LA-UR-16-23814

Analyzing PDV Spectrograms with Likelihood Methods

- J. Patrick Harding

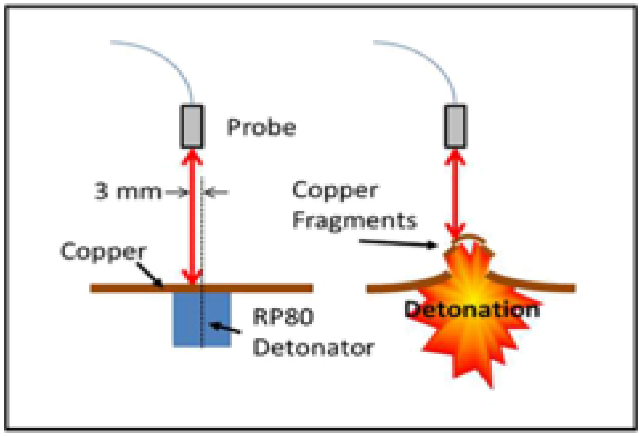
- 6/8/16

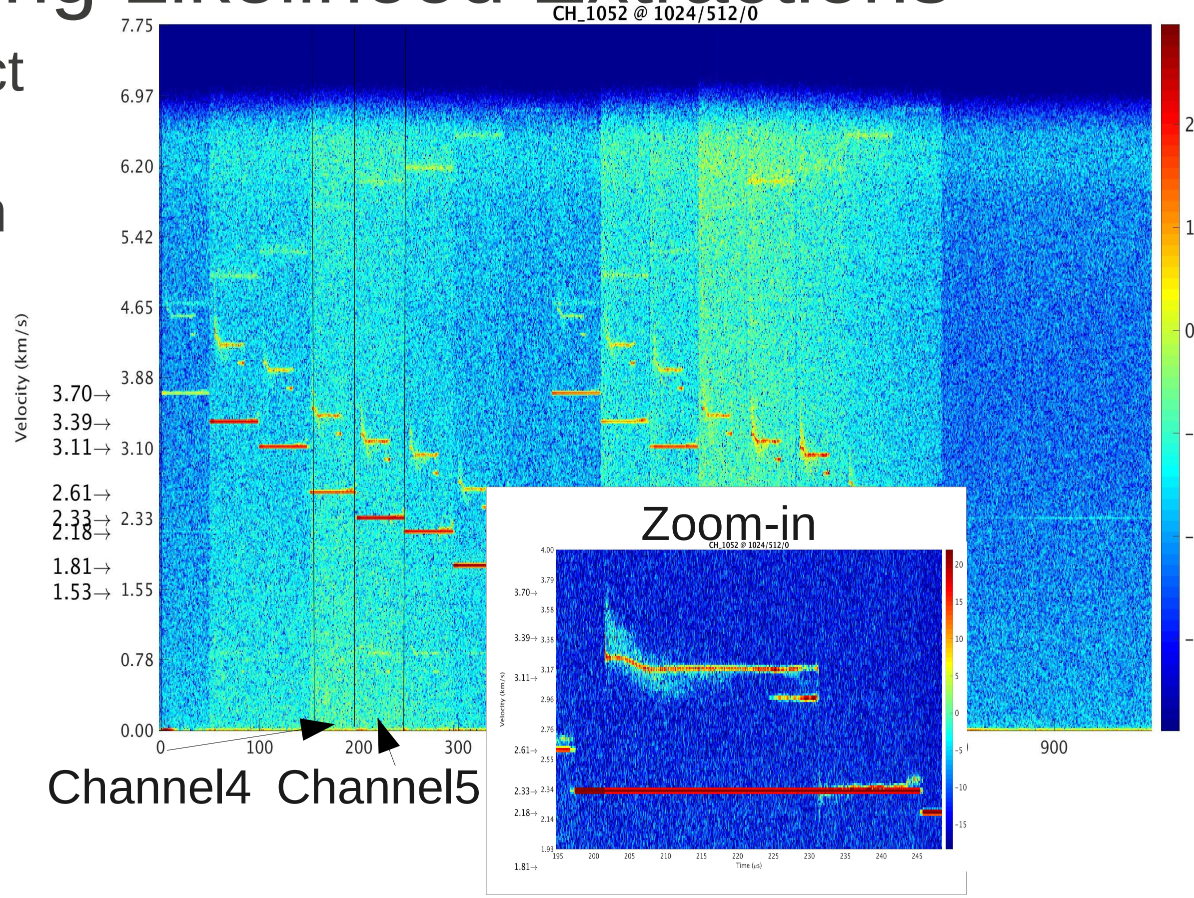
Contents

- Applying Likelihood Analysis to PDV Data
- Statistical Vetting of the Analysis
- Testing the Analysis Results with Synthetic Data

Data for Testing Likelihood Extractions

Test data has copper object moving at "single velocity" with all channels pointed in the same direction at the same place



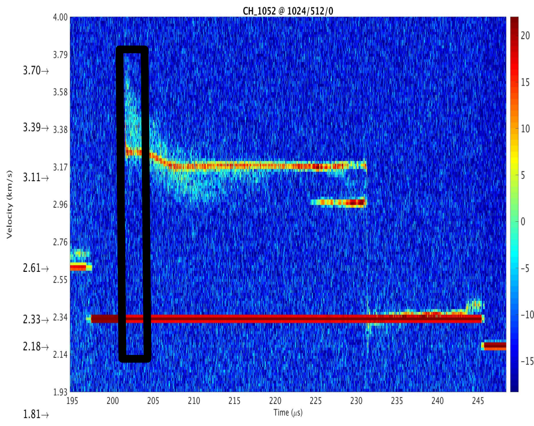


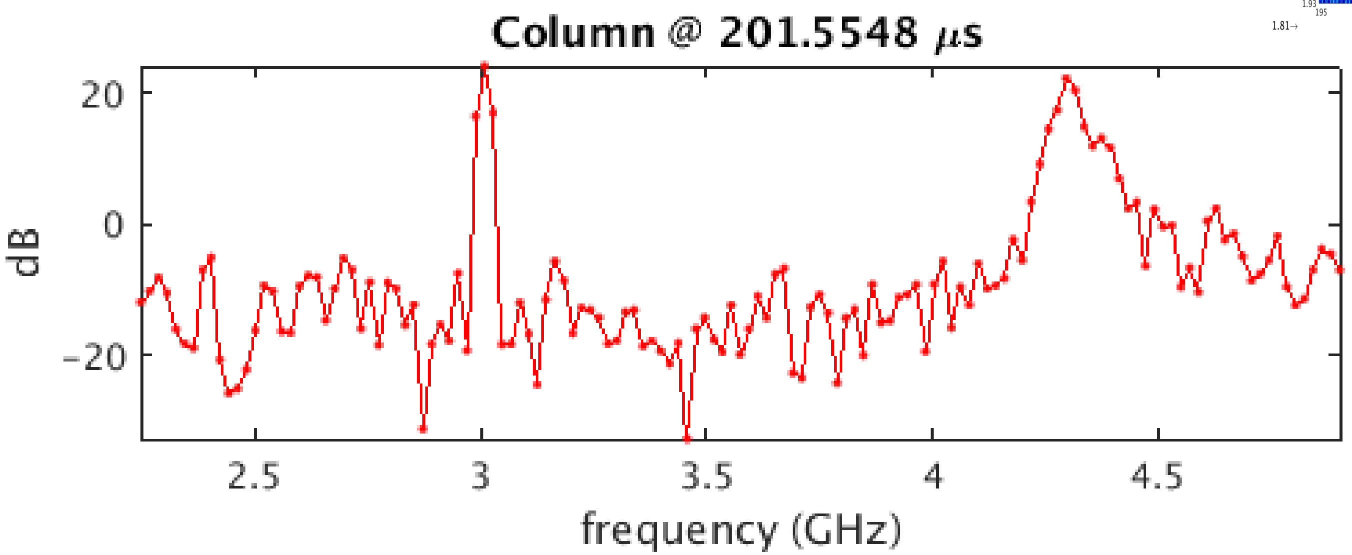
- Gen 3 Evaluation Shots, STL, May 2015 *
- Comparing the analysis of two channels will give a quantitative size of the statistical error bars of the analysis

* Credit: Ed Daykin, Chan Jung, Mike Pena, Abel Diaz, Marylesa Howard, Ben Valencia, Benji Stone, Kirk Miller

Sample Lineout from the Data

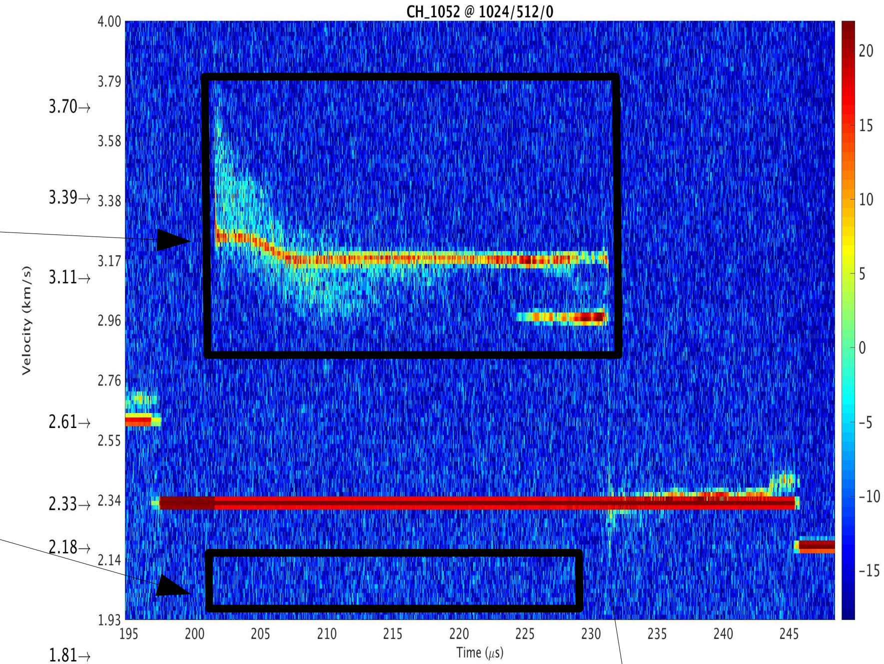
- Single slice in time
- Can have differing levels of noise
- Multiple peaks → Multiple surfaces
- Wide peaks → Ejecta

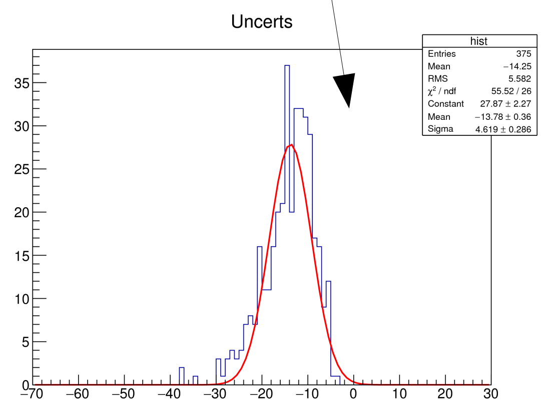




Likelihood Velocity Extractions

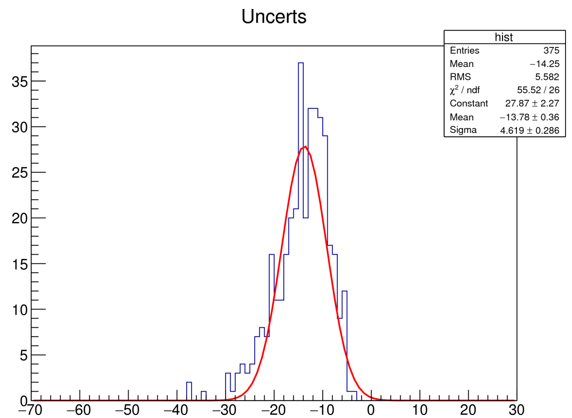
- Two ROIs per channel
  - Data region
    - All signals you want to test
    - Can be large
  - Background region
    - No signals
    - For statistics
- Assume Gaussian statistics
  - Fits background fluctuations well
- Model includes:
  - Free noise level
  - Multiple peaks (surfaces) with free widths





The Background

- Region far from any signal
- Fluctuations fit to a gaussian in dB-space
  - Width determines statistical significance of features in data region
  - Peak location determines nominal signal-free background level



Statistics with Maximum Likelihood (1)

- *If* background is Gaussain noise (and it does)
  - *Then* the log-likelihood test statistic (TS) is analogous to a chi-squared
  - Method works for arbitrary noise distribution, too
- The logarithm of the likelihood is

  $\ln(\mathrm{pdf}) = -\ln(\Sigma)-0.5*\ln(2\pi)-0.5*(dB_i-\mu(\nu_i))^2/\Sigma^2$

  summed over all frequency (velocity) bins I
  - for background with Gaussian width Σ
  - and for signal model μ(ν)
- Gaussian Statistics: $p(n_i;\mu_i,\Sigma)=\frac{1}{\sqrt{2\pi\Sigma^2}}exp\left[\frac{-(n_i-\mu_i)^2}{2\Sigma^2}\right]$

Statistics with Maximum Likelihood (2)

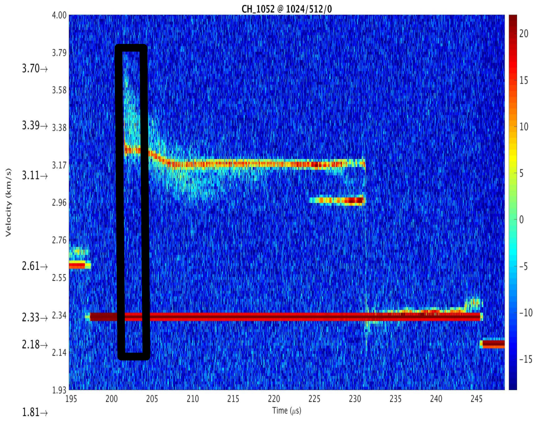
- Get LL for two models to compare (e.g. signal or not).
- Maximize each LL over the free parameters
- $TS^{max}=2*(LL_1^{max}-LL_0^{max})$
  - Related to significance that there is a velocity (or other parameter) to be extracted
  - If TS is large (considering the number of added parameters), then the parameters exist
  - Roughly, sqrt(TS) ~ # of sigmas of signal
- Lower TS threshold gives fainter features
  - But finds more “false peak” fluctuations, too
  - TS~10 balances these two extremes well

Statistics with Maximum Likelihood (2)

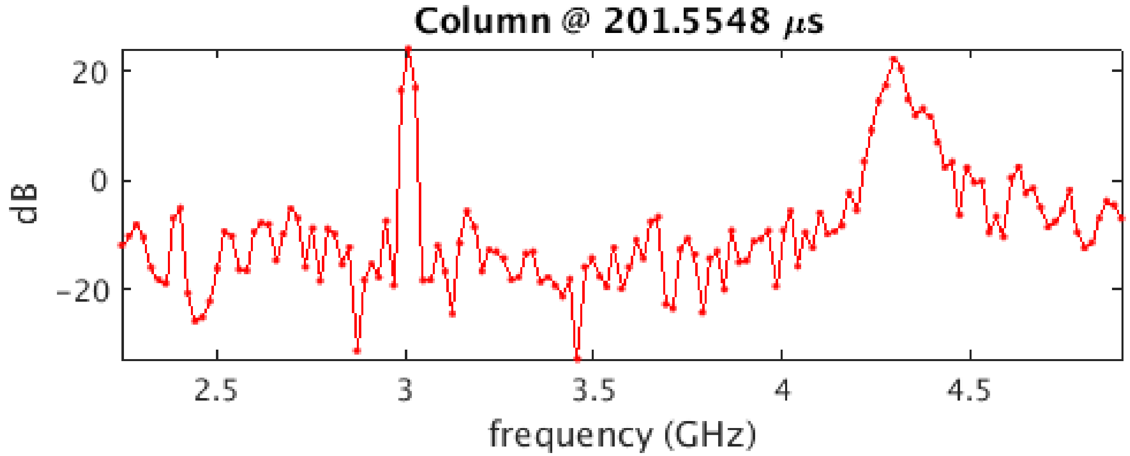
- Get LL for two models to compare (e.g. signal or not).
- Max[illegible]ize each LL over the free parameters
- TS [illegible] ^{max})
  - Relat[illegible]t there is a velocity (or other param[illegible]
  - If TS is large (consider[illegible]lded parameters), then the paramet[illegible]
  - Roughly, sqrt(TS) ~ # of sigmas of signa[illegible]
- Lower TS threshold gives fainter features
  - But finds more “false peak” fluctuations, too
  - TS~10 balances these two extremes well

Finds all model parameters
and uncertainties on all the parameters

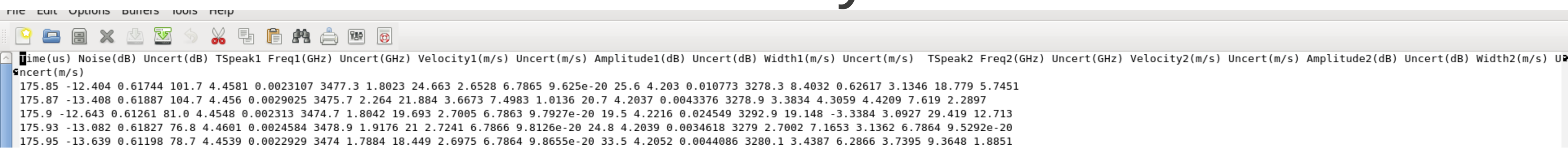
The Signal Model



- Each timeslice is considered separately
- Models fit as Gaussians in dB(frequency)
- Model checks:
  - Is there a true peak in frequency?
  - Is the peak wider than the baseline?
  - Is the noise level different than in background ROI?
  - Is there a second peak in frequency?



Extracted Velocity Information

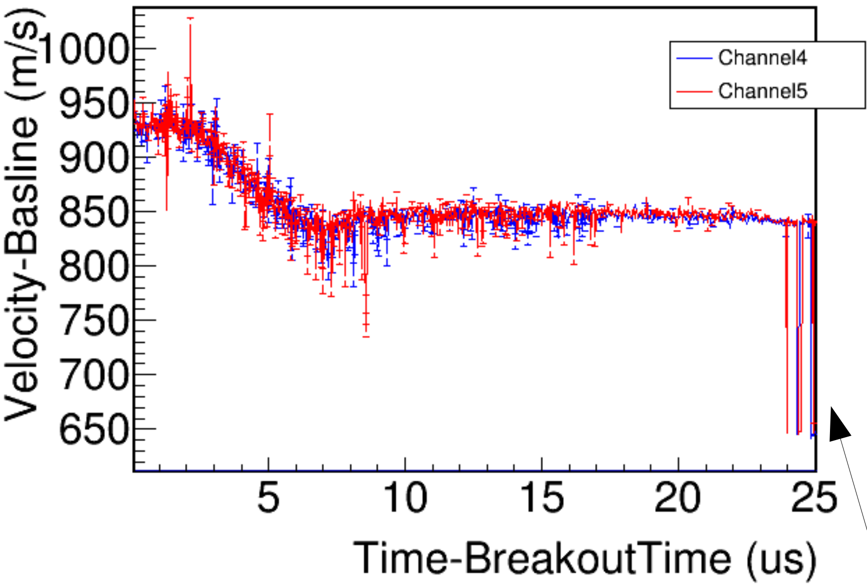


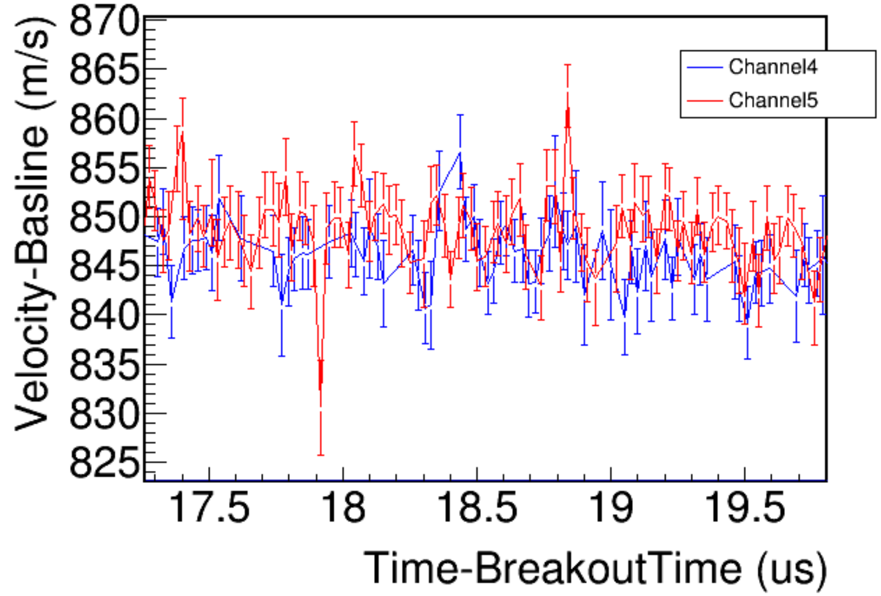
```
Time(us) Noise(dB) Uncert(dB) TSpeak1 Freq1(GHz) Uncert(GHz) Velocity1(m/s) Uncert(m/s) Amplitude1(dB) Uncert(dB) Width1(m/s) Uncert(m/s)  TSpeak2 Freq2(GHz) Uncert(GHz) Velocity2(m/s) Uncert(m/s) Amplitude2(dB) Uncert(dB) Width2(m/s) U
ncert(m/s)
175.85 -12.404 0.61744 101.7 4.4581 0.0023107 3477.3 1.8023 24.663 2.6528 6.7865 9.625e-20 25.6 4.203 0.010773 3278.3 8.4032 0.62617 3.1346 18.779 5.7451
175.87 -13.408 0.61887 104.7 4.456 0.0029025 3475.7 2.264 21.884 3.6673 7.4983 1.0136 20.7 4.2037 0.0043376 3278.9 3.3834 4.3059 4.4209 7.619 2.2897
175.9 -12.643 0.61261 81.0 4.4548 0.002313 3474.7 1.8042 19.693 2.7005 6.7863 9.7927e-20 19.5 4.2216 0.024549 3292.9 19.148 -3.3384 3.0927 29.419 12.713
175.93 -13.082 0.61827 76.8 4.4601 0.0024584 3478.9 1.9176 21 2.7241 6.7866 9.8126e-20 24.8 4.2039 0.0034618 3279 2.7002 7.1653 3.1362 6.7864 9.5292e-20
175.95 -13.639 0.61198 78.7 4.4539 0.0022929 3474 1.7884 18.449 2.6975 6.7864 9.8655e-20 33.5 4.2052 0.0044086 3280.1 3.4387 6.2866 3.7395 9.3648 1.8851
```

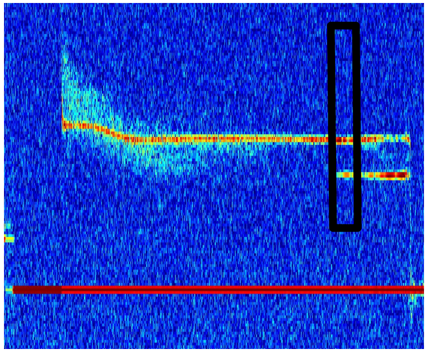
- My output includes (new features in **bold**):
  - Time
  - Noise level **and uncertainty**
  - **Probability that there is a first surface (freq. peak)**
  - Best-fit velocity (from frequency) and uncertainty
  - Amplitude **and uncertainty**
  - **Gaussian width and uncertainty (spread in velocities)**
  - **Probability that there is a second surface (if above 3-sigma)**
    - **All the same information about that surface**
  - **Probability for a third surface (if above 3-sigma)**
    - **etc**

Statistical Analysis on Sample Data

- Smaller error bars and more features



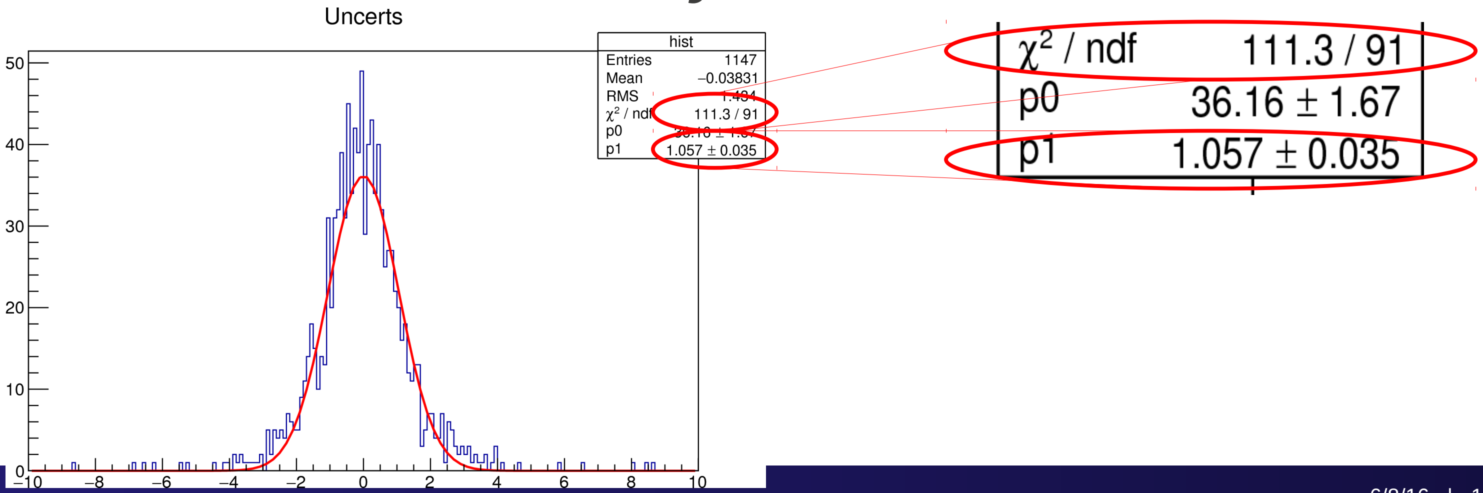




- Can use large Region of Interest (ROI)
  - Finds the second surface
- Error bars are similar size to differences between the channels

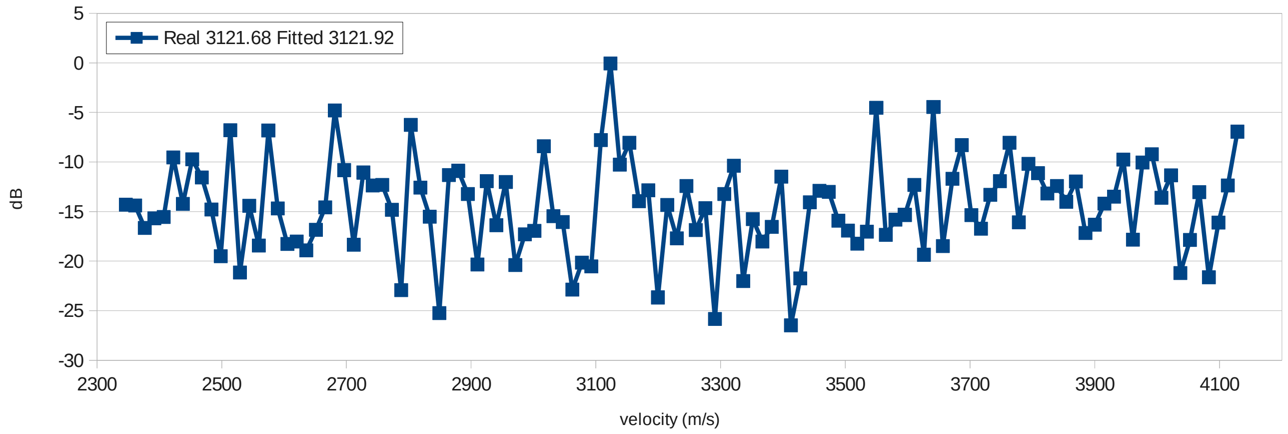
Likelihood Error Bars are Statistically Correct

- Histogrammed the difference between the channel4 and channel5 fits, divided by the uncertainty
- If the error bars are truly 68% containment (1-sigma), this histogram should be a *Gaussian* with width of *unity*



Synthetic Data Tests

- Made lineouts with noise randomly thrown on a Gaussian with mean -15dB and width 5dB
- Added peak on top of the noise at known frequency and amplitude



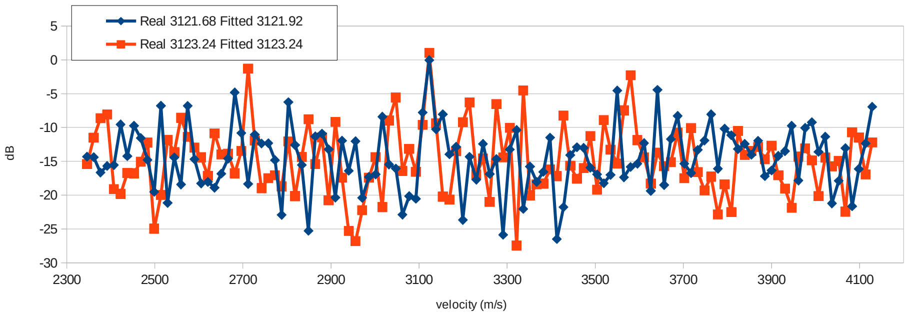
Synthetic Data – Consistency for a Dim Signal

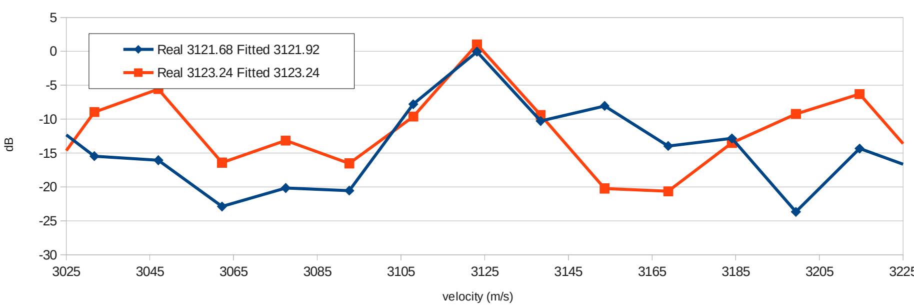
- Peak with width 7m/s, amplitude 0dB, peak frequency at 4GHz (~3-sigma)
  - Did 100 times to get mean, mean uncert, stdev
- Fitted noise level = **-15.3**dB, uncert = **0.41**dB
  - input = **-15.0**dB, stdev = **0.47**dB
- Fitted velocity = **3099**m/s, uncert = **5.0**m/s
  - input = **3100**m/s, stdev = **4.7**m/s
- Fitted amplitude = **-0.3**dB, uncert = **4.4**dB
  - input = **0.0**dB, stdev = **1.6**dB
- Fitted width = **9.8**m/s, uncert = **3.3**m/s
  - input = **7.0**m/s, stdev = **2.7**m/s

Note: Only 85/100 found the true peak. Others had a fluctuation which was higher that it found

Synthetic Data – Precision on a Ramped Velocity

- Peak with width 7m/s, amplitude 0dB (~3-sigma)
  - Ramped velocity from 3120-3134m/s in 1.56m/s steps
  - Single velocity bin is 14m/s wide





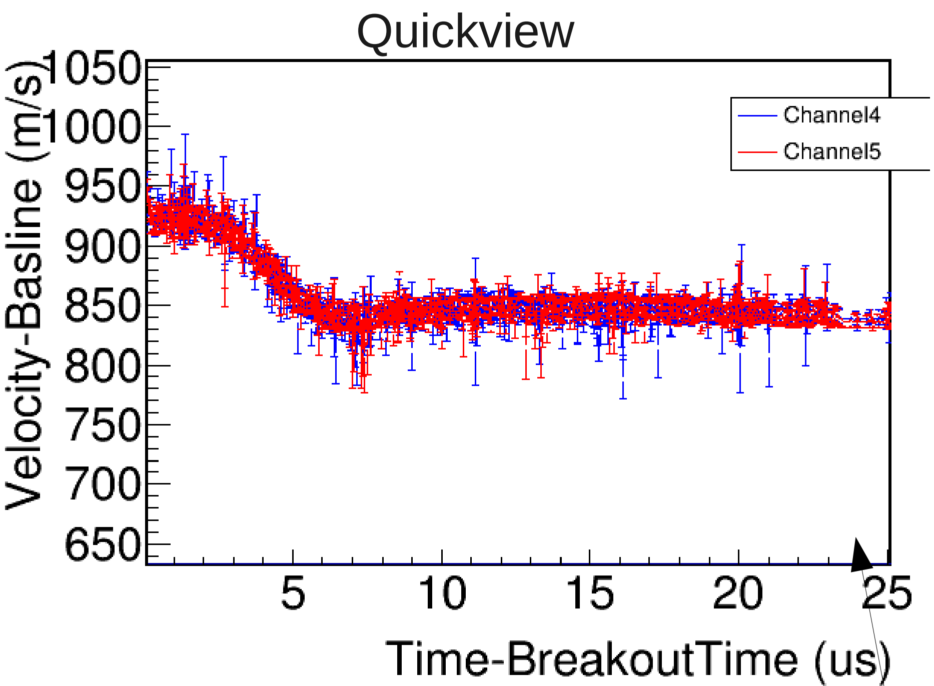
- Typical error in fitted value is ~2.3m/s
  - A few found false fluctuations, though
- Can find the peak to much better than pixel width
- Can find peak center to much better than peak width

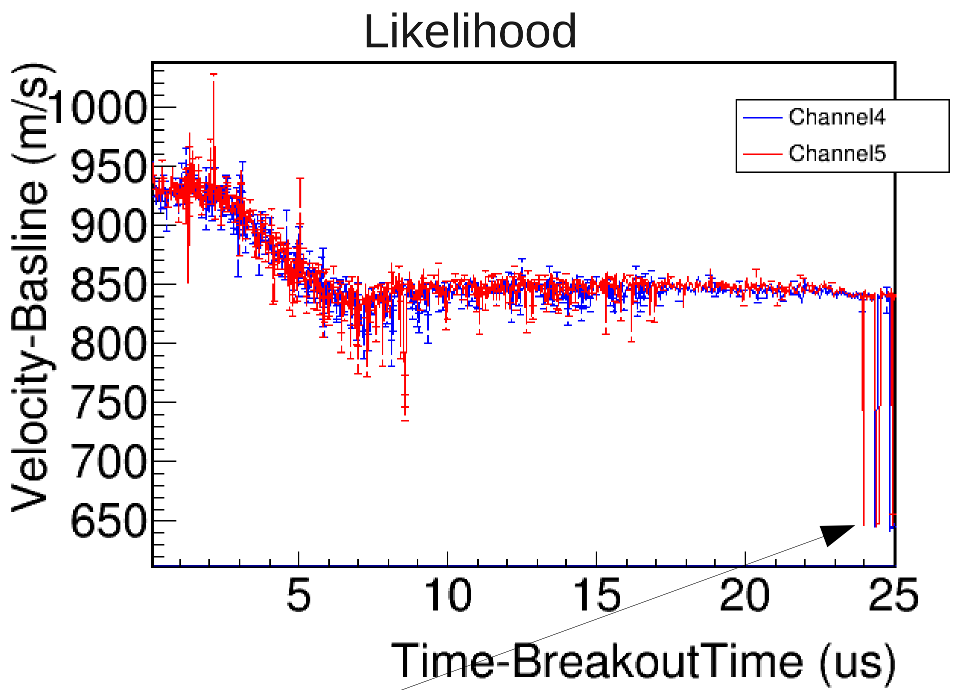
Summary

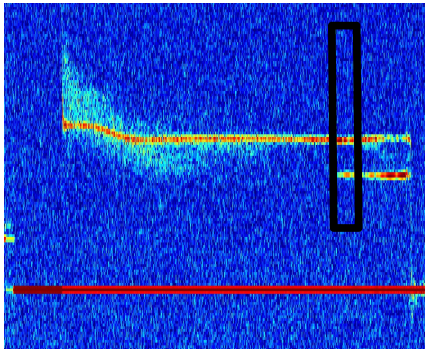
- Log-likelihood PDV velocity extractions will:
  - Remove sensitivity to ROI shape/size
  - Get statistically-validated error bars
  - Find features in the spectrogram beyond single surface
- Method is sensitive in noisy regions
  - to better than 5m/s
- Can improve sensitivity to fainter signals by using smaller ROIs (in production)
- Going to apply algorithm to Gemini-Leda-Lyra data

Backup

Likelihood vs Moments – New Features



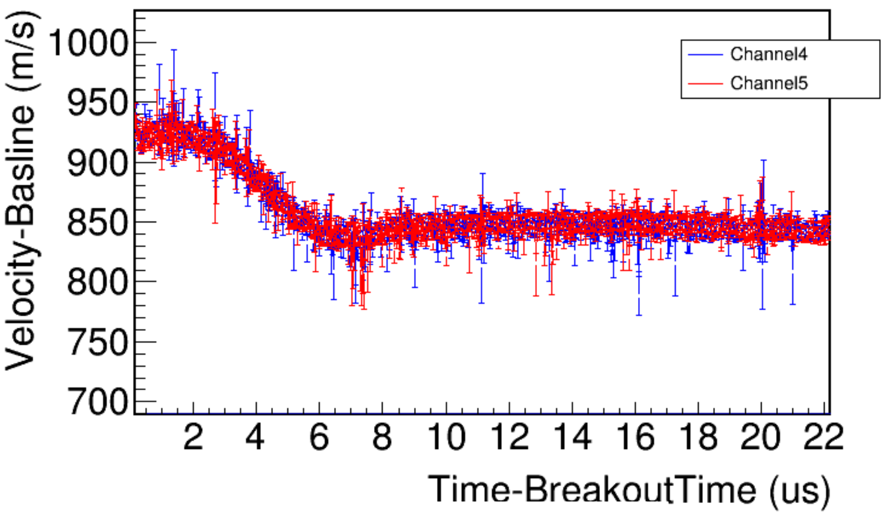




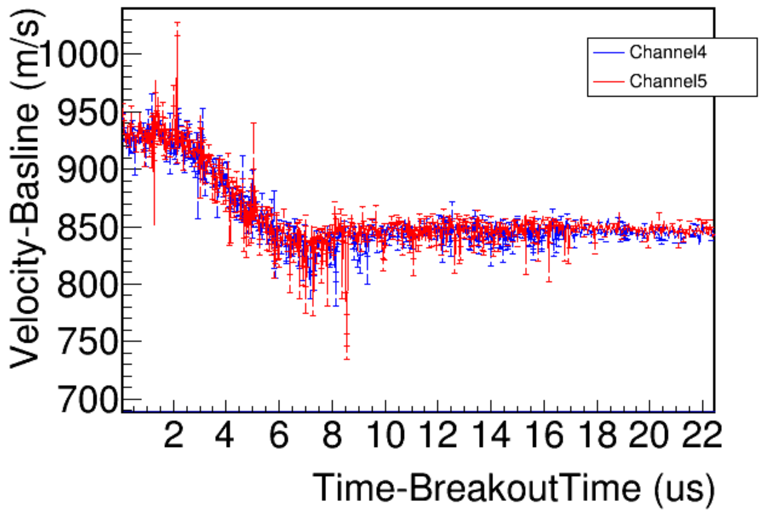
Second peak in this region
is sometimes brighter

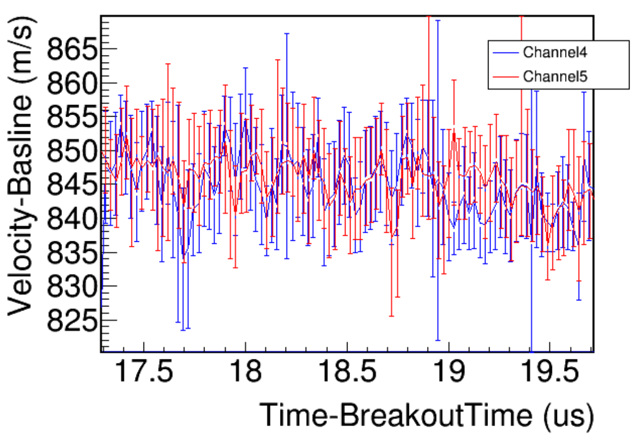
Likelihood vs Moments – Smaller Error Bars

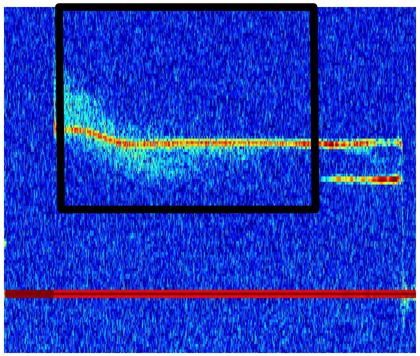
Quickview

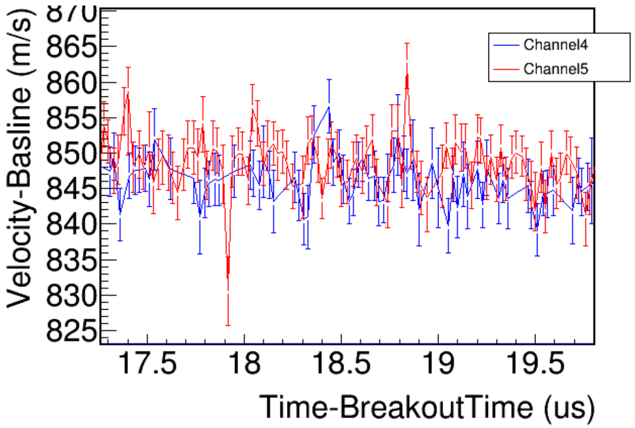


Likelihood



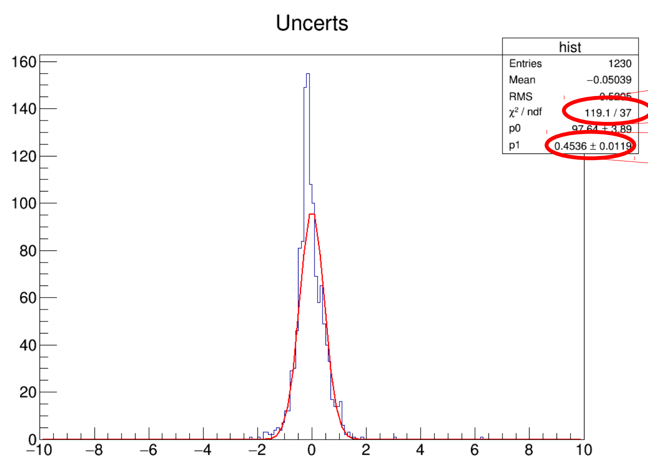


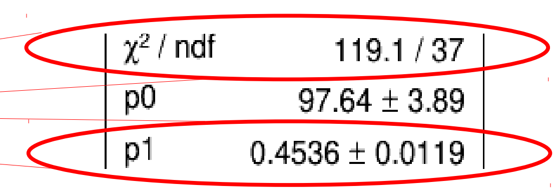




Error Bars for Moments Analysis

- Histogrammed the difference between the channel4 and channel5 fits, divided by the uncertainty
- If the error bars are truly 68% containment (1-sigma), this histogram should be a *Gaussian* with width of *unity*





- Doesn't fit well to a Gaussian
  - Errors are biased
- Errors are a factor of ~2 too large

Synthetic Data – Precision on a Ramped Velocity

- Marginal signal, found false peak 3/10 times
- Successful fits show clear linear trend upward
- Can determine velocity to much better than one spectrogram pixel, even for marginal signals