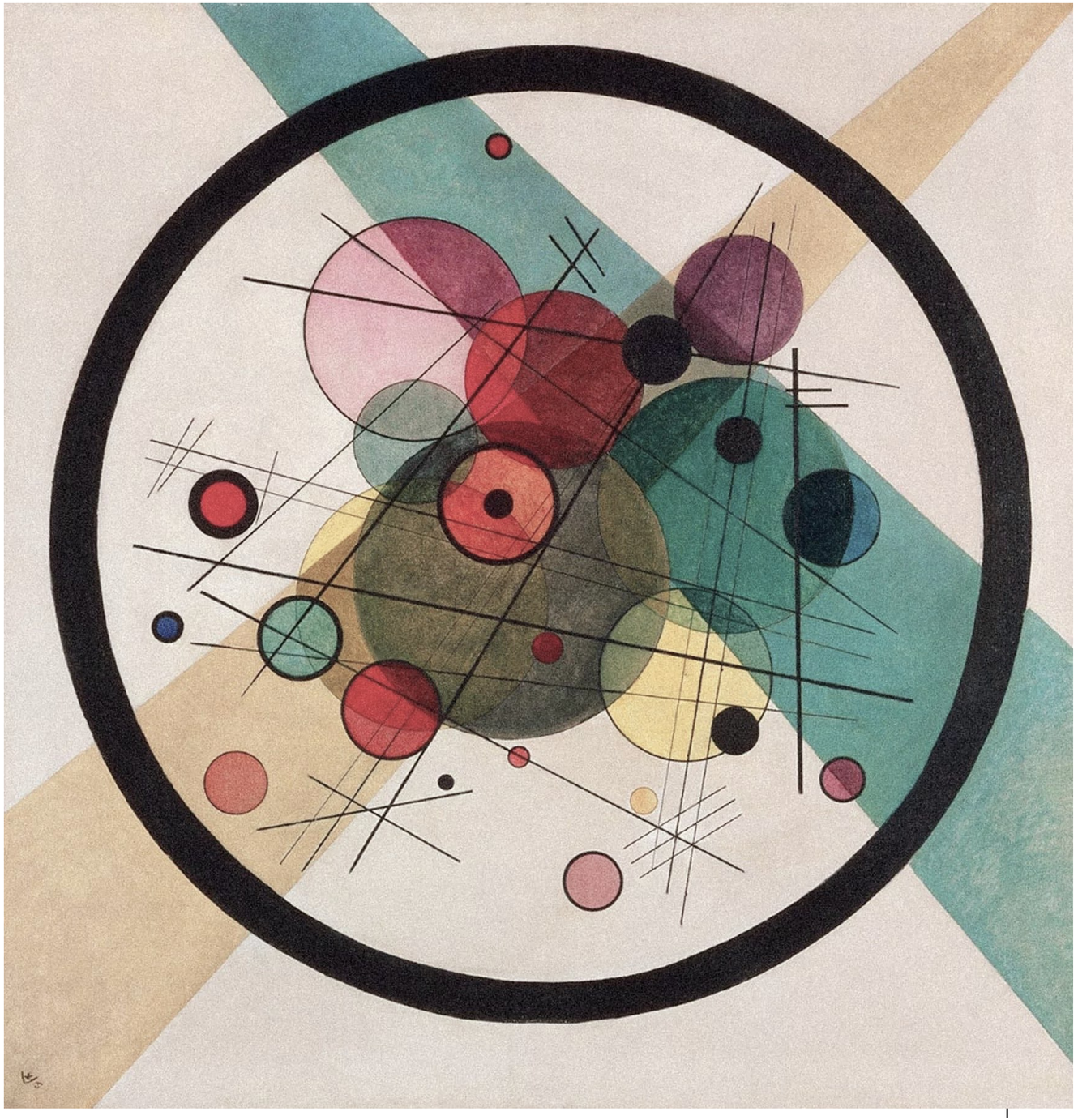


SCUOLA
NORMALE
SUPERIORE

using the new RDataFrame in a physics study:
a new W mass analysis

Elisabetta Manca

(Scuola Normale Superiore)

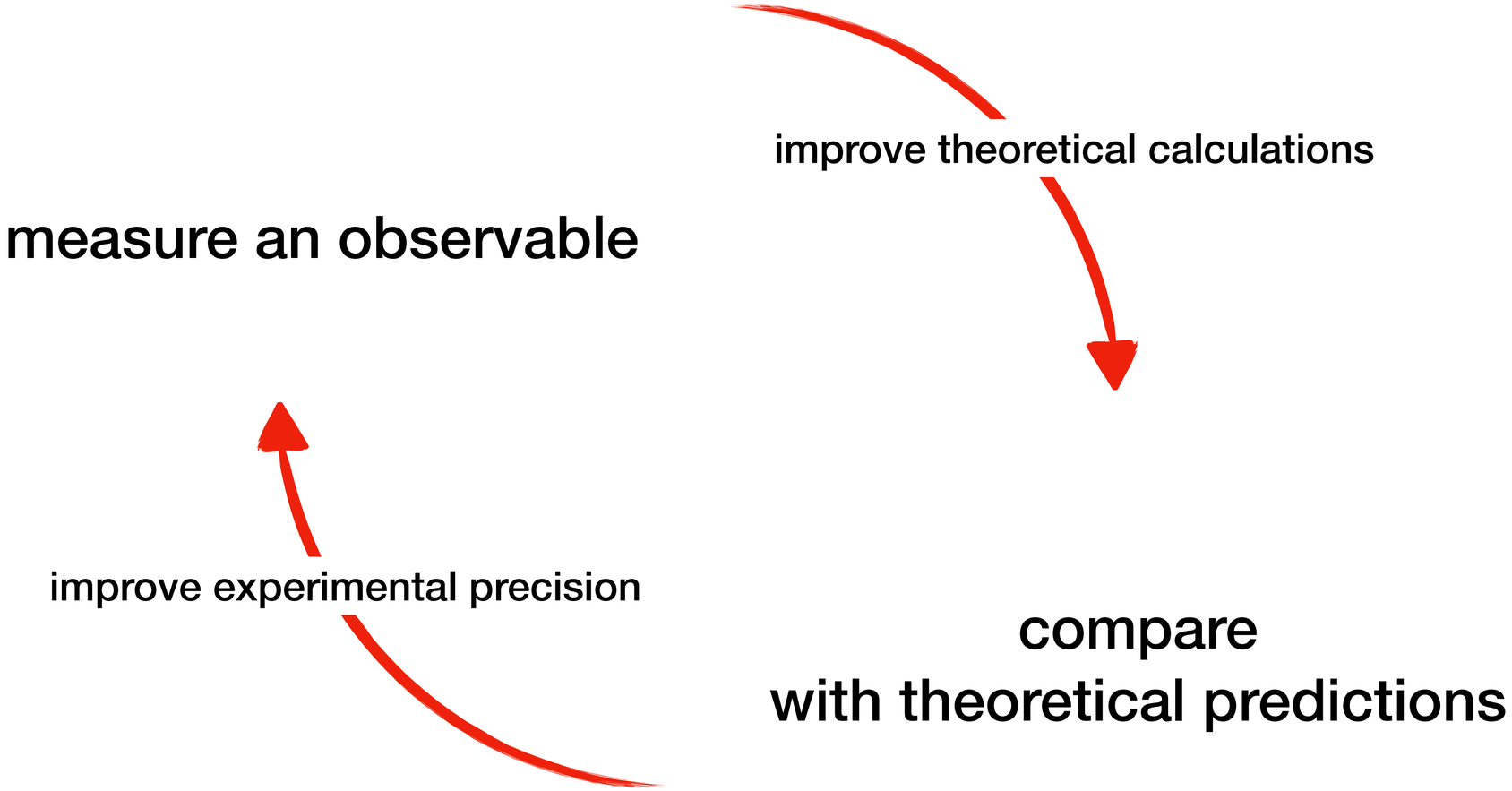
EP software seminar
CERN, 16th October 2019

advance in High Energy Physics using precision measurements

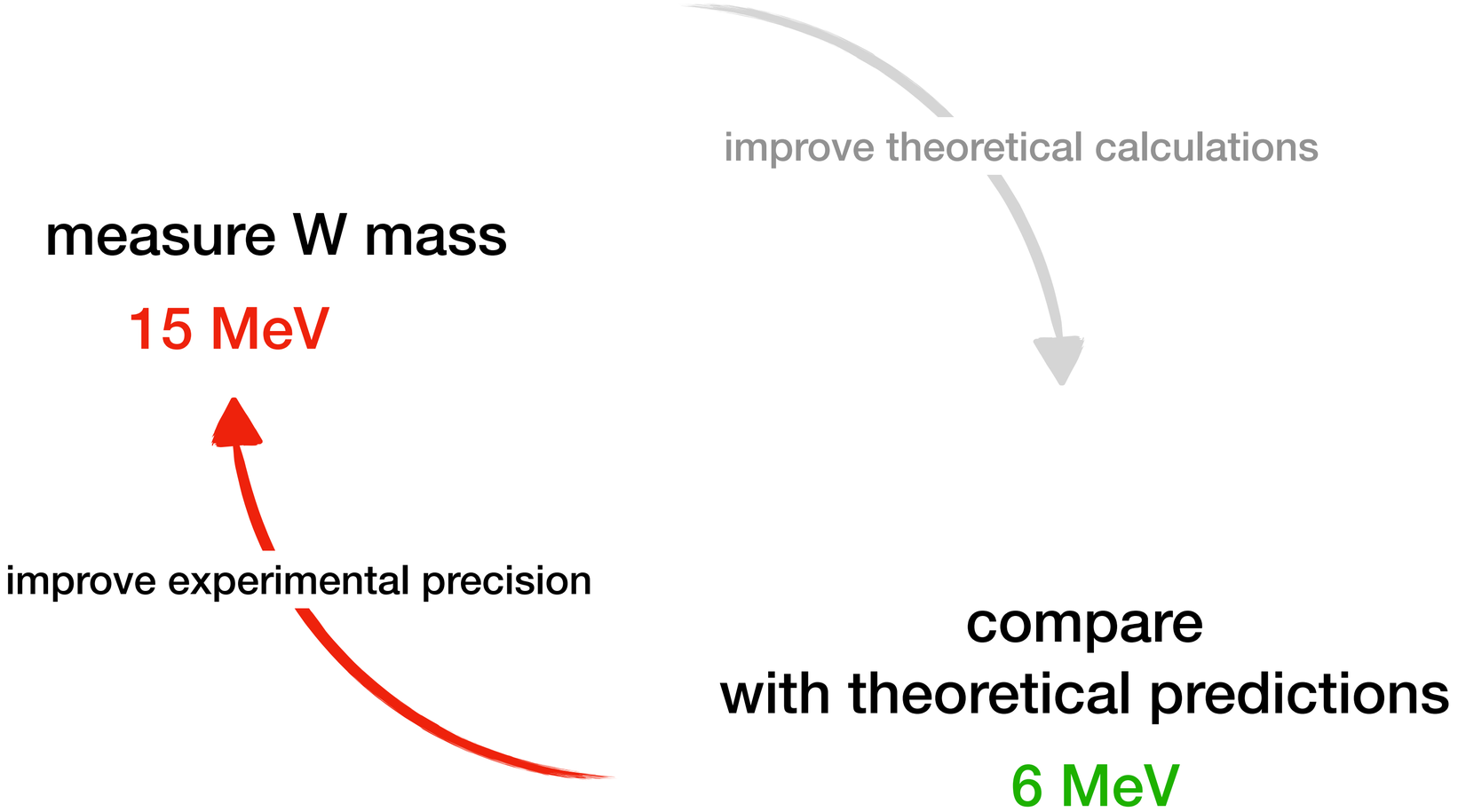
measure an observable

compare
with theoretical predictions

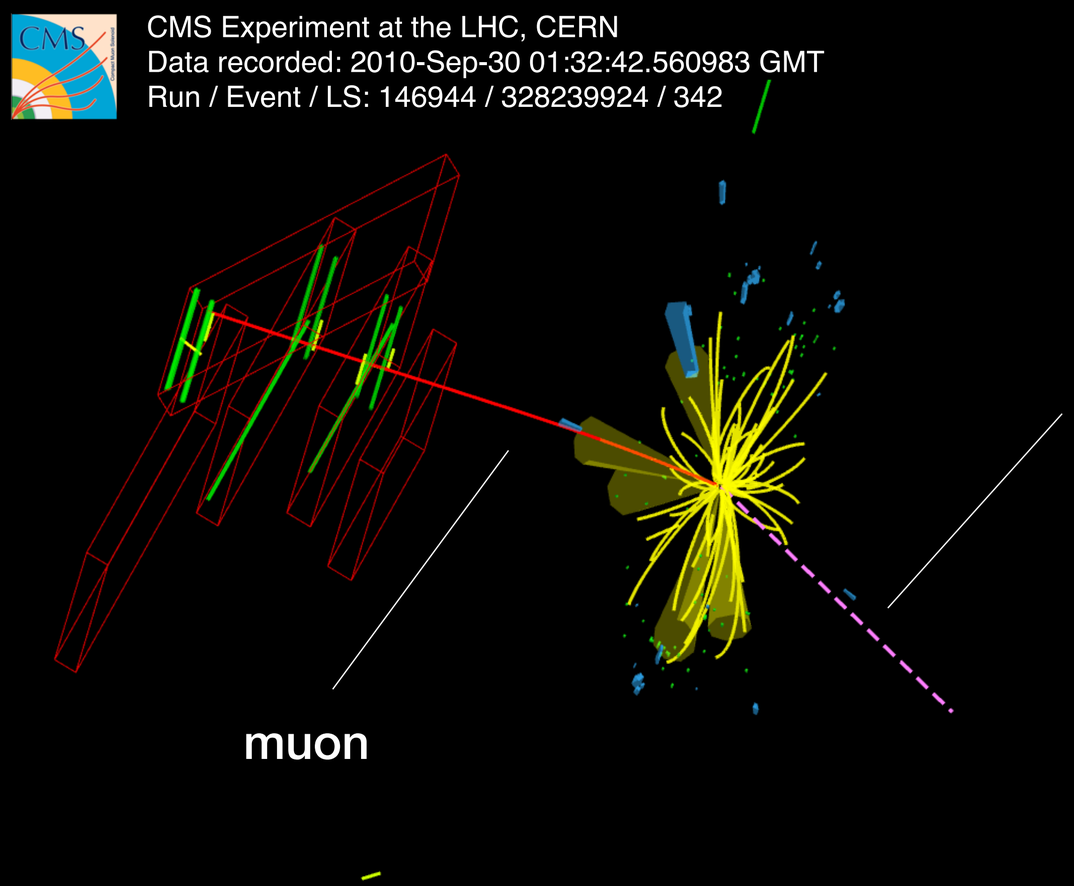
advance in High Energy Physics using precision measurements



advance in High Energy Physics using precision measurements

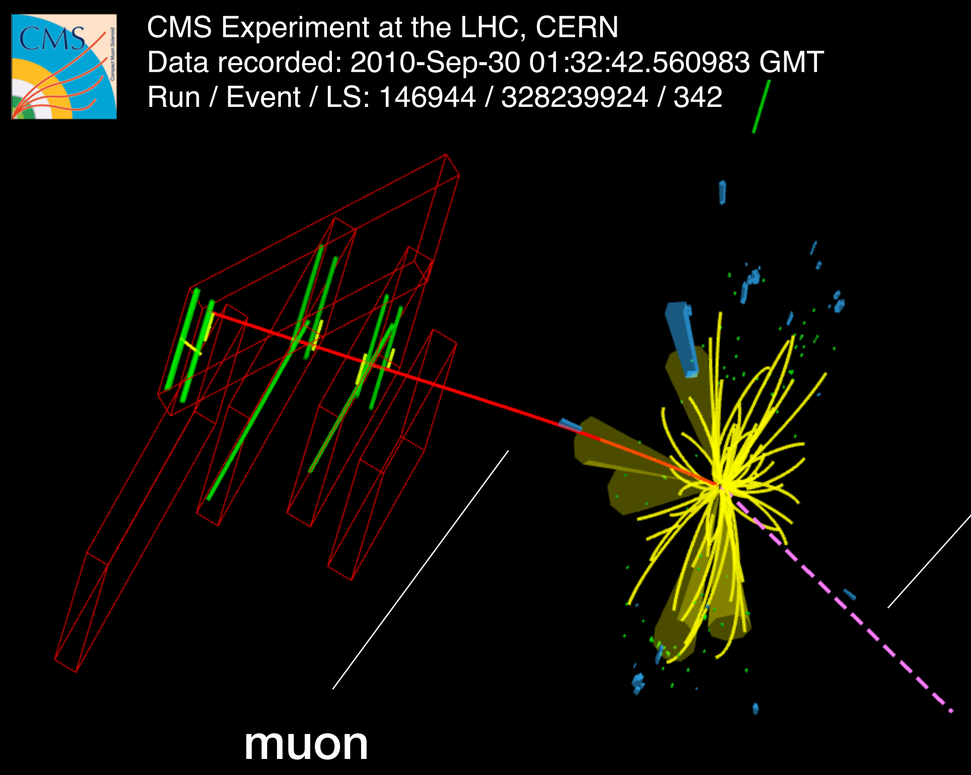


how do I measure the W boson mass?



missing transverse energy
proxy for neutrino p_T
with poor scale and resolution control

how do I measure the W boson mass?



muon

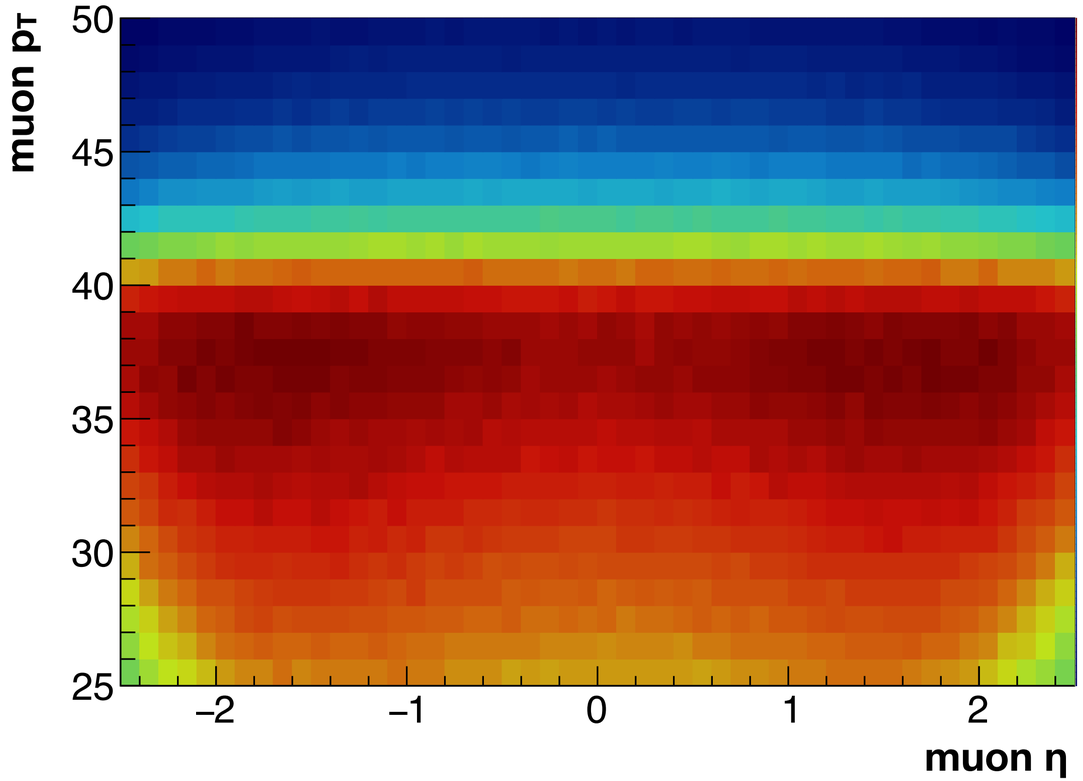
challenge: measure p_T scale with unprecedented precision

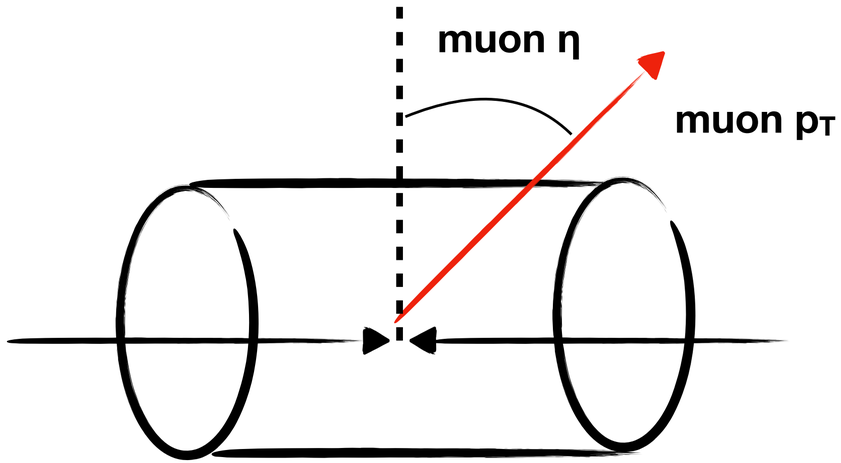
missing transverse energy
proxy for neutrino p_T
with poor scale and resolution control

transverse observables
do not give Lorentz invariants!

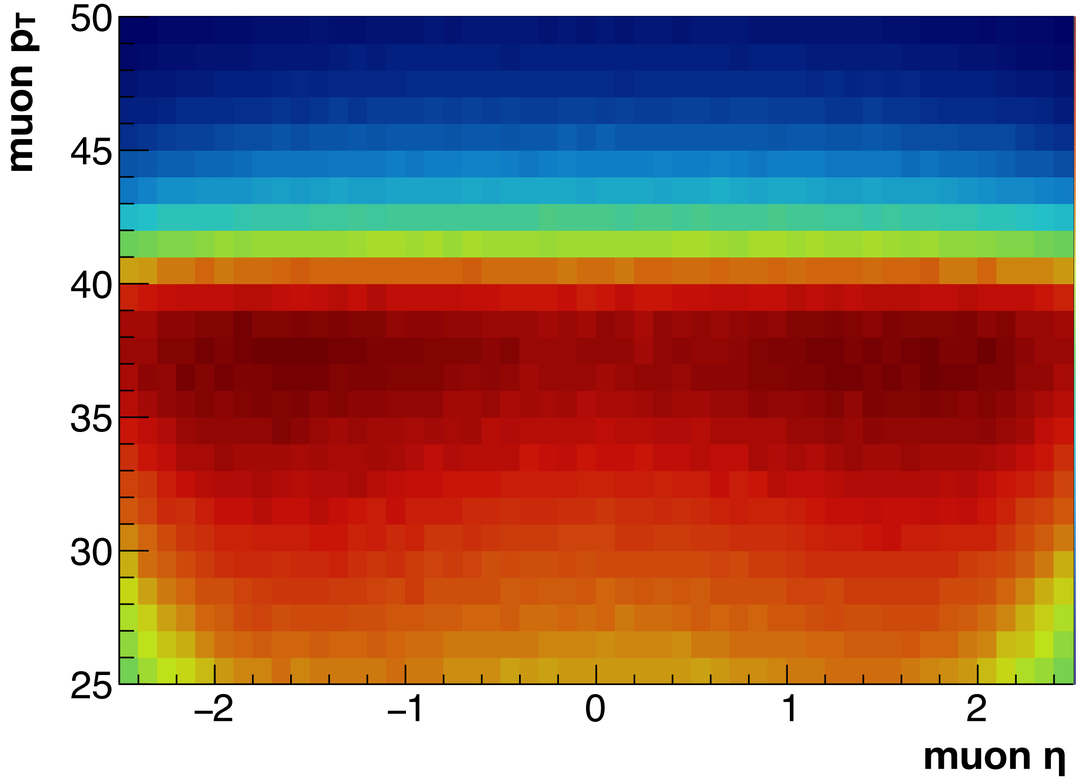
challenge: reduce systematic uncertainties due to this

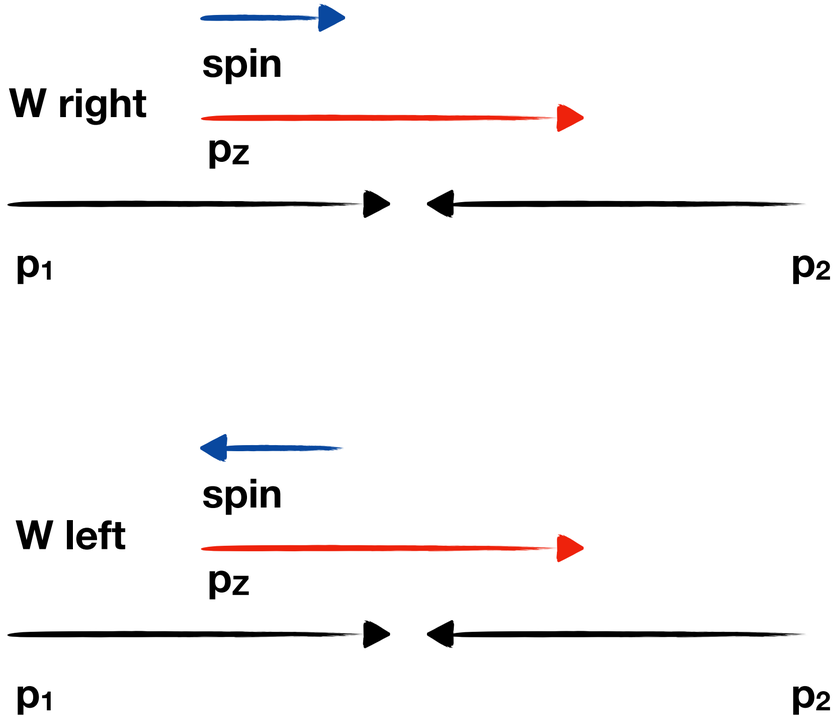
a simple
but powerful idea





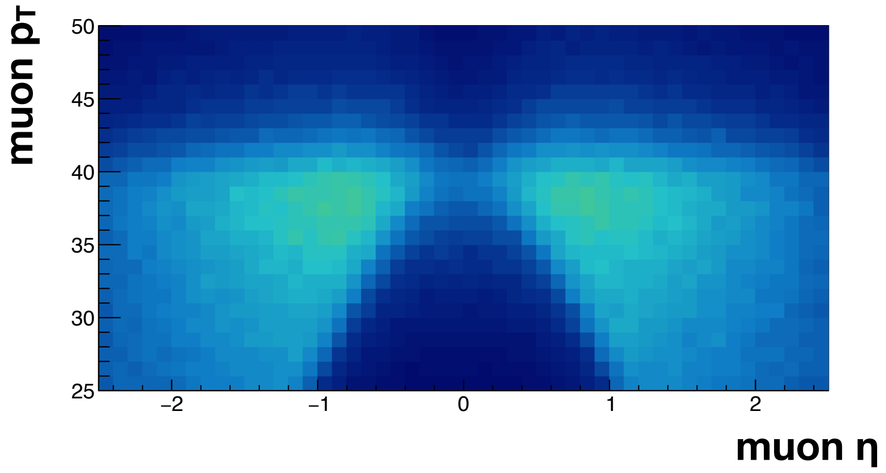
constrain production model directly from data

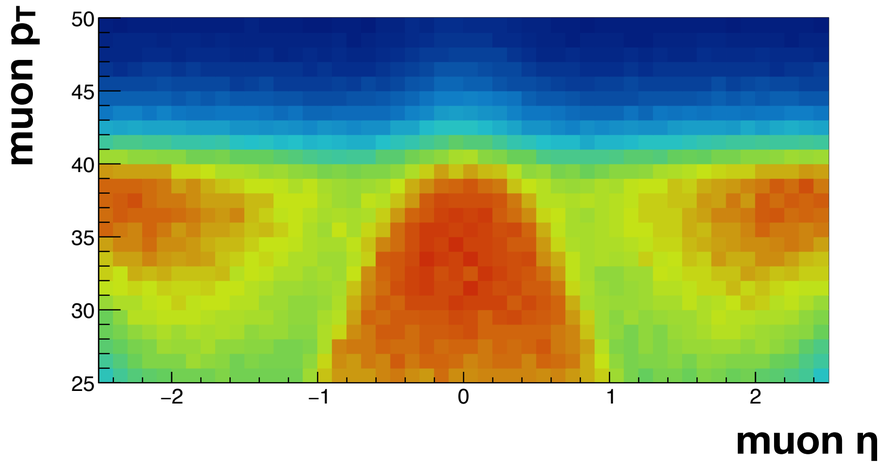


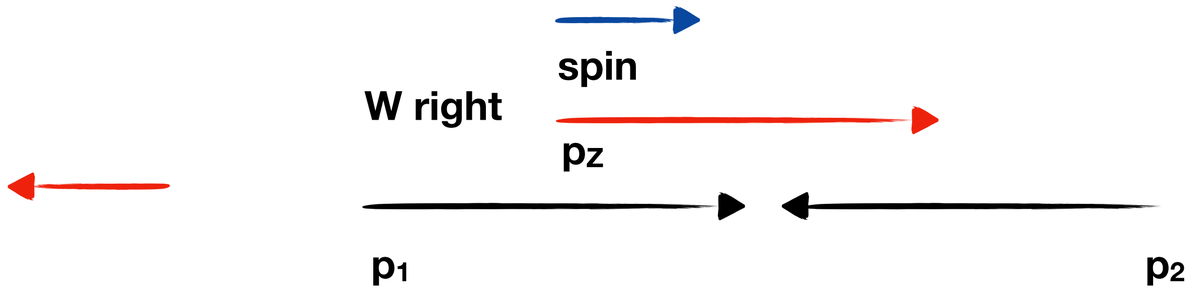


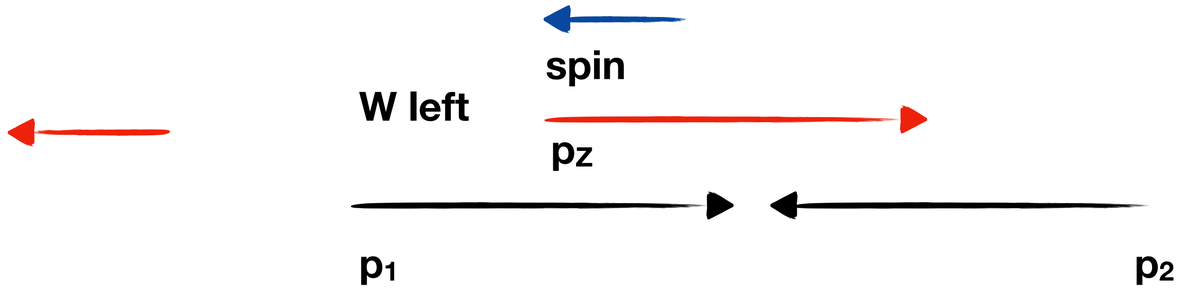
constrain production model directly from data

E.M. et al. J. High Energ. Phys. (2017) 2017: 130.

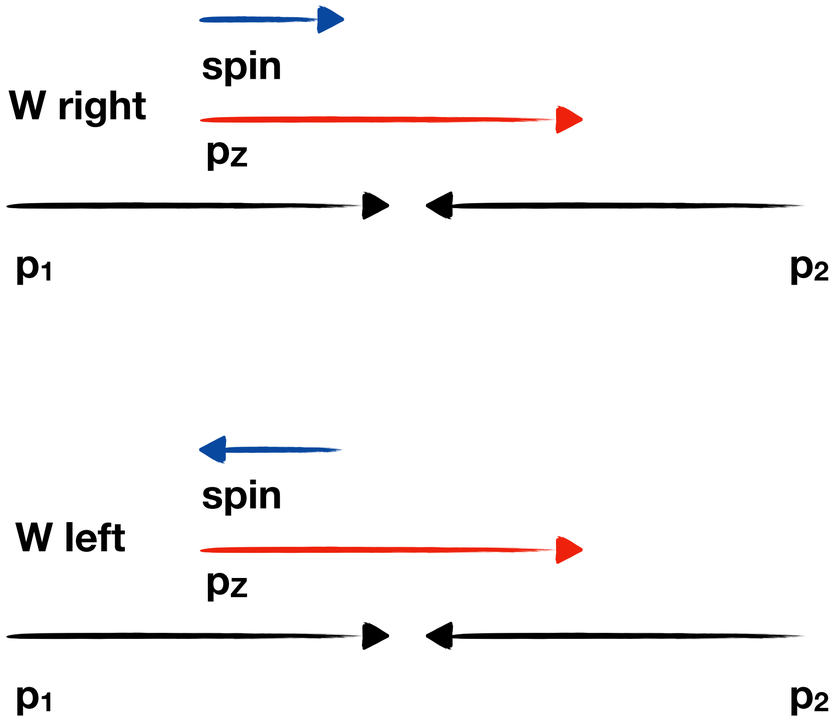




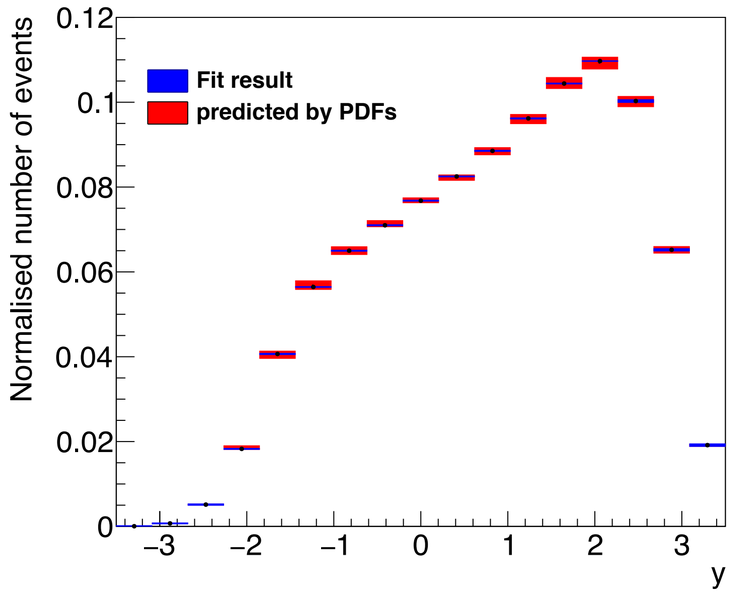




unfold rapidity and helicity distribution of W boson

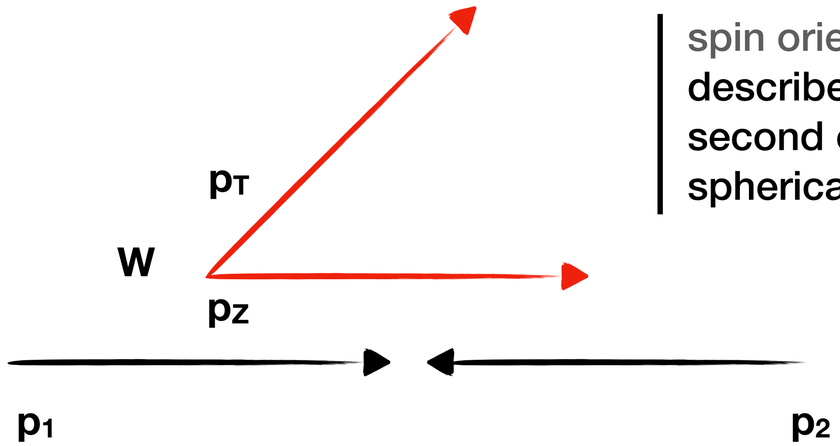


$$y = \frac{1}{2}\log\frac{E + p_z}{E - p_z}$$



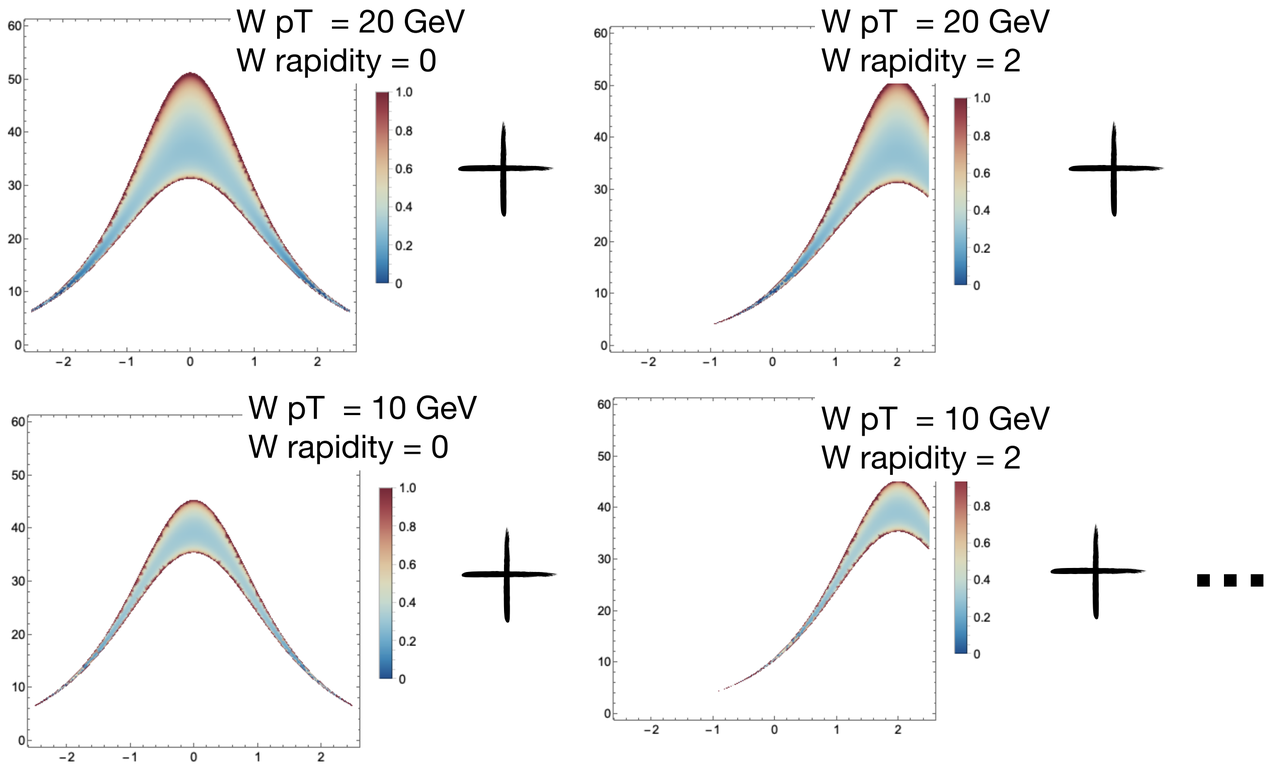
E.M. et al. J. High Energ. Phys. (2017) 2017: 130.

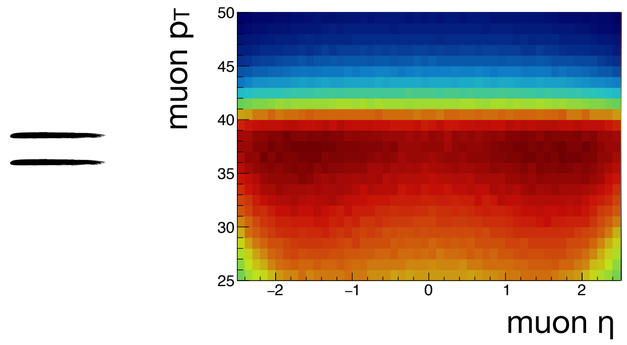
constrain production model directly from data

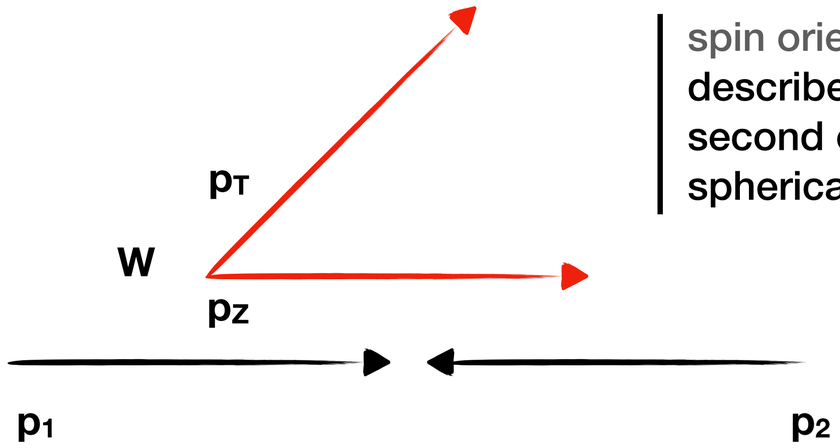


spin orientation described by second order spherical harmonics

constrain production model directly from data



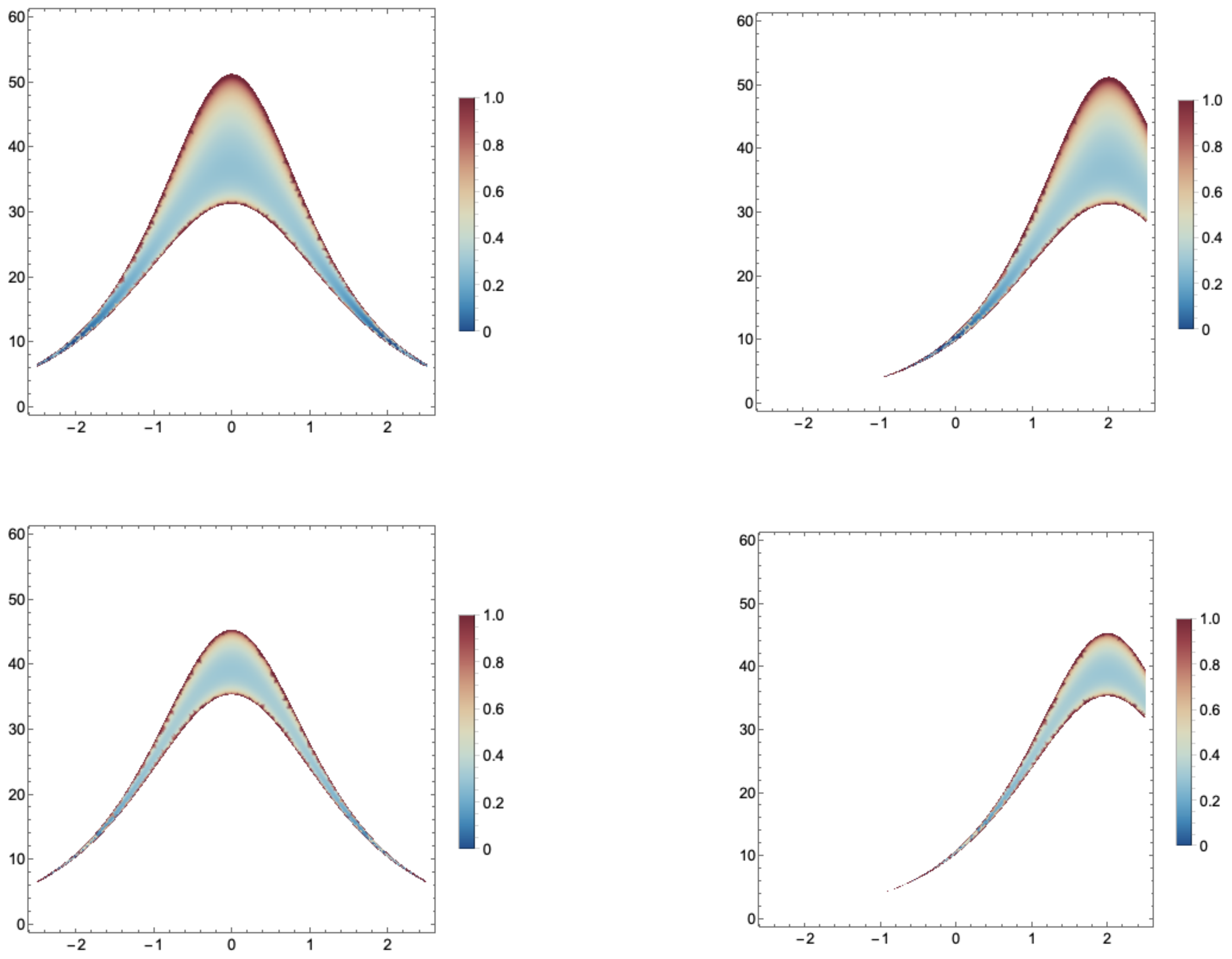




spin orientation described by second order spherical harmonics

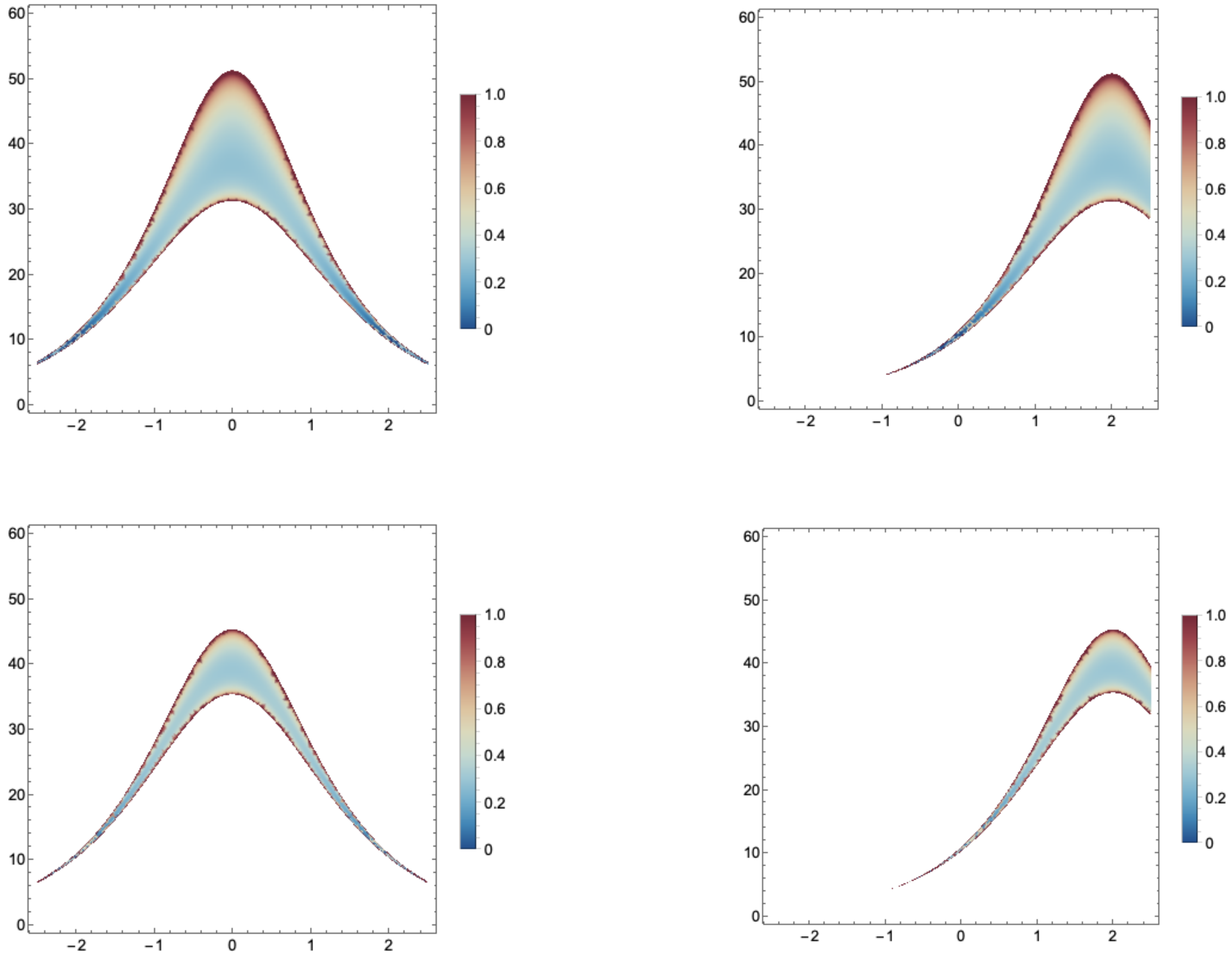
reconstruct the W production

from multiple copies of this plot



~10^9	events to analyse
~10^3	nominal 2D histograms
~10^3	variations to each histogram

reconstruct the W production
from multiple copies of this plot



$\sim 10^9$ events to analyse

$\sim 10^3$ nominal 2D histograms

$\sim 10^3$ variations to each histogram

5 MeV uncertainty on W mass (from toy studies)

requirements for an analysis
in a physicist's wishlist

fast

tidy

flexible

reusable

answer from the ROOT team: the new RDataFrame

fast **tidy** **flexible** **reusable**

parallelisable

transparent MT supported on multi-core machines

declarative

management of dependencies among objects without showing control-flow

clear workflow

graph-style organisation of the analysis

customisable

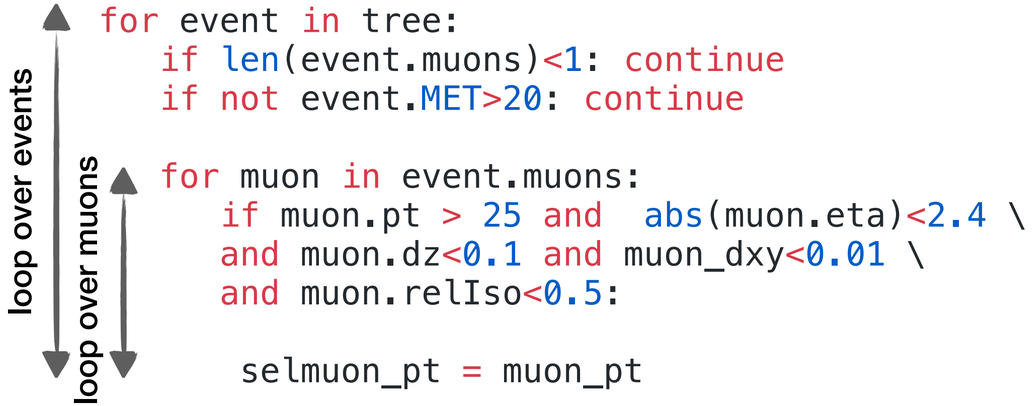
easy to write code in python or c++ to execute whatever action

optimised

loop on input file only when analysis is set

be prepared
for a massive change

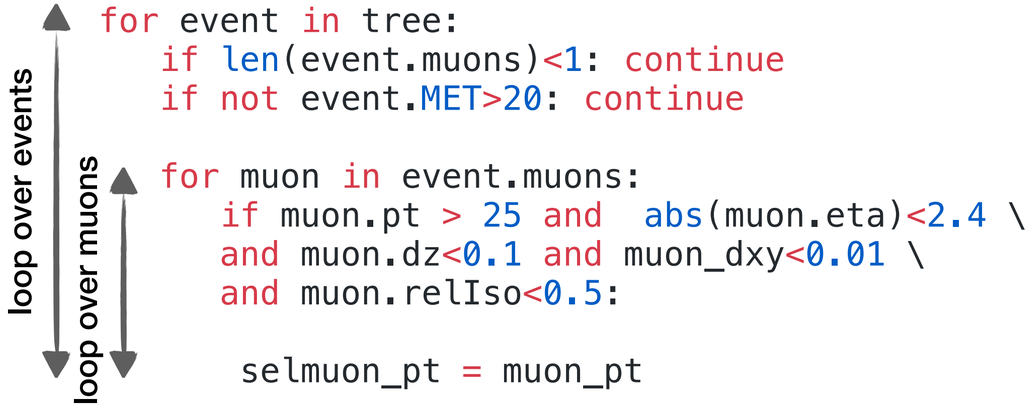
traditional way




```
import ROOT

fIn = ROOT.TFile.Open("file.root")
tree = fIn.tree

for event in tree:
   if len(event.muons)<1: continue
   if not event.MET>20: continue

   for muon in event.muons:
      if muon.pt > 25 and  abs(muon.eta)<2.4 \
      and muon.dz<0.1 and muon_dxy<0.01 \
      and muon.relIso<0.5:

       selmuon_pt = muon_pt
```

be prepared
for a massive change

traditional way




```
import ROOT

fIn = ROOT.TFile.Open("file.root")
tree = fIn.tree

for event in tree:
   if len(event.muons)<1: continue
   if not event.MET>20: continue

   for muon in event.muons:
      if muon.pt > 25 and  abs(muon.eta)<2.4 \
      and muon.dz<0.1 and muon_dxy<0.01 \
      and muon.relIso<0.5:

       selmuon_pt = muon_pt
```

introducing dataframes

```
import ROOT

ROOT.ROOT.EnableImplicitMT()

RDF = ROOT.ROOT.RDataFrame
d = RDF(treeName, inputFile)

d = d.Filter("nMuon>1 && MET>20”)\
     .Define(“SelMuon_pt”
     ,“Muon_pt[Muon_pt>25 \
     && abs(Muon_eta)<2.4 \
     && Muon_dz<0.1 && Muon_dxy<0.01 \
     && Muon_relIso<0.5]”)
```

be prepared
for a massive change

traditional way

```
import ROOT

fIn = ROOT.TFile.Open("file.root")
tree = fIn.tree

for event in tree:
   if len(event.muons)<1: continue
   if not event.MET>20: continue

   for muon in event.muons:
      if muon.pt > 25 and  abs(muon.eta)<2.4 \
      and muon.dz<0.1 and muon_dxy<0.01 \
      and muon.relIso<0.5:

       selmuon_pt = muon_pt
```

loop over events

loop over muons

parallelisation difficult - let alone in python

introducing dataframes

```
import ROOT

ROOT.ROOT.EnableImplicitMT()

RDF = ROOT.ROOT.RDataFrame
d = RDF(treeName, inputFile)

d = d.Filter("nMuon>1 && MET>20”)\
     .Define(“SelMuon_pt”
     ,“Muon_pt[Muon_pt>25 \
     & abs(Muon_eta)<2.4 \
     & Muon_dz<0.1 & Muon_dxy<0.01 \
     & Muon_relIso<0.5]”)
```

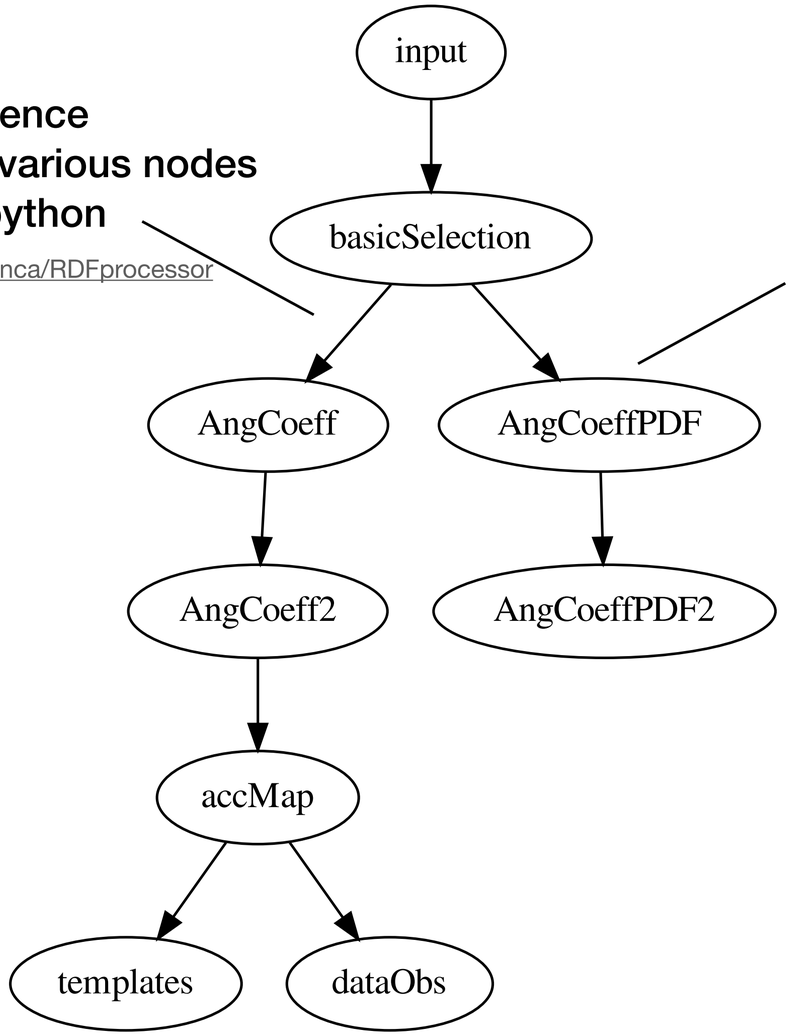
transparently parallelised

from the event loop to a graph-style analysis

each module is a python or C++ class that transforms a RDataFrame object

the dependence among the various nodes is done in python

https://github.com/emanca/RDFprocessor



each node contains a list of modules to be executed

RDF

cut events

define new columns

book histograms

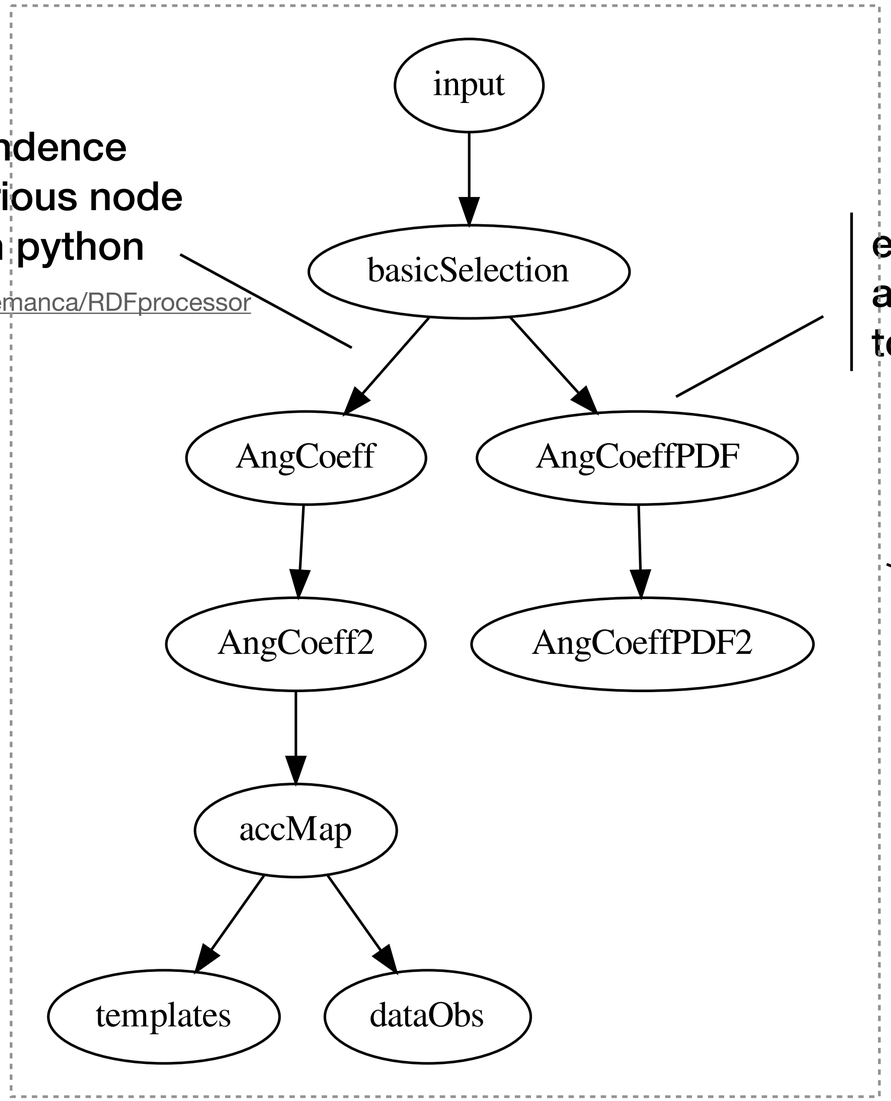
RDF

from the event loop to a graph-style analysis

each module is a python or C++ class that transforms a RDataFrame object

the dependence of the various node is done in python

https://github.com/emanca/RDFprocessor



each node contains a list of modules to be executed

RDataFrame provides optimised execution in a single parallelised event loop

RDF

cut events

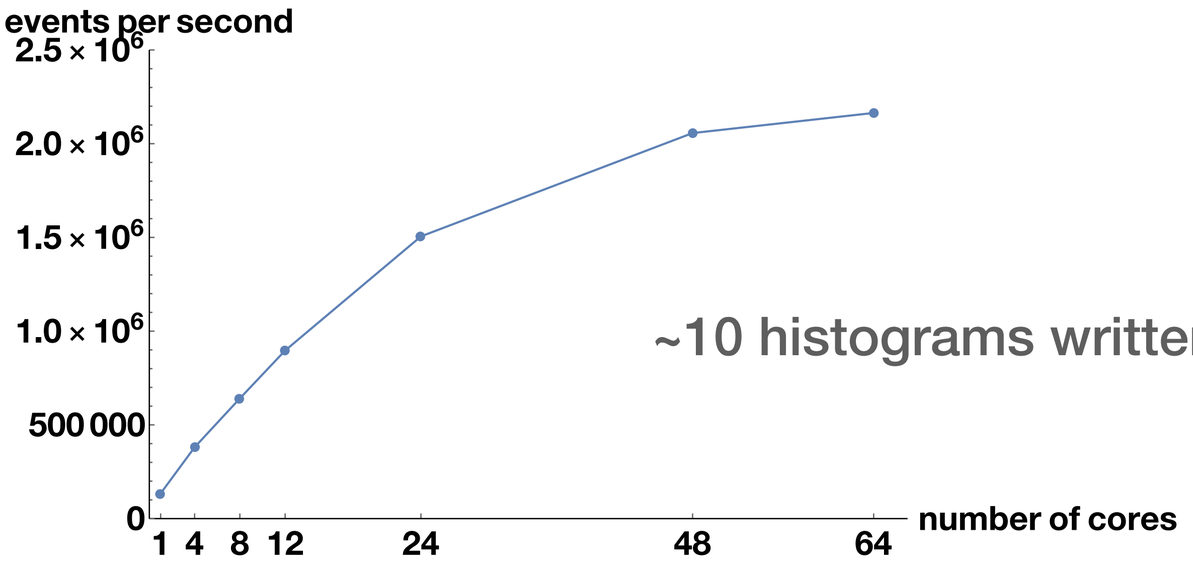
define new columns

book histograms

RDF

does it all scale?

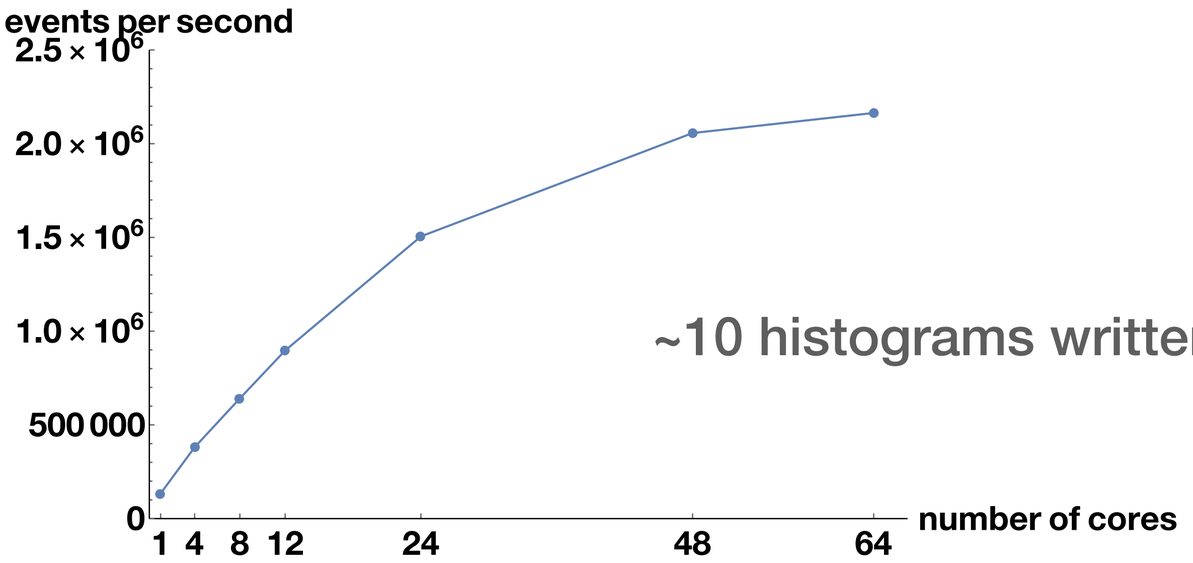
64 cores AMD machine, SSD, warm cache



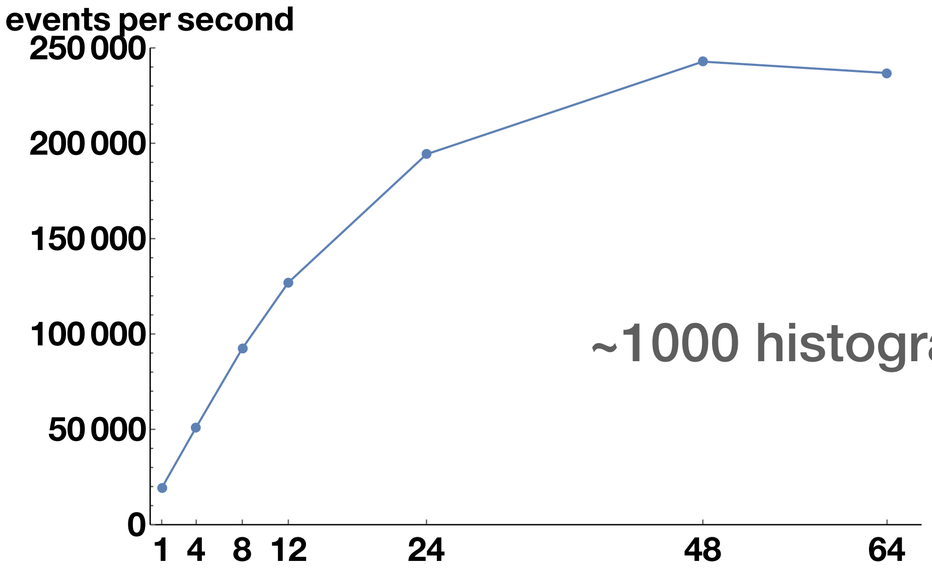
~10 histograms written

does it all scale?

64 cores AMD machine, SSD, warm cache

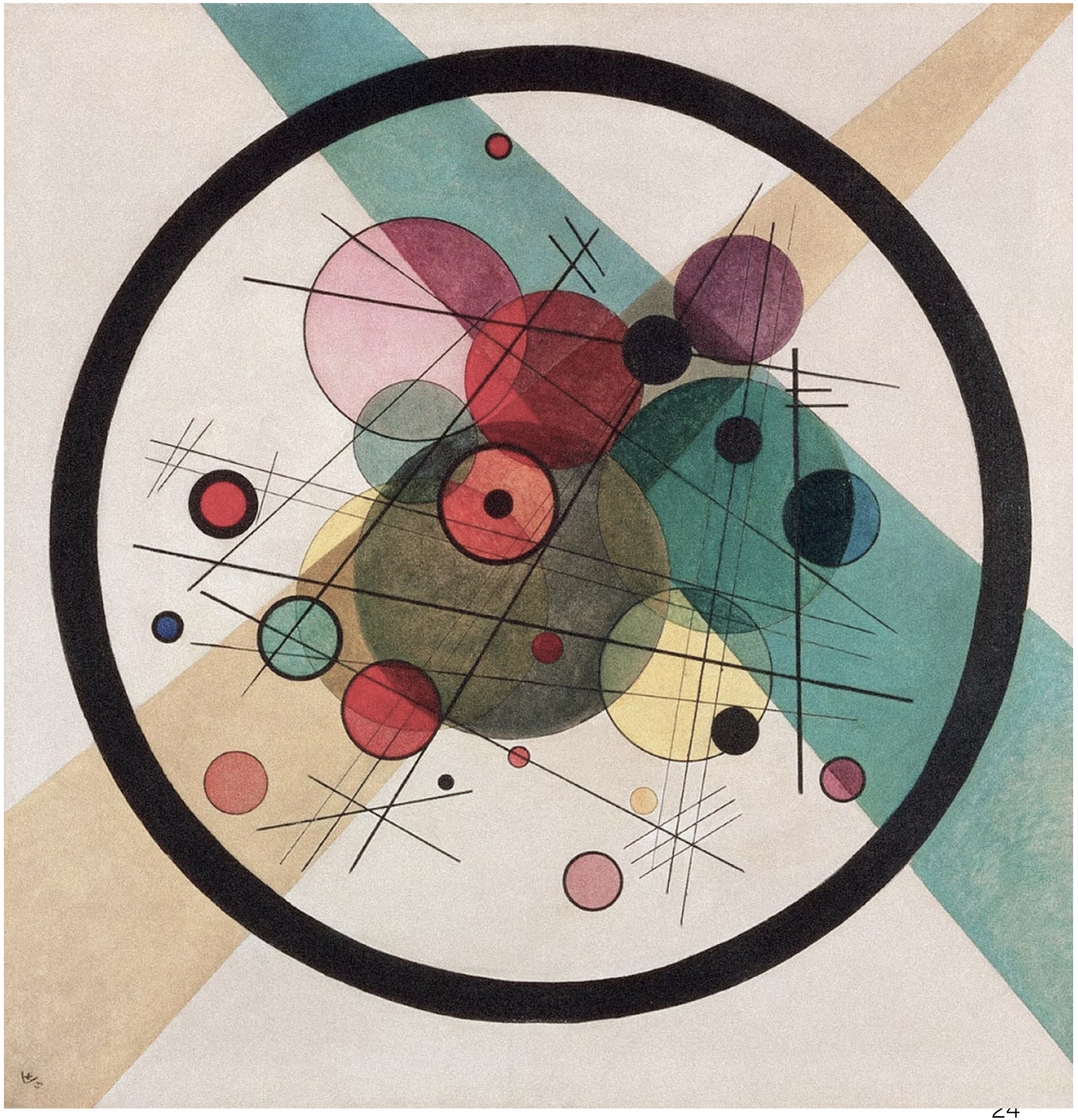


~10 histograms written



~1000 histograms written

roughly same scaling



SCUOLA
NORMALE
SUPERIORE

acknowledgments

ROOT
Data Analysis Framework

and especially
E. Guiraud