ARA Filter/Analysis Algorithm

Carl Pfendner

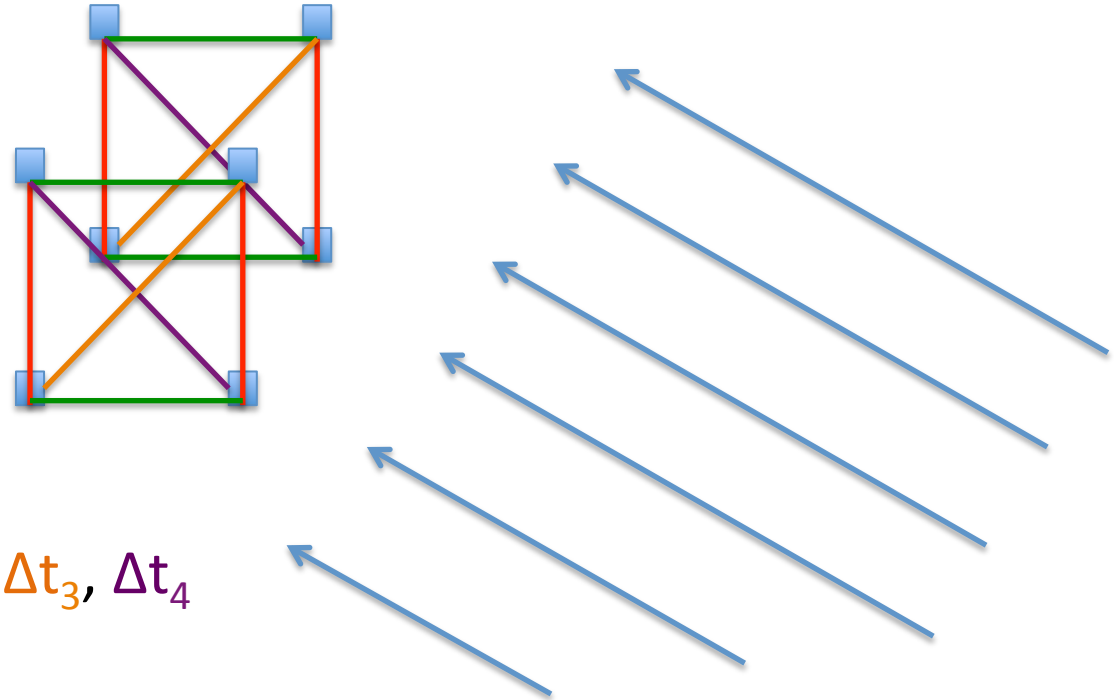
OSU

Supported by NSF CAREER Award 1255557 and
BigData Grant 1250720

Algorithm Goals

- Simple algorithm
- Decrease volume of data to then use more computationally complex analysis techniques (ray tracing, etc.)
- Single understandable output
- Easily differentiates between signal and thermal noise
- Include signal strength as a parameter

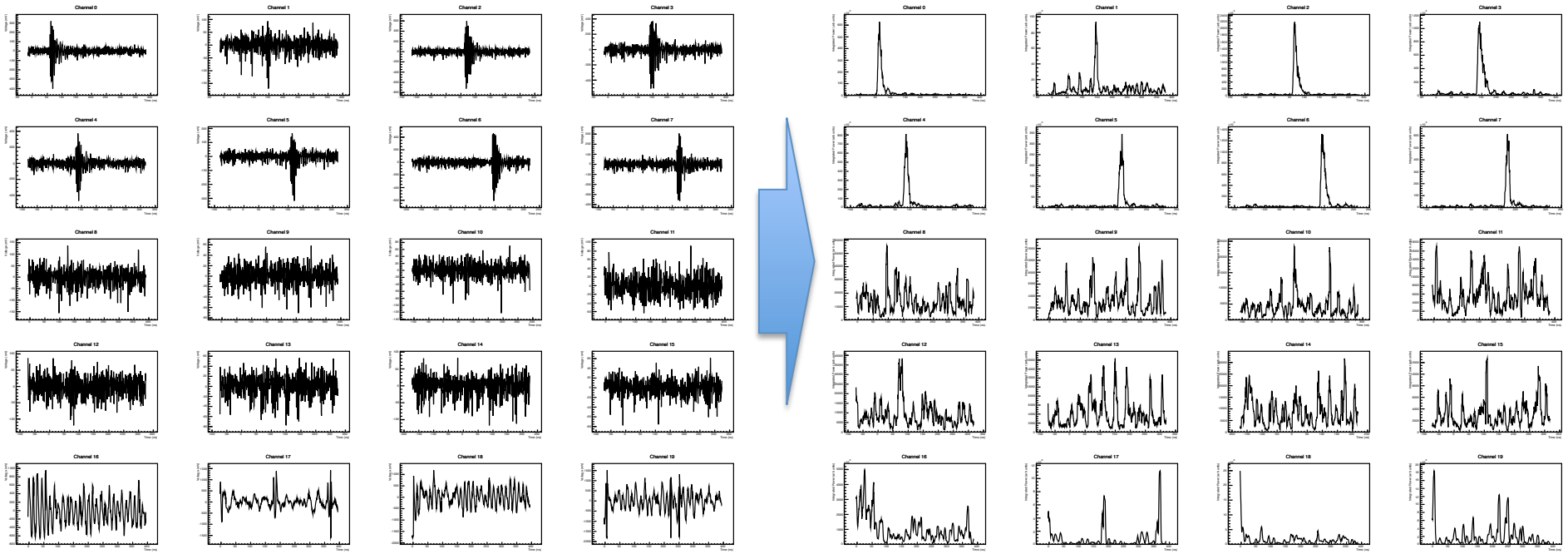
Plane-Wave-Like Geometry



Δt_1, Δt_2, Δt_3, Δt_4

- Take advantage of regular geometry
  - Find pairs of antennas that form parallel lines
- Delays in signal arrival times between similar pairs should be similar

Find “hit times”



Calibration pulser event

- Waveform a bit noisy on its own, try a scanning integrated power window
- Try ~5 ns window, maybe not “too narrow” and not “too broad”
- Find the peak, use this as a “hit time” for that channel

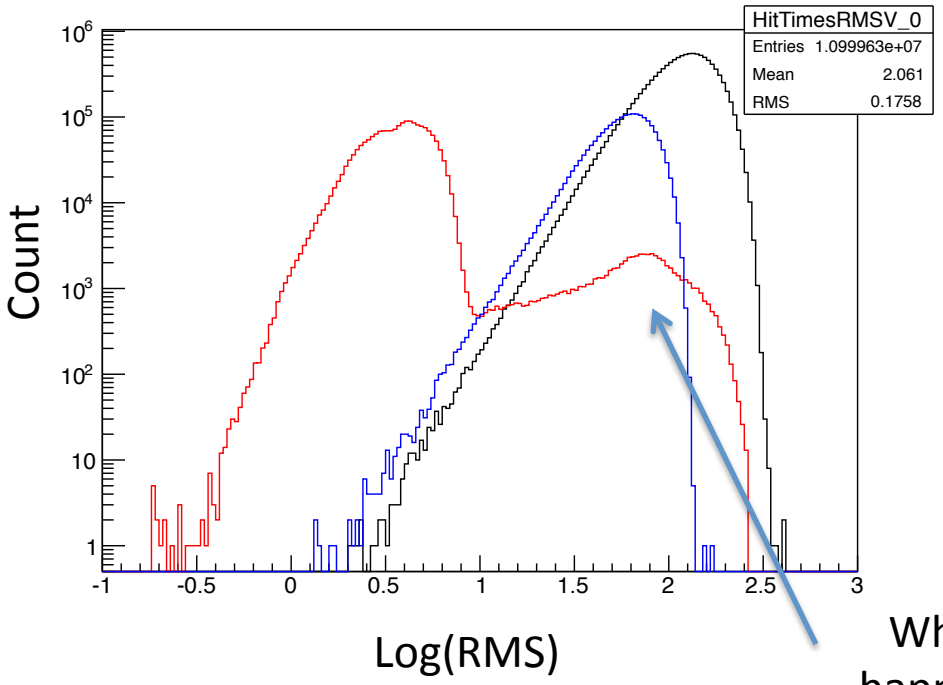
Algorithm Concept

- Check deviation using an
  - Signal: RMS → low, little variance in arrival times
  - Thermal noise: RMS → high, lots of variance independent of geometry
- For each set of parallel baselines:
  - Find average
  - Find variance of the pairs about that average
- Use a weighted RMS
  - Weights the stronger signals higher
    - Potentially decreases deviation due to lower signal strength channels
  - Can adjust the weights depending on the expected deviations in different pair types as well
- Sum the weighted $(t_{delay,i}-t_{avg})^2$ values for all pairs, divide by weight
- RMS $=\Sigma (w_i(t_{delay,i}-t_{avg})^2) / \Sigma w_i$

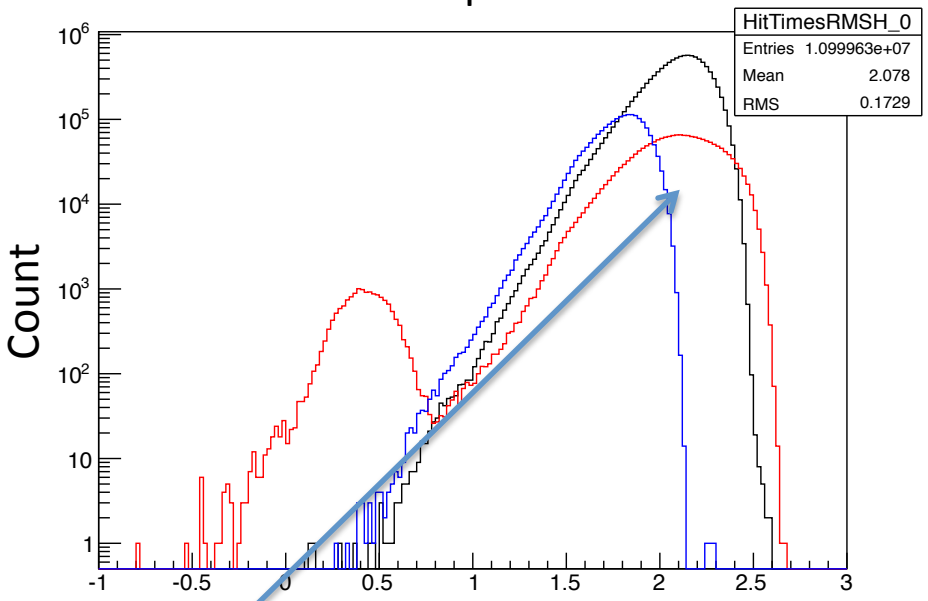
Initial Results – 2013 10% set

RF Triggered
Calibration Pulser
Soft Triggered

Vpol



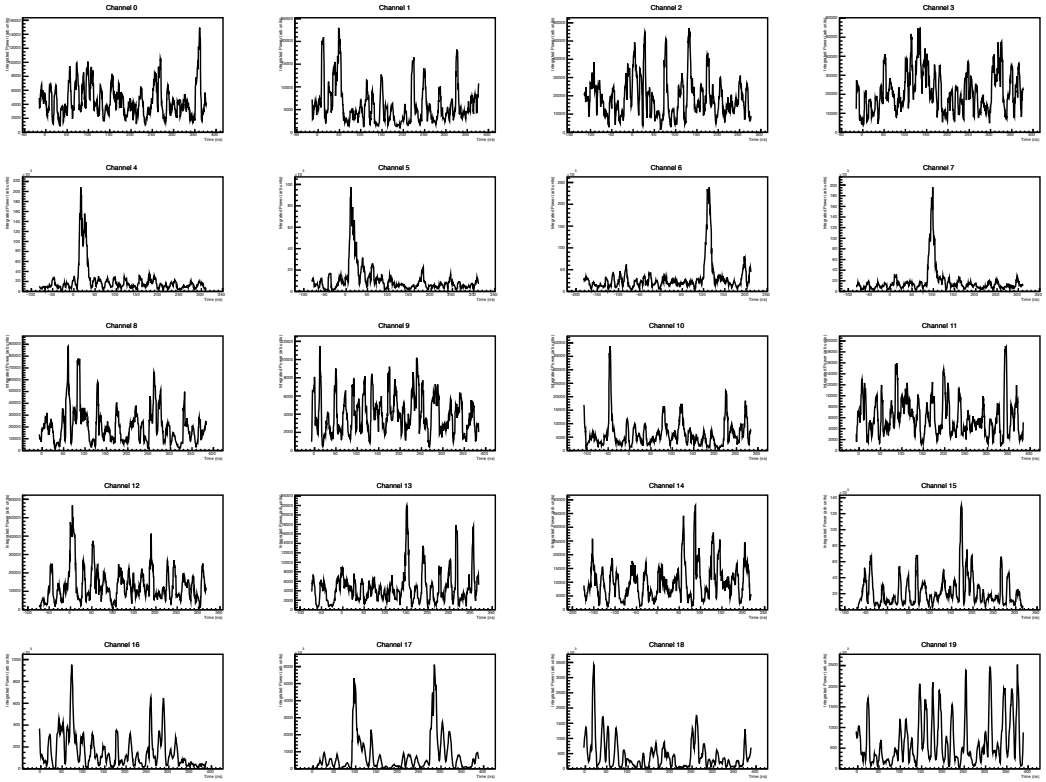
Hpol



What is happening here?

- Weight = power
- Some of the Vpol bump is due to Hpol-only calpulsers; look like thermal noise
- Most of Hpol bump is because most of the calpulsers only have Vpol signal
- Anything else?

Problems



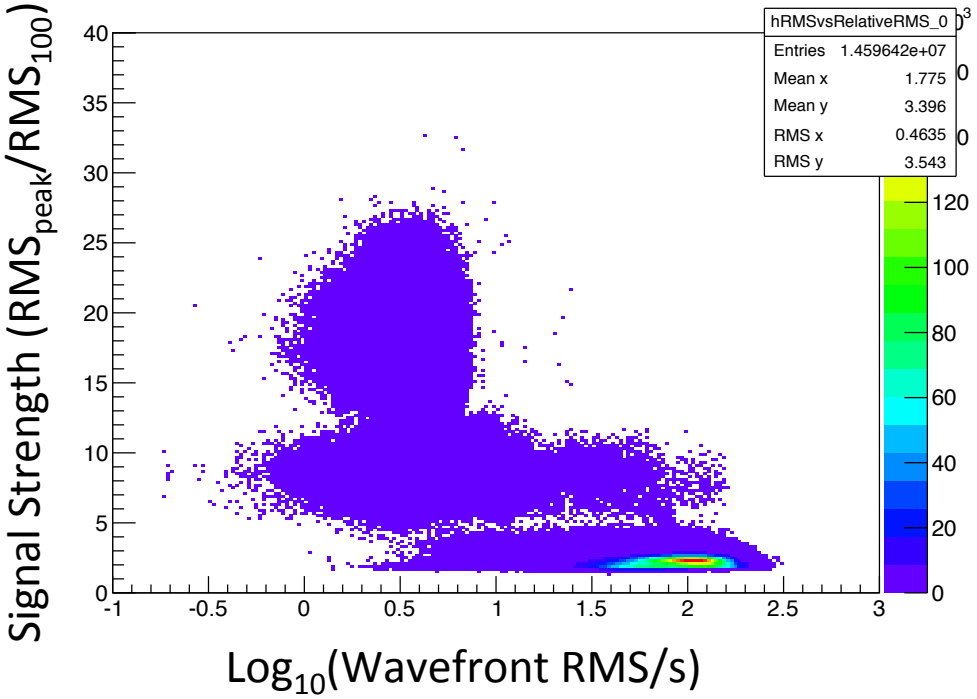
- Statistic has difficulty when signal not present in all channels
  - This event has high RMS in V and H

Signal Strength

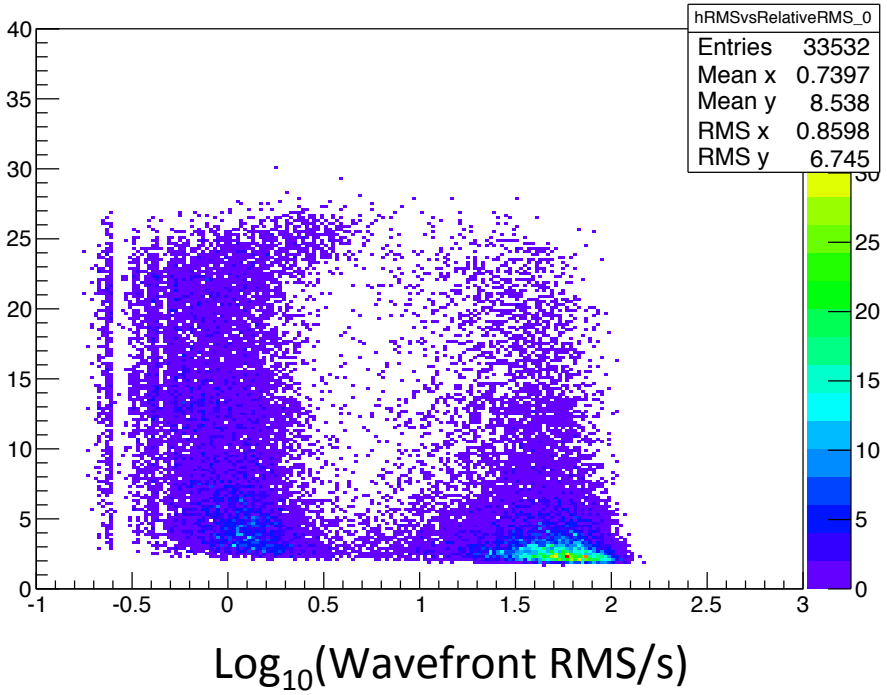
- Define signal strength based on waveform RMS
  - RMS_{peak} = RMS of the 10 bins used to define the peak
    - RMS = Sqrt(Power/10) so peak in power is the same as peak in RMS
  - RMS_{100} = RMS in the first 100 bins of the waveform
- Define "signal strength" as the third highest RMS_{peak}/RMS_{100}

Signal Strength vs Wavefront RMS

- All 2013 A2 data

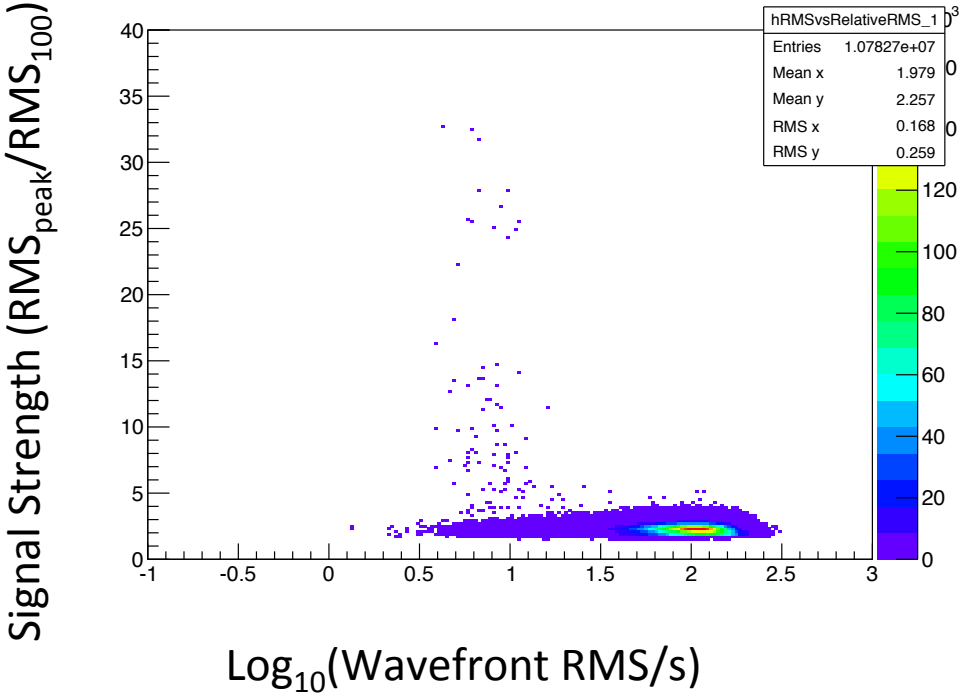


Simulation 10^{19} eV

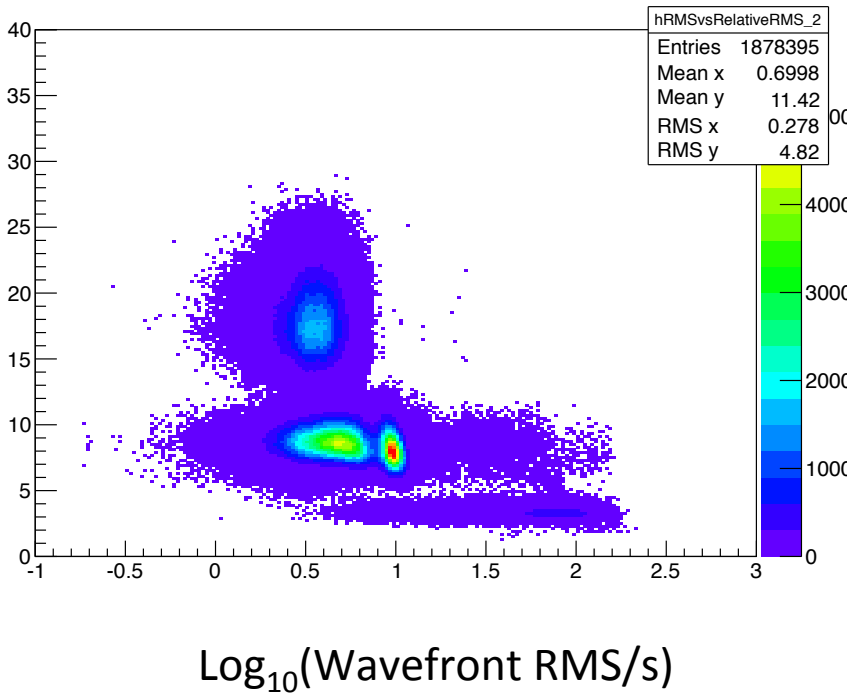


Signal vs RMS (cont.)

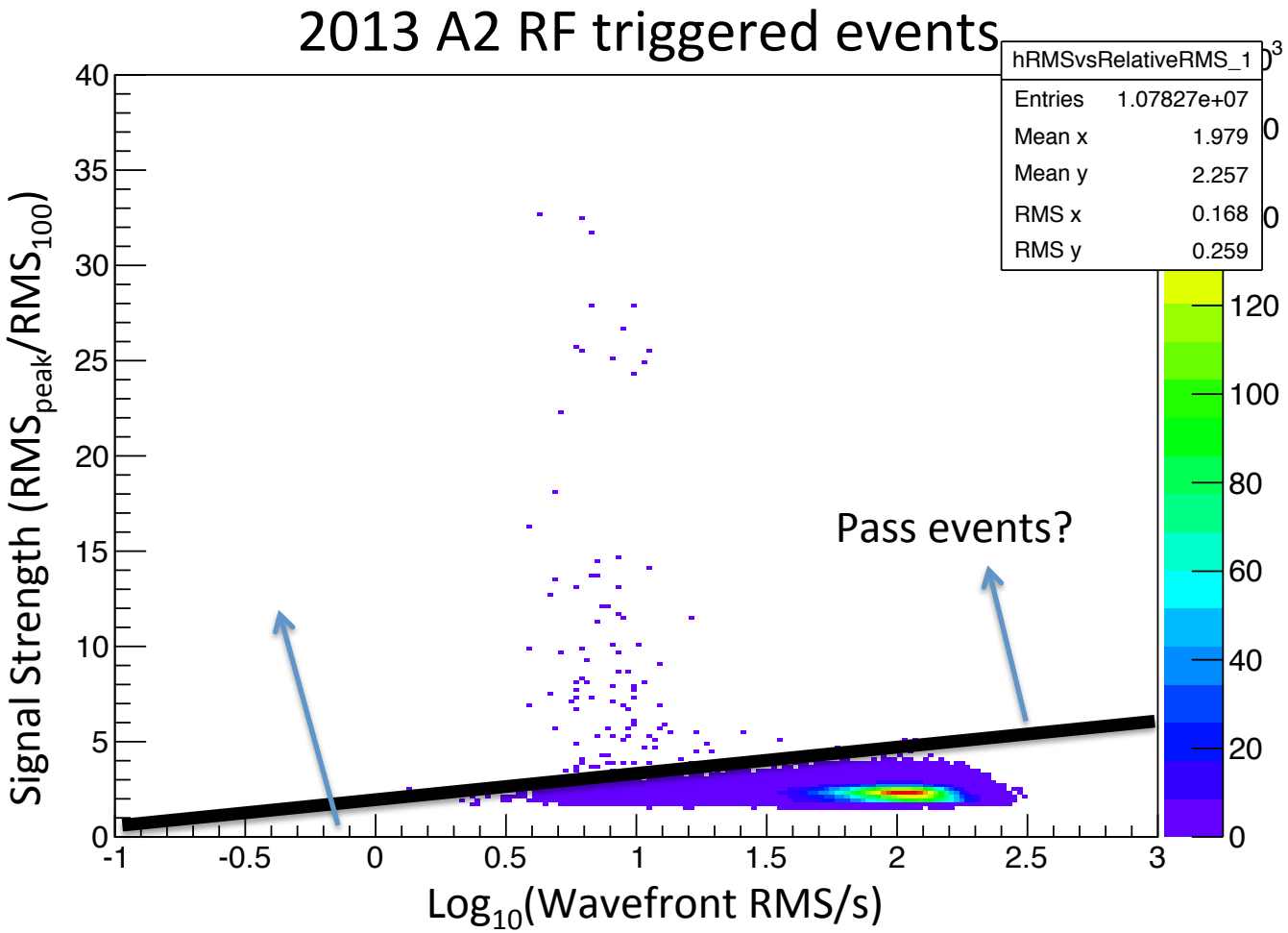
- 2013 A2 RF triggers



2013 A2 Calpulser



Combined statistic?



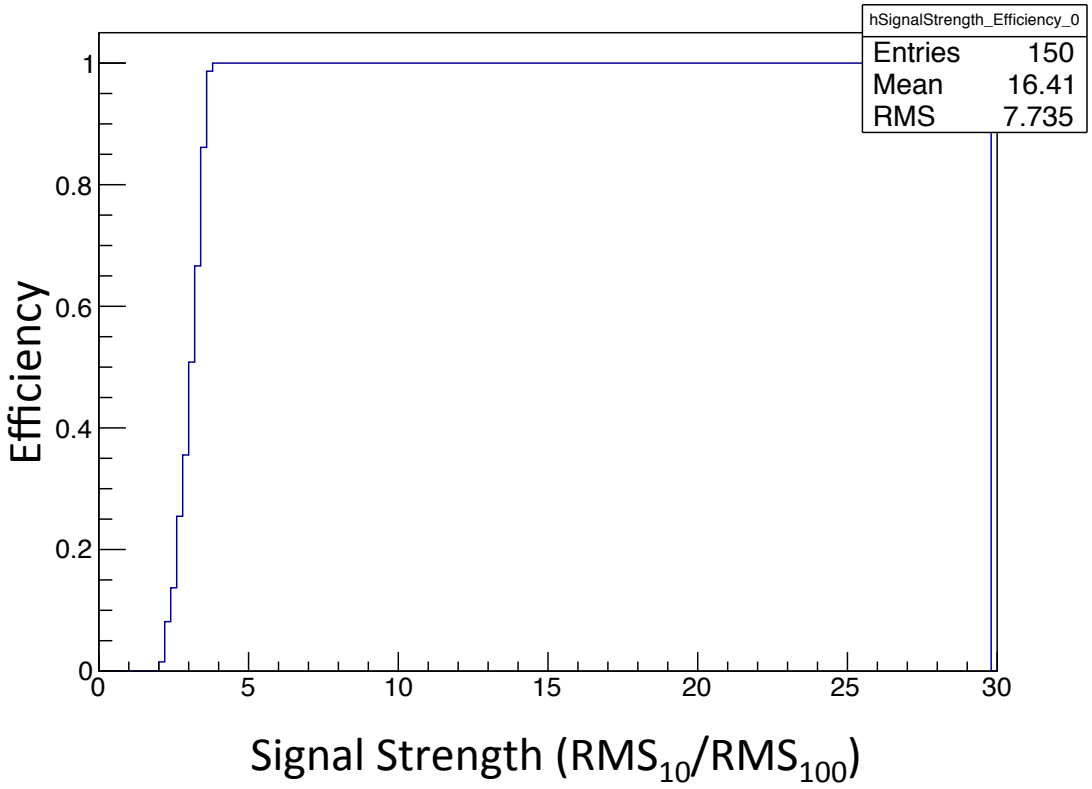
- Linear combination of wavefront RMS and signal strength?
- Currently working on "fitting" this for A2 and A3

Quick and Dirty Linear Combination

- $-1.25 > 1.25 * \mathrm{Log}_{10}(\mathrm{RMS}_{\mathrm{wavefront}}) - S_{\mathrm{signal}}$

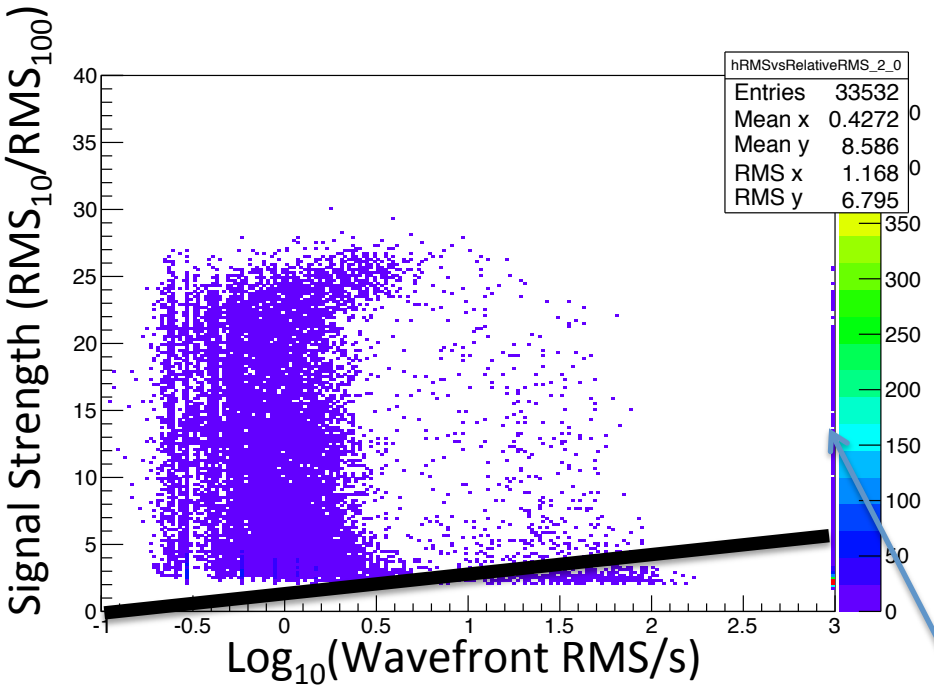
Overall efficiency ~60%

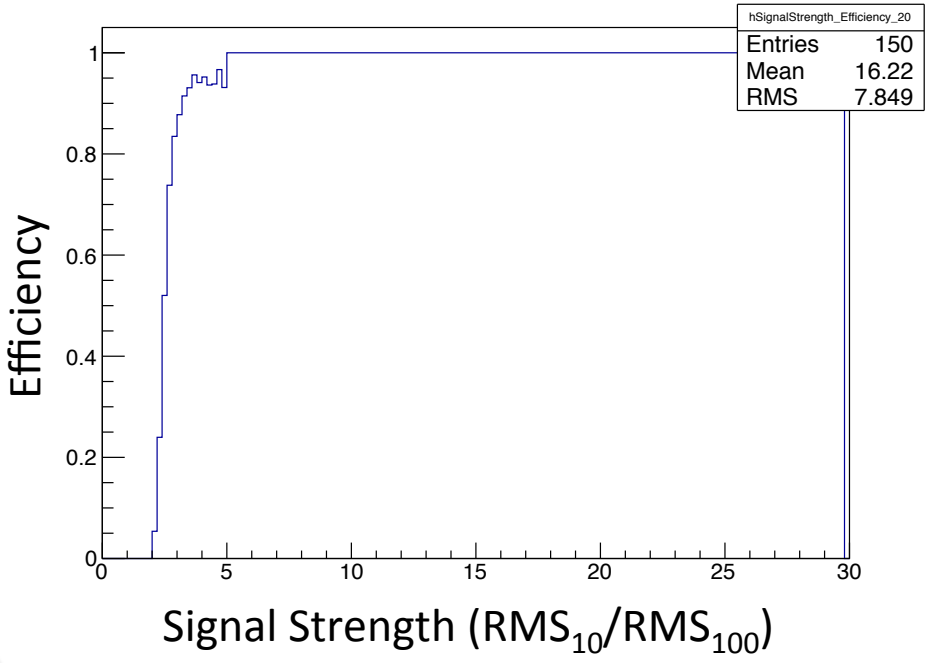
Remember signal here is a voltage/ voltage



Include a threshold?

- Require Signal Strength > 2.0 to be included in wavefront RMS
- If not enough channels included, set $RMS_{wavefront}$ to 1000
- Have not re-adjusted the cut line yet

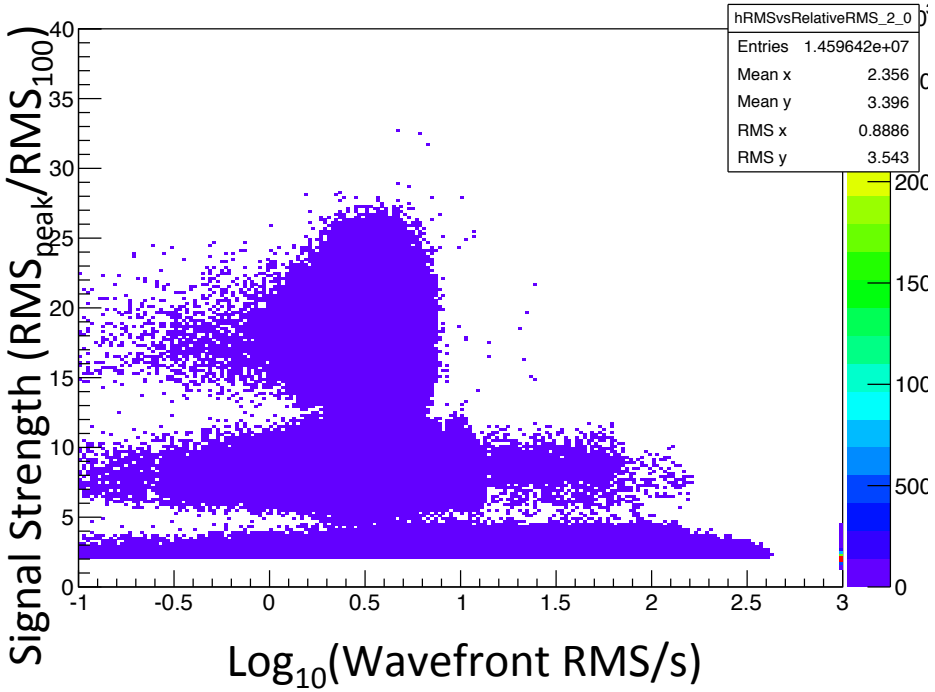




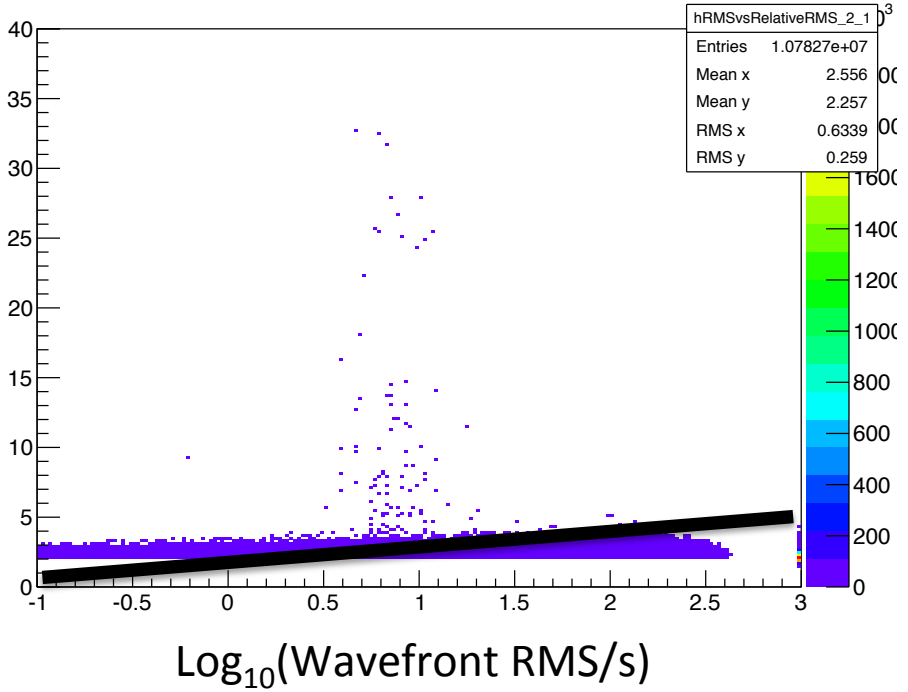
Events without enough channels

Threshold 2.0 Applied to Data

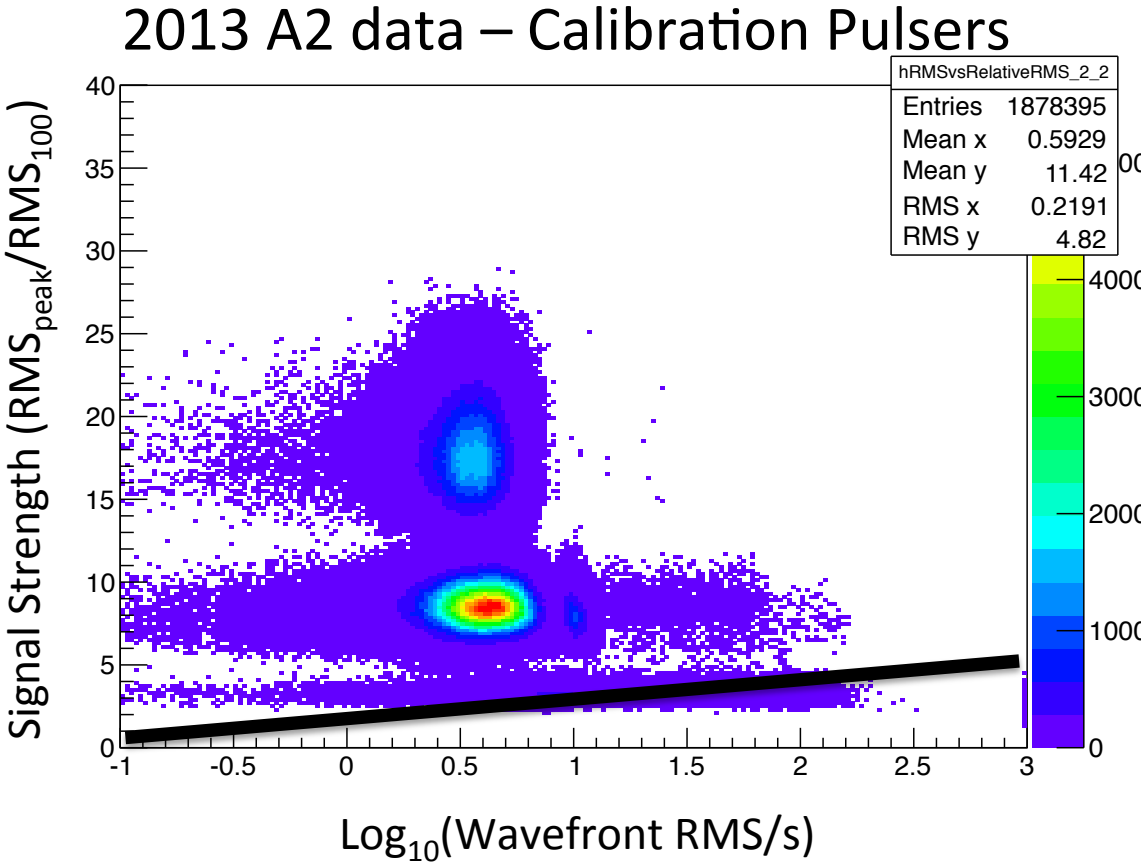
- All 2013 A2 data



2013 A2 data – RF Triggered



A2 Threshold data – cont.

- 

Options going forward

- Linear combination of signal strength and wavefront RMS?
  - Tunable to different pass rates, wavefront RMS vs signal strength distributions, i.e. these could change in understandable ways from station to station
- Other options:
  - Apply a threshold to eliminate noise channels in problematic events? Tried this to a limited degree earlier
    - Where to set threshold?
  - Find lowest wavefront RMS using combination of hit times?
    - Where to set threshold? Too many combinations?

Further Work

- Determine a way to best fit the 2-D distribution
- Need a good noise calibration for stations 2 and 3
  - Depends on valid thermal RMS distributions
- Better RMS baseline
  - Get from software triggered events?
- Can this handle events close to threshold?
  - Do we even care about these events?
- Need to examine the effect of this method on ultimate analysis efficiencies after reconstruction, etc.