Accuracy and Precision in PDV Data Analysis

Michael R. Furlanetto

P-23, LANL

Second Annual Heterodyne Workshop, LLNL, August 17, 2007

LA-UR-07-5610

Motivation

- **Given a PDV data set [A(t)], with what accuracy and precision can we measure a velocity/frequency?**
  - Comparison between data analysis methods
  - Comparison with other methods of velocimetry
  - Comparison with other diagnostics, calculations – *establish error bars*
  - Distinguish between *real physics* and *analysis artifacts*
- Questions:
  - **What are the fundamental limits on precision and accuracy?**
  - **Does precision matter?**
  - **How close can we get to those limits with Fourier analysis? Then what?**

Swept under the rug …

- Ignore everything before the digitizer
- Assume single time-varying velocity/frequency for the analysis
  - Ignore all resolution issues
  - Not required, though – just for ease of explanation
- Lots of signal processing I don't know

Typical PDV analysis: Short-time Fourier transform

- Choose a set of (more or less) noisy data points (typically a power of 2 in length)
- Multiply by a window function
- Calculate the Fourier transform
- Move to a new set of data points and repeat
- Generates a spectrogram

- More generally, start with a data set, choose a basis set and expand the signal in that basis

Time-frequency analysis for PDV

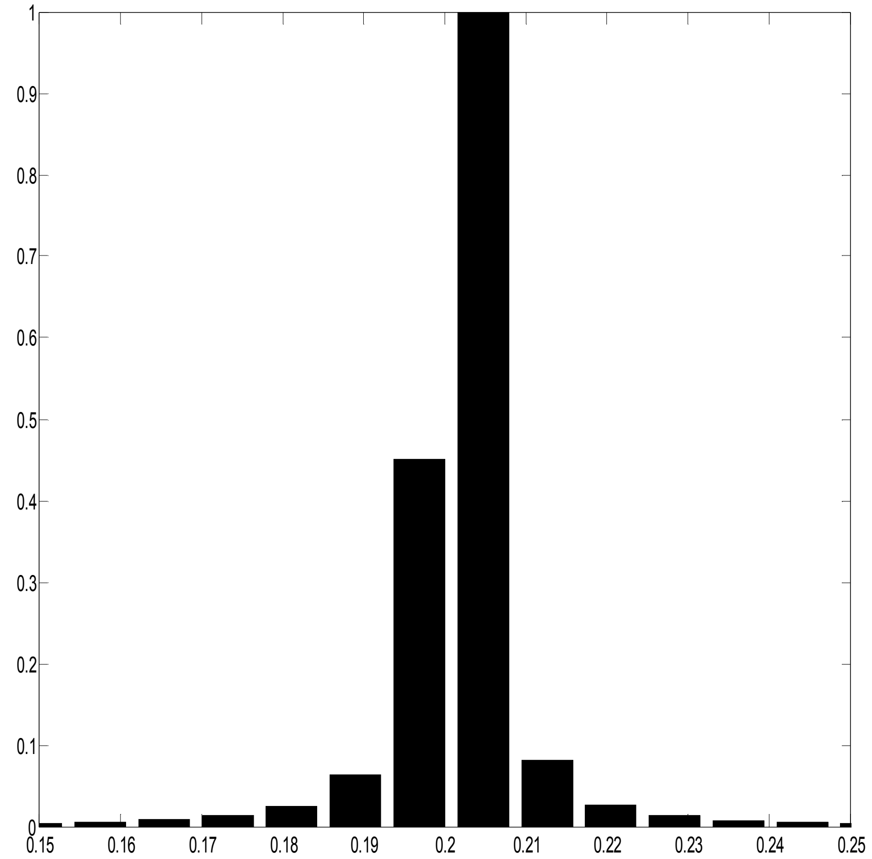
- We care about the actual velocities (frequencies)
- Discrete in time
  - have to choose locally-supported basis functions
  - e.g. window * continuous function
- Discrete in amplitude/frequency/velocity
  - “noise” from finite resolution of digitizer

Not all results from continuous (or singly discrete) theories apply!

- generally, discreteness makes things worse

Accuracy and precision in the velocity domain

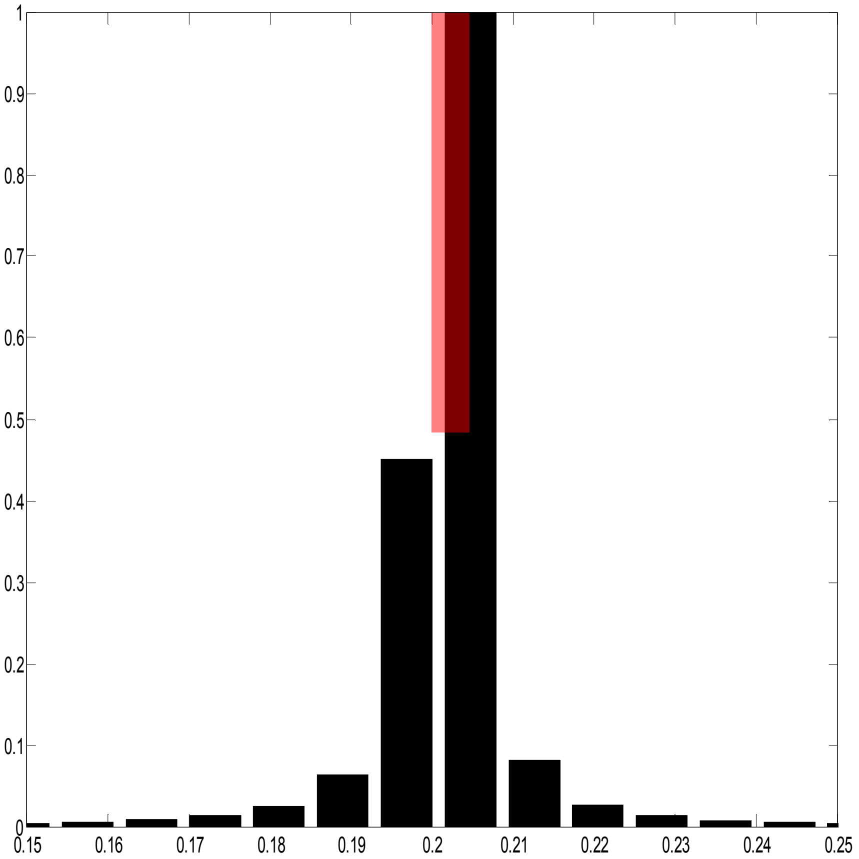
Assume the analysis procedure gives a distribution of velocities



Accuracy and precision in the velocity domain

Assume the analysis procedure gives a distribution of velocities

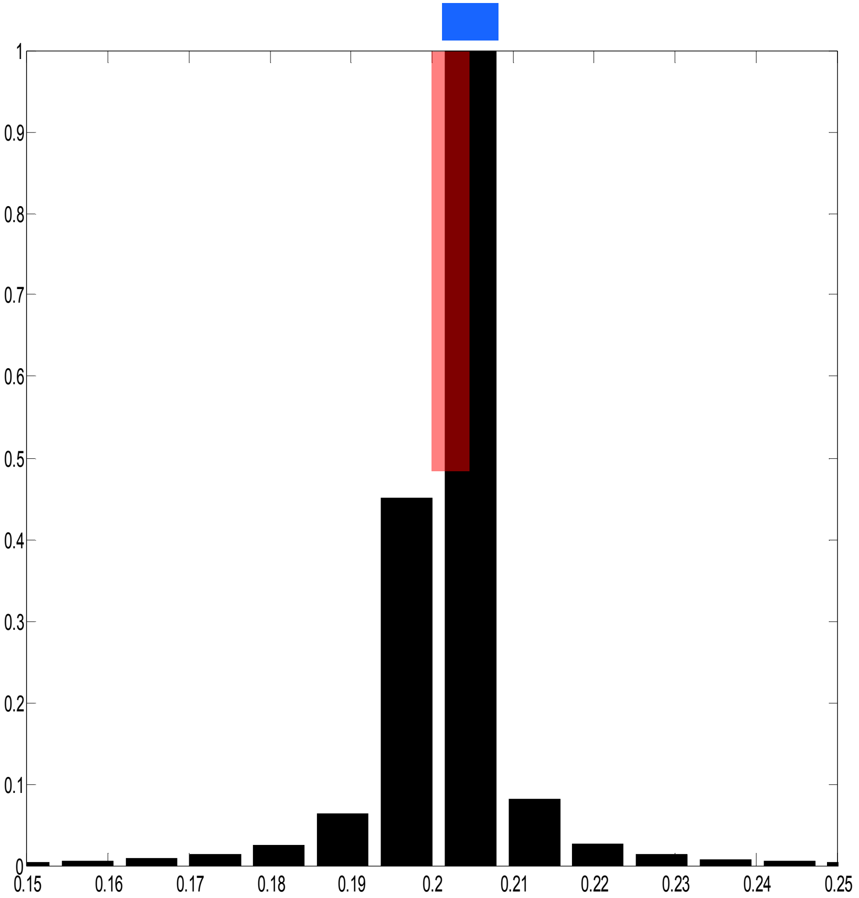
- **Accuracy (δv):** how well the "peak" of the distribution matches the actual instantaneous velocity



Accuracy and precision in the velocity domain

Assume the analysis procedure gives a distribution of velocities

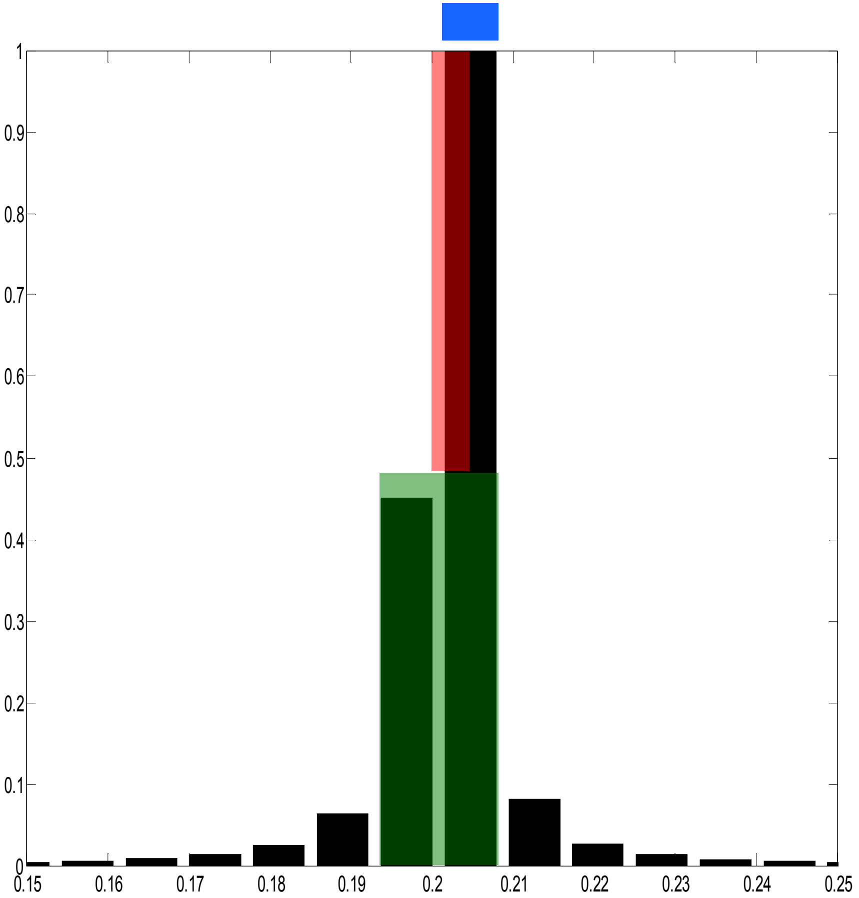
- **Accuracy (δv):** how well the "peak" of the distribution matches the actual instantaneous velocity
- **Binning precision (εf):** width of the peak of the distribution



Accuracy and precision in the velocity domain

Assume the analysis procedure gives a distribution of velocities

- **Accuracy (δv):** how well the "peak" of the distribution matches the actual instantaneous velocity
- **Binning precision (εf):** width of the peak of the distribution
- **Bandwidth precision (δf):** width of the distribution at half-maximum/-3 dB point



Accuracy and precision in the time domain?

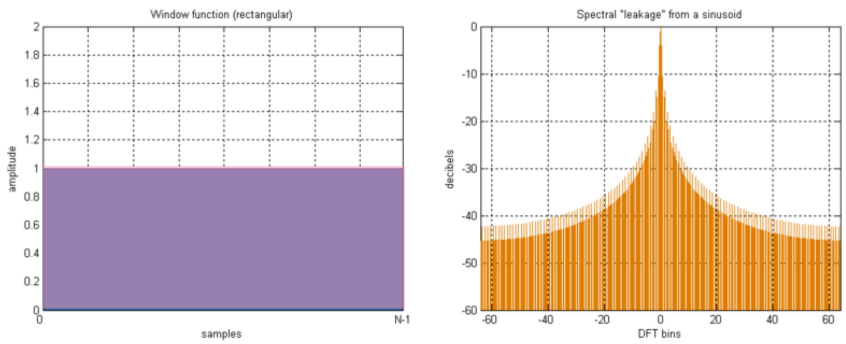
Fundamental difference between time and amplitude/frequency

→ timebase accuracy ≤1 ppm?

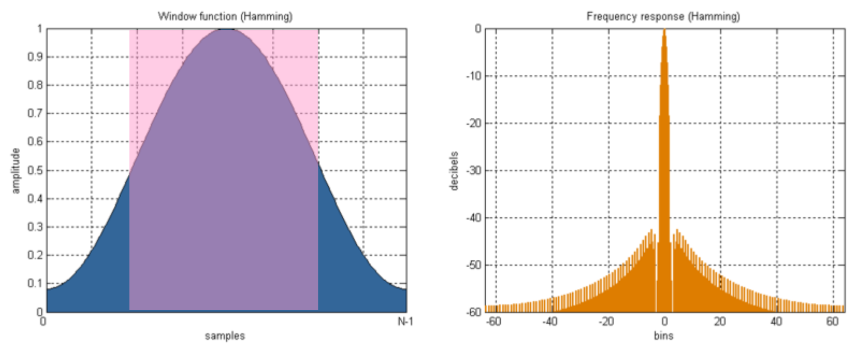
→ breakdown of analogy to quantum mechanics (Wigner-Ville and Cohen bilinear distributions)

- **Time width (t_w):** length of sequence of data under analysis between the half-maximum/-3 dB points of the window

Dirichelet



Hanning



en.wikipedia.org

Los Alamos NATIONAL LABORATORY EST.1943

Fundamental limits and uncertainty principles: a real limit

Cramér-Rao lower bound:

$$\sigma_f^2 \geq \frac{6f_s^2}{4\pi^2 * SNR * N * (N^2 - 1)}$$

- Derived from statistical estimation/information theory
  - Always valid
  - Based on SNR and number of data points
- Concerns δv and δf
- Consequences:
  - at 1 km/s, 20 Gs/s, 50 mV signal, 30 dB SNR (white noise), with 100 ps of data, can't do better than ± 5 m/s δv
  - Same conditions, 10 dB SNR, ± 50 m/s δv
  - same conditions, 62 ns of data, can't do better than ± 4 mm/s δv
  - **mainly important in noisy settings**

Fundamental limits and uncertainty principles: another real (Fourier) limit

Heisenberg-Pauli-Weyl uncertainty principle:

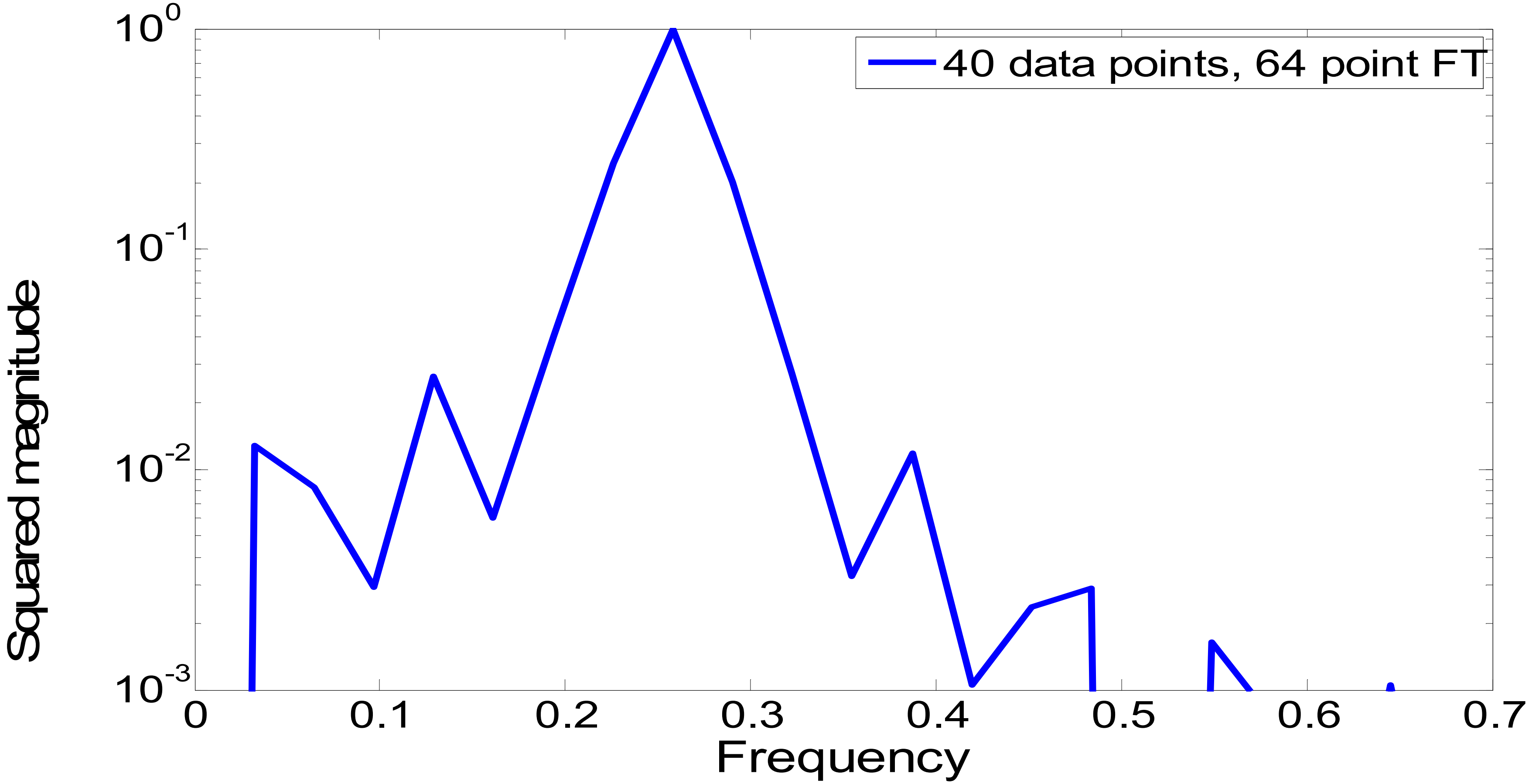
$$\sigma f \, \sigma t \geq 1/(4\pi)$$

- Derived for continuous Fourier transforms
  - More general versions: Donoho-Stark, etc.
  - Equality can't be achieved in a discrete environment
  - Similar forms are valid for all magnitude-based (non-phase) basis expansions
- Also depends on value of f – need ~1 complete period
- Concerns δf
- Consequences:
  - at 1 km/s, for 1 m/s δf, need ≥ 62 ns of data (JHRD, 2007)
  - same conditions with 100 ps of data, can't do better than ± 310 m/s δf

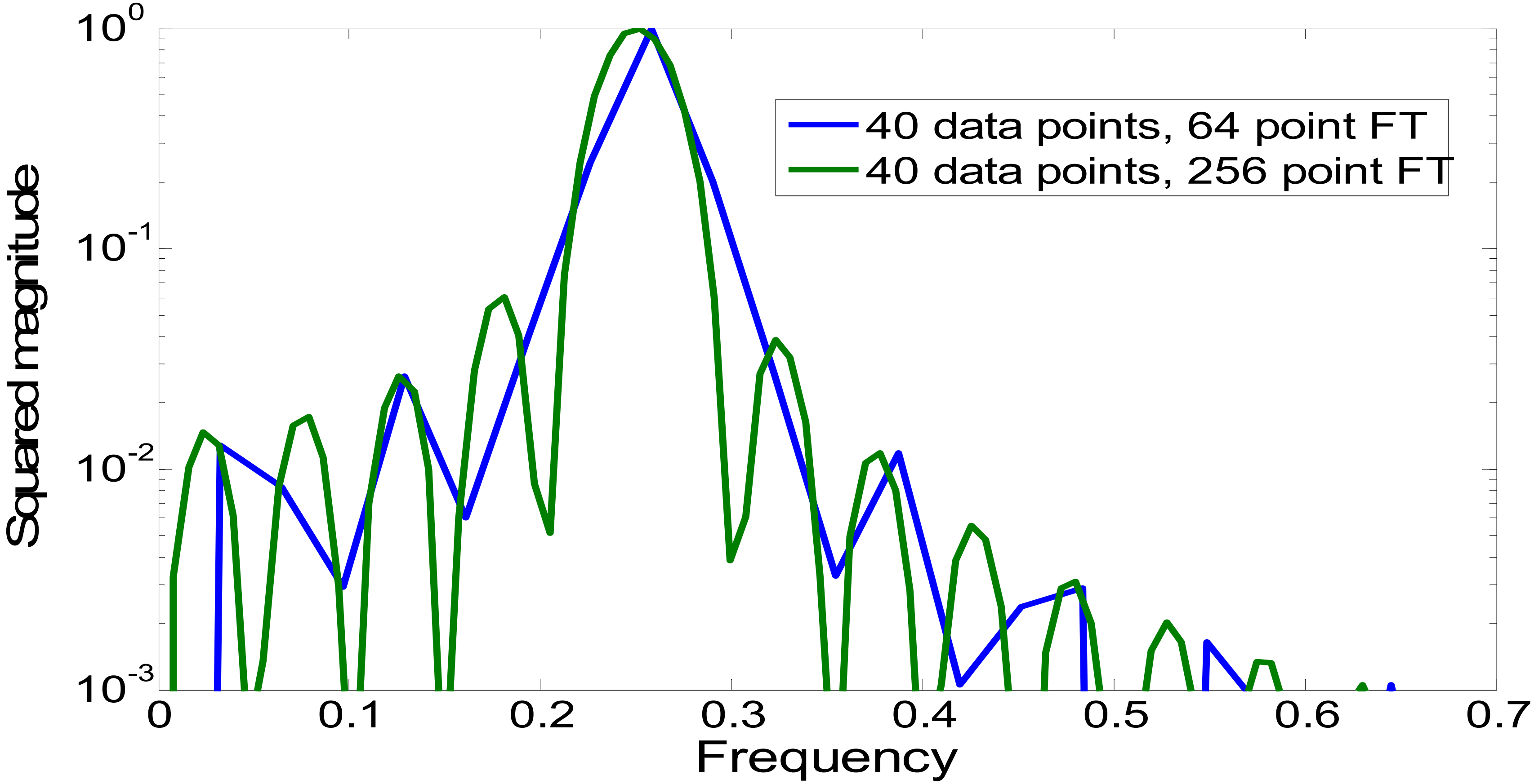
What this does not mean!

- A N_{FT}-point Fourier transform has ≤ N_{FT}/f_s span in time and frequency bins which are $f_s/(2*N_{FT})$ wide
- For a given N_d points of data you can use any $N_{FT} \geq N_d$ points for the Fourier transform
  - limited by information content (noise) of N_d, so more data points (at the same frequency) do give better frequency resolution
  - Rule of thumb: **$N_{FT} = 4 * N_d$**
  - Another rule of thumb: **use mean, not zero**
  - Already (sort of) doing this with window functions
  - Helps with εf, δf, and δv
- Anything about δv

Fourier transform length v. data length



Fourier transform length v. data length

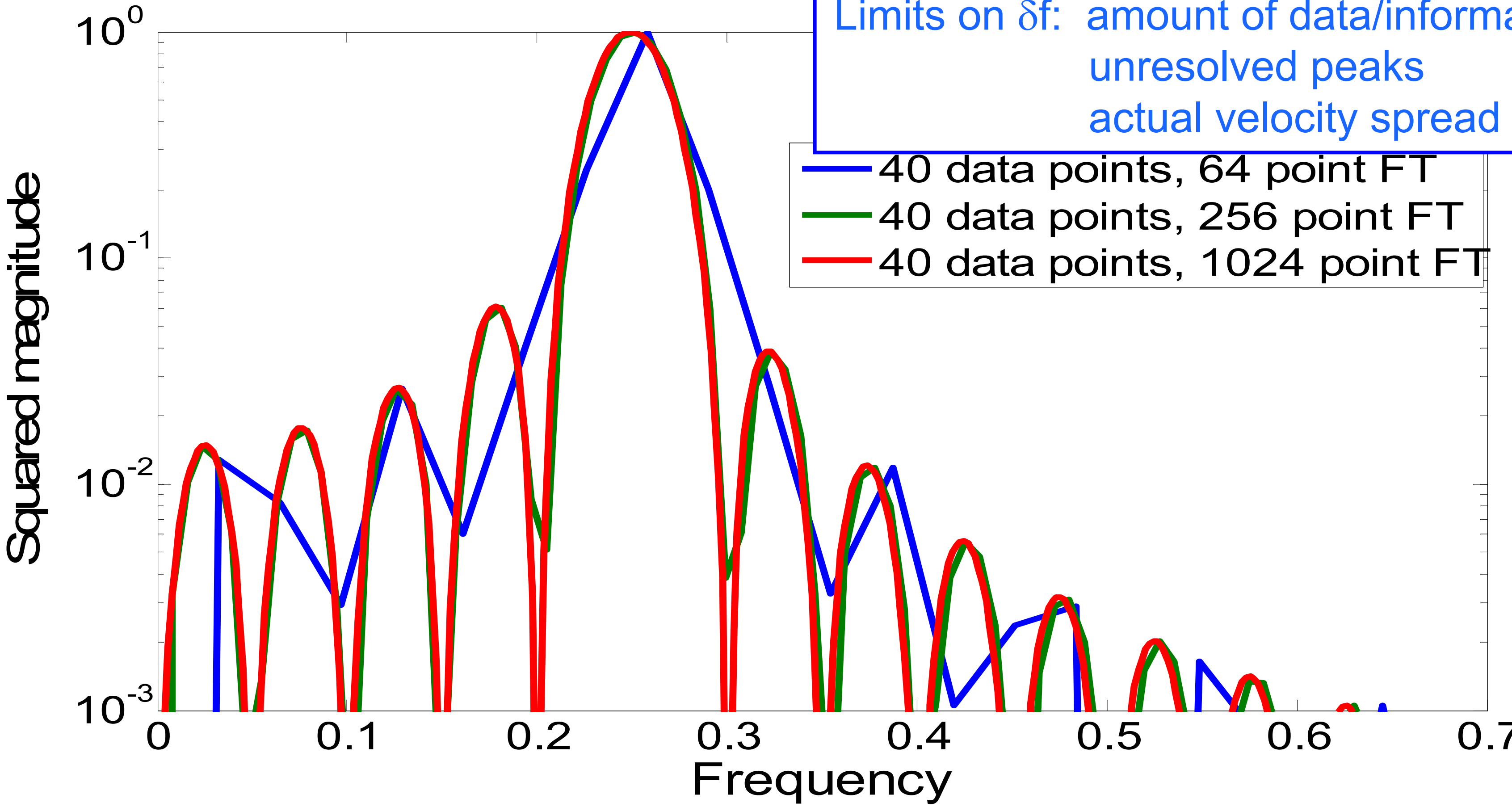


Los Alamos NATIONAL LABORATORY EST.1943

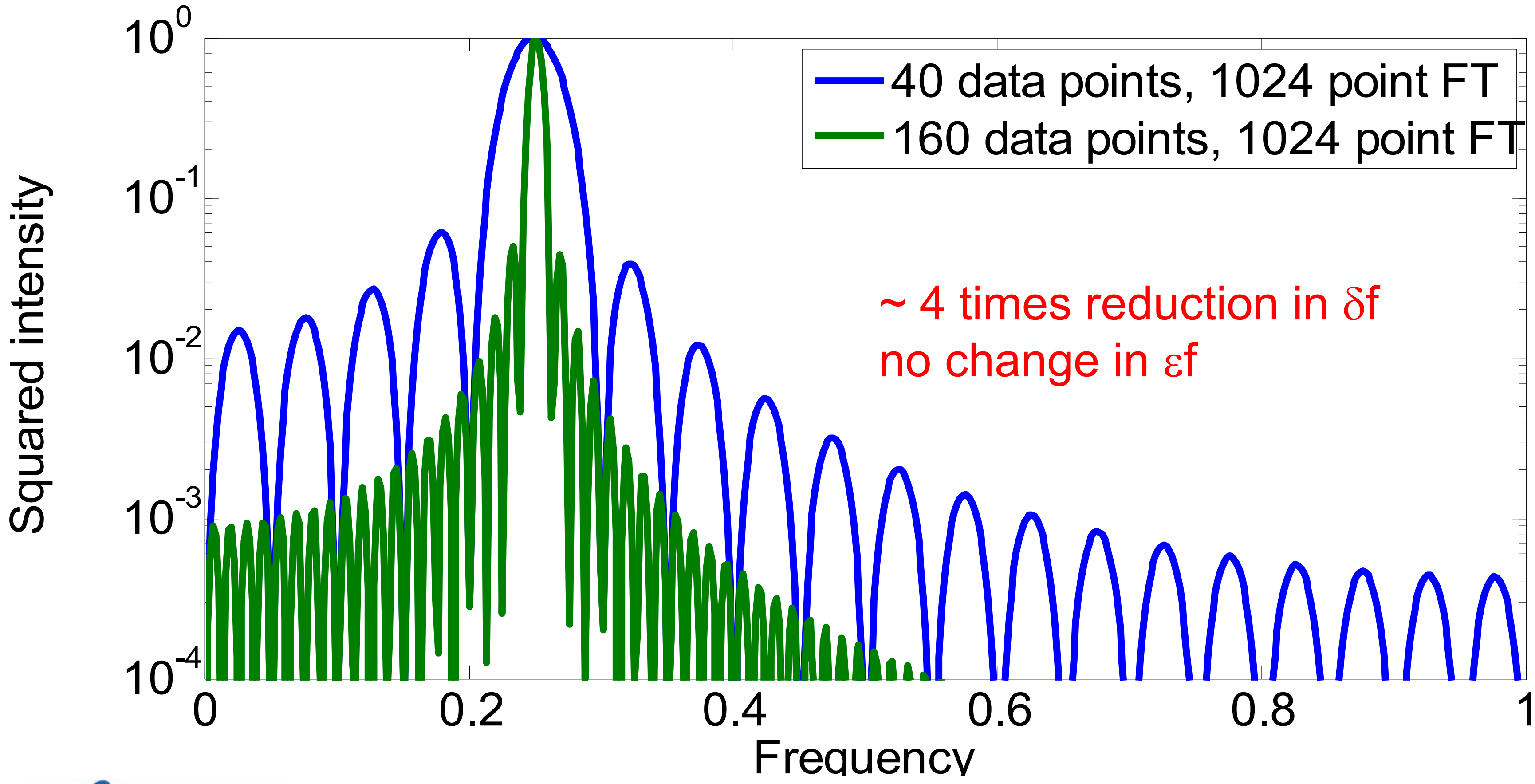
NNSA

Fourier transform length v. data length

Limits on δf: amount of data/information
unresolved peaks
actual velocity spread

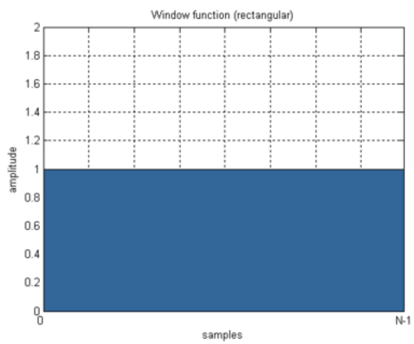


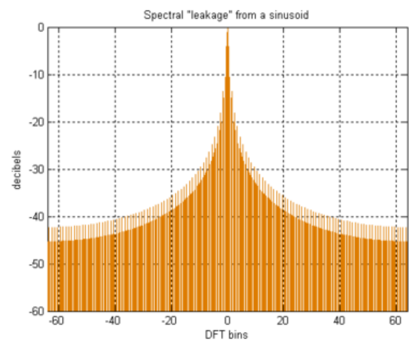
Number of data points in a Fourier transform

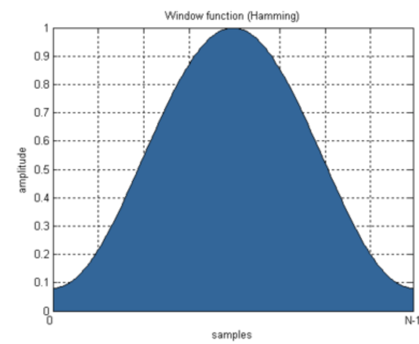


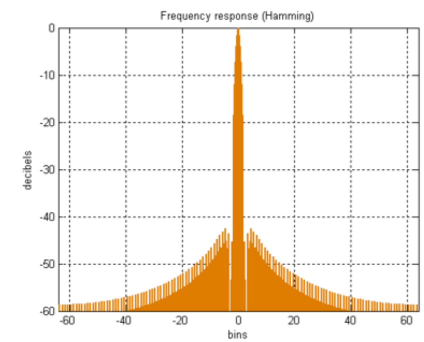
Window functions

Dirichelet | **Hanning**









Dirichelet

- $t_w = N_w * f_s$
- $\delta f = 2\pi/N_w$
- minimum δf of all windows, at a cost of -12 dB sidebands, poor dropoff

Hanning

- $t_w \sim 0.5 * N_w * f_s$
- $\delta f \sim 3\pi/N_w$
- good balance of δf, sideband amplitude, dropoff

en.wikipedia.org

Accuracy: Estimating the center of the distribution

- Center of the frequency bin with maximum value – depends on εf
- Fit a function (Gaussian, sinc, …) to the frequency distribution
- Adding more points to the distribution (followed by one of the above)
  - Fourier interpolation
  - Warped DFT, median marginal DFT, nonuniform DFT
- Correct for chirp – fractional Fourier transform?

A proposed algorithm

- Start by calculating a FT of a sample of the data (chosen by multiresolution FT?)
- Increase N_{FT} until the first sidebands are clearly resolved – minimize δf and εf
- Add or subtract N_d until the peak stops narrowing – minimize δf
- Find the center velocity by fitting and/or shifting – minimize δv and εf
- Test for more than one velocity (or a velocity distribution) in the central band by looking at the width of the peak; if multiple velocities are present, repeat steps 1-4 for each
- Look for other velocities outside of the central band and repeat 1-5 for each

Conclusions

- Cramér-Rao lower bound is a fundamental limit on δv, but mainly important in noisy environments
- Heisenberg-Pauli-Weyl uncertainty principle limits δf in Fourier-(amplitude-)based methods
- Rule of thumb: **$N_{FT} = 4 * N_d$** to minimize δf and εf
- Use adaptive window sizes to minimize δf
- Improve δv with more complicated Fourier transform
- Does precision matter?

Acknowledgments

- Mark Wilke, Andy Obst, Gus Sinnis, Todd Haines, and David Holtkamp (LANL)
- Jerry Stevens, Bruce Marshall, and Lynn Veeser (NSTec/STL)
- Tom Tunnell (NSTec/LAO)
- Campaigns 2 and 12 for funding