UPGRADING FROM REAL TO COMPLEX DATA: DETECTION OF VAUNIX AND CRES SIGNALS IN NOISE

N. Buzinsky, R. Zhang

June 25, 2025

1 Motivation

Cyclotron Radiation Emission Spectroscopy (CRES) infers electron kinematics from time–frequency spectrograms of the RF power emitted by a trapped electron. Until now, most CRES systems—including the Phase II setup in Project 8—have recorded only the magnitude-squared spectrogram, discarding the complex phase that accompanies each Fourier bin. This simplifies hardware and analysis but necessarily throws away information that could aid both detection and parameter estimation.

Thanks to recent firmware work on the ROACH platform by Goodson et al. we can now acquire zero-suppressed complex spectrograms (magnitude + phase) at high throughput. With this capability at hand, we attempt an early investigation of the potential benefits of using the complex phases for track and event reconstruction.

In principle, the real (magnitude-squared) data is a subset of the full (magnitude, phase) data, so reconstruction under this new paradigm must be strictly better than was achievable with the current spectrogram data.

Reconstruction performance can benefit in two main ways:

- Detection Efficiency: The fraction of detected events, for a given false alarm rate. Specifically, we should more efficiently detect short-duration or low SNR events than otherwise possible.
- Parameter Estimation: The accuracy with which we estimate the true start energy of observed betas.

The Vaunix signals are a natural launching point for investigating the benefits of complex data to CRES detection and parameter estimation, as (near) exact knowledge of the true signal exists in these cases:

1. Constant Tone Sinusoid
2. Linear Chirp
3. Pulsed Sinusoid

For item 1, we can evaluate the frequency resolution for real and complex data regimes, in data and in Monte Carlo. Cramer-Rao lower bounds (CRLBs) provide further theoretical support for the given reconstruction resolution being a true (ideal) lower bound.

This is extended to signals with non-zero slope in item 2, and to finite time and unknown start time in item 3. Item 3 has applications for both CRES and trap `ON/OFF` state estimation. In both cases, we are interested in the extent to which complex data reduces errors, and how this information should be incorporated in the analysis.

This leads naturally to questions for CRES signals. It is typical to treat CRES signals (locally) as linear chirps. There are 2 outstanding questions:

1. To what extent does complex data better cluster adjacent Fourier bins into a CRES event?

2. To what extent does complex data better cluster sidebands with mainbands?

2 Maximum Likelihood Estimation and Matched Filtering

In an additive white Gaussian noise (AWGN) model, our observed spectrogram data in a fixed time–frequency slice can be written as

$$\mathbf{y}_k = \mathbf{s}_k(\boldsymbol{\theta}) + \mathbf{n}_k, \quad \mathbf{n}_k \sim \mathcal{N}(0, \sigma^2), \tag{1}$$

where $s_k(\boldsymbol{\theta})$ is a deterministic template depending on unknown parameters $\boldsymbol{\theta}$ (e.g. start time, frequency, energy), and $\{\mathbf{n}_k\}$ are independent Gaussian samples.

2.1 Likelihood and Maximum-Likelihood Estimation

The *likelihood function* $\mathcal{L}(\boldsymbol{\theta}; \mathbf{y})$ is the probability of observing the data $\mathbf{y} = [y_0, \ldots, y_{N-1}]$ given parameters $\boldsymbol{\theta}$.

$$\mathcal{L}(\boldsymbol{\theta}; \mathbf{y}) = p(\mathbf{y} \mid \boldsymbol{\theta}) = \prod_{n=0}^{N-1} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\Big(-\tfrac{(y_n - s_n(\boldsymbol{\theta}))^2}{2\sigma^2}\Big). \tag{2}$$

Taking the log and discarding terms independent of $\boldsymbol{\theta}$ gives

$$\ln \mathcal{L}(\boldsymbol{\theta}; \mathbf{y}) = -\frac{1}{2\sigma^2} \sum_{n=0}^{N-1} \big(y_n - s_n(\boldsymbol{\theta})\big)^2 + \text{const.} \tag{3}$$

The *maximum likelihood estimate* (MLE), denoted $\hat{\boldsymbol{\theta}}$, is the value of $\boldsymbol{\theta}$ that makes the observed data the most probable under our model.

$$\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} \ln \mathcal{L} = \arg\max_{\boldsymbol{\theta}} \sum_{n=0}^{N-1} -\big\|y_n - s_n(\boldsymbol{\theta})\big\|^2. \tag{4}$$

Noting that the template energy $\sum s_n^2$ is independent of the data, this reduces to

$$\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} \sum_{n=0}^{N-1} y_n\, s_n. \tag{5}$$

i.e., one selects the template whose *inner product* with the data is largest. Intuitively, the MLE picks the template parameters that best "fit" the noisy data.

2.2 The Matched Filter

The statistic

$$T(\mathbf{y};\boldsymbol{\theta}) = \sum_{n=0}^{N-1} y_n\, s_n(\boldsymbol{\theta}) = \mathbf{y}\cdot\mathbf{s}(\boldsymbol{\theta}) \tag{6}$$

is known as the *matched filter*. Computing the exact MLE can be prohibitively expensive, so one approximates maximization by evaluating $T(\mathbf{y};\boldsymbol{\theta})$ over a discrete *template bank* and taking the value of $\boldsymbol{\theta}$ that gives the largest inner product.

In AWGN, the matched filter is the optimal linear filter for detecting a known signal in noise: it maximizes the output SNR and is equivalent to the likelihood-ratio test. In practice we approximate the continuous maximization over parameters $\boldsymbol{\theta}$ with a discrete *template bank*

$$\mathcal{B} = \{\mathbf{s}(\boldsymbol{\theta}_1), \ldots, \mathbf{s}(\boldsymbol{\theta}_M)\},$$

constructed on a grid in parameter space (start time, frequency, chirp rate, etc.). For a noisy observation $\mathbf{y}$, we evaluate a matched-filter score against each template and select the best match:

$$\hat{\boldsymbol{\theta}} \approx \arg\max_{\boldsymbol{\theta}_m \in \mathcal{B}} \; \mathbf{y}\cdot\mathbf{s}(\boldsymbol{\theta}_m).$$

The template-bank *granularity* is tunable: finer spacing reduces mismatch (better accuracy) but increases the number of templates M.

This computation is naturally a nearest-neighbor search problem. Interpreting each template as a vector $x_m \in \mathbb{R}^D$ and the observation as a query vector $q \in \mathbb{R}^D$, we seek the closest match under a metric derived from the matched-filter score. A brute-force solution is a linear scan over all templates, with per-query cost $O(MD)$, which becomes too slow at large M.

To accelerate retrieval, we use *approximate* nearest-neighbor (ANN) search: the returned neighbor is not guaranteed to be the exact best match, but search can be much faster while maintaining high recall. In particular, we propose Hierarchical Navigable Small World (HNSW) graphs [3], which support fast ANN queries via multi-layer graph traversal; in typical regimes, search time scales roughly like $O(\log M)$ rather than $O(M)$, with a tunable speed–accuracy trade-off.

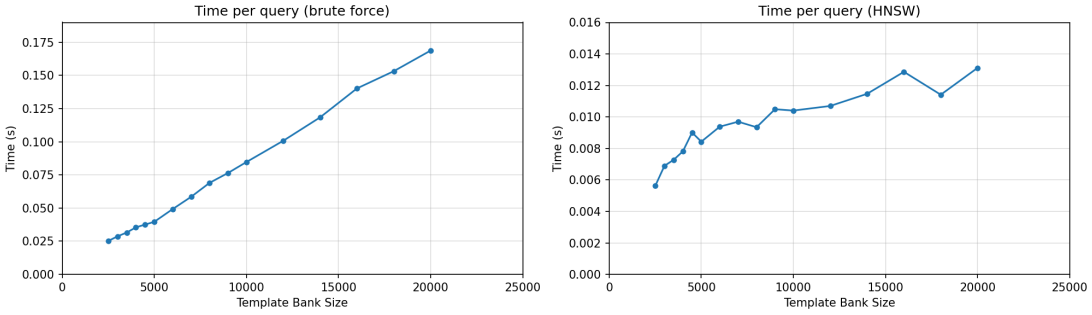

Figure 1: Measured time per query for template retrieval / matched-filter evaluation (benchmark).

3 Cramer-Rao Lower Bounds

3.1 General Description

Cramer-Rao lower bounds provide a lower limit on the minimal uncertainty of an unbiased parameter estimator. Their usage in CRES experiments was motivated by a pseudo-application of the Heisenberg/ Gabor limit ($\Delta f \Delta t \sim 1/2$), which was used to bound the energy resolution of a CRES experiment. Maximal likelihood frequency estimation can yield better frequency estimation than is expected by the Heisenberg/ Gabor limit (which has a different interpretation). The Cramer-Rao lower bounds, for instance, result in a $T^{-3/2}$ scaling in the frequency resolution as a function of the total track duration, observed numerically, as opposed to T^{-1}.

Cramer-Rao lower bounds are given by the diagonal elements of the inverse Fisher information matrix:

$$\mathrm{var}(\hat{\boldsymbol{\theta}}_{\boldsymbol{i}}) \geq [\mathbf{I}^{-1}(\boldsymbol{\theta})]_{ii} \tag{7}$$

where $\mathbf{I}(\boldsymbol{\theta})$ is defined as:

$$[\mathbf{I}(\boldsymbol{\theta})]_{ij} = \mathbb{E}\left[\left(\frac{\partial \log f(\mathbf{y} \mid \boldsymbol{\theta})}{\partial \theta_i}\right)\left(\frac{\partial \log f(\mathbf{y} \mid \boldsymbol{\theta})}{\partial \theta_j}\right)\right] \tag{8}$$

$f(\mathbf{y} \mid \boldsymbol{\theta})$ is the PDF of the data $\mathbf{y}$ as a function of the parameters of interest $\boldsymbol{\theta}$.

The Fisher information matrix quantifies the degree to which our measurement process depends on the parameters of interest (if what we observe doesn't depend on a parameter, we do not measure that parameter at all) and the degree to which changes from different parameters result in correlated changes in the resultant PDF.

The steps in computing the CRLBs are:

1. Write out the PDF for the measured quantity (e.g. electron energies, CRES Fourier bins) as a function of the parameter of interest (b_{fierz}, SNR).

2. Compute how much the PDF changes with changes of these parameters (derivatives), filling in the elements of the Fisher information matrix

3. Invert. Read off the diagonal entries.

3.2 Estimation of Signal Parameters with Additive White Gaussian Noise (Time-Domain)

Okay, let's consider parameter estimation of (real) time series signals in additive white Gaussian noise (AWGN):

$$\mathbf{y}(\boldsymbol{\theta}) = \mathbf{s}(\boldsymbol{\theta}) + \mathbf{n} \tag{9}$$

where $\mathbf{n}$ is our $\mathcal{N}(0, \sigma^2)$ time-series noise vector. Given observations of $\mathbf{y}$, we aim to estimate $\boldsymbol{\theta}$ as accurately as possible.

Consider a single observation of $\mathbf{y}$. It is distributed according to:

$$f(y \mid s, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y-s)^2}{2\sigma^2}} \tag{10}$$

which implies:

$$\log f(y \mid s, \sigma) = -\frac{1}{2}\log 2\pi\sigma^2 - \frac{(y-s)^2}{2\sigma^2} \tag{11}$$

We imagine that the signal s depends on a given parameter θ. Deriving $\log f$ with respect to θ yields:

$$\frac{\partial \log f(y \mid s, \sigma)}{\partial \theta} = \frac{(y-s)}{\sigma^2}\frac{\partial s}{\partial \theta}. \tag{12}$$

For the Fisher information matrix, we compute the expectation value of Equation 8 over all values of the data, y:

$$I_{i,j} = \frac{\partial s}{\partial \theta_i}\frac{\partial s}{\partial \theta_j}\,\mathbb{E}\left[\frac{(y-s)^2}{\sigma^4}\right] = \frac{1}{\sigma^2}\frac{\partial s}{\partial \theta_i}\frac{\partial s}{\partial \theta_j} \tag{13}$$

in which $\mathbb{E}$ is integrated over the Gaussian PDF (Eqn. 10), in which we use $\sigma^2 = \mathbb{E}\left[(y-\mu)^2\right]$.

If there are multiple observations as a time series ($\mathbf{y}$ is a vector), the Fisher information matrix becomes a sum of the time-domain of the signals. In vector notation, this is typically written as a dot product:

$$I_{i,j} = \frac{1}{\sigma^2}\left[\frac{\partial \mathbf{s}(t \mid \boldsymbol{\theta})}{\partial \theta_i} \cdot \frac{\partial \mathbf{s}(t \mid \boldsymbol{\theta})}{\partial \theta_j}\right]. \tag{14}$$

3.3 Estimation of Signal Parameters with Additive White Gaussian Noise (Spectrograms)

Tragically, He6-CRES does not have access to the full time-series. We instead use spectrograms, in which the Fourier transform is computed and the amplitude (squared) is written to disk. Clearly some information is lost when we write only the amplitude, as opposed to amplitude-plus-phase information. How do we compute the CRLB for signals in noisy data collected in such a manner?

We write our data model as:

$$\mathbf{y}(\boldsymbol{\theta}) = |\mathbf{s}(\boldsymbol{\theta}) + \mathbf{n}|^2 \tag{15}$$

We again consider a single data point before generalizing it onto the vector case. If n is still Gaussian distributed as $\mathcal{N}(0, \sigma^2)$, y is non-central χ^2-distribution with 2 degrees of freedom:

$$f(y \mid \lambda, \sigma) = \sigma\frac{e^{-\frac{\sigma(y+\lambda)}{2}}}{2} I_0(\sigma\sqrt{\lambda y}) \tag{16}$$

where $\lambda = 2s^2/\sigma^2$.

The PDF can describe both the Vaunix `ON` and `OFF` states by varying λ, with frequency bin 301 used as a representative example. It is noteworthy that both `ON` and `OFF` states result from the same amount of noise, though in the `ON` state this results in a much larger power variance. Intuitively, if we have some mean μ and some (small) variance σ, $(\mu+\sigma)^2 = \mu^2 + 2\mu\sigma + \sigma^2 \approx \mathcal{N}(\mu, \sqrt{2\mu\sigma})$. For high SNRs, the standard

deviation of these signals is proportional to the mean, and tiny fluctuations in σ cause larger fluctuations in the observed power $(\mu + \sigma)^2$.

Continuing on with the CRLBs, this expression for the log-likelihood implies:

$$\log f(y \mid \lambda, \sigma) = \log \sigma - \log 2 - \frac{\sigma(y + \lambda)}{2} + \log I_0(\sigma\sqrt{\lambda y}). \tag{17}$$

All signal parameters are

$$\frac{\partial \log f(y \mid \lambda)}{\partial \theta} = -\frac{\sigma}{2}\frac{\partial \lambda}{\partial \theta}\left[1 - \sqrt{\frac{y}{\lambda}}\frac{I_1(\sigma\sqrt{\lambda y})}{I_0(\sigma\sqrt{\lambda y})}\right] \tag{18}$$

in which we used $I_0'(x) = I_1(x)$.

$$I_{i,j} = \frac{\sigma^2}{4}\frac{\partial \lambda}{\partial \theta_i}\frac{\partial \lambda}{\partial \theta_j}\,\mathbb{E}\left[\left(1 - \sqrt{\frac{y}{\lambda}}\frac{I_1(\sigma\sqrt{\lambda y})}{I_0(\sigma\sqrt{\lambda y})}\right)^2\right] \tag{19}$$

or noting $\lambda = 2s^2/\sigma^2$:

$$I_{i,j} = \frac{\partial s^2}{\partial \theta_i}\frac{\partial s^2}{\partial \theta_j}\,\mathbb{E}\left[\left(1 - \sqrt{\frac{y}{\lambda}}\frac{I_1(\sigma\sqrt{\lambda y})}{I_0(\sigma\sqrt{\lambda y})}\right)^2\right]. \tag{20}$$

A question you may be asking yourself: what is going on? Equation 20 is analogous to Equation 13. There is an expectation value (over all possible noise values of the PDF), which we could analytically evaluate in the Gaussian case, but not here. This is fine. If we know σ (given our noise), we integrate over all possible y's, to get some function $w(\lambda)$, or effectively, a *weighting* function of the signal power $w(s^2)$.

Generalizing to observations over multiple time samples, we have a *weighted* dot product over our signal *power* vectors:

$$I_{i,j} = \sum_n w(\mathbf{s^2}(t_n))\frac{\partial \mathbf{s^2}(t_n)}{\partial \theta_i}\frac{\partial \mathbf{s^2}(t_n)}{\partial \theta_j} \tag{21}$$

where w is computed numerically from the expectation value in Equation 20. Contrast Equation 21 with Equation 14.

3.4 Constant-Tone Sinusoidal Signals

3.4.1 Time-Domain Data

In this section we apply the general AWGN CRLB machinery to the simplest possible signal: a constant-amplitude tone of known phase and frequency. Our goal is to see how much the variance of unbiased estimators of amplitude, phase, and frequency can be driven down by using time-series (complex) data versus spectrogram (power-only) data.

$$\mathbf{s}(t \mid \boldsymbol{\theta}) = Ae^{i\omega_0 t + i\phi_0} \tag{22}$$

Equation 22 defines our signal model as a single-tone complex sinusoid of amplitude A, angular frequency ω_0, and initial phase ϕ_0. To apply the CRLB formalism, we first compute the sensitivities of the signal to each parameter:

$$\frac{\partial \mathbf{s}}{\partial A} = \frac{\mathbf{s}}{A} \tag{23}$$

$$\frac{\partial \mathbf{s}}{\partial \phi} = i\mathbf{s} \tag{24}$$

$$\frac{\partial \mathbf{s}}{\partial \omega_0} = i\mathbf{t}\mathbf{s} \tag{25}$$

Substituting these into the general Fisher information expression in 14 and carrying out the time-sum yields the matrix shown in 26.

$$\mathbf{I}(\boldsymbol{\theta}) = \frac{2}{\sigma^2}\begin{bmatrix} \langle 0 \rangle & 0 & 0 \\ 0 & A^2\langle 0 \rangle & A^2\langle 1 \rangle \\ 0 & A^2\langle 1 \rangle & A^2\langle 2 \rangle \end{bmatrix} \tag{26}$$

where $\langle k \rangle = \sum_n^N t_n^k$.

Finally, inverting this 3×3 matrix and reading off its diagonal entries gives the CRLBs:

$$\mathrm{var}(\hat{A}) \geq \frac{\sigma^2}{2N} \tag{27}$$

$$\mathrm{var}(\hat{\omega}_0) \geq \frac{6\sigma^2}{A^2 t_s^2 N(N^2-1)} \sim \frac{6\tau_{\mathrm{SNR}}}{T^3} \tag{28}$$

$$\mathrm{var}(\hat{\phi}_0) \geq \frac{\sigma^2(2N-1)}{A^2 N(N+1)} \sim \frac{2\tau_{\mathrm{SNR}}}{T} \tag{29}$$

3.4.2 Spectrogram Data

Instead of the complete parameter estimation, which looks at the data in all frequency bins, we just consider the high-power bin. It is evident that ϕ_0 and ω_0 are completely unknown within the bin. The true distribution will be uniformly distributed within the Fourier bin, and in the range $[0, 2\pi]$, respectively. The minimal variance strategy is to choose the center of the frequency bin. The variance then corresponds to the uniform distribution.

$$\mathrm{var}\,\hat{\omega}_0 = \frac{(\Delta\omega)^2}{12} \tag{30}$$

$$\mathrm{var}\,\hat{\phi}_0 = \frac{4\pi^2}{12} \tag{31}$$

What about the amplitude estimation?

Let $P = s^2$.

$$\mathrm{var}\,\hat{P} \geq \frac{1}{Nw(P)} \tag{32}$$

Note for Gaussian noise, amplitude estimation mimics this result in the limit as $w(P) \approx 1/\sigma^2$.

3.5 Linear Chirp Signals

3.5.1 Time-Domain Data

For a linear chirp of the form

$$s(t \mid \boldsymbol{\theta}) = A \exp\Big\{ i \left(\omega_0 t + \frac{\alpha_0}{2} t^2 + \phi_0 \right) \Big\} \tag{33}$$

the additional chirp-rate parameter α_0 enters through

$$\frac{\partial \mathbf{s}}{\partial \alpha_0} = \frac{i \mathbf{t}^2 \mathbf{s}}{2} \tag{34}$$

Retaining the same derivatives with respect to A, ω_0, and ϕ_0, one evaluates the Fisher matrix and its inverse to arrive at the CRLBs:

$$\mathrm{var}(\hat{A}) \geq \frac{\sigma^2}{2N} \tag{35}$$

$$\mathrm{var}(\hat{\phi}_0) \geq \frac{\sigma^2}{8A^2} \frac{9N^2 - 21}{N(N^2 - 4)} \tag{36}$$

$$\mathrm{var}(\hat{\omega}_0) \geq \frac{6\sigma^2}{2A^2 t_s^2} \frac{1}{N(N^2 - 1)} \tag{37}$$

$$\mathrm{var}(\hat{\alpha}_0) \geq \frac{360\sigma^2}{A^2 t_s^4} \frac{1}{N(N^2 - 1)(N^2 - 4)} \tag{38}$$

3.5.2 Spectrogram Data

With only discrete spectrogram bins to trace the chirp, one typically fits a straight line to the bin-centers versus time. Standard least-squares then yields

$$\mathrm{var}\, \hat{\omega}_0 = \frac{(\Delta\omega)^2}{12N} \tag{39}$$

$$\mathrm{var}\, \hat{\alpha}_0 = \frac{(\Delta\omega)^2}{12N \left\langle t^2 \right\rangle} = \frac{(\Delta\omega)^2}{NT^2} \tag{40}$$

$$\mathrm{var}\, \hat{\phi}_0 = \frac{4\pi^2}{12} \tag{41}$$

$T = N\Delta t$.

The notable thing is that the scaling with the event length is different (worse) compared to the full time-series parameter estimation.

3.6 Pulsed Sinusoidal Signals

Finally, consider a train of K identical pulses of duration T_{on} spaced by T_{acq}:

$$s(t \mid \boldsymbol{\theta}) = A \sum_{k=0}^{K-1} g\left(\frac{t - t_0 - kT_{\mathrm{acq}}}{T_{\mathrm{on}}} \right) \tag{42}$$

Here t_0 is the unknown start time and $g(u)$ is the pulse-shape function. Differentiating with respect to each parameter gives

$$\frac{\partial \mathbf{s}}{\partial A} = \frac{\mathbf{s}}{A} \tag{43}$$

$$\frac{\partial \mathbf{s}}{\partial t_0} = -\frac{A}{T_{on}} \sum_{k}^{K-1} \dot{g}(u) \tag{44}$$

$$\frac{\partial \mathbf{s}}{\partial T_{acq}} = -\frac{A}{T_{on}} \sum_{k}^{K-1} k\, \dot{g}(u) \tag{45}$$

$$\frac{\partial \mathbf{s}}{\partial T_{on}} = -\frac{A}{T_{on}} \sum_{k}^{K-1} u\, \dot{g}(u) \tag{46}$$

$u = \frac{t - t_0 - kT_{\mathrm{acq}}}{T_{\mathrm{on}}}$

Substituting these into 14 (or its spectrogram analogue in 21) and summing over all t_n yields the Fisher matrix for the pulsed model. Inverting it then provides CRLBs for amplitude, timing, pulse period, and pulse-width estimation. As expected, the variance in t_0 and T_{acq} decreases with more pulses K, and the precision of T_{on} improves with both the number of pulses and the pulse-shape derivatives $\dot{g}$.

4 Results (Monte Carlo)

In this section we summarize the Monte Carlo experiments and data analyses carried out for each of our three test signals. For each case, we generated a large number of independent realizations of the model signal in additive white Gaussian noise, varied the SNR over a wide range, and applied two estimators:

- **Real-only estimator:** uses only the magnitude-squared spectrogram (no phase).
- **Complex estimator:** uses the full complex time-series (magnitude *and* phase).

We then computed the root-mean-square error (RMSE) of the parameter estimates at each SNR and compared both estimators to the theoretical Cramér–Rao lower bound (CRLB). All results are plotted on log–log axes, with RMSE on the vertical axis and SNR on the horizontal axis.

4.1 Constant-Tone Sinusoidal Signals

For the constant-tone case we fixed the amplitude A and phase ϕ_0 at their true values and estimated only the frequency ω_0. At each SNR value we ran 10^4 Monte Carlo trials, computed the estimated $\hat{\omega}_0$ via maximum-likelihood (i.e. matched-filter) fitting, and recorded the $\mathrm{RMSE}(\hat{\omega}_0) = \sqrt{\mathbb{E}\big[(\hat{\omega}_0 - \omega_0)^2\big]}$.

Figure 2 shows the RMSE curves for the real-only estimator, the complex estimator, and the CRLB.

A few key observations emerge:

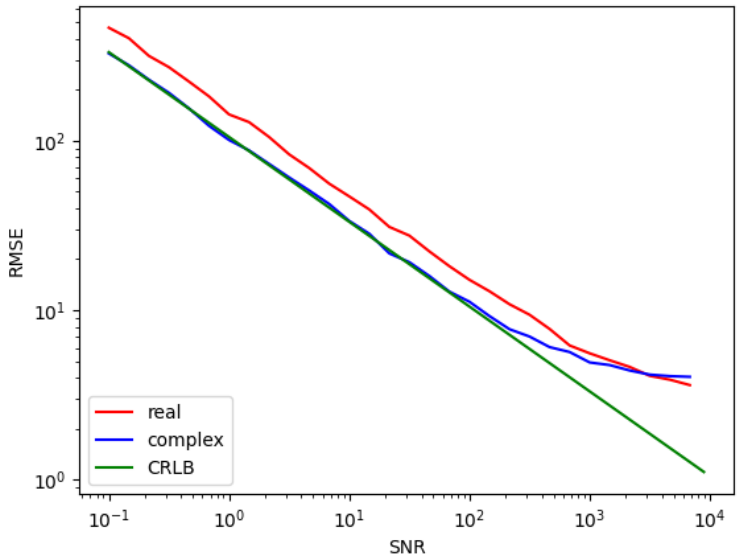

Figure 2: RMSE of frequency estimate vs. SNR for a constant-tone sinusoid. The complex estimator (blue) closely tracks the CRLB (green), while the real-only estimator (red) lies a constant factor $\sqrt{2}$ above. At high SNR all curves flatten due to finite-sample resolution.

1. All three curves slope downward roughly as $1/\mathrm{SNR}^{3/2}$, reflecting the $T^{-3/2}$ scaling of the CRLB in 28.

2. Over the low to mid SNR range, the complex-data estimator lies essentially on top of the CRLB, confirming that it achieves statistical efficiency, while the real-only estimator is uniformly higher by a factor of $\sqrt{2}$. This factor exactly matches the loss of one quadrature of information when the phase is discarded.

3. Above SNR $\sim 10^2$, both estimators level off at a common floor. This floor corresponds to the discrete-frequency bin resolution, template bank spacing, and to numerical precision limits in the solver.

These results validate our theoretical expectations for the constant-tone case: complex data fully realizes the CRLB, while magnitude-only data suffers a $\sqrt{2}$ penalty and cannot improve beyond the bin width at very high SNR.

4.2 Linear Chirp Signals

For the linear-chirp case we again fix the amplitude A and phase ϕ_0 at their true values, as well as the chirp rate α_0, estimating only the start frequency ω_0. Figure 3 shows the resulting RMSE curves.

4.3 Pulsed Sinusoidal Signals

Finally, for the pulsed sinusoid case we simulated a train of K equal-amplitude pulses, and applied matched-filter fitting to estimate only the pulse start time t_0, holding the

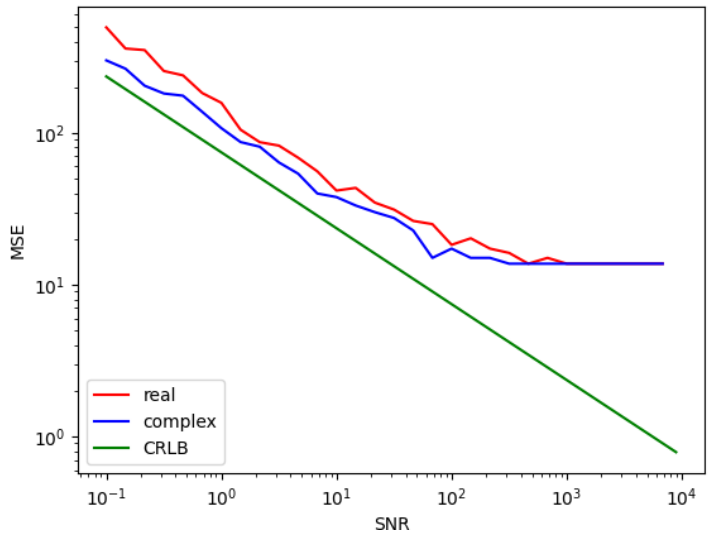

Figure 3: RMSE of start-frequency estimate vs. SNR for a linear chirp. The complex estimator (blue) lies slightly above the CRLB (green), while the real-only estimator (red) remains a factor of $\sqrt{2}$ above the complex curve.

pulse-on duration T_{on}, off-time T_{off}, carrier frequency ω_v, and phase ϕ_v at their true values. The results can be seen in Figure 4.

5 Summary

Across three archetypal signal models—constant-tone, linear chirp, and pulsed sinusoid—we have demonstrated quantitatively how retaining complex spectrogram data restores the full statistical information available. When the template exactly matches a stationary tone, the complex estimator realizes the CRLB floor and demonstrates a $\sqrt{2}$ factor improvement over the magnitude-only case. Allowing a time-varying frequency (linear chirp) decreases performance with respect to the CRLB, but the complex still outperforms magnitude-only by the same quadrature penalty. In the pulsed case, both methods deviate significantly from the CRLB, yet the $\sqrt{2}$ gap persists.

These results establish two takeaways for future CRES analysis.

- Wherever hardware permits, pipelines should transition to complex-data workflows: doing so fully recovers phase information, yielding significant improvements over the magnitude-only methods.
- The measured departures from the CRLB in the chirp and pulse scenarios point to opportunities for more sophisticated template design and iterative fitting algorithms that can close the remaining gap.

Building on this foundation, the next steps include applying complex-data matched filters to real experimental streams, integrating phase-aware clustering of adjacent bins and sidebands, and extending the framework to joint detection–estimation schemes. The benchmarks presented here will serve as a quantitative yardstick for assessing

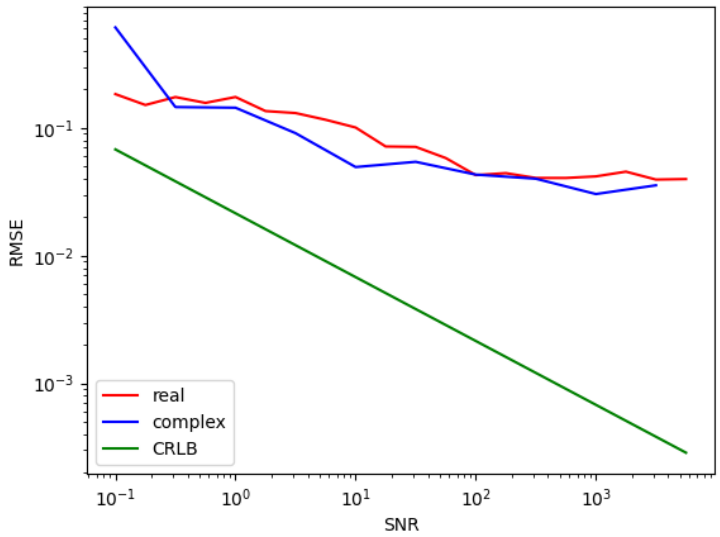

Figure 4: RMSE of start-time estimate t_0 vs. SNR for a pulsed sinusoid. The complex estimator (blue) lies a factor $\sqrt{2}$ below the real-only curve (red), even though both curves plateau compared to the CRLB (green).

those developments and for guiding both firmware upgrades and analysis-software enhancements.

References

[1] Nicholas Buzinsky. "Statistical Signal Processing and Detector Optimization in Project 8". Ph.D. thesis. PhD thesis. Massachusetts Institute of Technology, Sept. 2021.

[2] Hongya Ge and D. W. Tufts. "Cramer–Rao lower bounds on estimating the parameters of a filtered burst of sinusoid". In: *IEEE Transactions on Aerospace and Electronic Systems* 33.2 (1997), pp. 421–431. DOI: 10.1109/7.575875.

[3] Yu. A. Malkov and D. A. Yashunin. *Efficient and robust approximate nearest neighbor search using Hierarchical Navigable Small World graphs*. 2018. arXiv: 1603.09320 [cs.DS]. URL: https://arxiv.org/abs/1603.09320.