第1回データ解析勉強会

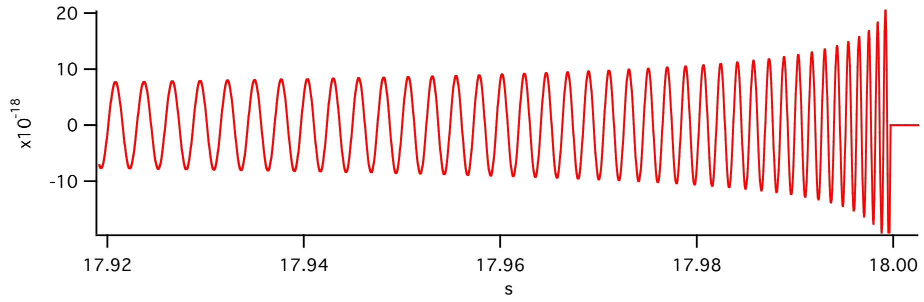
田越秀行

@大阪市大 2011年5月17日

インスパイラル解析の基礎

合体するコンパクト連星

コンパクト連星合体の基本的な値



	$(1M_{solar}, 1M_{solar})$	$(1.4M_{solar}, 1.4M_{solar})$	$(3M_{solar}, 3M_{solar})$
最大周波数	2198Hz	1570Hz	732Hz
100Hzでの軌道半径 (total mass単位)	47M	37M	22M
100Hzから合体までの時間	3.78 秒	2.15秒	0.60秒
100Hzから合体までの重力波のサイクル	605	345	97

観測周波数帯域: 100Hz 〜2kHz

TAMA最終結果(1)

	DT6	DT8	DT9
Observation time [hours]	876.6	1100	486.1
Threshold ζ^*	21.8	13.7	17.7
$N_{bg}^{(i)}$	0.1000	0.1255	0.0555
Detection probability	0.18	0.60	0.69
$(\delta R_i)_{\text{fluct}}$ [yr^{-1}]	+20.6	+2.52	+4.04
	−24.0	−2.82	−3.77
$(\delta R_i)_{\text{model}}$ [yr^{-1}]	+55.4	+4.18	+6.84
	−16.6	−1.53	−2.60
R_i [yr^{-1}]	130^{+59}_{-29}	$30^{+4.9}_{-4.6}$	$60^{+8.0}_{-4.6}$

TABLE II: Summary of the upper limit to the Galactic event rate. The errors for the upper limit are evaluated in section VI in detail.

TAMA最終結果(2)

結合した1つの上限値の誤差

$$\Rightarrow \quad R_{combined} = 17^{+3.02}_{-1.51} \quad [\mathrm{yr}^{-1}]$$

Conservative な上限値をえるために，大きい値をとる.

$$\boxed{R = 20 \quad [\mathrm{yr}^{-1}]}$$

c.f. LIGO S2 : 47 [yr^{-1}]
LIGO-TAMA S2-DT8 : 49 [yr^{-1}]

c.f. 連星パルサーの観測数に基づく推定値 : 8.3×10^{-5} [yr^{-1}]

(Kalogera et al., ApJ 601, L179(2004))

LIGO観測結果

- コンパクト連星合体レート上限値: NS-NS 1.5×10^{-2} yr^{-1} $MWEG^{-1}$
  BH-NS 3.7×10^{-3} yr^{-1} $MWEG^{-1}$
  BH-BH 7.5×10^{-4} yr^{-1} $MWEG^{-1}$

(銀河系相当の銀河１つあたり，１年間あたりでのレート)

LIGO(1) -ガンマ線バースト-

ガンマ線バースト: 1ミリ秒から1000秒という短時間だけ,
ガンマ線が1日1回程度地球に到達する現象.

長いバースト(Long Burst): 継続時間2秒以上

短いバースト(Short Burst): 継続時間2秒以下

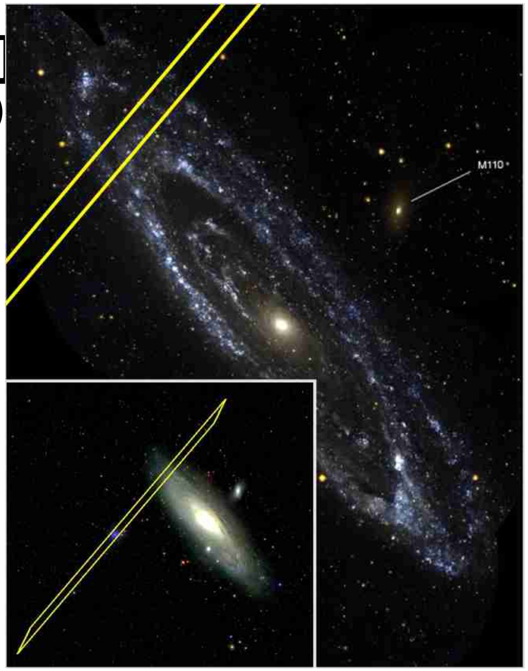
短いバーストの発生源(候補):連星中性子星, あるいは
中性子星ブラックホール連星の合体

GRB070201 短いバースト, 発生方向:M31の方向
(アンドロメダ星雲, 770kpc)
近い!

⇩

しかしながら, LIGOでのデータ解析の結果, この時刻には連星合体重力波は検出できない.

つまり, M31で起こったコンパクト連星合体現象ではない(信頼度99%以上で).



Astrophys. J. 681, 1419 (2008)

実際のデータ解析の基本概念(1)

•検出器からのデータ: $x(t) = s(t) + n(t)$

$s(t)$: 重力波　　$n(t)$: ノイズ

• $x(t)$より何らかの関数F[$x(t)$]をつくる.　F[$x(t)$]: statistic (統計量）

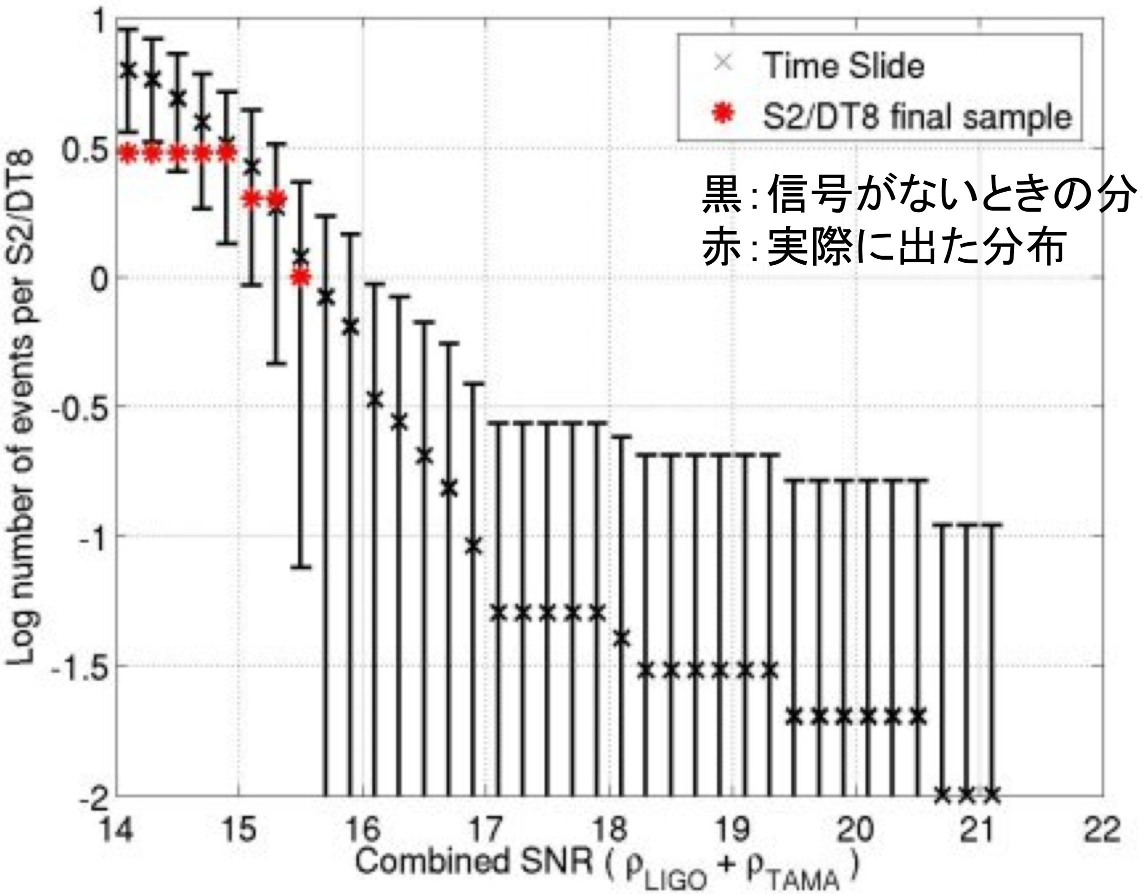
•F[x]は重力波信号の検出に有用である必要がある
($s(t)$ があるときとないときで, Fが違う値を取る)

•F[x]を実際に計算し, F[x]の分布を導出する.

実際のデータ解析の基本概念(2)

•x(t)に信号が含まれていないときのF[x]の分布を何らかの方法で評価する.

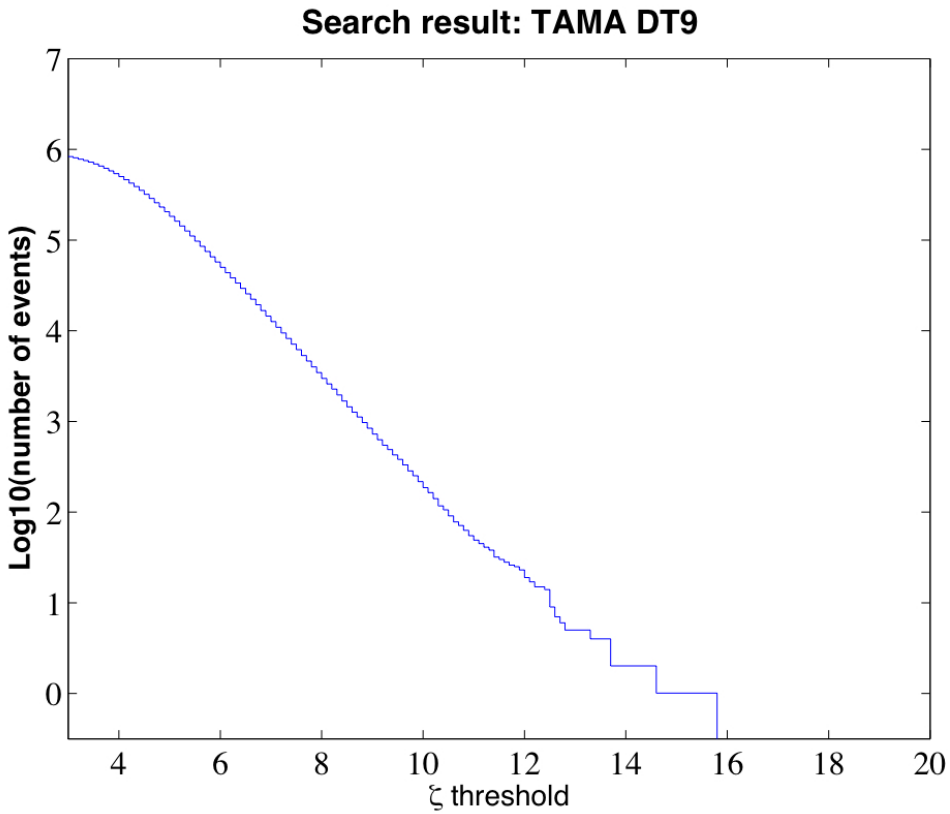
•x(t)に信号がない場合のF[x]の分布と，実際に出たF[x]の分布を比較し，重力波のある，なし，を判定する.



黒：信号がないときの分布
赤：実際に出た分布

実際のデータ解析の基本概念(3)

•x(t)に信号が含まれていないときのF[x]の分布を評価するのは簡単ではない.



F[x]

例: 検出器1台での観測

•仮定:重力波は稀にしか起こらない. 従って, F[x]の値が小さい部分での分布はほぼ全部ノイズによるものである, とする.

•ノイズによる分布から大きくはずれたものがあれば, 本物重力波候補となる.

(この方法では, 検出に関しては弱い主張しかできない.)

いずれにしても,
検出されなかった場合には, 重力波に関する上限値(イベント発生率, 重力波振幅)を導出する.

Matched filtering

- Detector outputs: $x(t) = As(t) + n(t)$

  h(*t*) : known gravitational waveform (template)

  n(*t*) : noise
- Matched filter : $\rho(t_c, M, \eta, \ldots) = 2\int_{-\infty}^{\infty} \frac{\tilde{x}(f)\tilde{s}^*(f)}{S_h(f)} df$

（ $S_n(f)$: 片側パワースペクトラム密度 ）

質量, 合体時刻, …., は事前には分からないので,

これらのパラメータについては最大のρを与えるものを探す

尤度比との関係

ノイズがガウシアンで, テンプレートが規格化されているとすれば

Log likelihood ratio: $\ln \Lambda = \frac{1}{2}\rho^2$

データ解析の基礎(1)

問題設定

観測データを $x_1,\ldots,x_n$ と書く.

このデータ中の重力波信号のある/なしを判定したい.

考え方 : 観測データ中に信号があるときとないときでは，データの従う確率分布は変わるであろう.

$$x \equiv (x_1,\ldots,x_n)^T$$

xは2種類の確率密度 $p_0(x), p_1(x)$ に従う.

$p_0(x)$: データに重力波信号が含まれていないときの分布

$p_1(x)$: データに重力波信号が含まれているときの分布

データ解析の基礎(2)

(a) ２種類の誤り

・**false alarm**: p_0 が正しいときに，p_1 であると主張する

・**false dismissal**: p_1 が正しいときに，p_0 であると主張する

(b) 検定とは，xの取りうる値を2つの領域に分けるルール

xが領域Rに含まれるとき，p_1 が正しいと主張する

それ以外の場合には，p_0 が正しいと主張する

従って，false alarmが起きる確率は $P_0(R) = \int_R p_0(x)dx$

false dismissalが起きる確率は $1 - P_1(R) = 1 - \int_R p_1(x)dx$

最適な領域Rの選択方法は？

データ解析の基礎(3)
- Likelihood ratio test -

False alarm 確率　$P_0(R) = \alpha$: significance level (有意水準)

Correct alarm 確率　$P_1(R)$:　power（検出効率）

あるfalse alarm probabilityの下で，検出効率を最大にする検定は何か？

☆ **Neyman-Peasonの基準**

$$\Lambda(x) \equiv \frac{p_1(x)}{p_0(x)}$$: likelihood ratio

与えられたfalse alarm probability α に対応して，$\Lambda(x)$ に対応する閾値 k を　$P_0[\Lambda(x) \geq k] = \alpha$　から選ぶ

すると，$\Lambda(x) \geq k$ のとき，重力波を検出したと認定するとすれば，最大の検出効率が得られる．

データ解析の基礎(4)

- ノイズがガウスノイズの時

$$x(t) = \begin{cases} n(t) \\ n(t) + s(t;\mu) \end{cases} \quad ;\ s(t;\mu): \text{signal},\quad \mu: \text{parameter}$$

Likelihood ratio

$$\Lambda = \frac{P(x \mid s(\mu))}{P(x \mid 0)}$$

$P(x \mid s(\mu))$:sが存在するとき, xを観測する確率

$P(x \mid 0)$:sが存在しないとき, xを観測する確率

$n(t) = x(t) - s(t;\mu)$ より$P(x \mid s) = P(x - s \mid 0)$ と書ける

よって, $P(x|0)$ を求めればよい.

ノイズ $n(t)$: 平均値=0のガウス過程とする

データ解析の基礎(5)

多変量ガウス分布より,

$$P(x|0) = \frac{\exp\left[-\frac{1}{2}(x,x)\right]}{[(2\pi)^N \det\|C_{n,ij}\|]} \qquad (1)$$

$C_n(\tau)$: correlation function

$C_{n,ij} = C_n[(i-j)\Delta t]$

$$(g,h) \equiv 2\int_{-\infty}^{\infty} df \, \frac{\tilde{g}(f)\tilde{h}^*(f)}{S_n(|f|)} \qquad (2)$$

従って

$$\Lambda = \frac{P(x|s(\mu))}{P(x|0)} = \frac{P(x-s(\mu)|0)}{P(x|0)}$$

$$= \frac{\exp\left[-\frac{1}{2}\big(x-s(\mu), x-s(\mu)\big)\right]}{\exp\left[-\frac{1}{2}(x,x)\right]} = \exp\left[\big(x, s(\mu)\big) - \frac{1}{2}\big(s(\mu), s(\mu)\big)\right] \qquad (3)$$

データ解析の基礎(6)

Maximum likelihood test

未知パラメータがあるとき

　未知パラメータ: μ

$$\max_{\mu} \Lambda(x,\mu) \geq k$$ のとき，検出とする.

対応するfalse alarm 確率は $$P_0[\max_{\mu} \Lambda(x,\mu) \geq k] = \alpha \qquad (1)$$

データ解析の基礎(7) 補足

Maximum likelihood estimation

$\Lambda(x,\mu)$ の最大値を与える $\mu=\hat{\mu}$ を μ の推定値とする.

Λ の指数部分（対数尤度比）

$$\lambda=\ln\Lambda(x,\mu)=\big(x,s(\mu)\big)-\frac{1}{2}\big(s(\mu),s(\mu)\big) \qquad (1)$$

$$0=\left.\frac{\partial L(\mu)}{\partial\mu^i}\right|_{\mu=\hat{\mu}}=\big(x,s_i(\hat{\mu})\big)-\big(s(\hat{\mu}),s_i(\hat{\mu})\big)=\big(x-s(\hat{\mu}),s_i(\hat{\mu})\big) \qquad (2)$$

$$\left(s_i(\mu)=\frac{\partial s(\mu)}{\partial\mu^i}\right)$$

$x(t)=n(t)+s(\tilde{\mu})$とする. $\tilde{\mu}$ は真の値

信号sは強く, Lは $\tilde{\mu}$ に強いピークを持つと仮定する.

$\delta\mu^i=\hat{\mu}^i-\tilde{\mu}^i$ とおき, $\delta\mu^i$ の1次まで評価.

データ解析の基礎(8) 補足

前ページからの続き

$$\delta\mu^j \Gamma_{ij}(\tilde{\mu}) = \left(n, s_i(\tilde{\mu})\right) \qquad \Gamma_{ij}(\tilde{\mu}) = \left(s_i(\tilde{\mu}), s_j(\tilde{\mu})\right)$$:Fisher行列 (1)

$$\delta\mu^i = \left(n, s_j(\tilde{\mu})\right)\left(\Gamma^{-1}\right)^{ij} \equiv \left(n, s^i(\tilde{\mu})\right)$$ (2)

$\left(n, s^i(\tilde{\mu})\right)$ はガウス変数

平均 $\overline{\left(n, s^i(\tilde{\mu})\right)} = 0$ (3)

共分散行列 $\overline{\left(n, s^i(\tilde{\mu})\right)\left(n, s^j(\tilde{\mu})\right)} = \left(s^i(\tilde{\mu}), s^j(\tilde{\mu})\right) = \left(\Gamma^{-1}\right)^{ij}$ (4)

従って $\delta\mu^i$ も同じく平均0,共分散 $\left(\Gamma^{-1}\right)^{ij}$ のガウス分布に従う

$$p(\delta\mu^i) = N e^{-\frac{1}{2}\Gamma_{ij}\delta\mu^i\delta\mu^j}$$ (5)

データ解析の基礎(9)

対数尤度比を振幅で最大化する

$$\lambda \equiv \ln \Lambda$$
$$= (x, s(\mu)) - \frac{1}{2}(s(\mu), s(\mu)) \qquad (1)$$

$$(s(\mu), s(\mu)) \equiv b^2$$
$$\hat{s}(\mu) \equiv s(\mu)/b \qquad (2)$$

$$\lambda = b(x, \hat{s}(\mu)) - \frac{1}{2}b^2$$
$$= -\frac{1}{2}(b - (x, \hat{s}(\mu)))^2 + \frac{1}{2}(x, \hat{s}(\mu))^2 \qquad (3)$$

$$\max_b \lambda = \frac{1}{2}\underline{(x, \hat{s}(\mu))}^2 \qquad (4)$$

：マッチドフィルターの(S/N)^2

インスパイラル波形

$$\tilde{s}(f) = NE(\phi, \theta, \psi, \epsilon) f^{-7/6} \exp i\Psi(f; t_c, \delta_c, m_1, m_2) \quad (1)$$

$$N = \left(\frac{5}{24}\right)^{1/2} \frac{M^{5/6}\eta^{1/2}\pi^{-2/3}}{r} \quad (2)$$

$$E = \left[\left(\frac{1+\cos^2\epsilon}{2}\right)^2 F_+^2 + \cos^2\epsilon F_\times^2\right]^{1/2} \quad (3)$$

$$\Psi = 2\pi f t_c - \delta_c - \frac{\pi}{4} + \frac{3}{128\eta}(\pi M f)^{-5/3} \sum \alpha_k (\pi M f)^{k/3} \quad (4)$$

$$\tilde{s}(f) = A(\tilde{s}_1(f)\cos\delta_c + \tilde{s}_2(f)\sin\delta_c) \quad (5)$$

$$\tilde{s}_1(f) = A f^{-7/6} \exp i\Psi(f; t_c, \delta_c = 0)$$
$$\tilde{s}_2(f) = -i\tilde{s}_1(f) \quad (6)$$
$$A = NE$$

インスパイラル波形，規格化

$$
\begin{aligned}
b^2 &= A^2(s_1 \cos\delta_c + s_2 \sin\delta_c, s_1 \cos\delta_c + s_2 \sin\delta_c) \\
&= N^2 E^2 s^2 \\
s^2 &\equiv (\tilde{s}_1, \tilde{s}_1) = (\tilde{s}_2, \tilde{s}_2)
\end{aligned}
$$

(1)

(2)

$$
\begin{aligned}
\hat{s}(f) &\equiv \tilde{s}(f)/b \\
&= \hat{s}_1(f) \cos\delta_c + \hat{s}_2(f) \sin\delta_c
\end{aligned}
$$

(3)

(4)

$$
\hat{s}_{1,2}(f) = \tilde{s}_{1,2}(f)/s = s^{-1} f^{-7/6} \exp i\Psi(f; t_c, \delta_c = 0)
$$

(5)

$$
(\hat{s}_1, \hat{s}_1) = (\hat{s}_2, \hat{s}_2) = 1
$$

$\hat{s}(f)$ はアンテナパターン関数，軌道傾斜角，距離によらない

振幅と位相で最大化

$$\begin{aligned}
\max_b \lambda &= \frac{1}{2}(x, \hat{s}(\mu))^2 \\
&= \frac{1}{2}(x, \hat{s})^2 \\
&= \frac{1}{2}(x, \hat{s}_1(f)\cos\delta_c + \hat{s}_2(f)\sin\delta_c)^2 \\
&= \frac{1}{2}[(x, \hat{s}_1)\cos\delta_c + (x, \hat{s}_2)\sin\delta_c]^2 \qquad (1)
\end{aligned}$$

$$2\max_{b,\delta_c} = (x, \hat{s}_1)^2 + (x, \hat{s}_2)^2 \equiv \rho^2 \qquad (2)$$

：マッチドフィルターの(S/N)^2

マッチドフィルター内積

テンプレートは次のようにt_cをくくりだして書ける

$$\begin{aligned}
\hat{s}_1 &= \frac{1}{s} f^{-7/6} e^{i\Psi(f;t_c;\delta_c=0)} \qquad (1) \\
&= \frac{1}{s} f^{-7/6} e^{2\pi i f t_c} e^{i\Psi(f;t_c=0;\delta_c=0)} \\
&= \hat{s}_1(f;t_c=0) e^{2\pi i f t_c} \qquad (2)
\end{aligned}$$

従って，マッチドフィルターの内積は

$$(x, \hat{s}_1) = 2 \int_{-\infty}^{\infty} \frac{x^*(f)\hat{s}_1(f;t_c=0)}{S_n(|f|)} e^{2\pi i f t_c} df \qquad (3)$$

逆フーリエ変換の形に書ける. (各t_cについてFFTで一度に計算可能)

Matched filter の性質(最大S/Nフィルター)

検出器出力: $x(t) = n(t) + s(t)$

filter: $q(t)$

フィルター出力

$$c(t) = \int dt' x(t') q(t'+t) dt' = \int_{-\infty}^{\infty} df \tilde{x}(f) \tilde{q}^*(f) e^{2\pi i f t} \quad (1)$$

(2)

$$\left(\frac{S}{N}\right)^2 = \frac{\langle c(t)\rangle^2}{\langle (c(t) - \langle c(t)\rangle)^2\rangle} = \frac{\left(\int df \tilde{s}(f) \tilde{q}^*(f) e^{2\pi i f t}\right)^2}{\int df S_n(f) |\tilde{q}(f)|^2} = \frac{\left(\tilde{s}(f)/S_n(f), \tilde{q}(f) e^{-2\pi i f t}\right)^2}{(\tilde{q}(f) e^{-2\pi i f t}, \tilde{q}(f) e^{-2\pi i f t})}$$

$$(a, b) \equiv \int S_n(f) a(f) b^*(f) df$$

(3)

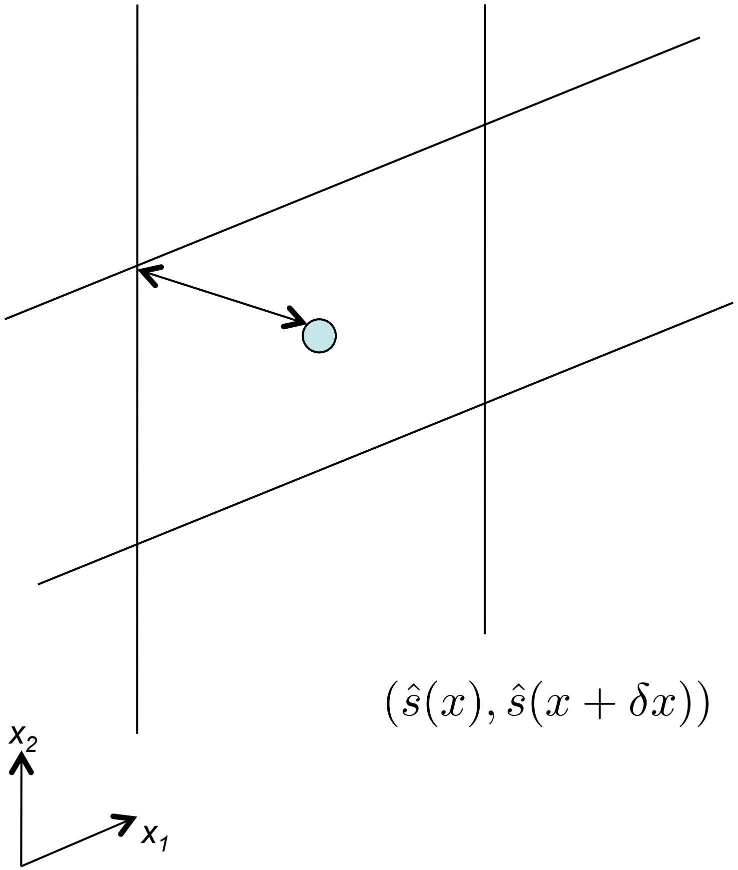
S/Nを最大にするのは, $\tilde{q} = \dfrac{\tilde{s}(f)}{S_n(f)} e^{2\pi i f t}$ (4) のとき. そのとき,

$$c(t) = \int_{-\infty}^{\infty} \frac{\tilde{x}(f) \tilde{s}^*(f)}{S_n(f)} \quad (5)$$

$$\left(\frac{S}{N}\right)^2_{\max} = \int_{-\infty}^{\infty} \frac{|\tilde{s}(f)|^2}{S_n(f)} \quad (6)$$

(両側スペクトラムなのでfactor 2違う)

パラメータ空間探査



massパラメータ空間にグリッドを張る
グリッド間隔は，グリッドからずれた点でのS/Nの最大減少率から決める
(例えば最大3%の減少にする)

パラメータ{x}={x_1, x_2,…}

$$
\begin{aligned}
(\hat{s}(x), \hat{s}(x+\delta x)) &= (\hat{s}(x), \hat{s}(x) + \partial_\alpha \hat{s}(x)\,\delta x^\alpha + \frac{1}{2}\partial_\alpha\partial_\beta \hat{s}(x)\,\delta x^\alpha \delta x^\beta) \quad (1) \\
&= 1 + (\hat{s}, \partial_\alpha \hat{s})\delta x^\alpha + \frac{1}{2}(\hat{s}, \partial_\alpha\partial_\beta \hat{s})\delta x^\alpha \delta x^\beta \\
&= 1 - \frac{1}{2}(\partial_\alpha \hat{s}, \partial_\beta \hat{s})\delta x^\alpha \delta x^\beta \equiv MM \quad \text{(Minimum Match)} \\
g_{\alpha\beta} &\equiv (\partial_\alpha \hat{s}, \partial_\beta \hat{s}) \quad \text{(metric)}
\end{aligned}
$$

テンプレート数

１台検出器，質量パラメータに関するテンプレート数と
必要な計算機パワー
(古いLCGT感度データに基づく)

minimal mass ($M_\odot$)	template number	CPU power (flops)	CPU power including χ^2 (flops)
0.2	2.6×10^7	5.4×10^{12}	4.3×10^{13}
0.4	1.5×10^6	2.8×10^{11}	2.2×10^{12}
0.6	1.4×10^6	2.6×10^{11}	2.1×10^{12}
0.8	6.4×10^5	1.1×10^{11}	9.4×10^{11}
1.0	3.7×10^5	6.7×10^{10}	5.3×10^{11}
0.2	8.4×10^6	1.5×10^{12}	1.2×10^{13}
1.0	9.7×10^4	1.5×10^{10}	1.2×10^{11}

表 1: 探査する最小質量 (1 つの星) とテンプレート数，データの長さと同じ時間で解析を終了させるための演算速度．最初の 5 行は $f_s = 10\mathrm{Hz}$，最後の 2 行は $f_s = 20\mathrm{Hz}$ としたもの．

おおざっぱなスケーリング

$$N_{temp} \sim (1 - MM)^{-1} m_{\min}^{-8/3} f_0^{-8/3}, \qquad P_{comp} \sim (1 - MM)^{-1} m_{\min}^{-8/3} f_0^{-5/3},$$

ガウス性

•パワースペクトラムを見て，ラインノイズ，バイオリンモードなどがない周波数帯を選ぶ.
•各ファイル毎に元データをフーリエ変換，伝達関数を作用させる

$$v(t) \to \tilde{v}(f_i) \to \tilde{x}(f_i)$$

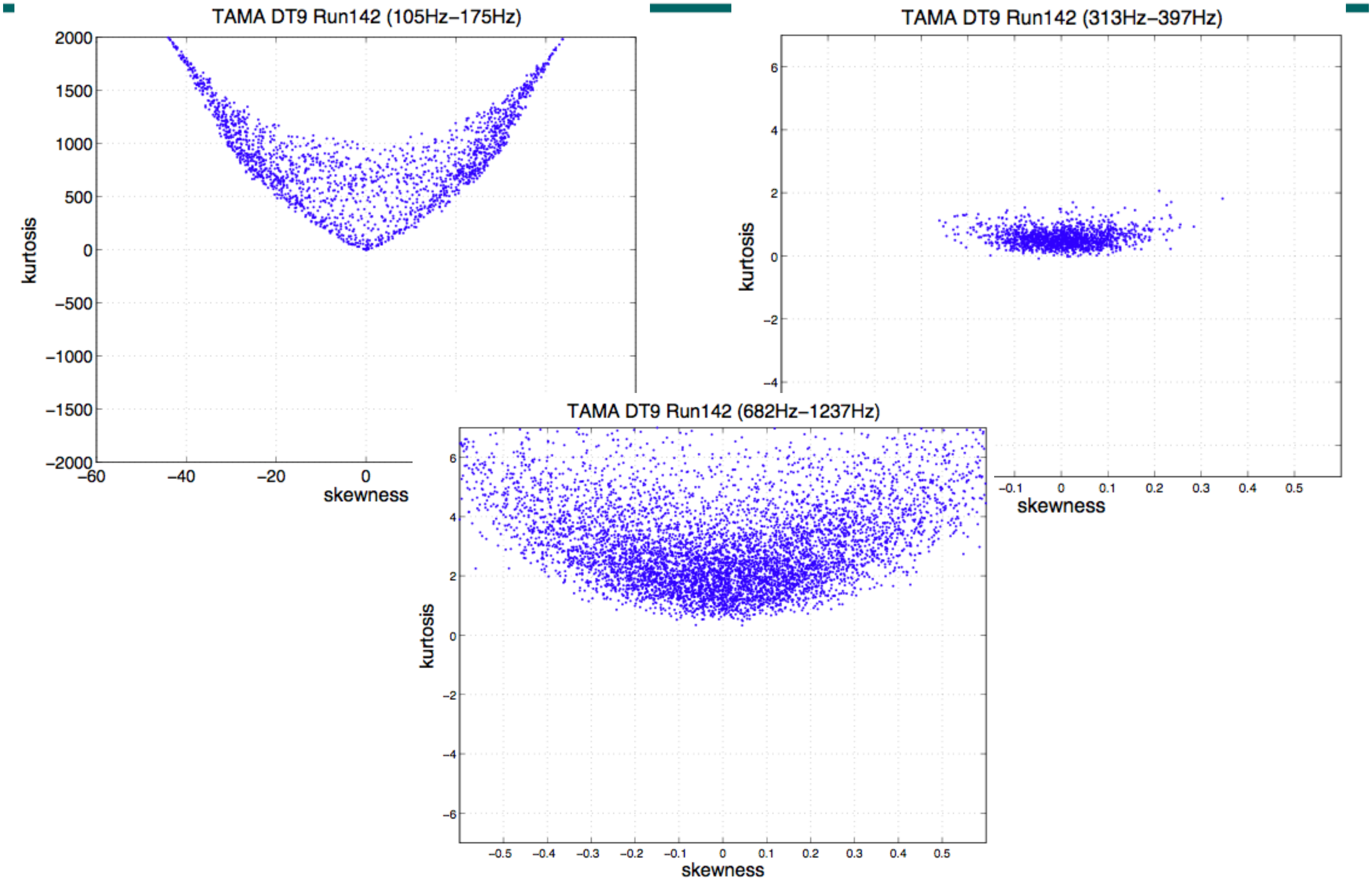
各周波数点f_i毎に$\tilde{x}(f_i)$の実部，虚部の統計性をみる.

•歪度(skewness) b_1，尖度(kurtosis) b_2 を計算する

$$\begin{array}{lrcl} \text{歪度} & b_1 & = & m_3/(m_2)^{3/2} \\ \text{尖度} & b_2 & = & m_4/(m_2)^2 - 3 \\ \text{m次モーメント} & m_r & \equiv & \sum_i (\underbrace{X_i}_{\text{データ}} - \underbrace{\mu}_{\text{平均}})^r / \underbrace{N}_{\text{データ数}} \end{array}$$

データの分布がガウス分布に従うとき歪度，尖度はほぼゼロになる

ガウス性評価 part I (2) TAMA 142



Chi square (1)

非ガウスノイズの取り扱いのために χ^2 を導入する

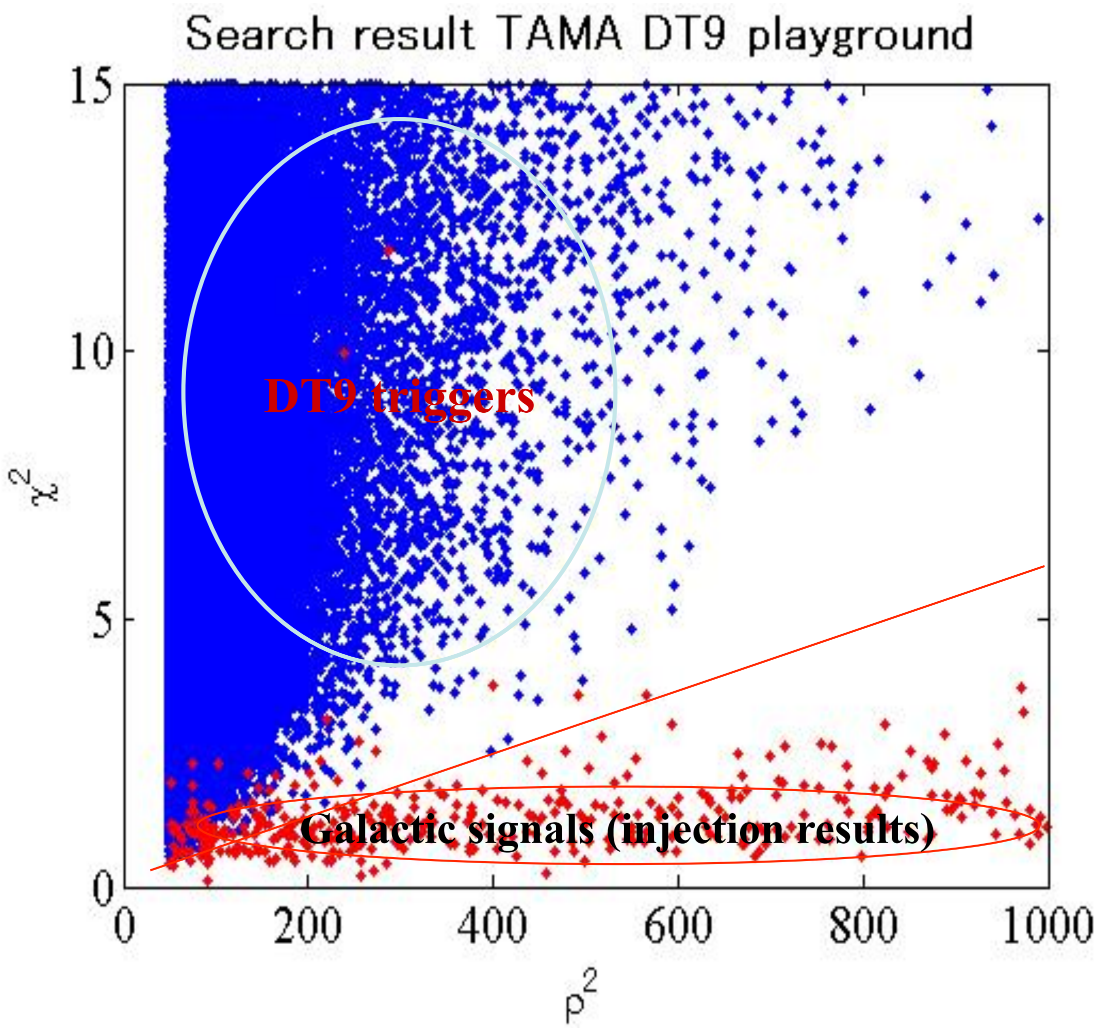
$$
\begin{array}{cccccccc}
 & \rho_1 & & \rho_2 & & \rho_3 & & \rho_4 & & \rho_5 & & \cdots & \\
f_{\min} & & f_1 & & f_2 & & f_3 & & f_4 & & f_5 & \cdots & f_{\max}
\end{array}
$$

$$\chi^2 \equiv \sum \frac{1}{\sigma_i{}^2}(\rho_i - \overline{\rho}_i)^2$$

$$\sigma_i{}^2 \equiv \left\langle (\rho_i - \overline{\rho}_i)^2 \right\rangle, \quad \overline{\rho}_i = \left\langle \rho_i \right\rangle$$:理論波形より分かっている

カイ2乗が小さい:本物らしい
カイ2乗が大きい:本物らしくない

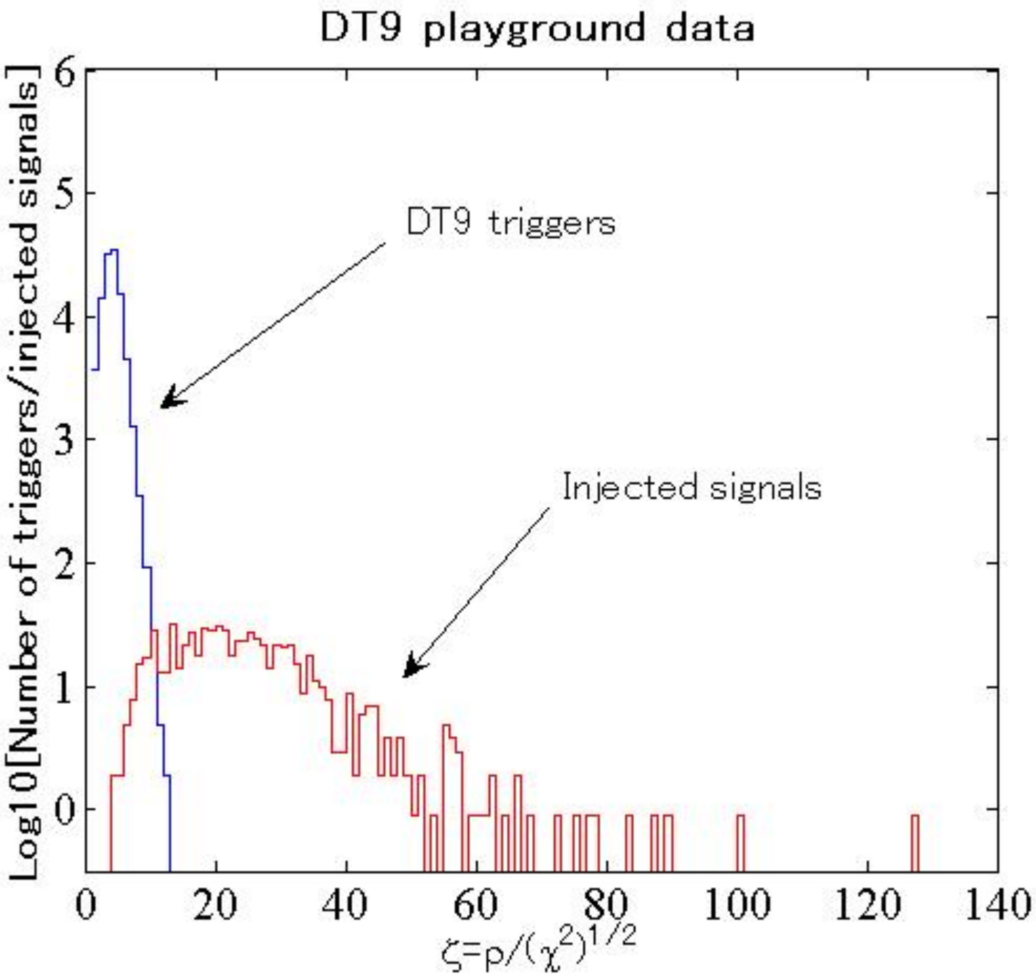
Chi square (2)



Search result TAMA DT9 playground
15
10
5
0
0
200
400
600
800
1000
χ^2
ρ^2
DT9 triggers
Galactic signals (injection results)

Chi square (3)

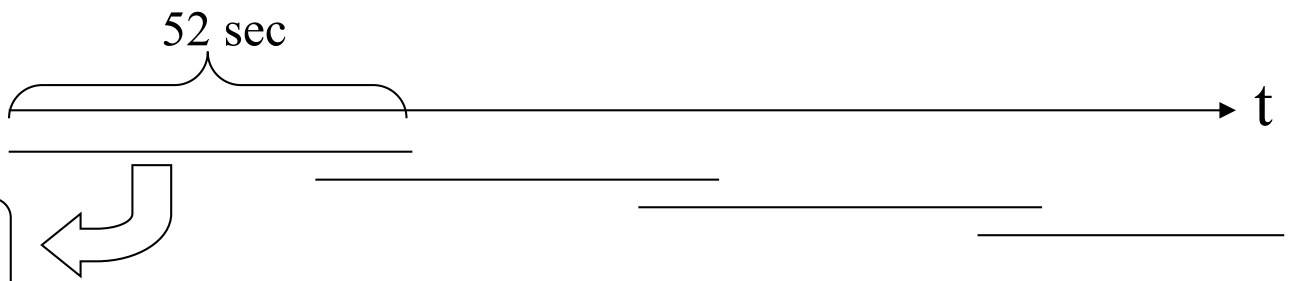
- ζ statistic -



We use $\boxed{\rho/\sqrt{\chi^2}\ (\equiv \zeta)}$ as the statistic to discriminate fake events from true signals. We set a threshold of ζ as $\zeta > \zeta^*$ where ζ^* is determined by the false alarm rate.
The chi square cut is automatically introduced by these procedures.

This statistic can accommodate large χ^2 signals which could occur due to mismatch between signals and templates.

解析手順



データ読み込み

データのFFT
伝達関数の適用
（変位データへ変換）

近傍データより
パワースペクトラム密度
$S_n(f)$

$$\rho(t_c, m_1, m_2, \ldots) = 2\int \frac{\tilde{s}(f)\tilde{h}^*(f)}{S_n(f)} df$$

$$\max_{t_c} \rho(t_c, M, \eta)$$

$\Delta t_c = 25\text{ms}$

$(if \quad \rho > 7)$

$$\max_{t_c, M, \eta} \rho(t_c, M, \eta), \quad \chi^2$$

新しい展開

連星合体重力波(CBC)探査の新しい展開

•スピンを取り入れる

•Inspiral-Merger-Ringdown波形モデルを使った探査

•パラメータ空間の探査方法（マルコフ連鎖モンテカルロ法，．．．）

•複数検出器での探査
コインシデンス解析 から Maximum likelihood detection strategy へ

終わり

データ解析勉強会の今後

- 連続波探査
- バースト探査
- 背景重力波探査
- Frameライブラリ
- キャリブレーションの方法，伝達関数
- LAL等のLIGO関係のライブラリ
- 新しい解析方法の話
- 統計の話（解析結果の解釈，上限値，…）
- …
- …