Phase-parameter marginalization: a new paradigm for continuous-wave searches? or Bayesian Ramblings on CW detection methods

Reinhard Prix

Albert-Einstein-Institut Hannover

GWPAW2012, June 2012, Hannover

LIGO-G1200438-v2

Outline

Outline

Toy CWs: Sinusoids in Gaussian Noise

Simplified CW Signal Model

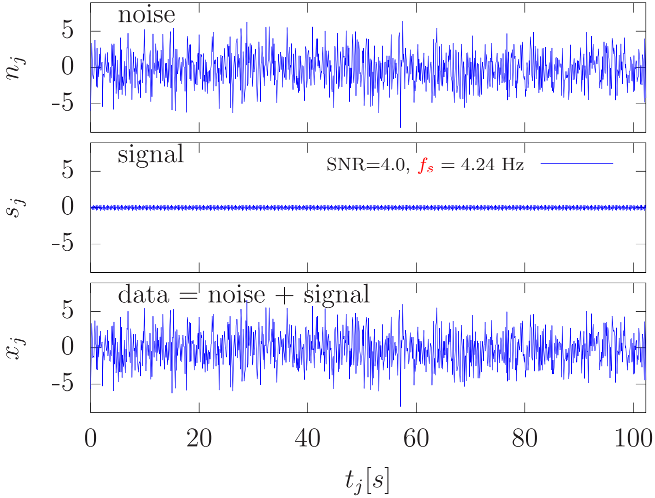
$$s(t;\mathcal{A},f) = \mathcal{A}_1 \sin(2\pi f\, t) + \mathcal{A}_2 \cos(2\pi f\, t)$$

"Amplitude parameters": $\mathcal{A} \equiv \{\mathcal{A}_1, \mathcal{A}_2\}$
"Phase parameters": $\lambda \equiv \{f\}$ $\qquad$ (CW : $\lambda = \{f, \dot{f}, \vec{n}, \vec{b} \ldots\}$)

- Measurement $x_j = n_j + s_j(\mathcal{A},\lambda)$, spanning $t \in [0, T]$
  - Sampling: $s_j \equiv s(t_j)$ where $t_j = j\,\Delta t$, $j = 1 \ldots N$
  - Gaussian noise n_j: $\mathrm{E}\left[n_j\right] = 0$, $\mathrm{E}\left[n_i\, n_j\right] = \frac{S_\mathrm{n}}{2\Delta t}\,\delta_{ij}$
- Signal-to-Noise ratio: $\mathrm{SNR}^2 \equiv (s|s) = \frac{2}{S_\mathrm{n}} \int_0^T s^2(t)\,dt$

  ☞ $\mathrm{SNR} = \frac{A\sqrt{T}}{\sqrt{S_\mathrm{n}}}$

Toy CWs: Sinusoids in Gaussian Noise



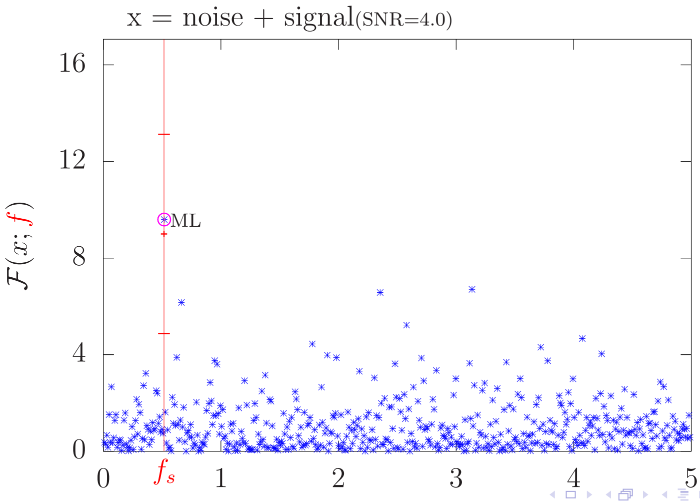
sampling:
$N = 1024$
$\Delta t = 0.1\,\mathrm{s}$
$T = 102.4\,\mathrm{s}$

signals:
$f_s \in [0, 5]\,\mathrm{Hz}$
fixed SNR

noise:
$S_\mathrm{n} = 1$

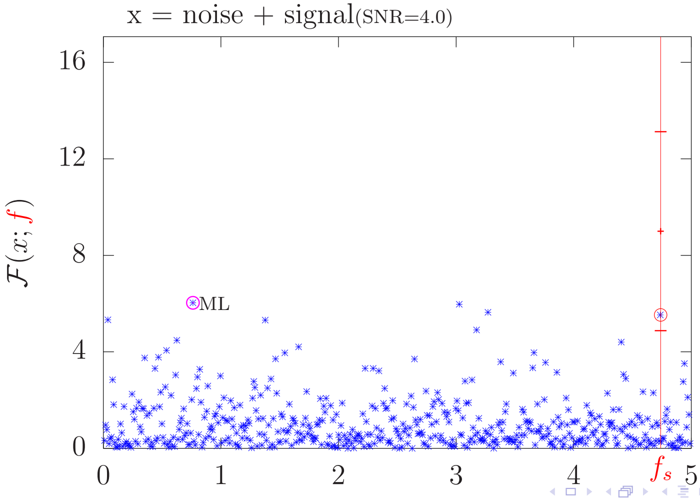
Fourier power: $\mathcal{F}(x; f) = \frac{2}{S_n T} \left|\widetilde{x}(f)\right|^2$, $E[\mathcal{F}] = 1 + \frac{\text{SNR}^2}{2}$

Example 1:



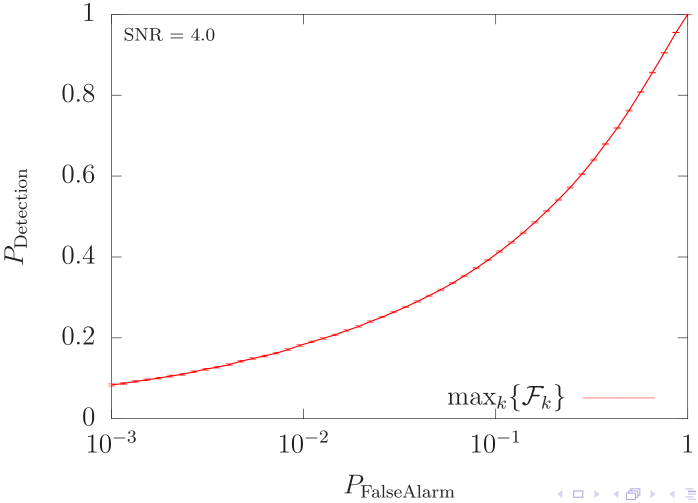
Fourier power: $\mathcal{F}(x; f) = \frac{2}{S_n T} \left|\widetilde{x}(f)\right|^2$, $\mathrm{E}\left[\mathcal{F}\right] = 1 + \frac{\mathrm{SNR}^2}{2}$

Example 2:



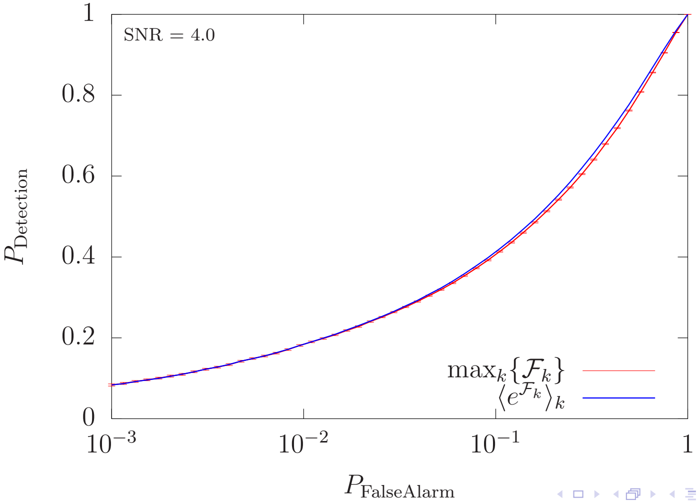
Is $\max_k\{\mathcal{F}_k\}$ (Neyman-Pearson) optimal?

$f_s \in [0, 5]$ Hz



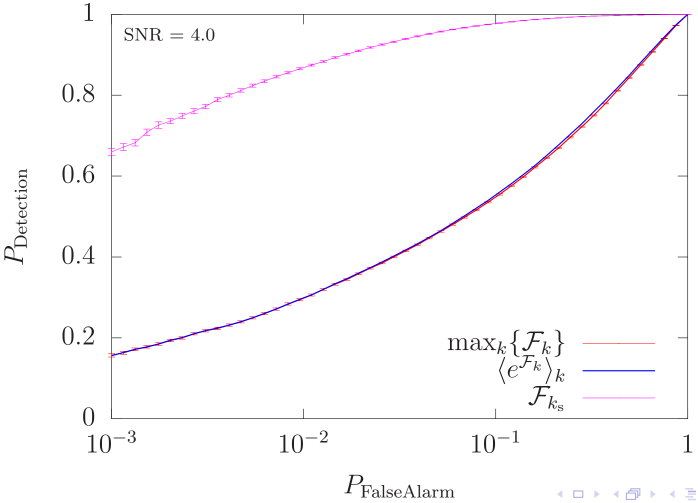
Is $\max_k\{\mathcal{F}_k\}$ (Neyman-Pearson) optimal?

$f_s \in [0, 5]\,\mathrm{Hz}$



Is $\max_k\{\mathcal{F}_k\}$ (Neyman-Pearson) optimal?

signal in DFT bins: $f_s \in \{f_k\}$



Outline

1. Toy CWs: Sinusoids in Gaussian noise
2. Optimal Signal Detection (unconstrained)
3. Cost-Constrained Optimal Signal Detection

Optimal Signal Detection I

Given data $x = \{x_j\}$, how to "optimally" decide between:
$\mathcal{H}_\mathrm{N} \equiv$ no signal: $x_j = n_j$
$\mathcal{H}_\mathrm{S} \equiv$ signal s: $x_j = n_j + s_j(\mathcal{A}, \lambda)$
Two parts to the answer. The better-known part:

Neyman-Pearson lemma for *simple* hypotheses

IF all signal parameters $\{\mathcal{A}_\mathrm{sig}, \lambda_\mathrm{sig}\}$ are *known*
☞ Likelihood ratio $\Lambda(x)$ is the "most powerful" test

$$\Lambda(x; \mathcal{A}_\mathrm{sig}, \lambda_\mathrm{sig}) \equiv \frac{P\left(x|\mathcal{H}_\mathrm{S}, \mathcal{A}_\mathrm{sig}, \lambda_\mathrm{sig}\right)}{P\left(x|\mathcal{H}_\mathrm{N}\right)} \in \mathbb{R}$$

$$\text{accept} \begin{cases} \mathcal{H}_\mathrm{S} & \text{if } \Lambda(x) > \Lambda^*(p_\mathrm{fA}) \\ \mathcal{H}_\mathrm{N} & \text{otherwise} \end{cases}$$

Optimal Signal Detection II

Less well-known: optimal statistic if $\{\mathcal{A}_{\rm sig}, \lambda_{\rm sig}\}$ *unknown*?

Most popular answer: "maximum-likelihood"

$\Lambda_{\rm ML}(x) \equiv \max_{\{\mathcal{A},\lambda\}} \Lambda(x;\mathcal{A},\lambda)$ ☞ intuitive, but *ad-hoc*

Neyman-Pearson lemma for *composite* hypotheses

If signal parameters have probability distribution $P(\mathcal{A},\lambda|\mathcal{H}_{\rm S})$
☞ Bayes factor (aka "marginal likelihood ratio")

$$\mathcal{B}(x) \equiv \frac{P(x|\mathcal{H}_{\rm S})}{P(x|\mathcal{H}_{\rm N})} = \int_{\mathbb{P}} \Lambda(x;\mathcal{A},\lambda)\, P(\mathcal{A},\lambda|\mathcal{H}_{\rm S})\; d\mathcal{A}\, d\lambda = \langle \Lambda \rangle_{\mathbb{P}}$$

is the most powerful test [A. Searle, arXiv:0804.1161v1].

☞ Claim "X is optimal" is usually wrong, unless $X = \mathcal{B}$

Application to Sinusoids

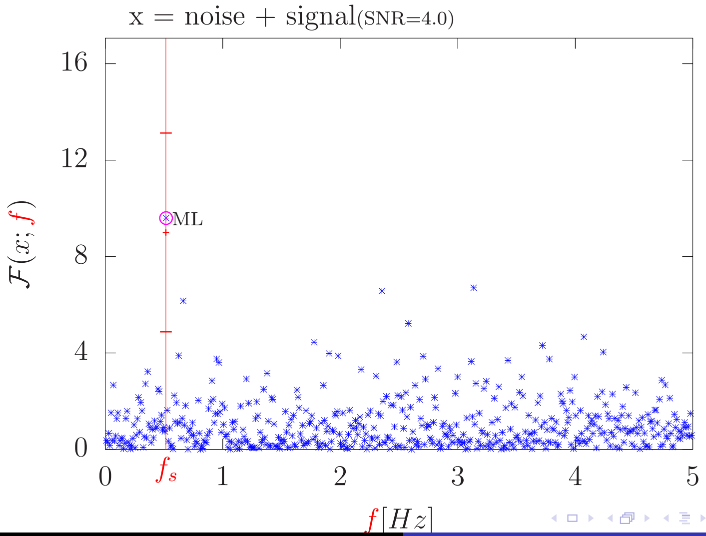
- maximum-likelihood: $\ln \Lambda_{\mathrm{ML}}(x) = \max_f \mathcal{F}(x; f)$
- Bayes-factor (flat prior): $\mathcal{B}(x) = \frac{1}{f_{\max}} \int_0^{f_{\max}} e^{\mathcal{F}(x;f)}\, df = \langle e^{\mathcal{F}} \rangle_f$

In practice: use DFT $\mathcal{F}(x; f_k)$ for $k = 1 \ldots \mathcal{N}$ "templates"

- nice to know the theoretical optimum, but
- not much gain in sensitivity ("intelligent design vs evolution")
- why does $\Lambda_{\mathrm{ML}}(x)$ work so well? (esp. for $p_{\mathrm{fA}} \ll 1$)

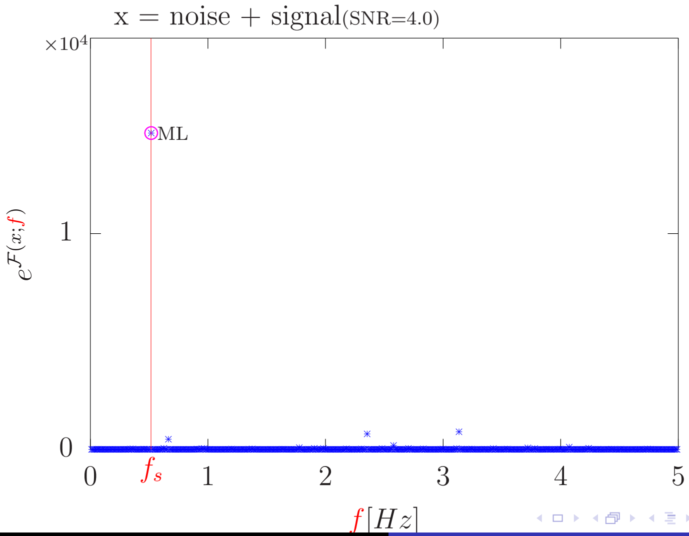
Application to Sinusoids

Example 1:



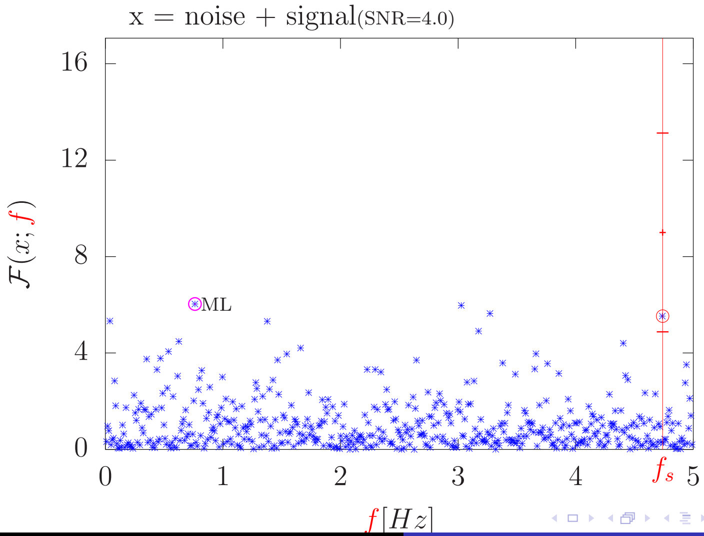
Application to Sinusoids

Example 1:



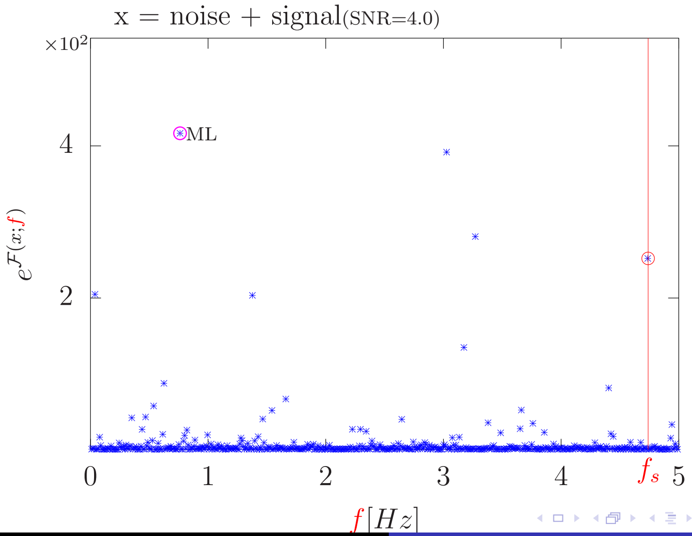
Application to Sinusoids

Example 2:



Application to Sinusoids

Example 2:



Application to Sinusoids

- maximum-likelihood: $\ln \Lambda_{\mathrm{ML}}(x) = \max_f \mathcal{F}(x; f)$
- Bayes-factor (flat prior): $\mathcal{B}(x) = \frac{1}{f_{\max}} \int_0^{f_{\max}} e^{\mathcal{F}(x;f)}\, df = \langle e^{\mathcal{F}} \rangle_f$

In practice: use DFT $\mathcal{F}(x; f_k)$ for $k = 1 \ldots \mathcal{N}$ "templates"

- nice to know the theoretical optimum, but
- not much gain in sensitivity ("intelligent design vs evolution")
- why does $\Lambda_{\mathrm{ML}}(x)$ work so well? (esp. for $p_{\mathrm{fA}} \ll 1$)

  If $\mathcal{F}_{\max} \gtrsim \langle \mathcal{F} \rangle$ ☞ $e^{\mathcal{F}_{\max}} \gg e^{\langle F \rangle} \Longrightarrow \mathcal{B}(x) \approx \frac{1}{\mathcal{N}} e^{\mathcal{F}_{\max}}$
- $\mathcal{B}(x)$ could detect *multiple* sub-threshold signals
- if $\mathbb{P} \Uparrow$, then $\mathrm{E}\left[\max\{\mathcal{F}\}\right]_{\mathrm{noise}} \Uparrow$ ☞ how many "independent" trials?
  while $\mathcal{B}_{\mathrm{noise}} \to \mathrm{E}\left[e^{\mathcal{F}}\right]_{\mathrm{noise}}$ ☞ $\mathcal{B}$ speaks for itself (incl "trials factor")
- posterior $P(f|x, \mathcal{H}_S) \propto e^{\mathcal{F}(x;f)}$ very informative!

Application to Sinusoids

- ☞ maximum-likelihood: $\ln \Lambda_{\mathrm{ML}}(x) = \max_f \mathcal{F}(x; f)$
- ☞ Bayes-factor (flat prior): $\mathcal{B}(x) = \frac{1}{f_{\max}} \int_0^{f_{\max}} e^{\mathcal{F}(x;f)}\, df = \langle e^{\mathcal{F}} \rangle_f$

In practice: use DFT $\mathcal{F}(x; f_k)$ for $k = 1 \ldots \mathcal{N}$ "templates"

- ❑ nice to know the theoretical optimum, but
- ❑ not much gain in sensitivity ("intelligent design vs evolution")
- ❑ why does $\Lambda_{\mathrm{ML}}(x)$ work so well? (esp. for $p_{\mathrm{fA}} \ll 1$)

  If $\mathcal{F}_{\max} \gtrsim \langle \mathcal{F} \rangle$ ☞ $e^{\mathcal{F}_{\max}} \gg e^{\langle F \rangle} \Longrightarrow \mathcal{B}(x) \approx \frac{1}{\mathcal{N}} e^{\mathcal{F}_{\max}}$
- ❑ $\mathcal{B}(x)$ could detect *multiple* sub-threshold signals
- ❑ if $\mathbb{P} \Uparrow$, then $\mathrm{E}\left[\max\{\mathcal{F}\}\right]_{\mathrm{noise}} \Uparrow$ ☞ how many "independent" trials?

  while $\mathcal{B}_{\mathrm{noise}} \to \mathrm{E}\left[e^{\mathcal{F}}\right]_{\mathrm{noise}}$ ☞ $\mathcal{B}$ speaks for itself (incl "trials factor")
- ❑ posterior $P\left(f | x, \mathcal{H}_{\mathrm{S}}\right) \propto e^{\mathcal{F}(x;f)}$ very informative!

Application to Sinusoids

- maximum-likelihood: $\ln \Lambda_{\mathrm{ML}}(x) = \max_f \mathcal{F}(x; f)$
- Bayes-factor (flat prior): $\mathcal{B}(x) = \frac{1}{f_{\max}} \int_0^{f_{\max}} e^{\mathcal{F}(x;f)}\, df = \left\langle e^{\mathcal{F}} \right\rangle_f$

In practice: use DFT $\mathcal{F}(x; f_k)$ for $k = 1 \ldots \mathcal{N}$ "templates"

- nice to know the theoretical optimum, but
- not much gain in sensitivity ("intelligent design vs evolution")
- why does $\Lambda_{\mathrm{ML}}(x)$ work so well? (esp. for $p_{\mathrm{fA}} \ll 1$)

  If $\mathcal{F}_{\max} \gtrsim \langle \mathcal{F} \rangle$ ☞ $e^{\mathcal{F}_{\max}} \gg e^{\langle F \rangle} \Longrightarrow \mathcal{B}(x) \approx \frac{1}{\mathcal{N}} e^{\mathcal{F}_{\max}}$
- $\mathcal{B}(x)$ could detect *multiple* sub-threshold signals
- if $\mathbb{P} \Uparrow$, then $\mathrm{E}\left[\max\{\mathcal{F}\}\right]_{\mathrm{noise}} \Uparrow$ ☞ how many "independent" trials?
  while $\mathcal{B}_{\mathrm{noise}} \to \mathrm{E}\left[e^{\mathcal{F}}\right]_{\mathrm{noise}}$ ☞ $\mathcal{B}$ speaks for itself (incl "trials factor")
- posterior $P\left(f|x, \mathcal{H}_{\mathrm{S}}\right) \propto e^{\mathcal{F}(x;f)}$ very informative!

Application to Sinusoids

- maximum-likelihood: $\ln \Lambda_{\mathrm{ML}}(x) = \max_f \mathcal{F}(x; f)$
- Bayes-factor (flat prior): $\mathcal{B}(x) = \frac{1}{f_{\max}} \int_0^{f_{\max}} e^{\mathcal{F}(x;f)}\, df = \langle e^{\mathcal{F}} \rangle_f$

In practice: use DFT $\mathcal{F}(x; f_k)$ for $k = 1 \ldots \mathcal{N}$ "templates"

- nice to know the theoretical optimum, but
- not much gain in sensitivity ("intelligent design vs evolution")
- why does $\Lambda_{\mathrm{ML}}(x)$ work so well? (esp. for $p_{\mathrm{fA}} \ll 1$)

  If $\mathcal{F}_{\max} \gtrsim \langle \mathcal{F} \rangle$ ☞ $e^{\mathcal{F}_{\max}} \gg e^{\langle F \rangle} \Longrightarrow \mathcal{B}(x) \approx \frac{1}{\mathcal{N}} e^{\mathcal{F}_{\max}}$
- $\mathcal{B}(x)$ could detect *multiple* sub-threshold signals
- if $\mathbb{P} \Uparrow$, then $\mathrm{E}\,[\max\{\mathcal{F}\}]_{\mathrm{noise}} \Uparrow$ ☞ how many "independent" trials?
  while $\mathcal{B}_{\mathrm{noise}} \to \mathrm{E}\left[e^{\mathcal{F}}\right]_{\mathrm{noise}}$ ☞ $\mathcal{B}$ speaks for itself (incl "trials factor")
- posterior $P(f|x, \mathcal{H}_S) \propto e^{\mathcal{F}(x;f)}$ very informative!

Application to Sinusoids

- ☞ maximum-likelihood: $\ln \Lambda_{\mathrm{ML}}(x) = \max_f \mathcal{F}(x; f)$
- ☞ Bayes-factor (flat prior): $\mathcal{B}(x) = \frac{1}{f_{\max}} \int_0^{f_{\max}} e^{\mathcal{F}(x;f)}\, df = \left\langle e^{\mathcal{F}} \right\rangle_f$

In practice: use DFT $\mathcal{F}(x; f_k)$ for $k = 1 \ldots \mathcal{N}$ "templates"

- ❑ nice to know the theoretical optimum, but
- ❑ not much gain in sensitivity ("intelligent design vs evolution")
- ❑ why does $\Lambda_{\mathrm{ML}}(x)$ work so well? (esp. for $p_{\mathrm{fA}} \ll 1$)

  If $\mathcal{F}_{\max} \gtrsim \langle \mathcal{F} \rangle$ ☞ $e^{\mathcal{F}_{\max}} \gg e^{\langle F \rangle} \Longrightarrow \mathcal{B}(x) \approx \frac{1}{\mathcal{N}} e^{\mathcal{F}_{\max}}$
- ❑ $\mathcal{B}(x)$ could detect *multiple* sub-threshold signals
- ❑ if $\mathbb{P} \Uparrow$, then $\mathrm{E}\left[\max\{\mathcal{F}\}\right]_{\mathrm{noise}} \Uparrow$ ☞ how many "independent" trials?
  while $\mathcal{B}_{\mathrm{noise}} \to \mathrm{E}\left[e^{\mathcal{F}}\right]_{\mathrm{noise}}$ ☞ $\mathcal{B}$ speaks for itself (incl "trials factor")
- ❑ posterior $P(f|x, \mathcal{H}_{\mathrm{S}}) \propto e^{\mathcal{F}(x;f)}$ very informative!

Outline

CW signal parameter-space size

Number of "templates" $\sim$ independent likelihood "cells":

$$\mathcal{N} \sim \int_{\mathbb{P}} \sqrt{\det g}\, d\theta \approx \sqrt{\det \bar{g}}\, V_{\mathbb{P}} \quad (\sim \text{"} T\,\Delta f \text{"})$$

where $V_{\mathbb{P}} = \int_{\mathbb{P}} d\theta$ is the coordinate volume, and g is the metric.

All-sky search for isolated NS: $V_{\mathbb{P}} \sim 10^3 \mathrm{Hz} \times 10^{-8} \frac{\mathrm{Hz}}{\mathrm{s}} \times 4\pi$

$$☞ \quad \mathcal{N}(T = 1\mathrm{y}) \sim \mathcal{O}\left(10^{30}\right)$$

- huge $\mathbb{P}$, and signals extremely *sparse* [Ra Inta's poster]
- impossible for covering, MCMC, MultiNest, NOMAD...
- Wanted: Optimal approximation to $\mathcal{B}(x)$ with limited cost

$$\mathcal{B}(x) \approx \frac{1}{\mathcal{N}} \sum_{k=1}^{\mathcal{N}} e^{\mathcal{F}(x;\lambda_k)}$$

Current approach: “semi-coherent” methods

“Coarse-graining”: $\Delta T = 1\mathrm{d} \Longrightarrow \Delta\mathcal{N} \equiv \mathcal{N}(1\mathrm{d}) \sim \mathcal{O}\left(10^{10}\right)$ ✓

Compute $\mathcal{F}(\Delta x; \lambda)$ over N_{seg} data-segments Δx of length $\Delta T = T/N_{\mathrm{seg}}$, then sum across segments:

$$\Sigma(x;\lambda) \equiv \sum_{l=1}^{N_{\mathrm{seg}}} \mathcal{F}(\Delta x_l; \lambda)$$

(“Hough”, “StackSlide”, “PowerFlux”, “Einstein@Home”,...)

- ✓ Reduced resolution due to coarse-graining $\Delta T \ll T$
- ✗ more permissive signal model $\Longrightarrow$ increased false-alarms
- ✗ non-hierarchical: information from first segment not used to reduce parameter space
- ✗ *ad-hoc*, no clear theoretical justification

☞ better methods might exist (but beware the “evolution” clause)

Simple-minded idea: 2-stage FFT

Ad-hoc attempt:

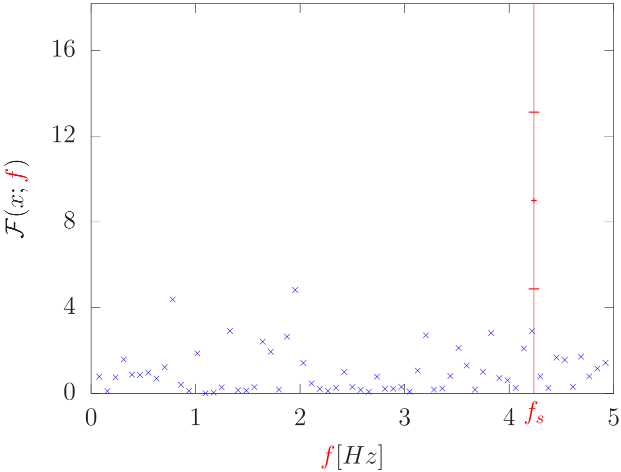
1. Compute coarse FFT $\mathcal{F}(\Delta x; f)$ on short segment ΔT:
(☞ posterior $P\left(f_k | \Delta x, \mathcal{H}_S\right) \propto e^{\mathcal{F}(\Delta x; f_k)}$)
2. pick $c = 1 \ldots \mathcal{N}_{\text{follow}}$ “loudest” $\mathcal{F}(\Delta x, f_{k_c})$
3. “zoom”: compute “fine” $\mathcal{F}(x, f_j)$ in each $f_{k_c} \pm \frac{1}{2\Delta T}$
4. approximate $\mathcal{B}(x) \approx \mathcal{B}_{\mathcal{H}}(x) \propto \left\langle e^{\mathcal{F}(x; f_j)} \right\rangle_{j=1\ldots\mathcal{N}'}$

(Relation to MIT's sparse-FFT?)

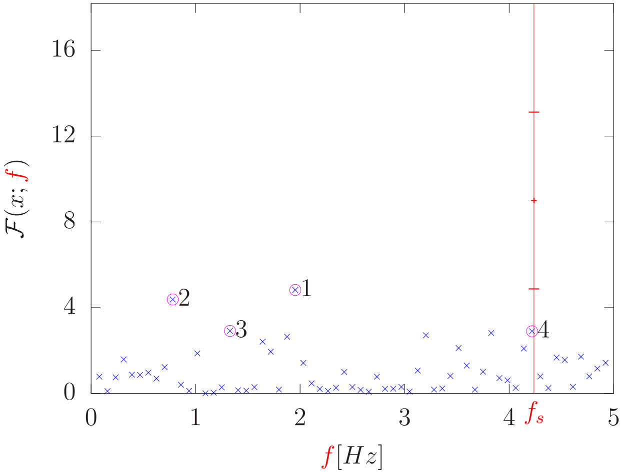
Would need to optimize this at fixed computing-cost ...

- $\mathcal{C}[\mathcal{B}] \sim \mathcal{O}\left(\mathcal{N} \log \mathcal{N}\right)$
- $\mathcal{C}[\mathcal{B}_{\mathcal{H}}] \sim \mathcal{O}\left(\Delta\mathcal{N} \log \Delta\mathcal{N} + \mathcal{N}_{\text{follow}} N_{\text{seg}} \log N_{\text{seg}}\right) \ll \mathcal{C}[\mathcal{B}]$

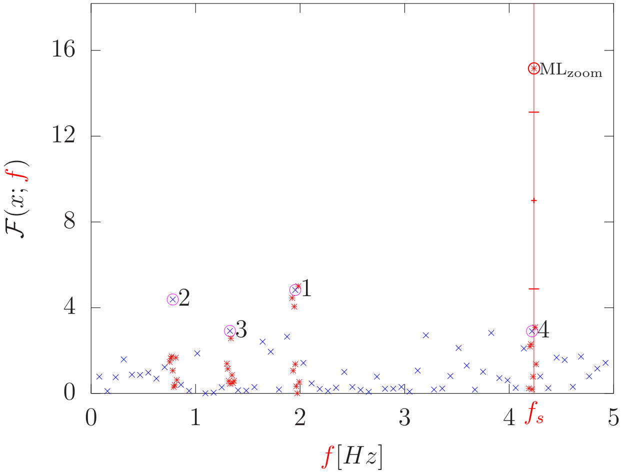
2-stage FFT Illustrated ($\Delta T = T/8$)



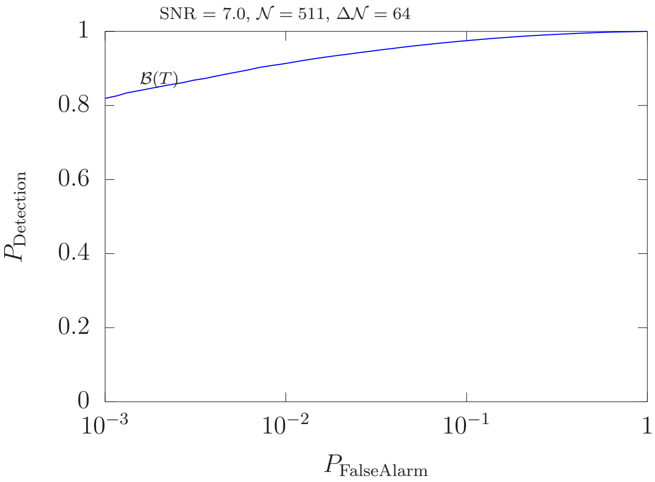
2-stage FFT Illustrated ($\Delta T = T/8$)



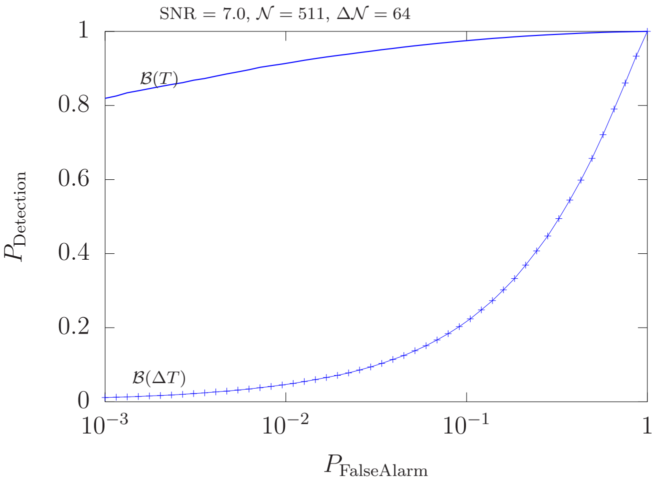
2-stage FFT Illustrated ($\Delta T = T/8$)



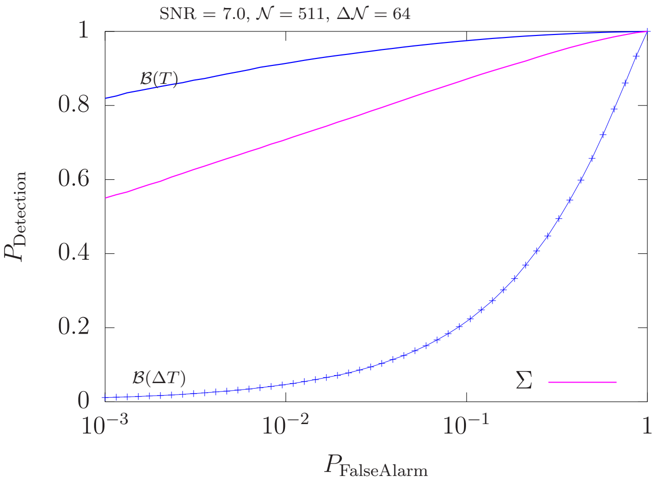
2-stage FFT ROC ($\Delta T = T/8$)



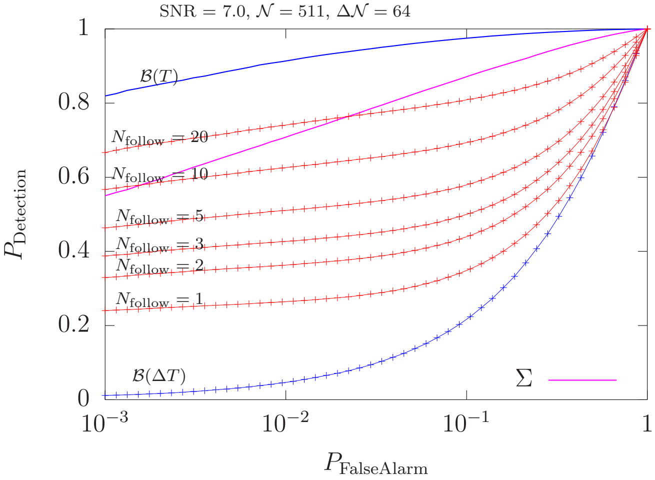
2-stage FFT ROC ($\Delta T = T/8$)



2-stage FFT ROC ($\Delta T = T/8$)



2-stage FFT ROC ($\Delta T = T/8$)



Warning: *not* comparing apples to apples! (different costs)

Conclusions

Current status:

- ✓ Known: Bayes factor $\mathcal{B}(x)$ is Neyman-Pearson optimal
  ☞ marginalize phase-parameters λ instead of maximize
- ❑ Unknown: *optimal approximation* to $\mathcal{B}(x)$ at limited cost
- ❑ Plausible: can we improve over “StackSlide”-type approach by using available information $P(f|\Delta x)$ to better distribute computing power over $\mathbb{P}$?