Data Volume Calculations

Josh Klein, Penn, December 3, 2018

Sources of Data (SP)

- Cosmic Rays (~4500/day in 10 ktonne)
- Beam Events
- Rock Muons
- Atmospheric νs
- Solar νs
- Radiologicals (^{39}Ar, ^{42}Ar, neutrons, ^{208}Tl,^{210}Tl,^{224}Rn…)
- Calibration Sources
- Noise and instrumental backgrounds
- Trigger Primitives

Items in red have been updated.

Parameters

Table 1:

Parameter	Value
ADC Dynamic Range	12 bits
ADC Sampling Rate	2 MHz
Event Readout Window	5.4 ms
Supernova Burst Readout Window	100 s
Total Number of Channels/Module	384000
Number of Channels/APA	2560
Size of a DUNE Event	6.22 GB

Operating Assumptions

For TDR Design:

Aim for zero bias at the (TPC) channel level and as low a bias as possible at trigger level.

Operationally this means initially no zero-suppression for majority of stored events other than canonical 5.4 ms "readout window," and reading entire detector over this period. [This is distinct from thresholds used for trigger primitives]

A standard DUNE "event" or "snapshot" is therefore a 5.4 ms slice of time roughly centered on a time-stamped trigger (at least when we start) across whole detector

384000 channels x 12 bits x 2 MHz x 5.4 ms= 49.77 Gb=6.22 GB/readout

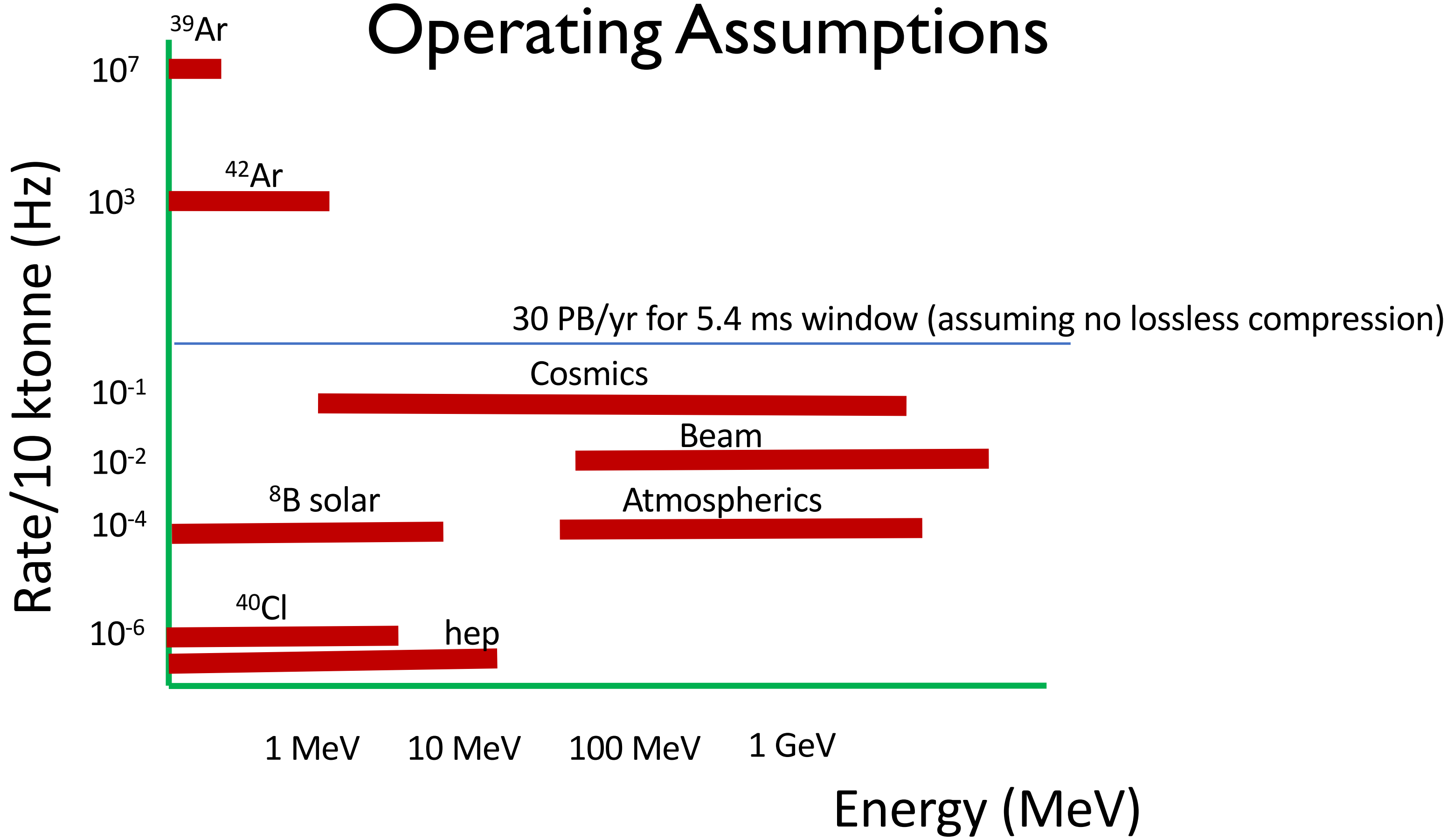
Operating Assumptions

Likely exceptions to no zero-suppression/localization paradigm:

- Electronics calibrations (run as pulse train anyway)
- Laser source calibrations (known, fixed tracks)
- Radioactive sources (known location)
- Random triggers for ^{39}Ar calibrations (known trigger type)
- Supernova bursts

These can be exceptions because we know time and position or that they are point-like deposits (ie, ^{39}Ar), or are so infrequent (SN bursts) and precious that we do something different.

Operating Assumptions



Threshold of 10 MeV on ***visible*** energy gets total data volume below 30 PB/year cap
(Supernova burst trigger candidates can have a lower threshold)

Beam-Coincident Events

For IDR, included only CDR ν_μ CC events (+rock muons+accidentals), which are overwhelmingly dominant. No reason not to include everything:

CDR Table 3.6 gives 10842 ν_μ CC in 150 kt-MW-years in 0.5 to 20.0 GeV or

773 per year in 10 ktonnes (for a 1.07 MW beam).

Adding in anti-nu events and other backgrounds, and appearance events and backgrounds, this becomes 773+153=**926 events/year/10 ktonnes.**

Beam-Coincident Events

Should include rock muons and accidental cosmics

Rock Muons

Richie Diurba (DocDB #6628) calculates rate including front (753) and sides (633).
Top and bottom should also be 633 so total is 2001 ~ 2000 events/year/10 ktonne.

The rounding is justified because this number has an error bar of probably 50% or more.

Accidental cosmics

$$4500\ \mu/\mathrm{day}/10\mathrm{ktonne} \times 1\ \mathrm{Hz} \times 10^7 \mathrm{s}/\mathrm{yr} = 2800/\mathrm{year}$$

Beam-Coincident Events

Adding everything together:

$$\boxed{5726/\text{year}/10\text{ktonne} \times 6.22 \text{ GB/event} = 35.6 \text{ TB/year}}$$

Beam-Coincident Events

Adding everything together:

4500/day x 365 days x 6.22 GB =10 PB/year

Atmospherics

From CDR:

41,777 for 350 kt-years or 1194 events/year/10 ktonnes

$$\boxed{1194/\text{year}/10\text{ktonne} \times 6.22\ \text{GB}/\text{event} = 7.4\ \text{TB}/\text{year}}$$

Solar Neutrinos

From CDR:
Only number listed is total event rate above 4.5 MeV (Q-value)
But those at 4.5 MeV create very little visible energy....
For now use this as an **upper limit—**
122 events/day/40 ktonnes ~ 30/day/10 ktonnes

$$30/\text{day}/10\text{ktonne} \times 6.22 \text{ GB/event} \times 365 \text{ days} = 68 \text{ TB/year}$$

THIS IS A HARD UPPER LIMIT—at 10 MeV visible energy threshold this is probably down by an order of magnitude.

Also: these are actual *solar neutrinos*---if threshold is low, majority of these triggers are radiological background (mostly neutrons).

Radiologicals and "Fake" Supernova Bursts

10 MeV threshold chosen to reduce individual radiological decays to negligible contribution, but we assume 12 fake SN bursts/year:

12/year/10 ktonne x 100 s x 1.5 B/sample x 2 MHz x 384000 channels = 1.4 PB

There are large uncertainties on the intrinsic (source) rate of radiologicals.

If we push lower to make a more interesting solar ν measurement, contribution from radiologicals will no longer be negligible.

Random Triggers

Putting aside ^{39}Ar calibrations, "reasonable" for this to be about 1% of total data volume, so 45/day is 1% of cosmic rate or 0.1 PB/year.

During commissioning we will certainly be running randoms at a much higher rate than this. But we get to dial this and even, if we want, throw out the data from that time period.

To do a ^{39}Ar calibration with a high (~ 1 Hz) rate of randoms will require some kind of back-end zero suppression or at least a fast turn-around analysis with the data deleted afterwards.

Front-End Calibrations

M. Worcester/Dave Christian

Test stand data is 10 ms pulse train (100 μs between pulses) for two gains and four shapings:

10 ms * 2 MHz * 2 gains * 4 shapings * 384000 channels = 61 GB/run/10 ktonne

Plus 4 points to determine linearity of ADC:
+ 10 ms * 2 MHz * 4 points * 384000 channels = 15 GB/run/10 ktonne

= 75 GB/run/10 ktonne

If this is done 1/week it is 4 TB/year.

This assumes DNL is good; worst case is 1000x more data.
Not clear yet how much needed for crosstalk measurement

Laser Calibrations

Calibration WG suggests 800k pulses/run

Laser can be tightly zero-suppressed, so

$$800,000/\mathrm{cal}/10\ \mathrm{ktonne}\times 100\mu\mathrm{s}\times 1.5\mathrm{Bytes/sample}\times 2\ \mathrm{MHz}\times 384000\ \mathrm{channels} = 92\ \mathrm{TB/cal/10ktonne}$$

If this is done twice/year it is **184 TB/year**

Radioactive Source Calibrations

UC Davis "neutron bomb" source is easy; 50k neutrons/pulse, run normally
Total data volume is neglible---don't need many pulses to get a lot of neutrons

Gamma source would require special handling.
We *assume* rate in detector is 10 Hz and it illuminates just 1 APA (2560 channels)
So we localize readout to just 1 APA. Calibration WG would like 8 hour run in 4 feedthroughs, so

$$8 \text{ hours} \times 4 \text{ FTs} \times 10\ Hz \times 1.5 \text{ Bytes} \times 2 \text{ MHz} \times 5.4 \text{ ms} \times 2560 \text{ channels} = 50 \text{ TB/run.}$$

If this is done 4x/year it is **200 TB/year**

Trigger Primitives

Trigger primitive storage:

- Provides a way of measuring trigger efficiency (in combination with randoms)
- May be an easy way of using radiologicals as calibration sources
- Allows us to look back at data in case of a later-discovered astrophysical event

Trigger Primitives

- Rate is dominated by ^{39}Ar = 100 Hz/collection wire at zero threshold
- Charge (summed ADC), time, time-over-threshold, ADC peak=20 B

100 Hz * 20 B/primitive*128,000 coll = 256 MB/s

=8 PB/year

This assumes "zero" threshold for ^{39}Ar; ¼ MIP=maybe x2 lower

We may also want longer timestamps, or bigger dynamic range for TPs

And this is likely compressible, perhaps x2.

Summary Table

Event Type	Data Volume PB/year	Assumptions
Beam-related Events	0.04	926 beam and 2000 dirt muons; 10 MeV threshold in coincidence with beam time; include 2800 accidental cosmics
Cosmics	10	10MeV threshold, anti-coincident with beam time
Front-end calibration	0.004	Existing test-stand scheme
Atmospheric *nus*	0.007	CDR interaction rates
Solar νs	0.07	*Upper limit* assuming rate above 4.5 MeV ν energy
Radioactive source calibration	0.2	Source rate $\leq$10Hz; single fragment readout; lossless readout, 4 times/year
Laser calibration	0.184	800,000 total laser pulses, lossy readout
Supernova candidates	1.4	100 seconds full readout, average once per month
Random triggers	0.1	45 per day
Trigger primitives	8	All three wire planes; 12 bits per primitive word; 4 primitive quantities; ^{39}Ar-dominated

Backups

Data Volumes Included in TP

Event Type	Data Volume PB/year	Assumptions
Beam interactions	0.03	800 beam and 800 dirt muons; 10 MeV threshold in coincidence with beam time; include cosmics
Cosmics and atmospherics	10	10 MeV threshold, anti-coincident with beam time
Front-end calibration	0.2	Four calibration runs per year, 100 measurements per point
Radioactive source calibration	0.1	Source rate $\leq$10 Hz; single fragment readout; lossless readout
Laser calibration	0.2	1×10^6 total laser pulses, lossy readout
Supernova candidates	0.5	30 seconds full readout, average once per month
Random triggers	0.06	45 per day
Trigger primitives	$\leq$6	All three wire planes; 12 bits per primitive word; 4 primitive quantities; ^{39}Ar-dominated