

MOCK DATA CHALLENGE

Ben Carls, Eric Church, Herb Greenlee, and
Steve Wolbers

Goals of the challenge

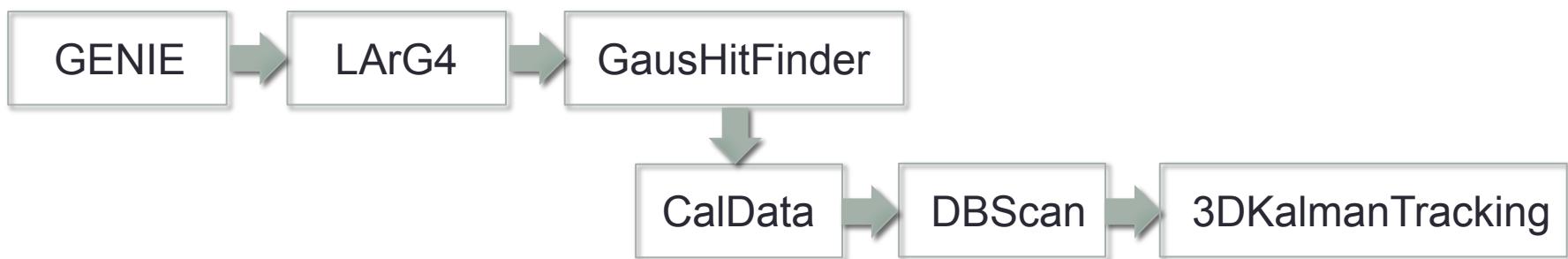
- Develop the framework necessary for large scale MC production
 - Write production scripts
 - Create a database of MC samples
 - Documentation
- Use the framework to generate a set of MC samples and provide them to the entire collaboration
- Practice running through the entire analysis chain (generation, simulation, reconstruction, and analysis)

First set of MC generation

- Process ran over the holidays
- Aimed to run a set of standard MC samples
 - Single particle
 - GENIE events
 - Other samples (cosmics and supernova)
- Events passed through the chain of generation, simulation, and reconstruction very smoothly, only minimal issues
- Generated 10,000 events for each sample, total of 6,090 jobs
 - 100 jobs for generation/simulation per sample
 - 100 jobs for reconstruction per sample
 - 10 jobs for merging per sample

MC Production chain

- We sought to include mature portions of the simulation and reconstruction (3D Kalman tracking, Gaussian hit finder)
- There are several stages we would like to add in the future (optical simulation and reconstruction)
- Dropped wires to reduce use of disk space



Tools used for generation

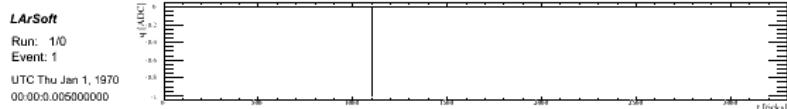
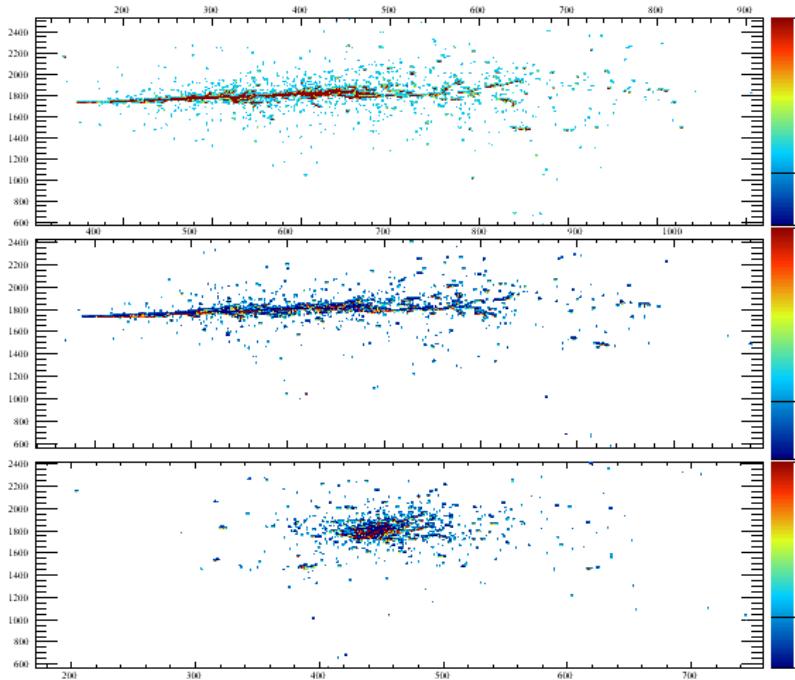
- We fully took advantage of the job management tools developed by Herb
- Information on the MicroBooNE software tools available here:
<https://cdcvn.fnal.gov/redmine/projects/ubooneoffline/wiki>
- Use of the project.py script made generation straightforward for the many samples
 - Allowed quick submission by supplying an XML file with job description
 - Could quickly check and resubmit jobs

```
project.py --xml myproject.xml --stage gen --submit  
project.py --xml myproject.xml --stage gen --check  
project.py --xml myproject.xml --stage reco --submit  
project.py --xml myproject.xml --stage reco --check
```

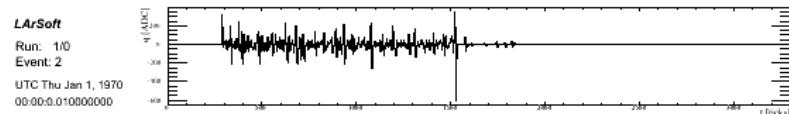
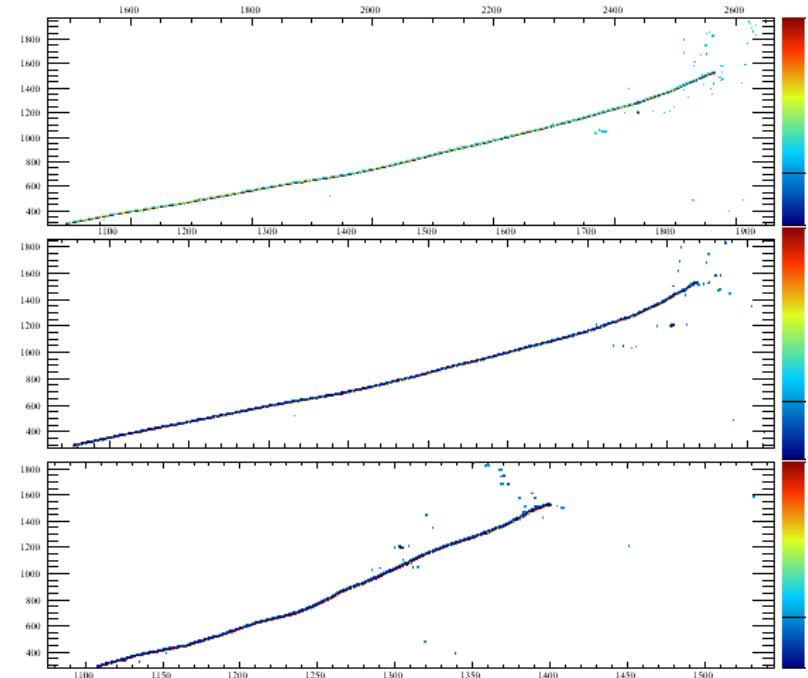
Single particle samples

- We generated a wide range of samples
 - Track-like ($\mu^\pm, \pi^\pm, K^\pm, p$)
 - Neutrals (n, Λ^0, K^0)
 - Shower-like (e^\pm, γ, π^0)
- Created both forward going samples ($< 25^\circ$) for μ^\pm and e^\pm and isotropic samples for everything
- Momentum varied for forward going and isotropic samples
 - Forward going 0.5-5.0 GeV
 - Isotropic 0.1-2.0 GeV

Single particle samples



Forward e^-



Isotropic μ^-

Single particle CPU time

Data Sample Name

Single Particles - Forward Going

prod_eminus_0.5-5.0GeV_25degf_t0_uboone	10000	767600	76.8	1261500	126.2
prod_eplus_0.5-5.0GeV_25degf_t0_uboone	10000	742100	74.2	1209000	120.9
prod_muminus_0.5-5.0GeV_25degf_t0_uboone	10000	200600	20.1	500400	50.0
prod_muplus_0.5-5.0GeV_25degf_t0_uboone	10000	179800	18.0	530100	53.0

<u>Events</u>	<u>Total</u>	<u>Avg</u>	CPU	CPU
			Simulation (secs)	Reconstruction (secs)

Single Particles – Isotropic

prod_eminus_0.1-2.0GeV_isotropic_t0_uboone	10000	253100	25.3	315900	31.6
prod_eplus_0.1-2.0GeV_isotropic_t0_uboone	10000	251100	25.1	299700	30.0
prod_gamma_0.1-2.0GeV_isotropic_t0_uboone	10000	226500	22.7	270500	27.1
prod_k0L_0.1-2.0GeV_isotropic_t0_uboone	10000	748200	74.8	195600	19.6
prod_k0S_0.1-2.0GeV_isotropic_t0_uboone	10000	576200	57.6	256500	25.7
prod_kminus_0.1-2.0GeV_isotropic_t0_uboone	10000	776700	77.7	275400	27.5
prod_kplus_0.1-2.0GeV_isotropic_t0_uboone	10000	346000	34.6	239000	23.9
prod_lambda_0.1-2.0GeV_isotropic_t0_uboone	10000	447700	44.8	171900	17.2
prod_muminus_0.1-2.0GeV_isotropic_t0_uboone	10000	111400	11.1	234000	23.4

<u>Events</u>	<u>Total</u>	<u>Avg</u>	CPU	CPU
			Simulation (secs)	Reconstruction (secs)

Single particle disk usage

<u>Data Sample Name</u>	<u>Data Sample Size (Bytes)</u>	<u>Sample Size (GB)¹</u>
<u>Single Particles - Forward Going</u>		
prod_eminus_0.5-5.0GeV_25degf_t0_uboone	38,379,125,603	35.7 GB
prod_eplus_0.5-5.0GeV_25degf_t0_uboone	38,553,512,243	35.9 GB
prod_muminus_0.5-5.0GeV_25degf_t0_uboone	16,031,446,847	14.9 GB
prod_muplus_0.5-5.0GeV_25degf_t0_uboone	15,918,341,533	14.8 GB
<u>Single Particles – Isotropic</u>		
prod_eminus_0.1-2.0GeV_isotropic_t0_uboone	13,038,329,050	12.1 GB
prod_eplus_0.1-2.0GeV_isotropic_t0_uboone	13,037,027,943	12.1 GB
prod_gamma_0.1-2.0GeV_isotropic_t0_uboone	11,729,688,210	10.9 GB
prod_k0L_0.1-2.0GeV_isotropic_t0_uboone	11,328,631,008	10.6 GB
prod_k0S_0.1-2.0GeV_isotropic_t0_uboone	13,509,942,387	12.6 GB
prod_kminus_0.1-2.0GeV_isotropic_t0_uboone	12,347,820,301	11.5 GB
prod_kplus_0.1-2.0GeV_isotropic_t0_uboone	8,949,957,827	8.3 GB
prod_lambda_0.1-2.0GeV_isotropic_t0_uboone	7,690,347,162	7.2 GB
prod_muminus_0.1-2.0GeV_isotropic_t0_uboone	7,176,055,038	6.6 GB

Divide by 10,000 to get per event size:

- The forward going e^- sample takes 3.57 MByte per event
- The forward going μ^- sample takes 1.49 MByte per event

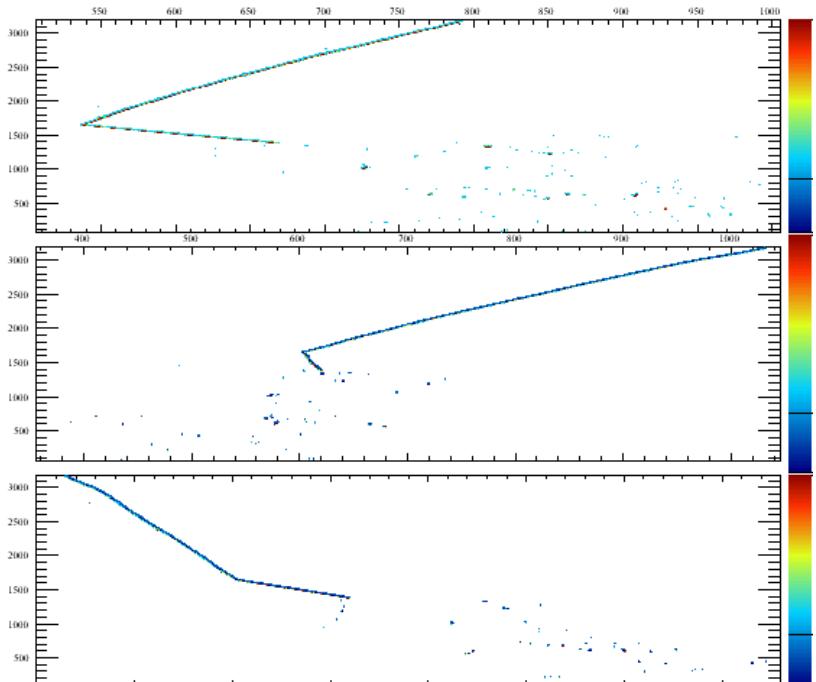
GENIE

- Also wide range of samples generated
 - For CC: ν_μ , $\bar{\nu}_\mu$, ν_e , $\bar{\nu}_e$
 - For NC: ν_μ , $\bar{\nu}_\mu$
 - Also a BNB sample
- For all samples, used BNB flux

Other

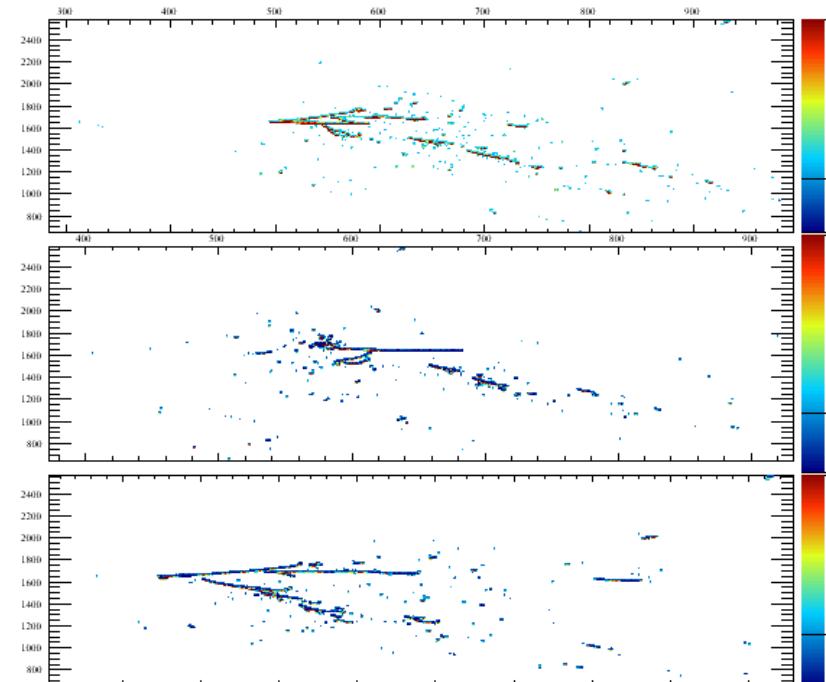
- Supernova
- Cosmics

GENIE



LArSoft
Run: 1/0
Event: 1
UTC Thu Jan 1, 1970
00:00:0.005000000

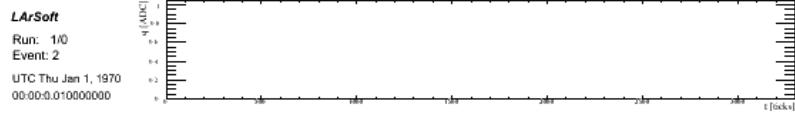
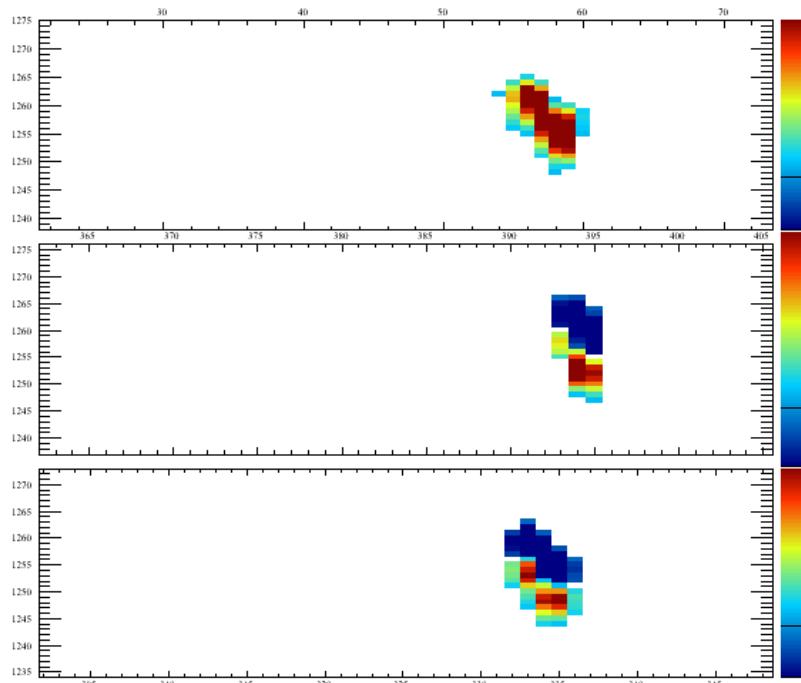
ν_μ CC



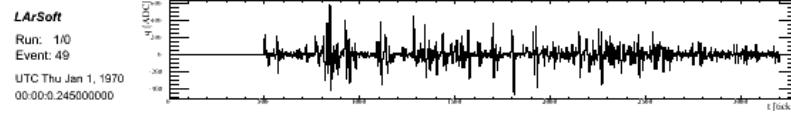
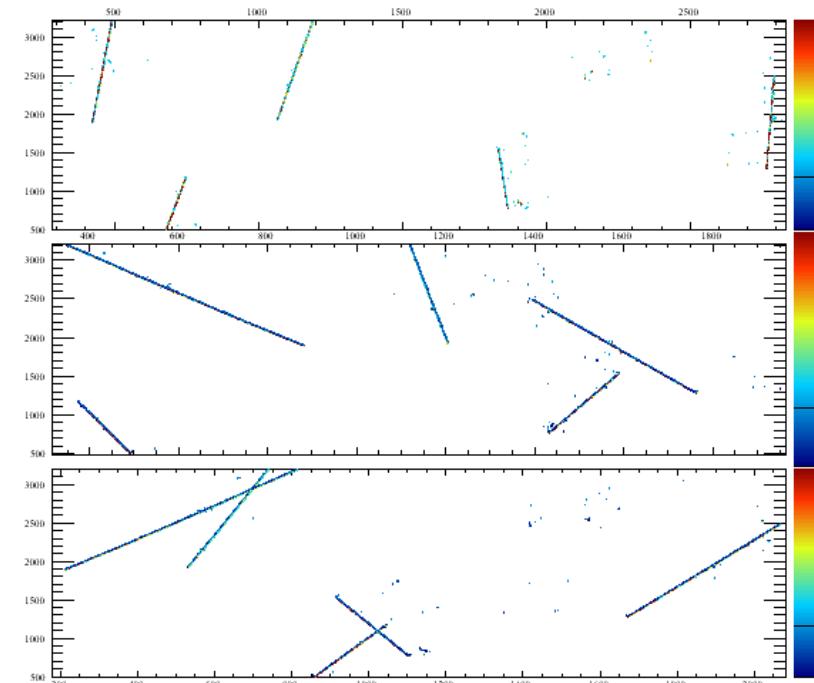
LArSoft
Run: 1/0
Event: 4
UTC Thu Jan 1, 1970
00:00:0.020000000

ν_e CC

Other



Supernova



Cosmics

GENIE and others CPU time

Data Sample Name

Genie

prodgenie_anue_cc_uniform_flux_t0_uboone
prodgenie_anumu_cc_uniform_flux_t0_uboone
prodgenie_anumu_nc_uniform_flux_t0_uboone
prodgenie_bnb_nu_t0_uboone
prodgenie_nue_cc_uniform_flux_t0_uboone
prodgenie_numu_cc_uniform_flux_t0_uboone
prodgenie_numu_nc_uniform_flux_t0_uboone

Events	CPU Simulation (secs)	CPU		CPU Reconstruction (secs)	
		Total	Avg	Total	Avg
10000	1618100	161.8	2074700	207.5	
10000	1104400	110.4	1207000	120.7	
10000	896200	89.6	343300	34.3	
10000	1622600	162.3	6359700	636.0	
10000	1605000	160.5	1528300	152.8	
10000	1230400	123.0	1160900	116.1	
10000	1096100	110.0	449900	45.0	
10000	2770000	277.0	648900	64.9	
10000	101000	10.1	145600	14.6	

Other

prodcosmics_uboone
prod_supernova_uboone

GENIE and others disk usage

<u>Data Sample Name</u>	<u>Data Sample Size (Bytes)</u>	<u>Sample Size (GB)^{*1}</u>
<u>Genie</u>		
prodgenie_anue_cc_uniform_flux_t0_uboone	47,603,378,081	44.3 GB
prodgenie_anumu_cc_uniform_flux_t0_uboone	35,136,953,782	33.5 GB
prodgenie_anumu_nc_uniform_flux_t0_uboone	13,503,472,457	12.6 GB
prodgenie_bnb_nu_t0_uboone	41,845,324,984	39.0 GB
prodgenie_nue_cc_uniform_flux_t0_uboone	42,611,631,438	39.7 GB
prodgenie_numu_cc_uniform_flux_t0_uboone	32,637,617,206	30.4 GB
prodgenie_numu_nc_uniform_flux_t0_uboone	18,108,343,875	16.9 GB
<u>Other</u>		
prodcosmics_uboone	267,974,090,653	249.6 GB
prod_supernova_uboone	1,308,082,385	1.2 GB

Divide by 10,000 to get per event size:

- The ν_μ CC sample takes 3.04 MByte per event
- The cosmics sample takes 25.0 MByte per event

Status and future plans

- The first run of mock data went very well
 - Took roughly a week to complete all samples
 - Samples take up roughly 700GB of disk space
- We hope to add in more as algorithms improve
 - Optical simulation and reconstruction is high on the list
- Aim to perform two more this year
- Locations of the samples:
http://www-microboone.fnal.gov/at_work/AnalysisTools/mc/fall2012/
- Documentation (where the CPU time and disk usage is in its entirety) available in DocDB 2323