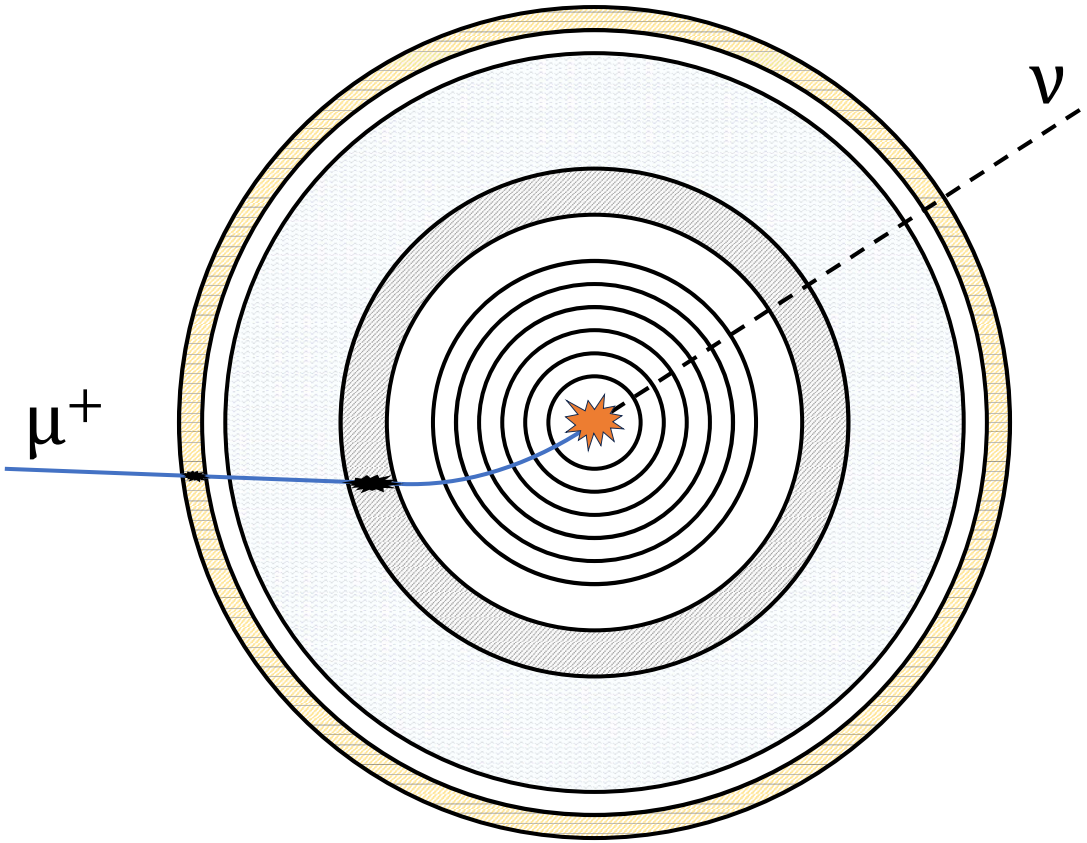
Detectors: Quick Quiz



Detectors: Quick Quiz



Detectors: Quick Quiz



Detectors: Calorimetry

- Calorimeters utilised to determine incident particle energy
- Many different designs utilising a range of different materials

Detectors: Calorimetry

- Calorimeters utilised to determine incident particle energy
- Many different designs utilising a range of different materials
- Two common types
  - Homogeneous - Detector is an absorber and a signal source
  - Sampling - Alternating layers of absorber and sampler material

Detectors: Calorimetry

- Calorimeters utilised to determine incident particle energy
- Many different designs utilising a range of different materials
- Two common types
  - Homogeneous - Detector is an absorber and a signal source
  - Sampling - Alternating layers of absorber and sampler material
- In EM calorimeters, detect electromagnetic showers
  - Bremsstrahlung, pair production

Detectors: Calorimetry

- Calorimeters utilised to determine incident particle energy
- Many different designs utilising a range of different materials
- Two common types
  - Homogeneous - Detector is an absorber and a signal source
  - Sampling - Alternating layers of absorber and sampler material
- In EM calorimeters, detect electromagnetic showers
  - Bremsstrahlung, pair production
- Hadronic Calorimeters, detect hadronic showers
  - Generally broader and longer than EM showers
  - More interactions possible, more complex

Detectors: Calorimetry

- Calorimeters utilised to determine incident particle energy
- Many different designs utilising a range of different materials
- Two common types
  - Homogeneous - Detector is an absorber and a signal source
  - Sampling - Alternating layers of absorber and sampler material
- In EM calorimeters, detect electromagnetic showers
  - Bremsstrahlung, pair production
- Hadronic Calorimeters, detect hadronic showers
  - Generally broader and longer than EM showers
  - More interactions possible, more complex
- Many parameters to consider -
  - Energy resolution
  - Timing Resolution
  - Radiation length/nuclear interaction length
  - Sampling fraction
  - Position resolution
  - ...

Detectors: Tracking

- Want **charged particles** to leave distinct points
- Want energy deposition per point... but a large signal!

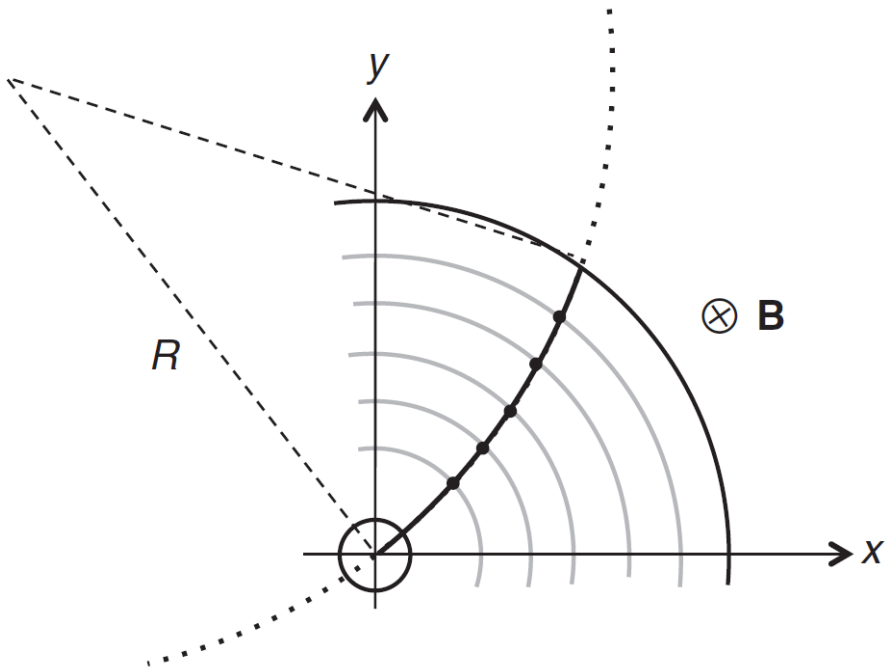
Detectors: Tracking

- Want **charged particles** to leave distinct points
- Want energy deposition per point... but a large signal!
- Many technologies available, two broad types:
  - Gas filled detectors
  - Semiconductor detectors

Detectors: Tracking

- Want **charged particles** to leave distinct points
- Want energy deposition per point... but a large signal!
- Many technologies available, two broad types:
  - Gas filled detectors
  - Semiconductor detectors

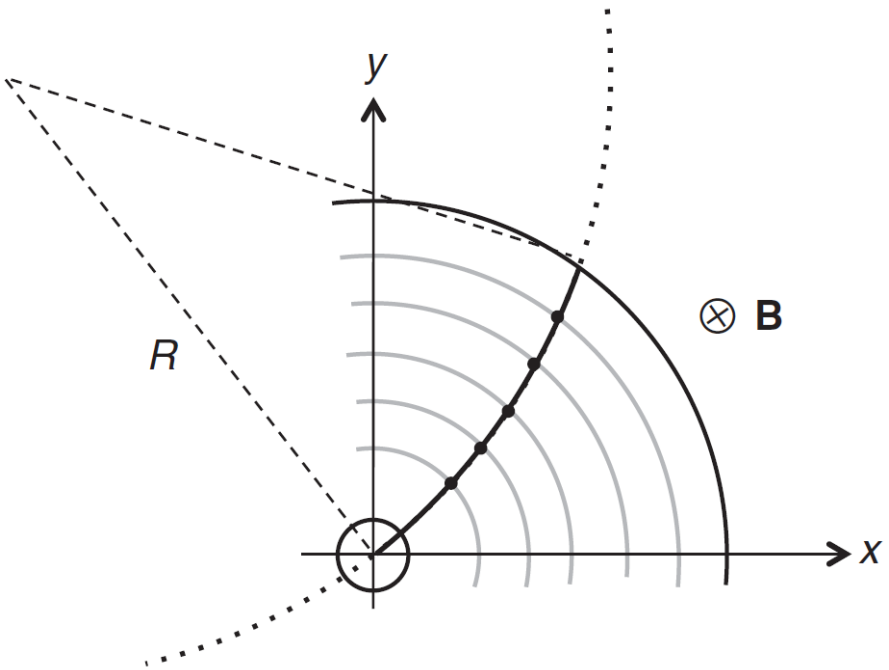
- Use software to reconstruct tracks
- **Vertex** and **momentum** determination



Detectors: Tracking

- Want **charged particles** to leave distinct points
- Want energy deposition per point... but a large signal!
- Many technologies available, two broad types:
  - Gas filled detectors
  - Semiconductor detectors

- Use software to reconstruct tracks
- **Vertex** and **momentum** determination
- Very non-trivial - usually lots of possible tracks!



Detectors: PID

- Dedicated PID detectors to distinguish between particles
- For example, pions, kaons and protons

Detectors: PID

- Dedicated PID detectors to distinguish between particles
- For example, pions, kaons and protons
- Cherenkov detectors are commonly used for this purpose

Detectors: PID

- Dedicated PID detectors to distinguish between particles
- For example, pions, kaons and protons
- Cherenkov detectors are commonly used for this purpose

- Utilise Cherenkov radiation
  - "Sonic boom" of light

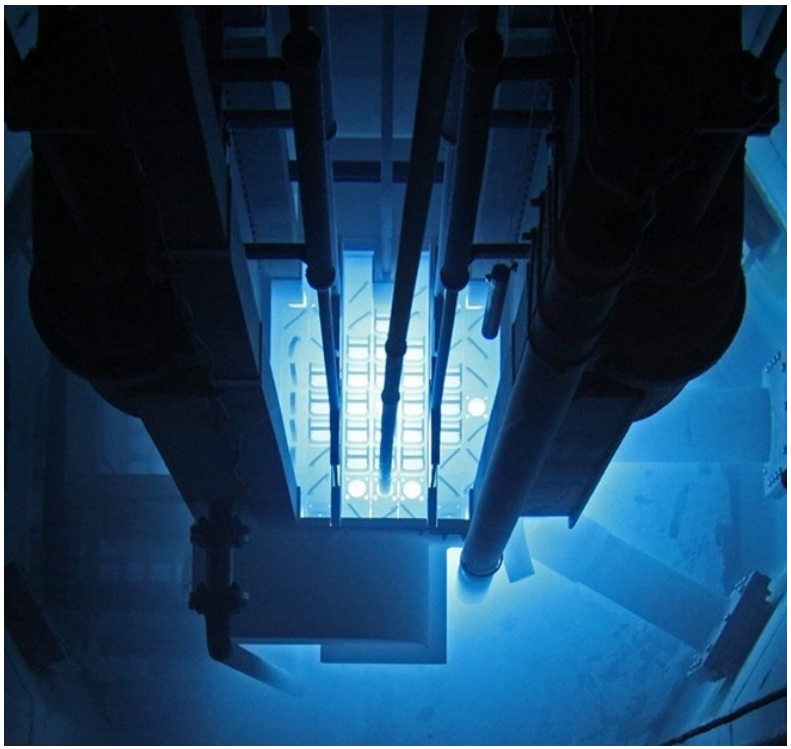


Image - University of Massachusetts, Lowell, public domain

Detectors: PID

- Dedicated PID detectors to distinguish between particles
- For example, pions, kaons and protons
- Cherenkov detectors are commonly used for this purpose

- Utilise Cherenkov radiation
  - "Sonic boom" of light
- Many types of Cherenkov detector
- **R**ing **I**maging **CH**erenkov, RICH, image ring of light directly

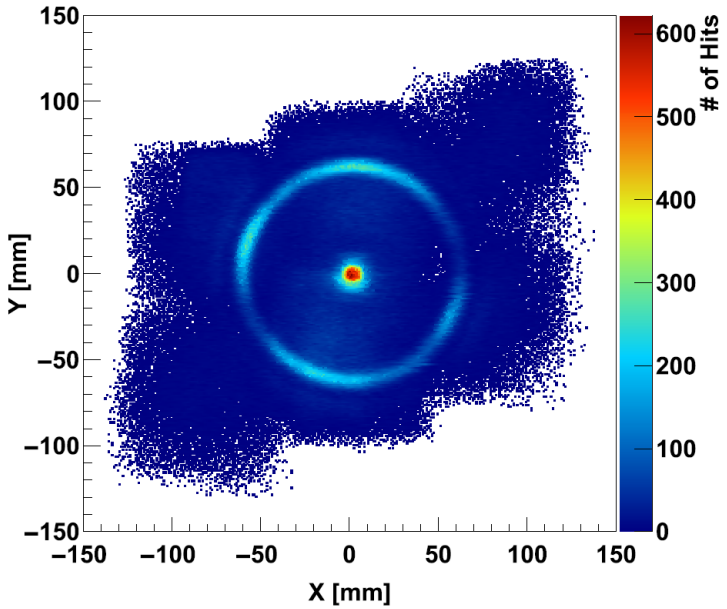
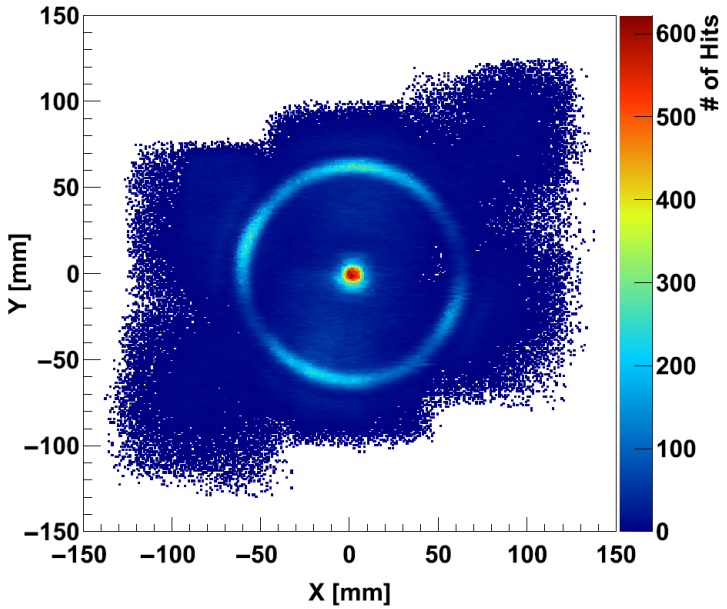


Image - S. Iwata et. al.

Detectors: PID

- Dedicated PID detectors to distinguish between particles
- For example, pions, kaons and protons
- Cherenkov detectors are commonly used for this purpose

- Utilise Cherenkov radiation
  - "Sonic boom" of light
- Many types of Cherenkov detector
- **R**ing **I**maging **CH**erenkov, RICH, image ring of light directly
- PID can also come from combining information across systems



Detectors: Combining Information

- Rare that we look at one detector in isolation
- Can combine information across systems for further PID

Detectors: Combining Information

- Rare that we look at one detector in isolation
- Can combine information across systems for further PID
- E.g. combine tracking and calorimetry information
  - The calorimeter cluster position is another "hit" we could use as a sanity check

Detectors: Combining Information

- Rare that we look at one detector in isolation
- Can combine information across systems for further PID
- E.g. combine tracking and calorimetry information
  - The calorimeter cluster position is another "hit" we could use as a sanity check

- Look at 2D correlations such as:
  - E vs dE
  - P vs β (or $1/\beta$)
  - Can have dedicated detectors to determine dE or β too

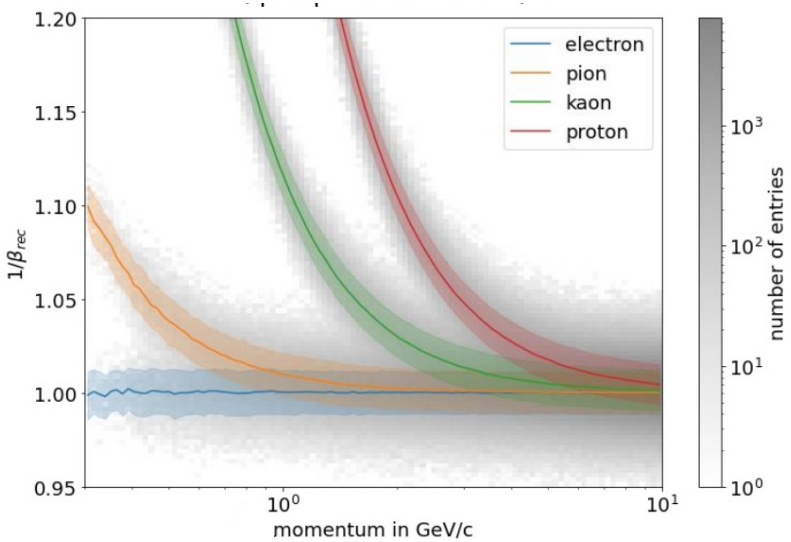
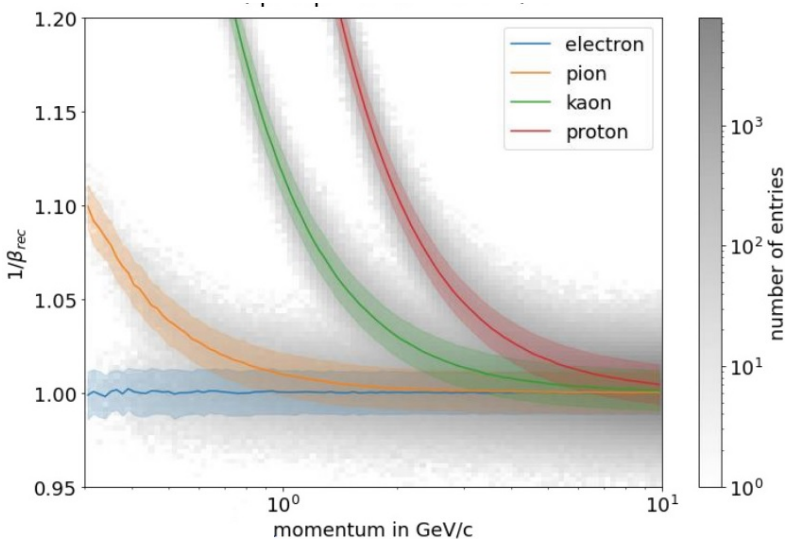


Image - Chandradoy Chatterjee, Hadron 2023

Detectors: Combining Information

- Rare that we look at one detector in isolation
- Can combine information across systems for further PID
- E.g. combine tracking and calorimetry information
  - The calorimeter cluster position is another "hit" we could use as a sanity check

- Look at 2D correlations such as:
  - E vs dE
  - P vs β (or $1/\beta$)
  - Can have dedicated detectors to determine dE or β too
- Can also be binary yes/no combinatorics as we did by eye earlier



Readout: Signals

- Hardware starts to interplay with software here
- Our various detectors output some electrical signal

Readout: Signals

- Hardware starts to interplay with software here
- Our various detectors output some electrical signal
- We typically process the raw output in various ways to interpret it
  - Amplification
  - Filtering
  - Shaping
  - ...

Readout: Signals

- Hardware starts to interplay with software here
- Our various detectors output some electrical signal
- We typically process the raw output in various ways to interpret it
  - Amplification
  - Filtering
  - Shaping
  - ...
- Signal then digitised for readout
  - ... usually, still have analog readout too

Readout: Signals

- Hardware starts to interplay with software here
- Our various detectors output some electrical signal
- We typically process the raw output in various ways to interpret it
  - Amplification
  - Filtering
  - Shaping
  - ...
- Signal then digitised for readout
  - ... usually, still have analog readout too
- Software to digitise our processed signal
- Could read it out now... or pass it to a trigger

Readout: Triggering

- Typically get a **lot** of data very quickly in HEP experiments
- EIC will have bunch crossings every 10 *ns*
  - Many interactions per bunch crossing possible!

Readout: Triggering

- Typically get a **lot** of data very quickly in HEP experiments
- EIC will have bunch crossings every 10 *ns*
  - Many interactions per bunch crossing possible!
- Typically, HEP experiments utilise a trigger (or trigger**s**)
- Fast logic to determine whether to keep a signal or not

Readout: Triggering

- Typically get a **lot** of data very quickly in HEP experiments
- EIC will have bunch crossings every 10 *ns*
  - Many interactions per bunch crossing possible!
- Typically, HEP experiments utilise a trigger (or trigger**s**)
- Fast logic to determine whether to keep a signal or not

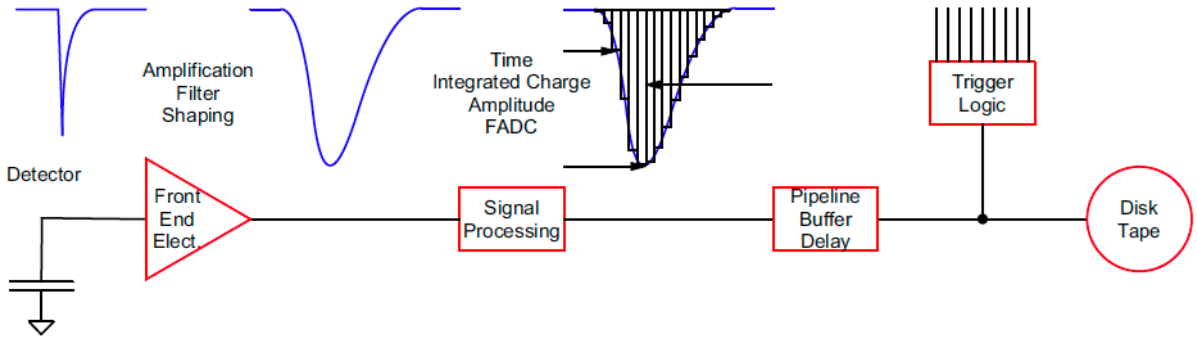
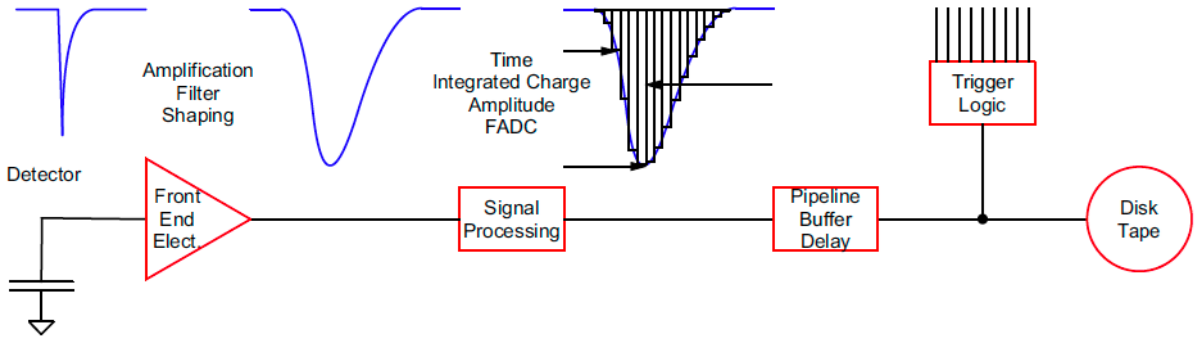


Image - Marco Battaglieri, CFNS Summer School 2024

Readout: Triggering

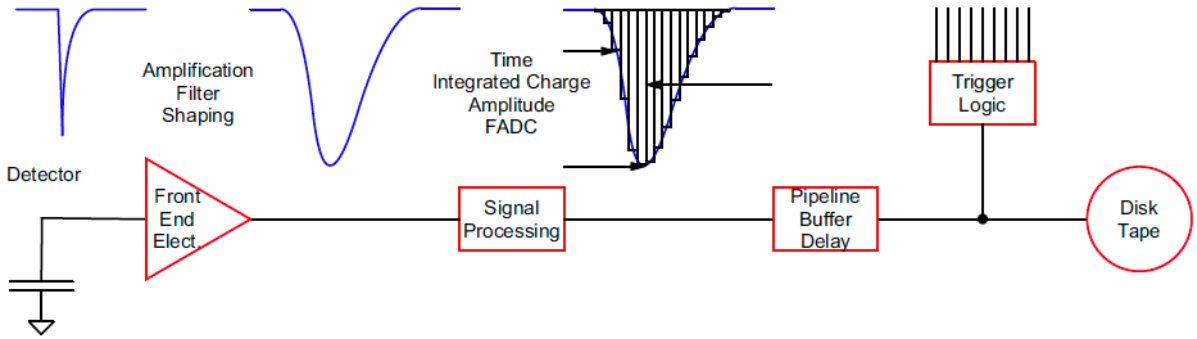
- Typically get a **lot** of data very quickly in HEP experiments
- EIC will have bunch crossings every 10 *ns*
  - Many interactions per bunch crossing possible!
- Typically, HEP experiments utilise a trigger (or trigger**s**)
- Fast logic to determine whether to keep a signal or not



- Trigger could be simple, or quite complex
  - Require X particles of type Y within Z second window

Readout: Triggering

- Typically get a **lot** of data very quickly in HEP experiments
- EIC will have bunch crossings every 10 *ns*
  - Many interactions per bunch crossing possible!
- Typically, HEP experiments utilise a trigger (or trigger**s**)
- Fast logic to determine whether to keep a signal or not



- Trigger could be simple, or quite complex
  - Require X particles of type Y within Z second window
- **Almost inevitable that a trigger will "lose" some information**

Readout: Triggerless

- ePIC is a general purpose detector
- I.e. we don't want to arbitrarily exclude things
  - So a selective trigger is a problem...

Readout: Triggerless

- ePIC is a general purpose detector
- I.e. we don't want to arbitrarily exclude things
  - So a selective trigger is a problem...
- Use triggerless, streaming readout instead

Readout: Triggerless

- ePIC is a general purpose detector
- I.e. we don't want to arbitrarily exclude things
  - So a selective trigger is a problem...
- Use triggerless, streaming readout instead

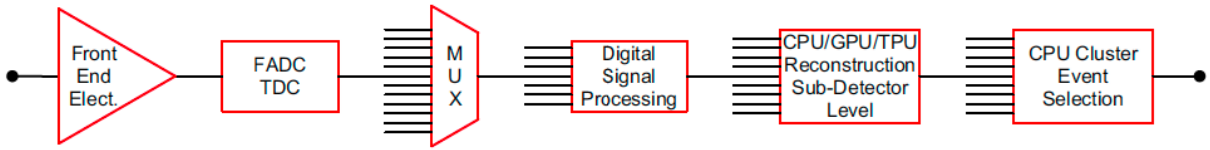
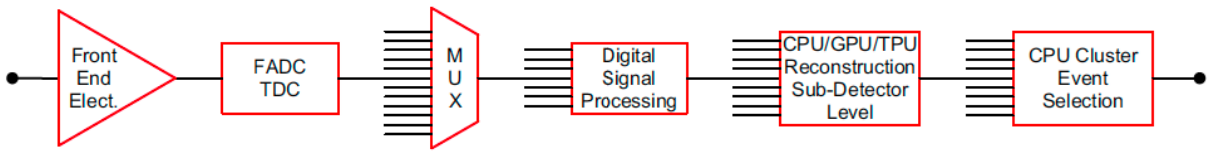


Image - Marco Battaglieri, CFNS Summer School 2024

Readout: Triggerless

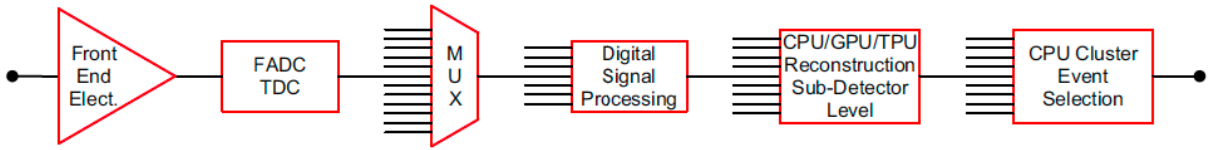
- ePIC is a general purpose detector
- I.e. we don't want to arbitrarily exclude things
  - So a selective trigger is a problem...
- Use triggerless, streaming readout instead



- Effectively, a lot of the "triggering" is shunted to software

Readout: Triggerless

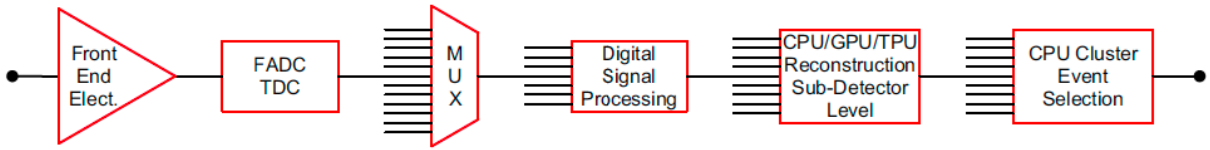
- ePIC is a general purpose detector
- I.e. we don't want to arbitrarily exclude things
  - So a selective trigger is a problem...
- Use triggerless, streaming readout instead



- Effectively, a lot of the "triggering" is shunted to software
- The concept of a distinct **event** is lost
- Instead, deal with **timeframes**
  - 1 timeframe - Max 2^{16} bunch crossings (660 μs)
  - $\sim$ 5 MB per timeframe

Readout: Triggerless

- ePIC is a general purpose detector
- I.e. we don't want to arbitrarily exclude things
  - So a selective trigger is a problem...
- Use triggerless, streaming readout instead



- Effectively, a lot of the "triggering" is shunted to software
- The concept of a distinct **event** is lost
- Instead, deal with **timeframes**
  - 1 timeframe - Max 2^{16} bunch crossings (660 μs)
  - $\sim$ 5 MB per timeframe
  - I'll let you extrapolate from those figures...

Reconstruction: Calorimetry

- Software needed to reconstruct events in detector systems
- Challenge is different for each type of system

Reconstruction: Calorimetry

- Software needed to reconstruct events in detector systems
- Challenge is different for each type of system
- For calorimeters, task is usually to identify clusters
  - Calorimeters typically segmented
  - Need to group together hits into clusters

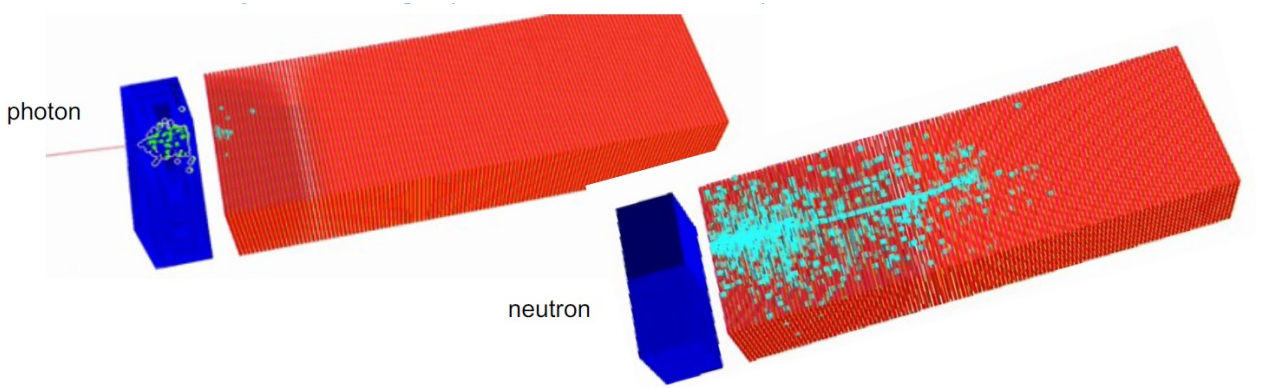
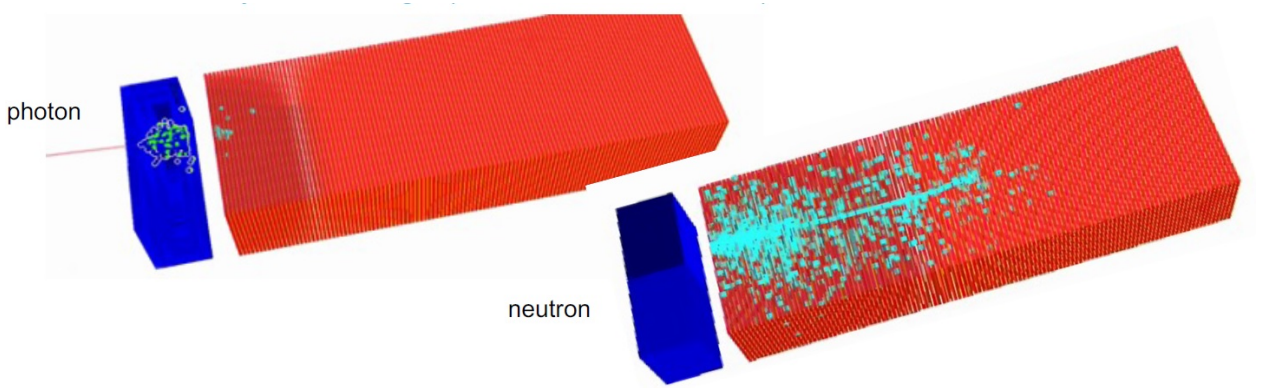


Image - A. Jentsch, BNL

Reconstruction: Calorimetry

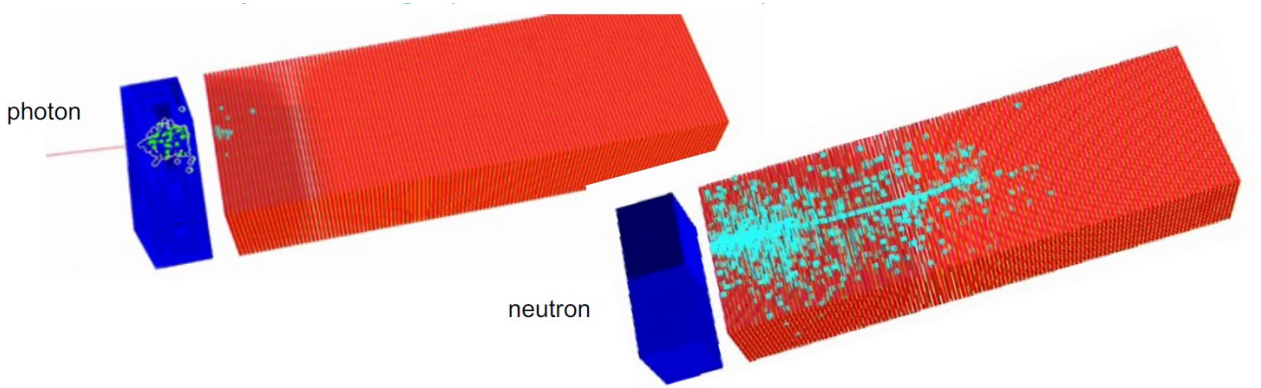
- Software needed to reconstruct events in detector systems
- Challenge is different for each type of system
- For calorimeters, task is usually to identify clusters
  - Calorimeters typically segmented
  - Need to group together hits into clusters



- Key cluster parameters: Cluster energy and Cluster position

Reconstruction: Calorimetry

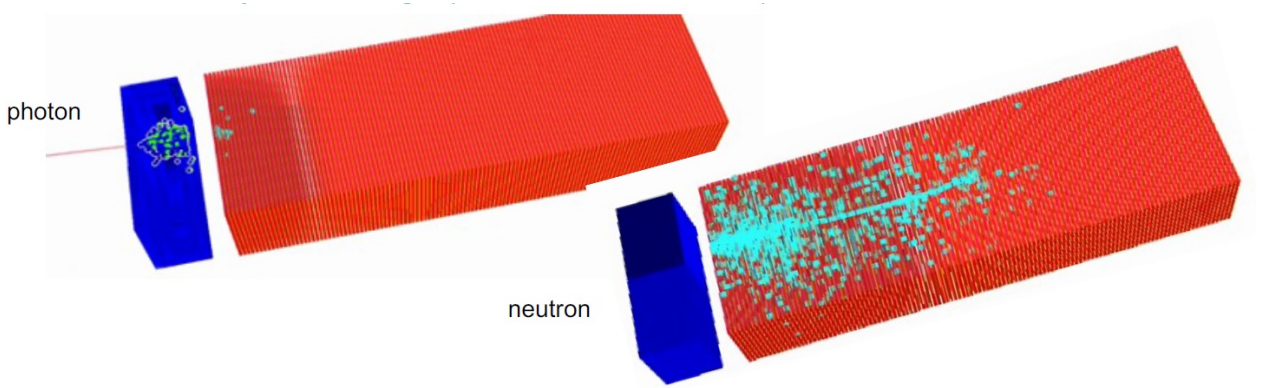
- Software needed to reconstruct events in detector systems
- Challenge is different for each type of system
- For calorimeters, task is usually to identify clusters
  - Calorimeters typically segmented
  - Need to group together hits into clusters



- Key cluster parameters: Cluster energy and Cluster position
- Range of algorithms can be used to reconstruct clusters

Reconstruction: Calorimetry

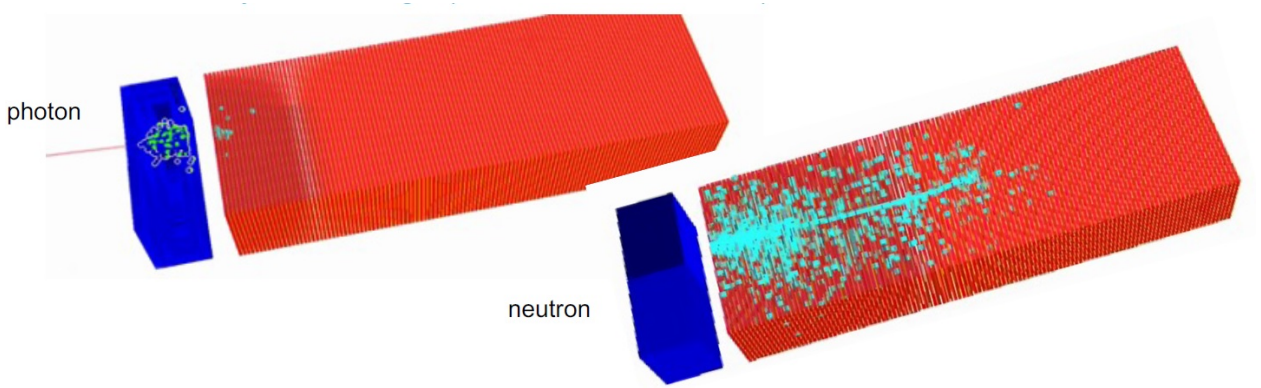
- Software needed to reconstruct events in detector systems
- Challenge is different for each type of system
- For calorimeters, task is usually to identify clusters
  - Calorimeters typically segmented
  - Need to group together hits into clusters



- Key cluster parameters: Cluster energy and Cluster position
- Range of algorithms can be used to reconstruct clusters
- Shape of cluster can also be indicative of incident particle

Reconstruction: Calorimetry

- Software needed to reconstruct events in detector systems
- Challenge is different for each type of system
- For calorimeters, task is usually to identify clusters
  - Calorimeters typically segmented
  - Need to group together hits into clusters



- Key cluster parameters: Cluster energy and Cluster position
- Range of algorithms can be used to reconstruct clusters
- Shape of cluster can also be indicative of incident particle
  - Potential AI/ML application

Reconstruction: Tracks

- Idea is easy, connect the dots and make a track
- Reality is more painful

Reconstruction: Tracks

- Idea is easy, connect the dots and make a track
- Reality is more painful
- Many hits per event $\rightarrow$ **a lot of possible tracks**
- Geometry of tracking layers may not be straightforward
- Need tracking algorithms to filter through junk and decide on which tracks are "good"

Reconstruction: Tracks

- Idea is easy, connect the dots and make a track
- Reality is more painful
- Many hits per event $\rightarrow$ **a lot of possible tracks**
- Geometry of tracking layers may not be straightforward
- Need tracking algorithms to filter through junk and decide on which tracks are "good"

- Find and fit tracks
- More detailed summaries here and here

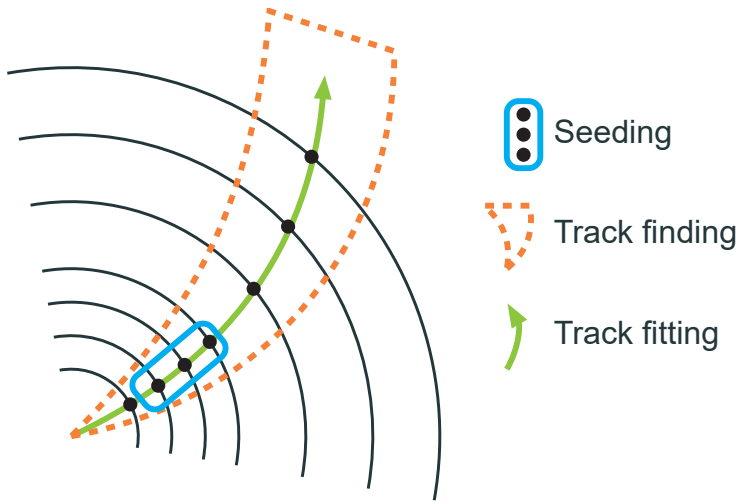
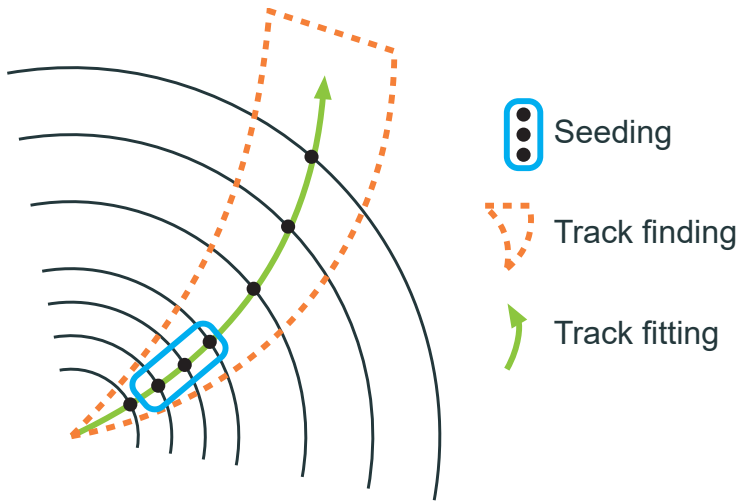


Image - ACTS Project

Reconstruction: Tracks

- Idea is easy, connect the dots and make a track
- Reality is more painful
- Many hits per event $\rightarrow$ **a lot of possible tracks**
- Geometry of tracking layers may not be straightforward
- Need tracking algorithms to filter through junk and decide on which tracks are "good"

- Find and fit tracks
- More detailed summaries here and here
- ePIC utilises ACTS for track identification and reconstruction



Reconstruction: PID

- As with tracking, straightforward in principle
- Once again, reality isn't so nice

Reconstruction: PID

- As with tracking, straightforward in principle
- Once again, reality isn't so nice
- Could have contradictory information
- Noise, background events
- Cannot be 100% certain of PID assignment

Reconstruction: PID

- As with tracking, straightforward in principle
- Once again, reality isn't so nice
- Could have contradictory information
- Noise, background events
- Cannot be 100% certain of PID assignment
- Detector might have multiple overlapping signals
  - E.g. in a RICH, multiple rings

Reconstruction: PID

- As with tracking, straightforward in principle
- Once again, reality isn't so nice
- Could have contradictory information
- Noise, background events
- Cannot be 100% certain of PID assignment
- Detector might have multiple overlapping signals
  - E.g. in a RICH, multiple rings
- As with tracking and calorimetry, often need to utilise reconstruction algorithms to interpret information from dedicated PID detectors
- Can try to quantify confidence of assignment in some way

Analysis: Combining Our Reconstructed Info

- Clearly, we need software for various reconstruction tasks
- Ideally, all of our reconstruction software should provide us with output in a common format
- We can then further analyse and interpret this

Analysis: Combining Our Reconstructed Info

- Clearly, we need software for various reconstruction tasks
- Ideally, all of our reconstruction software should provide us with output in a common format
- We can then further analyse and interpret this
- For the EIC/ePIC, EICrecon is the software we utilise to reconstruct detector output
- Once we reconstruct the detector output, **we can analyse it to extract our events of interest**

Analysis: Combining Our Reconstructed Info

- Clearly, we need software for various reconstruction tasks
- Ideally, all of our reconstruction software should provide us with output in a common format
- We can then further analyse and interpret this
- For the EIC/ePIC, EICrecon is the software we utilise to reconstruct detector output
- Once we reconstruct the detector output, **we can analyse it to extract our events of interest**

- We might reconstruct a lot of particles in our event
- But we may only care about certain particle combinations

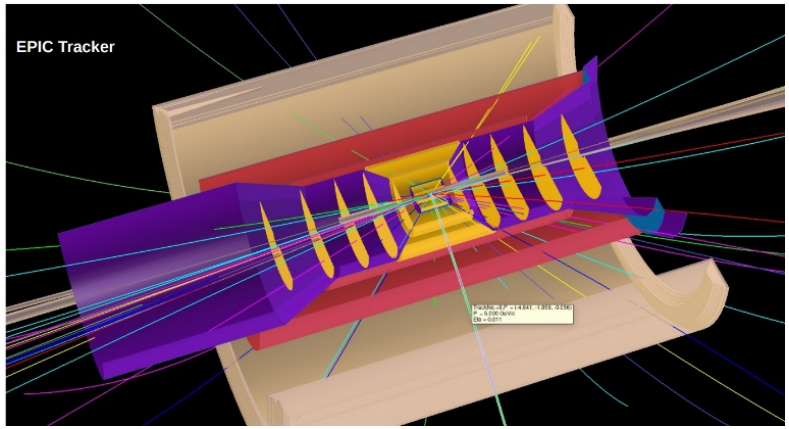
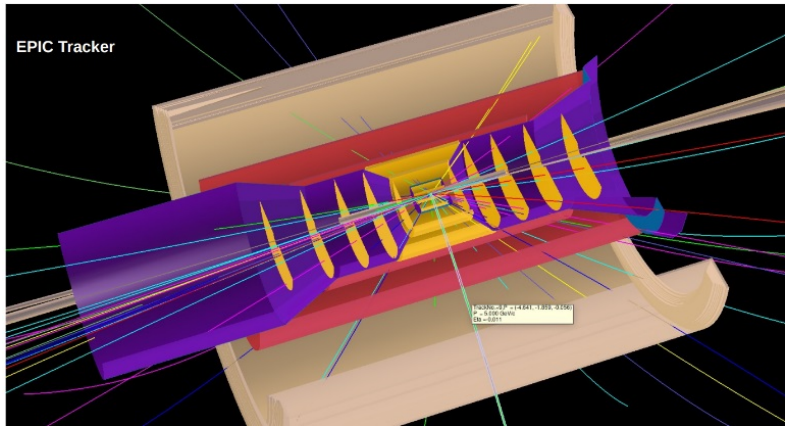


Image - ePIC/EICrecon

Analysis: Combining Our Reconstructed Info

- Clearly, we need software for various reconstruction tasks
- Ideally, all of our reconstruction software should provide us with output in a common format
- We can then further analyse and interpret this
- For the EIC/ePIC, EICrecon is the software we utilise to reconstruct detector output
- Once we reconstruct the detector output, **we can analyse it to extract our events of interest**

- We might reconstruct a lot of particles in our event
- But we may only care about certain particle combinations
- Or perhaps only care about certain kinematics



Processing Data

- As we saw from the readout discussion, we have a lot of data!
- Typically use a distributed data processing model
- ePIC is no different

Processing Data

- As we saw from the readout discussion, we have a lot of data!
- Typically use a distributed data processing model
- ePIC is no different

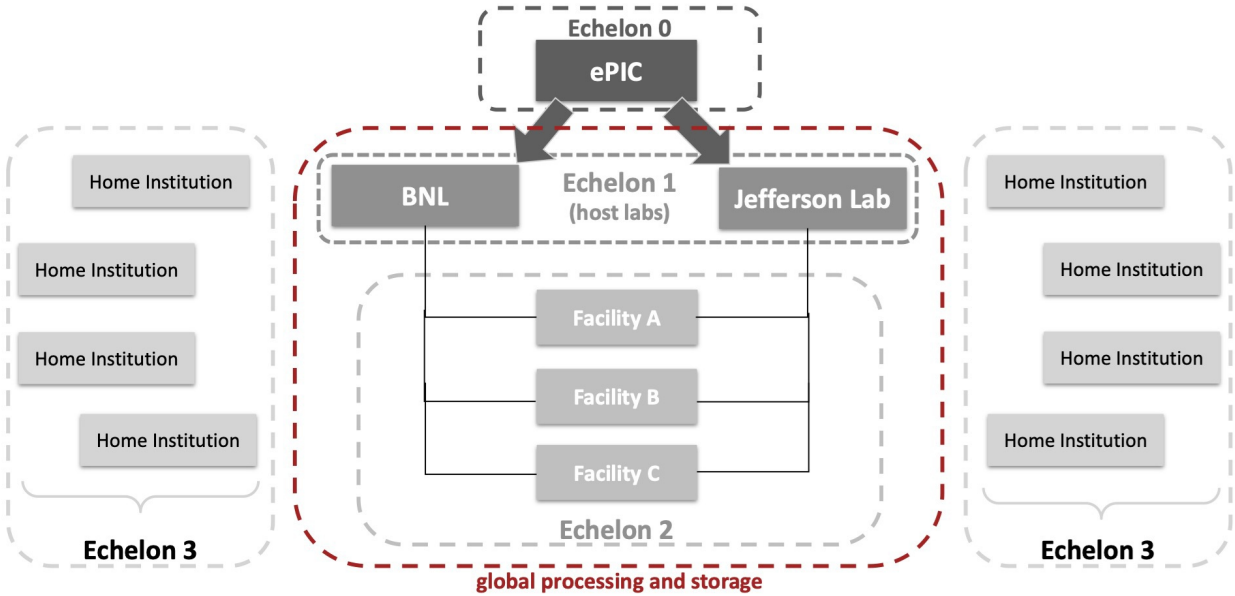
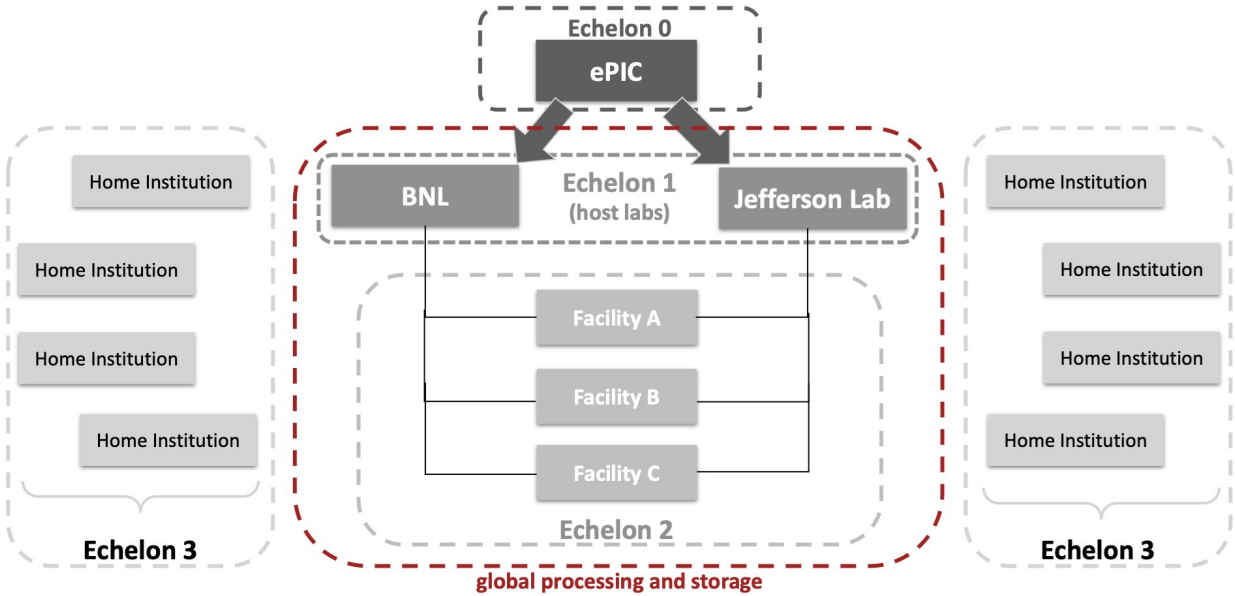


Image - The ePIC Streaming Computing Model v2

Processing Data

- As we saw from the readout discussion, we have a lot of data!
- Typically use a distributed data processing model
- ePIC is no different



- Once we eventually get to our reconstructed data, need tools to analyse our output

Processing Data: Analysis Tools

- Analysing HEP data is not a "new" problem
- Lots of existing tools and frameworks to utilise

Processing Data: Analysis Tools

- Analysing HEP data is not a "new" problem
- Lots of existing tools and frameworks to utilise
- CERN ROOT is very commonly used

- Multiple ways to utilise ROOT

Image - CERN ROOT

Processing Data: Analysis Tools

- Analysing HEP data is not a "new" problem
- Lots of existing tools and frameworks to utilise
- CERN ROOT is very commonly used

- Multiple ways to utilise ROOT
  - Use ROOT directly
  - PyRoot,
  - Python/Uproot
  - RDataFrames

Working with Simulated Data

- In addition to real detector data, we might want to look at simulations for various reasons
  - Aid in design and development
  - Compare output to models/ideal case

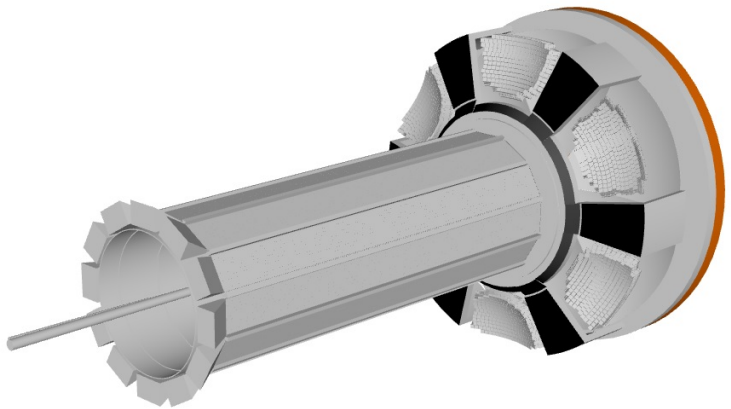
Working with Simulated Data

- In addition to real detector data, we might want to look at simulations for various reasons
  - Aid in design and development
  - Compare output to models/ideal case
- Geant4 is a very common physics simulation framework

Working with Simulated Data

- In addition to real detector data, we might want to look at simulations for various reasons
  - Aid in design and development
  - Compare output to models/ideal case
- Geant4 is a very common physics simulation framework

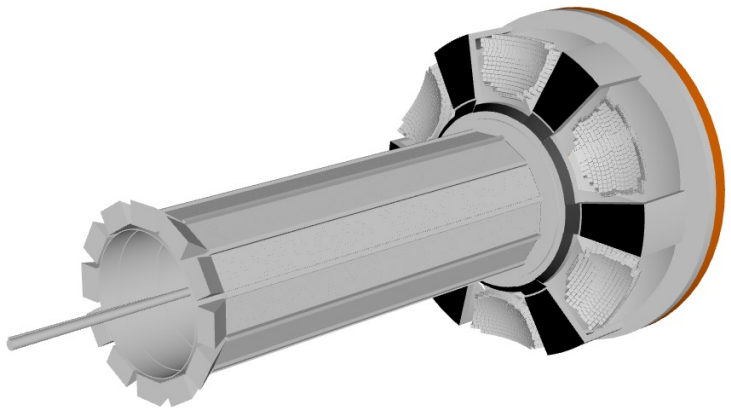
- Can build and model full detector simulations
- Reconstruct and analyse data as though it is real
- Direct comparison to "real" data (if it is available)



Working with Simulated Data

- In addition to real detector data, we might want to look at simulations for various reasons
  - Aid in design and development
  - Compare output to models/ideal case
- Geant4 is a very common physics simulation framework

- Can build and model full detector simulations
- Reconstruct and analyse data as though it is real
- Direct comparison to "real" data (if it is available)
- Also allows us to develop reconstruction/analysis before a detector is built



Working with Simulated Data: MC Truth

- When analysing real data, we don't know with 100% certainty what actually happened in our event
- Key difference with simulated data though

Image - The Simpsons, Season 9 Episode 9, "Realty Bites"

Working with Simulated Data: MC Truth

- When analysing real data, we don't know with 100% certainty what actually happened in our event
- Key difference with simulated data though
- We *do* **know** the actual truth
- We know exactly what we have fed in to the simulation

Image - The Simpsons, Season 9 Episode 9, "Realty Bites"

Working with Simulated Data: MC Truth

- When analysing real data, we don't know with 100% certainty what actually happened in our event
- Key difference with simulated data though
- We *do* **know** the actual truth
- We know exactly what we have fed in to the simulation

- If we process simulated data in the same way as "real" data, we can compare it to the truth
- Look at efficiency, resolution, acceptance...

Working with Simulated Data: MC Truth

- When analysing real data, we don't know with 100% certainty what actually happened in our event
- Key difference with simulated data though
- We *do* **know** the actual truth
- We know exactly what we have fed in to the simulation

- If we process simulated data in the same way as "real" data, we can compare it to the truth
- Look at efficiency, resolution, acceptance...
- **Not** something we can do for real data of course, but a vital tool for projections

Workshop Overview

- Sadly, ePIC and the EIC don't exist, yet
- However, we have a very advanced simulation of how things are likely to be

Workshop Overview

- Sadly, ePIC and the EIC don't exist, yet
- However, we have a very advanced simulation of how things are likely to be
- So in our workshop, we'll be using this simulation
- We will cover:

Workshop Overview

- Sadly, ePIC and the EIC don't exist, yet
- However, we have a very advanced simulation of how things are likely to be
- So in our workshop, we'll be using this simulation
- We will cover:
  - Wednesday: The **output of our simulation** and how we can **analyse** it

Workshop Overview

- Sadly, ePIC and the EIC don't exist, yet
- However, we have a very advanced simulation of how things are likely to be
- So in our workshop, we'll be using this simulation
- We will cover:
  - Wednesday: The **output of our simulation** and how we can **analyse** it
  - Thursday: A closer look at the **ePIC detector** and how we reconstruct events, **focus on PID**

Workshop Overview

- Sadly, ePIC and the EIC don't exist, yet
- However, we have a very advanced simulation of how things are likely to be
- So in our workshop, we'll be using this simulation
- We will cover:
  - Wednesday: The **output of our simulation** and how we can **analyse** it
  - Thursday: A closer look at the **ePIC detector** and how we reconstruct events, **focus on PID**
  - Friday: **Physics analysis**, what we do need to consider for in a physics analysis? How can we structure and write an analysis code?

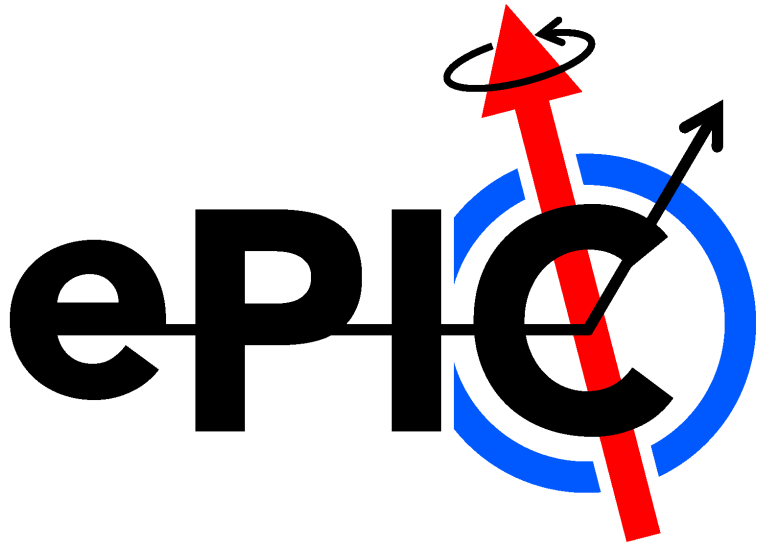
Workshop Overview

- Sadly, ePIC and the EIC don't exist, yet
- However, we have a very advanced simulation of how things are likely to be
- So in our workshop, we'll be using this simulation
- We will cover:
  - Wednesday: The **output of our simulation** and how we can **analyse** it
  - Thursday: A closer look at the **ePIC detector** and how we reconstruct events, **focus on PID**
  - Friday: **Physics analysis**, what we do need to consider for in a physics analysis? How can we structure and write an analysis code?
  - Saturday: Working time on an analysis task

Workshop Overview

- Sadly, ePIC and the EIC don't exist, yet
- However, we have a very advanced simulation of how things are likely to be
- So in our workshop, we'll be using this simulation
- We will cover:
  - Wednesday: The **output of our simulation** and how we can **analyse** it
  - Thursday: A closer look at the **ePIC detector** and how we reconstruct events, **focus on PID**
  - Friday: **Physics analysis**, what we do need to consider for in a physics analysis? How can we structure and write an analysis code?
  - Saturday: Working time on an analysis task
- Give you the tools and the techniques to analyse ePIC data

Thanks for listening, any questions before we move on?



stephen.kay@york.ac.uk

Ok then, time to try setting up our environment for the tutorials!