

# LCGT-E@Hに必要な計算機資源について

伊藤 洋介

平成 21 年 7 月 17 日

## 目次

|          |  |          |
|----------|--|----------|
| <b>1</b> | <b>この文章の目的</b>                               | <b>2</b> |
| <b>2</b> | <b>Isolated pulsars search</b>               | <b>2</b> |
| 2.1      | Sensitivity . . . . .                        | 2        |
| 2.2      | All-sky wide-frequency-band search . . . . . | 3        |
| 2.2.1    | Coherent search . . . . .                    | 3        |
| 2.2.2    | Incoherent search . . . . .                  | 4        |
| <b>3</b> | <b>Einstein@Home についてのメモ</b>                 | <b>5</b> |
| 3.1      | E@H current configuration . . . . .          | 5        |
| 3.2      | E@H データ量 . . . . .                           | 6        |
| 3.3      | 計算機速度 . . . . .                              | 7        |
| 3.4      | E@H の仕事 . . . . .                            | 7        |
| 3.4.1    | Code preparation . . . . .                   | 7        |
| 3.4.2    | SFT . . . . .                                | 7        |
| 3.4.3    | Data preparation . . . . .                   | 7        |
| 3.4.4    | Coherent search . . . . .                    | 8        |
| 3.4.5    | Results validator . . . . .                  | 8        |
| 3.4.6    | Post-processing . . . . .                    | 8        |
| 3.5      | LSC computing cluster . . . . .              | 9        |
| 3.6      | 電気代 . . . . .                                | 10       |
| 3.7      | E@H サーバーの仕事 . . . . .                        | 10       |
| 3.8      | E@H server specification . . . . .           | 10       |

# 1 この文章の目的

LCGT データを使って連続重力波探索を Einstein@Home 上でおこなう時に必要な計算機資源について考える。

## 2 Isolated pulsars search

探索パラメータはパルサーの位置 (2 つ:  $\alpha, \delta$ ) とスピンドウンパラメータ (n 個:  $\dot{f}, \ddot{f}, \dots$ )、重力波の周波数 ( $f$ )。

### 2.1 Sensitivity

Triaxial なパルサーからの重力波の振幅 [1]:

$$h_0 = \frac{16\pi^2 G \epsilon I_{zz} f^2}{c^4 r} \simeq 7.7 \times 10^{-26} \left( \frac{\epsilon}{10^{-6}} \right) \left( \frac{r}{10\text{kpc}} \right)^{-1} \left( \frac{f}{1\text{kHz}} \right)^2 \quad (1)$$

$$\epsilon \equiv \frac{I_{xx} - I_{yy}}{I_{zz}} \quad (2)$$

$f$  は重力波周波数、 $I_{xx,yy,zz} \sim 10^{45} \text{gcm}^2$  は moment of inertia.  $\epsilon \lesssim 10^{-7}$  for neutron stars and  $\epsilon \lesssim 10^{-6}$  for possible strange stars.

Cutler [3] によるとパルサーの birth rate が 1 個/30 年程度で銀河中に一樣に分布し、生まれたときには大きな自転角運動量を持ち、かつすべてのパルサーが LIGO の周波数帯域では重力波によってスピンドウンするなら地球で期待されるパルサーからの重力波振幅は  $4 \times 10^{-24}$  程度。(Populatio に関する議論から、地球から 40pc ぐらいの距離に現在電磁波では見つからないパルサーがいて LIGO 周波数帯域内で重力波をだしていてもおかしくはないだろうということ。)

False alarm rate 1%, False dismissal rate 10 %, F statistic, targeted coherent search の sensitivity [1]:

$$\langle h_0 \rangle = 11.4 \sqrt{\frac{S_n(f)}{T_{\text{coh}}}} \quad (3)$$

$$\simeq 2 \times 10^{-26} \left( \frac{\sqrt{S_n(f)}}{10^{-23}} \right) \left( \frac{T_{\text{coh}}}{1\text{yr}} \right)^{-1/2} \quad (4)$$

Template の少ない Targeted search ならこれぐらいはできるということ。

Incoherent search(例えば Hough transform[8]) では False alarm rate 1%, False dismissal rate 10 %, F statistic, targeted search の sensitivity は

$$\langle h_0 \rangle = \frac{8.92}{N^{1/4}} \sqrt{\frac{S_n(f)}{T_{\text{coh}}}} \quad (5)$$

$$\simeq 7 \times 10^{-26} \left(\frac{N}{365}\right)^{-1/4} \left(\frac{\sqrt{S_n(f)}}{10^{-23}}\right) \left(\frac{T_{\text{coh}}}{1\text{day}}\right)^{-1/2} \quad (6)$$

ここで  $N$  は coherent integration をするセグメントの数 (stack slide なら stack するセグメントの数)。

実際にはテンプレートの数が多すぎて 1 年間もの coherent all-sky wide-frequency-range spin-down-parameters search はできない (Sec. 2.2.1 参照)。Blind search では Incoherent search (短時間の coherent search の結果を incoherent に足し合わせる search。LSC では Hough, Stack-Slide, Power-Flux など) をすることになる。

Blind search では template の数に合わせて False dismissal rate を決める必要がある。たとえば Template の数が  $10^{12}$  個もあると False dismissal rate 1% では Follow-up すべき Candidate event の数が  $10^{10}$  個にもなってしまって、計算機能力にもよるが現実的では無い。False alarm rate  $\alpha$ , False dismissal rate  $\beta$  のとき [2]

$$\langle h_0 \rangle = \frac{5.34\mathcal{S}^{1/2}}{N^{1/4}} \sqrt{\frac{S_n(f)}{T_{\text{coh}}}} \quad (7)$$

$$\mathcal{S} = \text{erfc}^{-1}(2\alpha) + \text{erfc}^{-1}(2\beta). \quad (8)$$

E@H Early S5[4] では 30 時間 coherent 積分、840 時間 incoherent 積分 ( $N=28$ ) で 120 Hz 付近で  $2 \times 10^{-24}$  (90% detection probability)。

## 2.2 All-sky wide-frequency-band search

### 2.2.1 Coherent search

パルサーの位置は、Doppler phase modulation

$$f(t) = f_0 \left(1 + \frac{\vec{v} \cdot \vec{n}}{c}\right) \quad (9)$$

における周波数シフト  $\Delta f = f_0 \frac{\vec{v} \cdot \vec{n}}{c}$  に影響を与える。逆にシフトが周波数分解能以上 ( $1/t_{\text{obs}}$ ) であるような空の二つの方向は分解でき、これでパル

サーの位置を決定できる。  $\vec{v} \cdot \vec{n} \sim v \sin(\omega t) \sim v \omega t$  より (Eq. (7.151) of [9])

$$\begin{aligned} \Delta\theta &< \frac{1}{f_0(v_{orb}/c)\omega_{orb}t_{orb}^2} \\ &\simeq 0.1 \text{arcseconds} \left( \frac{10^7 \text{s}}{t_{obs}} \right)^2 \left( \frac{1 \text{kHz}}{f} \right) \end{aligned} \quad (10)$$

逆に  $\Delta\theta$  以上異なる2つの方向からの重力波は Doppler modulation によって位相が大きくずれる。したがって、位置についてのテンプレート数は、

$$\begin{aligned} \mathcal{N}_{\text{sky}} &= 4\pi f_{\text{max}}^2 \left( \frac{v_{orb}}{c} \right)^2 \omega_{orb}^2 t_{orb}^4 \\ &\sim 10^{14} \left( \frac{f_{\text{max}}}{1.5 \text{kHz}} \right)^2 \left( \frac{t_{obs}}{10^7 \text{s}} \right)^4 \end{aligned} \quad (11)$$

$2f_{\text{max}}t_{obs} = 3 \times 10^{10} (f_{\text{max}}/1.5 \text{kHz})(t_{obs}/10^7 \text{s})$  点のデータ (240Gbyte) を FFT するのに TFLOPS 計算機で 1 秒。  $10^{14}$  点の sky grid を探索すると  $10^7 \text{yr}$ 。これにさらにスピンドアウンパラメータの探索で数倍の時間がかかる。

したがって、All-sky coherent search で長時間積分は不可能。この結論は metric を使った計算でも変わらない。Optimal な方法はおそらく以下のようなになる。例えば1年間のデータを1日程度のデータセグメントに分けて、それぞれのセグメントで all-sky coherent search をおこない、これを incoherent に1年間積分する。Incoherent search で得た候補天体について探索パラメータ領域を制限して Coherent な follow-up をおこなう。

## 2.2.2 Incoherent search

E@H のように coherent search を参加者のホスト計算機にしてもらい Incoherent search をサーバー側でおこなう場合、計算量は比較的少なくてすむ。Incoherent search method として stack slide をとる。計算量は [6]

$$3f_{\text{max}}T_{\text{coh}}(N-1)N_p(T_{\text{coh}}, \mu_{\text{max}}, N) \quad (12)$$

ここで、 $N$  は stack するセグメントの数、 $T_{\text{coh}}$  はセグメントの長さ、 $\mu_{\text{max}}$  は maximal mismatch。  $\text{SNR}^2$  は mismatch 分だけ減る：

$$\eta = (1 - \mu_{\text{max}}) \frac{4}{25} \frac{h_0^2 T}{S_n(f)} \quad (13)$$

(ただし、template grid が cubic lattice のとき平均の loss は  $1 - \mu_{\text{max}}$  ではなく  $1 - \mu_{\text{max}}/3$ 。)

1年ちかい incoherent 積分をしたときのテンプレート数を数えた論文は見当たらない。1年間より十分短い場合は metric を使った計算がある [6]。大雑把に言って [2]

$$\delta f = \frac{1}{T_{\text{coh}}} \quad (14)$$

$$\delta \dot{f} = \frac{\delta f}{t_{\text{obs}}} = \frac{1}{T_{\text{coh}} t_{\text{obs}}} \quad (15)$$

$$\delta \theta = \frac{c \delta f}{v f} \quad (16)$$

とすると

$$N_p \sim \frac{\Delta f \Delta \dot{f}}{\delta f \delta \dot{f}} \frac{4\pi}{(\delta \theta)^2} \quad (17)$$

なので、coherent search に比べて  $(T_{\text{coh}}/t_{\text{obs}})^4 = 1/N^4$  ( $N$  は stack の数) だけ少ない template 数になる。 $N \sim 100$  として coherent search だと  $10^7 \times N_{\text{spin-down,YR}}$  だけかかったものが、incoherent search を使うと  $0.01 \times N_{\text{spin-down,YR}}$  だけかかることになる。

LSC S2 Hough[2] では minimum spin down age を  $\sim 10^4$  years とし、30分長 SFT を約 1000個使用したときテンプレートの数は  $\sim 10^9$  個 (200-400Hz)。計算量は

$$8 \times 10^{17} \left( \frac{f_{\text{max}}}{400\text{Hz}} \right) \left( \frac{t_{\text{obs}}}{2 \times 10^6\text{s}} \right) \left( \frac{N_p}{10^9} \right) \quad (18)$$

10GFLOPS の計算機 100台 (total 1TFLOPS) で1週間程度 (これは overestimate で低周波数ではテンプレートは少ないため、実際はもっと速く終わった)。

## 3 Einstein@Home についてのメモ

### 3.1 E@H current configuration

Early S5 では total 840時間のデータを使っている。840時間を28個の30時間セグメントにわけ、各セグメントをE@H参加者がF statistic(JKS[7])を使って all-sky wide-frequency-band spindown parameter coherent search をおこなう。結果は University of Wisconsin-Milwaukee にあるサーバー (と AEI などにあるミラーサーバー) で集める。結果を incoherent に足し合わせる。

### 3.2 E@H データ量

倍精度：8byte。周波数  $f_{max}$  までの探索には  $2f_{max}$ Hz までのデータが必要で観測時間を  $t_{obs}$  とするとデータ量は  $8\text{byte} \times 2f_{max}t_{obs}$

$$757\text{Gbyte} \left( \frac{f_{max}}{1.5\text{kHz}} \right) \left( \frac{t_{obs}}{3.1536 \times 10^7\text{s}} \right) \quad (19)$$

一つのファイルサーバーは一回の science run について 3Tbyte(time domain/frequency domain/結果) 程度は必要。

E@H のように  $T_{\text{SFT}} = 30$  分の Short Fourier Transform (SFT) を作成するなら 1SFT ファイルあたり

$$43.2\text{Mbyte} \left( \frac{f_{max}}{1.5\text{kHz}} \right) \left( \frac{T_{\text{SFT}}}{1800\text{s}} \right) \quad (20)$$

現在の E@H は周波数帯域を区切った SFT ファイル  $M$  個をユーザーに送信している。ホストは受け取った帯域内で位置・スピンドウンパラメータ template について Coherent search をおこなう。このとき、各ホストがおこなう計算時間がおなじぐらいになるように周波数帯域幅を決めている。

SFT ファイル  $M = T_{\text{coh}}/T_{\text{SFT}}$ 、周波数帯域幅  $\Delta f$  のデータ量  $2 \times 8\text{byte} \times M\Delta f T_{\text{SFT}}$

$$1\text{kbyte} \left( \frac{M}{60} \right) \left( \frac{T_{\text{SFT}}}{1800\text{s}} \right) \left( \frac{\Delta f}{6 \times 10^{-4}\text{Hz}} \right) \quad (21)$$

ただし、コヒーレント積分時間を  $T_{\text{coh}} = 30$  h とし、帯域幅は地球自転による Doppler modulation 幅の 3 倍にとった。SFT ファイルに加えてサーチコード (~1kbyte)、地球自転・公転の ephemeris ファイル、サーチパラメータファイルを圧縮して送る。

E@H Early S5 では 1 0 0 0 個の候補 event をサーバーに戻す。各候補は 5 つの値で特徴付けられる (right ascension, declination,  $f, \dot{f}, F$ )、 $5 \times 1000 \times 8 = 40\text{kbyte}$ 。(detection statistic  $F$  については JKS[7] 参照。)

10 万台 ([http://einstein.phys.uwm.edu/server\\_status.php](http://einstein.phys.uwm.edu/server_status.php)) のホスト計算機が活動。一つのワークユニットは、1 日から 1 週間で作業を終えられるように設計している。 $40\text{kbyte} \times 10^5/1\text{d} = 50\text{kbyte/s}$  のネットワークトラフィック。ただしデータ量は設定する false dismissal にもよる。参加国は 2 0 0 国にもものぼるので、参加者側のネットワークの状況を考える必要あり。

結果のチェックのため、同一のテンプレートに対して 3 回 (S4) もしくは 2 回 (Early S5) 同じ計算を異なるホストにしてみよう。

3日で20Hz分の探索が終わるとすると、1.5kHz探索に225日かかる。これにさらに重力波振幅の upper limit を決めるなら Monte Carlo が必要。

### 3.3 計算機速度

- Core 2 Quad. 102.4GFLOPS (QX9775) (wikipedia)
- Einstein@Home 190 TFLOPS (2009 July 11) (<http://einstein.phys.uwm.edu/>)

### 3.4 E@Hの仕事

#### 3.4.1 Code preparation

コードはLSC Continuous Wave Upper Limit (CWUL) group が使っていた `lalapps_ComputeFstastic` がベース。このコードは主にテンプレートグリッド生成・F statisticの計算をおこなう。

Windows/mac/linuxなどのOSまたCPUの違い(Intel AMD x86/IBM Power etc)に対応する必要がある(endianの問題など)。スクリーンセーバーなどincentiveの開発、参加者が計算機を使用したらメモリ上のコードとデータをHDDに退避させるようにする設計など細かいこと多数。

#### 3.4.2 SFT

現在のE@HではSFTは30分長(～45Mbyte)。 $f_{\max} = 2\text{kHz}$ とするとデータ点数は $N_d = 2 \times 2000\text{Hz} \times 1800\text{s} = 7.2 \times 10^6$ で、計算量は $3N_d \log_2 N_d = 5 \times 10^8$ 。1年間でSFTの数は $M = 1\text{yr}/20\text{min} = 17520$ より、総計算量は $10^{13}$ 。100GFlopsの計算機なら100 seconds。多分ファイルI/Oのほうが重要。

F statisticはノイズパワースペクトラムで規格化した量なので、SFTの絶対値にはよらない。つまり検出器の感度は参加者に知らせる必要は無い。LSC-E@Hでは参加者には適当に再規格化したSFTを配布している。参加者がデータを見て検出器の感度を知ることはできないようにしている。

#### 3.4.3 Data preparation

Server側でwok-unit generatorがband limited SFTを生成、search parameter file, ephemeris file, template grid fileを圧縮してホストに送る。

### 3.4.4 Coherent search

ホストが計算。E@H S3 当時の coherent integration コードは FFT を使っていなかったため (F statistic の計算に  $n \log n$  ではなく、 $n^2$  の計算をしていた)、計算量は一つのテンプレートあたり、640 M 浮動小数点演算 (ここで M は SFT file の数。30 分 SFT で 30 時間 coherent 積分なら  $M=60$ )。

Krishnan らの metric based の評価 ([5]LIGO-T080340) によると sky grid の数は

$$N_{\text{sky}} = 0.1 \left( \frac{T_{\text{coh}}}{1\text{h}} \right)^{2.41} \left( \frac{f}{100\text{Hz}} \right)^2 \left( \frac{0.5}{m} \right) \quad (22)$$

これは  $T_{\text{coh}}=30$  時間、 $f=100$  Hz で 1000 ぐらい。スピンドウンパラメータの数は周波数帯域  $\Delta f$ 、スピンドウンパラメータ  $\Delta \dot{f}$  Hz/s をサーチするとして、

$$N_{\text{spindown}} = \frac{\pi^2 \Delta f \Delta \dot{f}}{24 \sqrt{15m}} T_{\text{coh}}^3 \quad (23)$$

ただし、 $\mu_{\text{max}} \simeq 1.25m$  は maximal mismatch。

$$\eta = (1 - \mu_{\text{max}}) \frac{4 h_0^2 T}{25 S_n(f)} \quad (24)$$

(ただし、template gridd が cubic lattice のとき平均の loss は  $1 - \mu_{\text{max}}$  ではなく  $1 - \mu_{\text{max}}/3$ 。)

E@H S4 Revision 2 では各 work unit は  $\sim 3 \times 10^8$  程度のテンプレートに対して 30 時間の coherent search をしている。 $640 \times 60 \times 3 \times 10^8 = 12 \times 10^{12}$  Floating Point Operations。100MFLOPS の計算機なら 30 時間。

### 3.4.5 Results validator

計算結果の sanity check (file corruption、参加者が勝手に「重力波」を入れていないかどうかのチェック、など)。同じデータを複数の人に解析してもらい、結果の一致をチェックするのが基本。

### 3.4.6 Post-processing

Implementation dependent。

E@HS3 では 2GHz/2Gbyte memory の計算機で 1 時間程度で全データ解析可能。重力波振幅の Upper limit を決めるのに Monte carlo が必要。

例えば S2 では 1Hz band ごとに 5000 回の software signal injection を 20 の異なる重力波振幅でおこなっている。S2 では  $f_{\max} = 1.5\text{kHz}$  まで 10 時間のデータを作成し、upper limit を求めるのに当時のコードで 2 週間かかった。当時は software signal generation を time domain でやっていたため非常に時間がかかった。frequency domain でやれば時間は短縮できるはず (Greg. Mendall がコードを作っていたのでおそらく現在は frequency domain で signal generation をしている)。

Early S5 までの E@H のように帯域幅を制限したファイルをホストに解析してもらう場合、FFT を使った方がよいのかどうかは不明。

### 3.5 LSC computing cluster

LSC はデータ解析のための計算機クラスターを UWM, AEI, Birmingham などに持つ。Upper limit を決めるための Monte carlo などに必要。E@H ホストに Monte Carlo もやってもらうなら不要。

Morgane cluster (AEI:<http://gw.aei.mpg.de/>) specification

- \* Master nodes:
    - o dual-core Opteron 185
    - o 2 GB RAM
    - o 160 GB RAID-1 system disk
    - o 2 Gigabit ethernet interfaces (public and private network)
  - \* Compute nodes:
    - o dual-core Opteron 185
    - o 2 GB RAM
    - o 160 GB system and scratch disk
    - o Gigabit ethernet interface (private network)
  - \* Storage servers:
    - o dual-core Opteron 185
    - o 2 GB RAM
    - o Areca-1160 RAID controller
    - o 16\*500 GB disks in RAID-6 configuration
    - o two volumes: one small for system & scratch, one large (6.8 TB) for data storage
- which will be combined into one large namespace using a "parallel" or "cluster file system"

The total computing power of the 1230 CPU cores has been estimated

to be more than 6 Tflops peak, the data storage capacity is about 100 TB.

### 3.6 電気代

個人用ワークステーションの電源は 1kW クラス。常時稼働として年間約 20 万円。ファイルサーバー (x2) とバックアップサーバーで計 4 台。これに加えて冷房代。

### 3.7 E@H サーバーの仕事

web server (宣伝、参加者の登録、結果報告), BOINC server, work-unit generator, result validator (ホストから戻ってきた結果ファイルが正しいかどうかをチェック) など。

### 3.8 E@H server specification

unknown.

## 参考文献

- [1] B. Abbott et al. Setting upper limits on the strength of periodic gravitational waves using the first science data from the GEO 600 and LIGO detectors. *Phys. Rev.*, D69:082004, 2004.
- [2] B. Abbott et al. First all-sky upper limits from LIGO on the strength of periodic gravitational waves using the Hough transform. *Phys. Rev.*, D72:102004, 2005.
- [3] B. Abbott et al. Coherent searches for periodic gravitational waves from unknown isolated sources and Scorpius X-1: Results from the second LIGO science run. *Phys. Rev.*, D76:082001, 2007.
- [4] B. Abbott et al. All-sky LIGO Search for Periodic Gravitational Waves in the Early S5 Data. *Phys. Rev. Lett.*, 102:111102, 2009.
- [5] R. Prix B. Krishnan, B. J. Owen and A. M. Sintes. Searching for isolated pulsars using Einstein@Home. *LIGO Technical document*, T080340, 2008.

- [6] Patrick R. Brady and Teviet Creighton. Searching for periodic sources with LIGO. II: Hierarchical searches. *Phys. Rev.*, D61:082001, 2000.
- [7] P. Jaranowski, A. Królak, and B. F. Schutz. Data analysis of gravitational-wave signals from spinning neutron stars: The signal and its detection. *Phys. Rev. D.*, 58(6):063001–+, September 1998.
- [8] Badri Krishnan et al. The Hough transform search for continuous gravitational waves. *Phys. Rev.*, D70:082001, 2004.
- [9] Michele Maggiore. *Gravitational Waves. Vol. 1: Theory and Experiments*. Oxford University Press, 2007.