Interaction point simulations: Output format

Tom Blackburn
Department of Physics, University of Gothenburg

23 Feb 2021
Simulation and Analysis Task Force

UNIVERSITY OF GOTHENBURG

Interaction point simulations

Output format

- Present status:
  - IPstrong ".out": plain-text, metadata header + columns of particle properties
  - Event information in a separate file in stdhep/HEPmc format.

```
#-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------
#  Final state particles.
#-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------
#  First interacting species: electron     Second interacting species: laser
#  First initial particle energy = 16.5000 +/- .16 GeV,  Sigma_xyz = 5.00 5.00 24.00  microns,  Emit_xy = 1.401.40 mm mrad
#  Laser Intensity = 2.55 x 10^18 W/cm^2,  Wavelength = 800.00 nm,  pulse length = 25.00 fs,  spot size = 1256.64 micron^2
#  Pulse peak xi =.7765, Pulse peak chi =.1375, Misalignment = .0000 microns
#-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------
#    E (GeV)          x (um)          y (um)          z(um)              beta_x              beta_y              beta_z              PDG_NUM      MP_Wgt      MP_ID
#-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------
   16.496183       -5.3290756       -2.3010402       100.85535      0.2602675254308950E-05 -0.5944696053919820E-05  0.9999999949916235071        11      750000.          1
   16.510355        4.1228638        1.4070765       67.531998      0.1743879115798838E-05  0.1220677456487837E-04  0.9999999944501909166        11      750000.          2
   16.515006        7.1969290       -2.5056124       12.002699      0.1037741131022853E-05  0.1246274409033604E-04  0.9999999944311360433        11      750000.          3
   15.633417       -1.4690765       -5.0552488       69.123344      0.9653832384162547E-05  0.2120576224343157E-05  0.9999999941695546851        11      750000.          4
   16.491223       -2.0644660        3.0609196       102.14609      0.3606096816450605E-05 -0.3520621637173127E-05  0.9999999950723099759        11      750000.          5
   16.502842        11.296187       -0.65213668       59.888678      0.7112123243721278E-05 -0.1147111657874488E-04  0.9999999942952174883        11      750000.          6
   16.493390        8.5975999        1.6663475       42.337080      0.3714799915420995E-05 -0.3234152734504097E-06  0.9999999951310435348        11      750000.          7
   16.493896        2.6756580       -2.4684185       44.241572      0.4350125248144688E-06 -0.2066262416805113E-05  0.9999999951785660302        11      750000.          8
   14.947983       -6.2654336       -5.4633779       33.392538     -0.4712590953603720E-06 -0.4425118503112067E-06  0.9999999941547914771        11      750000.          9
   16.504829       -4.9861585       -2.7277981       86.387792      0.1453250219267275E-04 -0.1710384153669591E-04  0.9999999926885404367        11      750000.         10
   16.495772       -2.6027415       0.16137765       43.910537      0.1628948169464526E-04  0.2473660058680029E-04  0.9999999908157182312        11      750000.         11
   16.505128        3.0853489       -3.3995191       32.793661      0.1195349621790757E-04  0.6719874506058525E-05  0.9999999942671755017        11      750000.         12
   13.455472       0.81972461        2.5497898       53.966603      0.1801050516811686E-05  0.6381423032405565E-05  0.9999999925688927633        11      750000.         13
   16.496408       -4.6651470       -1.6045006       26.002043      0.4983738370420851E-05 -0.8658561263677554E-05  0.9999999947032792405        11      750000.         14
   16.492309       -4.7901101        2.4687461       65.535619      0.2198781253331845E-04 -0.5925053622456324E-05  0.9999999926070849823        11      750000.         15
   16.499131        3.4297124        9.7689837       85.144129      0.1088220214963627E-04  0.8411757271211285E-06  0.9999999946082550204        11      750000.         16
   16.487409       0.40003768       0.46133209       87.411781      0.2656076966815602E-05 -0.4100623794214353E-05  0.9999999950777332831        11      750000.         17
   16.507074        1.1067653        4.0147561       93.411639      0.7459114854059220E-05 -0.1130996500365574E-04  0.9999999942907506355        11      750000.         18
   16.517828        5.5873724        1.6667491       25.979863      0.4610362954265061E-05 -0.4191946859220493E-05  0.9999999950206165460        11      750000.         19
   16.503257        2.6265917        3.0300659       39.361477      0.5322922030232208E-05 -0.8565250899524718E-05  0.9999999946978173867        11      750000.         20
```

UNIVERSITY OF GOTHENBURG

Interaction point simulations

Output format

- Things to consider in a replacement
  - Binary saves space, easier to store at higher precision, faster to read/write
  - Reproducibility means knowing exactly how data was generated, which means storing lots more metadata
  - Ideal format would be platform/language independent, straightforward to pickup
  - How easy is it to pipe data into the next code in the sequence?

UNIVERSITY OF GOTHENBURG

Interaction point simulations

Output format

- Things to consider in a replacement
  - Binary saves space, easier to store at higher precision, faster to read/write
  - Reproducibility means knowing exactly how data was generated, which means storing lots more metadata
  - Ideal format would be platform/language independent, straightforward to pickup
  - How easy is it to pipe data into the next code in the sequence?

- HDF5 as the "minimal viable" option
  - Time taken to finalize this is time that could be spent on implementing physics processes
  - Accessible by most software packages, e.g. Matlab, and libraries available for, e.g. python, C/C++
  - Hierarchical data format (filesystem within a file), with "groups" (folders) and "datasets" (files/tables)
  - Data "attributes" allow for storage of, e.g. units and definitions, together with the data itself
  - Additional milestone: transfer of output format from plain-text to custom HDF5 [March 2020].

- Proposed structure:
  - At top level: `/build`, `/config`, `/final-state` and `/event`

UNIVERSITY OF GOTHENBURG

Interaction point simulations

Proposed HDF5 structure

- /build
  - Anything needed to rebuild the executable: code version, commit hash, compilation options...
- /config
  - Anything need to reconstruct the run: a copy of the input file, values of laser intensity, electron energy...
- /final-state
  - /electron, /photon, /positron
  - Each containing a dataset (i.e. a table) of position and momentum, both four-vectors
- /event
  - Format not decided – should store enough information on individual events to reconstruct the interaction, e.g. vertex position, incoming and outgoing momenta etc – subdivided by type?
  - Standard formats, e.g. HEPmc3, are all plain-text?
  - Any thoughts welcome

UNIVERSITY OF GOTHENBURG

Interaction point simulations

Python example

[3]
```
import h5py
import matplotlib.pyplot as plt
import numpy as np
```

[4]
```
file = h5py.File('0.5x1d_particles.h5', 'r')
```

[3]
```
list(file.keys())
```
```
['build', 'config', 'final-state']
```

[6]
```
list(file['build'].keys())
```
```
['branch', 'commit-hash', 'features', 'version']
```

[11]
```
print(file['build/branch'].value)
print(file['build/version'].value)
```
```
feature/hdf5-output
0.5.2
```

[7]
```
list(file['config'].keys())
```
```
['beam', 'control', 'input-file', 'laser', 'mpi-tasks', 'output']
```

[12]
```
file['config/laser/a0'].value
```
```
0.5
```

[13]
```
test=file['config/laser/polarization']
print(test.dtype.metadata)
print(test.value)
```
```
{'enum': {'Linear': 0, 'Circular': 1}}
1
```

UNIVERSITY OF GOTHENBURG

Interaction point simulations

Python example

[5]

```
p=file['final-state/photon/momentum'].value
p
```

```
array([(1046.39615101, -0.02735217,  0.00301574, -1046.39615065),
       ( 938.56174897,  0.02246415, -0.02779876,  -938.56174829),
       (  15.26210519,  0.0030487 ,  0.00917001,   -15.26210213), ...,
       ( 123.95959235, -0.02401781, -0.01048612,  -123.95958958),
       ( 992.59664286,  0.01383645,  0.02797869,  -992.59664237),
       (1055.07636518, -0.01997857, -0.02183526, -1055.07636477)],
      dtype=[('0', '<f8'), ('1', '<f8'), ('2', '<f8'), ('3', '<f8')])
```

[6]

```
energies=p[:]['0']
```

[7]

```
plt.hist(energies, density=True, bins=100)
plt.ylabel('density')
plt.xlabel('energy (MeV)')
plt.show()
```