University of Texas Rio Grande Valley

# ScholarWorks @ UTRGV

# Einstein@Home search for periodic gravitational waves in LIGO S4 data

B. Abbott

R. Abbott

R. Adhikari

P. Ajith

B. Allen

*See next page for additional authors*

## Recommended Citation

## Authors

B. Abbott, R. Abbott, R. Adhikari, P. Ajith, B. Allen, G. Allen, R. Amin, D. P. Anderson, S. B. Anderson, W. G. Anderson, M. A. Arain, M. Araya, H. Armandula, P. Armor, Y. Aso, S. Aston, P. Aufmuth, C. Aulbert, S. Babak, S. Ballmer, H. Bantilan, B. C. Barish, C. Barker, D. Barker, B. Barr, P. Barriga, M. A. Barton, M. Bastarrika, K. Bayer, and J. Betzwieser

# Einstein@Home search for periodic gravitational waves in LIGO S4 data

B. Abbott,[16] R. Abbott,[16] R. Adhikari,[16] P. Ajith,[2] B. Allen,[2,55] G. Allen,[32] R. Amin,[20] D. P. Anderson,[41] S. B. Anderson,[16] W. G. Anderson,[55] M. A. Arain,[42] M. Araya,[16] H. Armandula,[16] P. Armor,[55] Y. Aso,[10] S. Aston,[40] P. Aufmuth,[15] C. Aulbert,[2] S. Babak,[1] S. Ballmer,[16] H. Bantilan,[8] B. C. Barish,[16] C. Barker,[18] D. Barker,[18] B. Barr,[43] P. Barriga,[54] M. A. Barton,[43] M. Bastarrika,[43] K. Bayer,[17] J. Betzwieser,[16] P. T. Beyersdorf,[28] I. A. Bilenko,[23] G. Billingsley,[16] R. Biswas,[55] E. Black,[16] K. Blackburn,[16] L. Blackburn,[17] D. Blair,[54] B. Bland,[18] T. P. Bodiya,[17] L. Bogue,[19] R. Bork,[16] V. Boschi,[16] S. Bose,[56] P. R. Brady,[55] V. B. Braginsky,[23] J. E. Brau,[48] M. Brinkmann,[2] A. Brooks,[16] D. A. Brown,[33] G. Brunet,[17] A. Bullington,[32] A. Buonanno,[44] O. Burmeister,[2] R. L. Byer,[32] L. Cadonati,[45] G. Cagnoli,[43] J. B. Camp,[24] J. Cannizzo,[24] K. Cannon,[16] J. Cao,[17] L. Cardenas,[16] T. Casebolt,[32] G. Castaldi,[51] C. Cepeda,[16] E. Chalkley,[43] P. Charlton,[9] S. Chatterji,[16] S. Chelkowski,[40] Y. Chen,[1,6] N. Christensen,[8] D. Clark,[32] J. Clark,[43] T. Cokelaer,[7] R. Conte,[50] D. Cook,[18] T. Corbitt,[17] D. Coyne,[16] J. D. E. Creighton,[55] T. D. Creighton,[36] A. Cumming,[43] L. Cunningham,[43] R. M. Cutler,[40] J. Dalrymple,[33] K. Danzmann,[2,15] G. Davies,[7] D. DeBra,[32] J. Degallaix,[1] M. Degree,[32] V. Dergachev,[46] S. Desai,[34] R. DeSalvo,[16] S. Dhurandhar,[14] M. Díaz,[36] J. Dickson,[4] A. Dietz,[7] F. Donovan,[17] K. L. Dooley,[42] E. E. Doomes,[31] R. W. P. Drever,[5] I. Duke,[17] J.-C. Dumas,[54] R. J. Dupuis,[16] J. G. Dwyer,[10] C. Echols,[16] A. Effler,[18] P. Ehrens,[16] G. Ely,[8] E. Espinoza,[16] T. Etzel,[16] T. Evans,[19] S. Fairhurst,[7] Y. Fan,[54] D. Fazi,[16] H. Fehrmann,[2] M. M. Fejer,[32] L. S. Finn,[34] K. Flasch,[55] N. Fotopoulos,[55] A. Freise,[40] R. Frey,[48] T. Fricke,[16,49] P. Fritschel,[17] V. V. Frolov,[19] M. Fyffe,[19] J. Garofoli,[18] I. Gholami,[1] J. A. Giaime,[19,20] S. Giampanis,[49] K. D. Giardina,[19] K. Goda,[17] E. Goetz,[46] L. Goggin,[16] G. González,[20] S. Gossler,[2] R. Gouaty,[20] A. Grant,[43] S. Gras,[54] C. Gray,[18] M. Gray,[4] R. J. S. Greenhalgh,[27] A. M. Gretarsson,[11] F. Grimaldi,[17] R. Grosso,[36] H. Grote,[2] S. Grunewald,[1] M. Guenther,[18] E. K. Gustafson,[16] R. Gustafson,[46] B. Hage,[15] J. M. Hallam,[40] D. Hammer,[55] C. Hanna,[20] J. Hanson,[19] J. Harms,[2] G. Harry,[17] E. Harstad,[48] K. Hayama,[36] T. Hayler,[27] J. Heefner,[16] I. S. Heng,[43] M. Hennessy,[32] A. Heptonstall,[43] M. Hewitson,[2] S. Hild,[40] E. Hirose,[33] D. Hoak,[19] D. Hosken,[39] J. Hough,[43] S. H. Huttner,[43] D. Ingram,[18] M. Ito,[48] A. Ivanov,[16] B. Johnson,[18] W. W. Johnson,[20] D. I. Jones,[52] G. Jones,[7] R. Jones,[43] L. Ju,[54] P. Kalmus,[10] V. Kalogera,[26] S. Kamat,[10] J. Kanner,[44] D. Kasprzyk,[40] E. Katsavounidis,[17] K. Kawabe,[18] S. Kawamura,[25] F. Kawazoe,[25] W. Kells,[16] D. G. Keppel,[16] F. Ya. Khalili,[23] R. Khan,[10] E. Khazanov,[13] C. Kim,[26] P. King,[16] J. S. Kissel,[20] S. Klimenko,[42] K. Kokeyama,[25] V. Kondrashov,[16] R. K. Kopparapu,[34] D. Kozak,[16] I. Kozhevatov,[13] B. Krishnan,[1] P. Kwee,[15] P. K. Lam,[4] M. Landry,[18] M. M. Lang,[34] B. Lantz,[32] A. Lazzarini,[16] M. Lei,[16] N. Leindecker,[32] V. Leonhardt,[25] I. Leonor,[48] K. Libbrecht,[16] H. Lin,[42] P. Lindquist,[16] N. A. Lockerbie,[53] D. Lodhia,[40] M. Lormand,[19] P. Lu,[32] M. Lubinski,[18] A. Lucianetti,[42] H. Lück,[2,15] B. Machenschalk,[2] M. MacInnis,[17] M. Mageswaran,[16] K. Mailand,[16] V. Mandic,[47] S. Márka,[10] Z. Márka,[10] A. Markosyan,[32] J. Markowitz,[17] E. Maros,[16] I. Martin,[43] R. M. Martin,[42] J. N. Marx,[16] K. Mason,[17] F. Matichard,[20] L. Matone,[10] R. Matzner,[35] N. Mavalvala,[17] R. McCarthy,[18] D. E. McClelland,[4] S. C. McGuire,[31] M. McHugh,[22] G. McIntyre,[16] G. McIvor,[35] D. McKechan,[7] K. McKenzie,[4] T. Meier,[15] A. Melissinos,[49] G. Mendell,[18] R. A. Mercer,[42] S. Meshkov,[16] C. J. Messenger,[2] D. Meyers,[16] J. Miller,[16,43] J. Minelli,[34] S. Mitra,[14] V. P. Mitrofanov,[23] G. Mitselmakher,[42] R. Mittleman,[17] O. Miyakawa,[16] B. Moe,[55] S. Mohanty,[36] G. Moreno,[18] K. Mossavi,[2] C. MowLowry,[4] G. Mueller,[42] S. Mukherjee,[36] H. Mukhopadhyay,[14] H. Müller-Ebhardt,[2] J. Munch,[39] P. Murray,[43] E. Myers,[18] J. Myers,[18] T. Nash,[16] J. Nelson,[43] G. Newton,[43] A. Nishizawa,[25] K. Numata,[24] J. O'Dell,[27] G. Ogin,[16] B. O'Reilly,[19] R. O'Shaughnessy,[34] D. J. Ottaway,[17] R. S. Ottens,[42] H. Overmier,[19] B. J. Owen,[34] Y. Pan,[44] C. Pankow,[42] M. A. Papa,[1,55] V. Parameshwaraiah,[18] P. Patel,[16] M. Pedraza,[16] S. Penn,[12] A. Perreca,[40] T. Petrie,[34] I. M. Pinto,[51] M. Pitkin,[43] H. J. Pletsch,[2] M. V. Plissi,[43] F. Postiglione,[50] M. Principe,[51] R. Prix,[2] V. Quetschke,[42] F. Raab,[18] D. S. Rabeling,[4] H. Radkins,[18] N. Rainer,[2] M. Rakhmanov,[30] M. Ramsunder,[34] H. Rehbein,[2] S. Reid,[43] D. H. Reitze,[42] R. Riesen,[19] K. Riles,[46] B. Rivera,[18] N. A. Robertson,[16,43] C. Robinson,[7] E. L. Robinson,[40] S. Roddy,[19] A. Rodriguez,[20] A. M. Rogan,[56] J. Rollins,[10] J. D. Romano,[36] J. Romie,[19] R. Route,[32] S. Rowan,[43] A. Rüdiger,[2] L. Ruet,[17] P. Russell,[16] K. Ryan,[18] S. Sakata,[25] M. Samidi,[16] L. Sancho de la Jordana,[38] V. Sandberg,[18] V. Sannibale,[16] S. Saraf,[29] P. Sarin,[17] B. S. Sathyaprakash,[7] S. Sato,[25] P. R. Saulson,[33] R. Savage,[18] P. Savov,[6] S. W. Schediwy,[54] R. Schilling,[2] R. Schnabel,[2] R. Schofield,[48] B. F. Schutz,[1,7] P. Schwinberg,[18] S. M. Scott,[4] A. C. Searle,[4] B. Sears,[16] F. Seifert,[2] D. Sellers,[19] A. S. Sengupta,[16] P. Shawhan,[44] D. H. Shoemaker,[17] A. Sibley,[19] X. Siemens,[55] D. Sigg,[18] S. Sinha,[32] A. M. Sintes,[1,38] B. J. J. Slagmolen,[4] J. Slutsky,[20] J. R. Smith,[33] M. R. Smith,[16] N. D. Smith,[17] K. Somiya,[1,2] B. Sorazu,[43] L. C. Stein,[17] A. Stochino,[16] R. Stone,[36] K. A. Strain,[43] D. M. Strom,[48] A. Stuver,[19] T. Z. Summerscales,[3] K.-X. Sun,[32] M. Sung,[20] P. J. Sutton,[7] H. Takahashi,[1] D. B. Tanner,[42] R. Taylor,[16] R. Taylor,[43] J. Thacker,[19] K. A. Thorne,[34] K. S. Thorne,[6] A. Thüring,[15] K. V. Tokmakov,[43] C. Torres,[19] C. Torrie,[43] G. Traylor,[19] M. Trias,[38] W. Tyler,[16] D. Ugolini,[37] J. Ulmen,[32] K. Urbanek,[32] H. Vahlbruch,[15] C. Van Den Broeck,[7] M. van der Sluys,[26] S. Vass,[16] R. Vaulin,[55] A. Vecchio,[40] J. Veitch,[40]

P. Veitch,[39] S. Vigeland,[8] A. Villar,[16] C. Vorvick,[18] S. P. Vyachanin,[23] S. J. Waldman,[16] L. Wallace,[16] H. Ward,[43] R. Ward,[16] M. Weinert,[2] A. Weinstein,[16] R. Weiss,[17] S. Wen,[20] K. Wette,[4] J. T. Whelan,[1] D. M. Whitbeck,[34] S. E. Whitcomb,[16] B. F. Whiting,[42] C. Wilkinson,[18] P. A. Willems,[16] H. R. Williams,[34] L. Williams,[42] B. Willke,[2,15] I. Wilmut,[27] W. Winkler,[2] C. C. Wipf,[17] A. G. Wiseman,[55] G. Woan,[43] R. Wooley,[19] J. Worden,[18] W. Wu,[42] I. Yakushin,[19] H. Yamamoto,[16] Z. Yan,[54] S. Yoshida,[30] M. Zanolin,[11] J. Zhang,[46] L. Zhang,[16] C. Zhao,[54] N. Zotov,[21] M. Zucker,[17] and J. Zweizig[16]

(LIGO Scientific Collaboration)[*]

[1]*Albert-Einstein-Institut, Max-Planck-Institut für Gravitationsphysik, D-14476 Golm, Germany*
[2]*Albert-Einstein-Institut, Max-Planck-Institut für Gravitationsphysik, D-30167 Hannover, Germany*
[3]*Andrews University, Berrien Springs, Michigan 49104, USA*
[4]*Australian National University, Canberra, 0200, Australia*
[5]*California Institute of Technology, Pasadena, California 91125, USA*
[6]*Caltech-CaRT, Pasadena, California 91125, USA*
[7]*Cardiff University, Cardiff, CF24 3AA, United Kingdom*
[8]*Carleton College, Northfield, Minnesota 55057, USA*
[9]*Charles Sturt University, Wagga Wagga, NSW 2678, Australia*
[10]*Columbia University, New York, New York 10027, USA*
[11]*Embry-Riddle Aeronautical University, Prescott, Arizona 86301, USA*
[12]*Hobart and William Smith Colleges, Geneva, New York 14456, USA*
[13]*Institute of Applied Physics, Nizhny Novgorod, 603950, Russia*
[14]*Inter-University Centre for Astronomy and Astrophysics, Pune—411007, India*
[15]*Leibniz Universität Hannover, D-30167 Hannover, Germany*
[16]*LIGO—California Institute of Technology, Pasadena, California 91125, USA*
[17]*LIGO—Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA*
[18]*LIGO Hanford Observatory, Richland, Washington 99352, USA*
[19]*LIGO Livingston Observatory, Livingston, Louisiana 70754, USA*
[20]*Louisiana State University, Baton Rouge, Louisiana 70803, USA*
[21]*Louisiana Tech University, Ruston, Louisiana 71272, USA*
[22]*Loyola University, New Orleans, Louisiana 70118, USA*
[23]*Moscow State University, Moscow, 119992, Russia*
[24]*NASA/Goddard Space Flight Center, Greenbelt, Maryland 20771, USA*
[25]*National Astronomical Observatory of Japan, Tokyo 181-8588, Japan*
[26]*Northwestern University, Evanston, Illinois 60208, USA*
[27]*Rutherford Appleton Laboratory, Chilton, Didcot, Oxon OX11 0QX, United Kingdom*
[28]*San Jose State University, San Jose, California 95192, USA*
[29]*Sonoma State University, Rohnert Park, California 94928, USA*
[30]*Southeastern Louisiana University, Hammond, Louisiana 70402, USA*
[31]*Southern University and A&M College, Baton Rouge, Louisiana 70813, USA*
[32]*Stanford University, Stanford, California 94305, USA*
[33]*Syracuse University, Syracuse, New York 13244, USA*
[34]*The Pennsylvania State University, University Park, Pennsylvania 16802, USA*
[35]*The University of Texas at Austin, Austin, Texas 78712, USA*
[36]*The University of Texas at Brownsville and Texas Southmost College, Brownsville, Texas 78520, USA*
[37]*Trinity University, San Antonio, Texas 78212, USA*
[38]*Universitat de les Illes Balears, E-07122 Palma de Mallorca, Spain*
[39]*University of Adelaide, Adelaide, SA 5005, Australia*
[40]*University of Birmingham, Birmingham, B15 2TT, United Kingdom*
[41]*University of California at Berkeley, Berkeley, California 94720, USA*
[42]*University of Florida, Gainesville, Florida 32611, USA*
[43]*University of Glasgow, Glasgow, G12 8QQ, United Kingdom*
[44]*University of Maryland, College Park, Maryland 20742, USA*
[45]*University of Massachusetts, Amherst, Massachusetts 01003, USA*
[46]*University of Michigan, Ann Arbor, Michigan 48109, USA*
[47]*University of Minnesota, Minneapolis, Minnesota 55455, USA*
[48]*University of Oregon, Eugene, Oregon 97403, USA*
[49]*University of Rochester, Rochester, New York 14627, USA*
[50]*University of Salerno, 84084 Fisciano (Salerno), Italy*

[51]*University of Sannio at Benevento, I-82100 Benevento, Italy*
[52]*University of Southampton, Southampton, SO17 1BJ, United Kingdom*
[53]*University of Strathclyde, Glasgow, G1 1XQ, United Kingdom*
[54]*University of Western Australia, Crawley, Western Australia 6009, Australia*
[55]*University of Wisconsin-Milwaukee, Milwaukee, Wisconsin 53201, USA*
[56]*Washington State University, Pullman, Washington 99164, USA*

A search for periodic gravitational waves, from sources such as isolated rapidly spinning neutron stars, was carried out using 510 h of data from the fourth LIGO science run (S4). The search was for quasimonochromatic waves in the frequency range from 50 to 1500 Hz, with a linear frequency drift $\dot{f}$ (measured at the solar system barycenter) in the range $-f/\tau < \dot{f} < 0.1 f/\tau$, where the minimum spin-down age $\tau$ was 1000 yr for signals below 300 Hz and 10 000 yr above 300 Hz. The main computational work of the search was distributed over approximately 100 000 computers volunteered by the general public. This large computing power allowed the use of a relatively long coherent integration time of 30 h, despite the large parameter space searched. No statistically significant signals were found. The sensitivity of the search is estimated, along with the fraction of parameter space that was vetoed because of contamination by instrumental artifacts. In the 100 to 200 Hz band, more than 90% of sources with dimensionless gravitational-wave strain amplitude greater than $10^{-23}$ would have been detected.

## I. INTRODUCTION

Gravitational waves are a fundamental prediction of Einstein's general theory of relativity [1,2]. But these waves are very weak, so although there is compelling indirect evidence for their existence [3], direct detection has so far not been possible.

In the past decade, advances in lasers, optics, and control systems have enabled construction of a new generation of gravitational-wave (GW) detectors [4] that offer the first realistic promise of a direct detection. The Laser Interferometer Gravitational-Wave Observatory (LIGO) [5,6] is currently the most sensitive of these instruments. LIGO consists of three kilometer-scale instruments. Two are located in a common vacuum envelope in Hanford, Washington, USA and the other is located in Livingston, Louisiana, USA.

This paper reports on the results of the Einstein@Home search for "continuous wave" (CW) sources in the data from the fourth LIGO science run (S4). The configuration of the LIGO detectors during the S4 run is described in a separate instrumental paper [7].

### A. Continuous wave sources and detection methods

"Continuous waves" are quasimonochromatic gravitational-wave signals whose duration is longer than the observation time. They have a well-defined frequency on short time scales, which can vary slowly over longer times. These types of waves are expected, for example, from spinning neutron stars with nonaxisymmetric deformations. If the system is isolated, then it loses angular momentum due to the radiation. The spinning motion slows down, and the gravitational-wave frequency de-

creases. Gravitational acceleration towards a large nearby mass distribution can also produce such a frequency drift (of either sign). Many possible emission mechanisms could lead the to the emission of such waves by spinning neutron stars [8–16].

If there were no acceleration between the LIGO detectors and the GW sources, then it would be possible to search for CW signals from unknown sources using only "standard" computing resources, such as a high-end workstation or a small computing cluster. In this case the analysis technique would be simple: compute the fast Fourier transform (FFT) [17,18] of the original time-series data, and search along the frequency axis for peaks in the power spectrum. Time-domain resampling or similar techniques could be used to compensate for the effects of a linear-in-time frequency drift.

However, this simple analysis is not possible because of the terrestrial location of the LIGO detectors: signals that are purely sinusoidal at the source are Doppler-modulated by the Earth's motion and thus are no longer sinusoidal at the detector. The Earth's rotation about its axis modulates the signal frequency at the detector by approximately one part in $10^6$, with a period of one sidereal day. In addition, the Earth's orbit about the Sun modulates the signal frequency at the detector by approximately one part in $10^4$, with a period of 1 yr. These two modulations, whose exact form depends upon the precise sky location of the source, greatly complicate the data analysis when searching for unknown sources. The search becomes even more complicated if the CW emitter is part of a binary star system, since the orbital motion of the binary system introduces additional modulations into the waveform.

The "brute force" approach to the data analysis problem would employ matched filtering, convolving all available data with a family of template waveforms corresponding to

---

*http://www.ligo.org/

all possible putative sources. The resulting search statistic is called the $\mathcal{F}$-statistic and was first described in a seminal paper of Jaranowski, Królak, and Schutz [19]. But even for isolated neutron stars (i.e. which are not in binary systems) the parameter space of possible sources is four-dimensional, with two parameters required to describe the source sky position using standard astronomical equatorial coordinates $\alpha$ (right ascension) and $\delta$ (declination), and additional coordinates $(f, \dot{f})$ denoting the intrinsic frequency and frequency drift. To achieve the maximum possible sensitivity, the template waveforms must match the source waveforms to within a fraction of a cycle over the entire observation time (with current detectors this is months or years). So one must choose a very closely spaced grid of templates in this four-dimensional parameter space, and the computational cost exceeds all available computing resources on the planet [20]. Thus the direct approach is not possible in practice.

More efficient and sensitive methods for this type of search have been studied for more than a decade and are under development [21–25]. In this paper, the frequency-domain method described in [26,27] is used to calculate the $\mathcal{F}$-statistic. In order to maximize the possible integration time, and hence achieve a more sensitive coherent search, the computation was distributed among approximately $10^5$ computers belonging to $\sim 5 \times 10^4$ volunteers in $\sim 200$ countries. This distributed computation project, called Einstein@Home [28], follows the model of a number of other well-known volunteer distributed computing projects such as SETI@home [29] and Folding@home [30].

Other methods have also been employed for the CW search of the S4 data [31,32] and searches for other signal types (burst, inspiral, stochastic background) have also been carried out [33–38] with this data set. The results of these searches are all upper bounds, with no detections reported.

### B. Outline of this paper

The outline of this paper is as follows. Sections II and III describe the overall construction of the search, including the data set preparation, regions of parameter space searched, and the choices of thresholds and sensitivities. Section IV describes the post-processing pipeline. The level of sensitivity of the search is estimated in Sec. V. Section VI describes the vetoing of instrumental line artifacts and the fraction of parameter space that was therefore excluded. Section VII describes the end-to-end validation of the search and the post-processing pipeline, which was done by injecting simulated CW signals into the detector hardware. Section VIII describes the final results of the search, and is followed by a short conclusion.

### II. DATA SELECTION AND PREPARATION

The data for the S4 run was collected between February 22, 2005 and March 23, 2005. The data analyzed

consisted of 300 h of data from the LIGO Hanford 4-km (H1) detector and 210 h of data from the LIGO Livingston 4-km (L1) detector.

The search method used here (explained in detail in Sec. III) consists of computing a coherent $\mathcal{F}$-statistic over data segments of 30 h each, and combining these results via an incoherent coincidence scheme. However, the 30-h segments have time gaps, and the number of templates needed for the coherent $\mathcal{F}$-statistic step grows rapidly as the gaps get longer. For this reason, the start and end times of the data segments were selected based on the criteria that the gaps totaled no more than ten hours: each data segment contains 30 h of science-mode data and lies within a total time span of less than 40 h. Here and in the following the term "segment" is always used to refer to one of these time stretches of data, each of which contains exactly $T_{\text{obs}} = 30$ h of data. The total time spanned by the data in segment $j$ is written $T_{\text{span},j}$; 30 h < $T_{\text{span},j}$ < 40 h.

The data segments consist of *uninterrupted* blocks of 1800 s of *contiguous* science-mode data. This is for technical reasons: the $\mathcal{F}$-statistic code uses short Fourier transforms (SFTs) over $T_{\text{SFT}} = 1800$ s as input data (this data format is described in [39]). To produce these SFTs, the data is first calibrated in the time domain using the method described in [40,41]. Then the data is windowed in 1800 s intervals using a Tukey window with a characteristic turn-on/turn-off time of 500 ms, followed by a FFT.

Applying the above constraints to the S4 data set yielded a total of $N_{\text{seg}} = 17$ data segments (10 from H1, 7 from L1), labeled by $j = 1, \ldots, 17$. This paper uses Global Positioning System (GPS) time coordinates [42] through-

TABLE I. Segments of S4 data used in this search, in order of decreasing sensitivity at 141.3 Hz for H1 and at 135.3 Hz for L1. The columns are the data segment index $j$, the GPS start time $t_j$, the GPS end time $t_j^{\text{end}}$, and the time spanned $T_{\text{span},j} = t_j^{\text{end}} - t_j$.

| $j$ | Detector | $t_j$ [s] | $t_j^{\text{end}}$ [s] | $T_{\text{span},j}$ [s] |
|---|---|---|---|---|
| 1 | H1 | 794 461 805 | 794 583 354 | 121 549 |
| 2 | H1 | 794 718 394 | 794 840 728 | 122 334 |
| 3 | H1 | 795 149 016 | 795 285 470 | 136 454 |
| 4 | H1 | 793 555 944 | 793 685 616 | 129 672 |
| 5 | H1 | 795 493 128 | 795 626 713 | 133 585 |
| 6 | H1 | 793 936 862 | 794 078 332 | 141 470 |
| 7 | H1 | 794 885 406 | 795 015 166 | 129 760 |
| 8 | H1 | 794 244 737 | 794 378 322 | 133 585 |
| 9 | H1 | 794 585 154 | 794 714 794 | 129 640 |
| 10 | H1 | 793 766 877 | 793 910 817 | 143 940 |
| 11 | L1 | 795 553 209 | 795 678 679 | 125 470 |
| 12 | L1 | 795 115 986 | 795 246 307 | 130 321 |
| 13 | L1 | 795 408 715 | 795 545 555 | 136 840 |
| 14 | L1 | 794 625 269 | 794 762 685 | 137 416 |
| 15 | L1 | 794 053 883 | 794 197 272 | 143 389 |
| 16 | L1 | 794 206 397 | 794 328 337 | 121 940 |
| 17 | L1 | 794 875 424 | 795 016 053 | 140 629 |

out. The GPS start time of segment $j$ is denoted $t_j$; the values are listed in Table I.

The maximum Doppler modulation (from the Earth's motion about the Sun) is about one part in $10^4$. Over the length of S4, and in the parameter range considered, the frequency changes due to intrinsic spin-down are comparable or smaller. This means that the CW signals searched for here always stays within a narrow frequency band, drifting no more than about $\pm 0.15$ Hz from some fiducial frequency. For this reason the input data, spanning the frequency range of 50 to 1500 Hz, is partitioned in the frequency domain into 5800 "slices" of 0.5 Hz plus wings of 0.175 Hz on either side (1531 Fourier bins). The size of one such input data slice is 7 368 000 bytes for H1 (con-

TABLE II. Instrumental lines replaced with Gaussian noise in the input data. The three columns show the frequency of the fundamental harmonic $f_{\mathrm{Line}}$, the number of harmonics $N$, and the bandwidth $\Delta f_{\mathrm{Line}}$ removed on either side of the central frequency (total bandwidth removed per harmonic is $2\Delta f_{\mathrm{Line}}$). In total 77.285 556 Hz of H1 data (139 114 Fourier bins) and 142.137 778 Hz of L1 data (255 848 Fourier bins) have been excluded *ab initio* from the frequency range $50$ Hz $\leq f \leq$ 1500 Hz. If $\Delta f_{\mathrm{Line}} = 0$ then the line-cleaning algorithm replaces a single Fourier bin with the average of bins on either side. The spacing between Fourier bins is $1/1800$ Hz.

| | H1 | |
|---|---|---|
| $f_{\mathrm{Line}}$ [Hz] | $N$ | $\Delta f_{\mathrm{Line}}$ [Hz] |
| 1.0 | 1451 | 0.0006 |
| 60.0 | 1 | 3.0 |
| 60.0 | 25 | 1.0 |
| 346.0 | 1 | 4.0 |
| 392.365 | 1 | 0.01 |
| 393.1 | 1 | 0.0 |
| 393.835 | 1 | 0.01 |
| 688.5 | 1 | 2.0 |
| 694.75 | 1 | 1.25 |
| 973.3 | 1 | 0.0 |
| 1030.55 | 1 | 0.1 |
| 1032.18 | 1 | 0.04 |
| 1032.58 | 1 | 0.1 |
| 1033.7 | 1 | 0.1 |
| 1033.855 | 1 | 0.05 |
| 1034.6 | 1 | 0.4 |
| 1042.5 | 1 | 1.5 |
| 1143.5672 | 1 | 0.2 |
| 1144.3 | 1 | 0.0 |
| 1145.0328 | 1 | 0.2 |
| 1373.75 | 1 | 0.1 |
| 1374.44 | 1 | 0.1 |
| 1377.14 | 1 | 0.1 |
| 1378.75 | 1 | 0.1 |
| 1379.52 | 1 | 0.1 |
| 1389.06 | 1 | 0.06 |
| 1389.82 | 1 | 0.07 |
| 1391.5 | 1 | 0.5 |

TABLE II. (*Continued*)

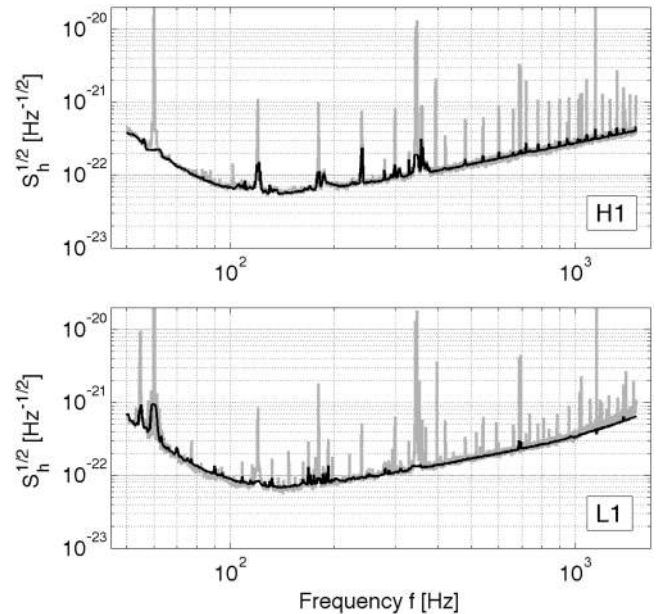| | L1 | |
|---|---|---|
| $f_{\mathrm{Line}}$ [Hz] | $N$ | $\Delta f_{\mathrm{Line}}$ [Hz] |
| 1.0 | 1451 | 0.0006 |
| 36.8725 | 39 | 0.8 |
| 54.7 | 1 | 0.0 |
| 60.0 | 25 | 1.0 |
| 345.0 | 1 | 10.0 |
| 396.7 | 1 | 0.0 |
| 686.5 | 1 | 1.0 |
| 688.83 | 1 | 0.5 |
| 693.7 | 1 | 0.7 |
| 1029.5 | 1 | 0.25 |
| 1031.0 | 1 | 0.5 |
| 1033.6 | 1 | 0.2 |
| 1041.0 | 1 | 1.0 |
| 1151.5 | 1 | 0.0 |
| 1372.925 | 1 | 0.075 |
| 1374.7 | 1 | 0.1 |
| 1375.2 | 1 | 0.1 |
| 1378.39 | 1 | 0.1 |
| 1387.4 | 1 | 0.05 |
| 1388.5 | 1 | 0.5 |



FIG. 1. Strain amplitude spectral densities $\sqrt{S_h(f)}$ of the S4 data from the LIGO detectors H1 (top) and L1 (bottom). The gray curves are medians of the *entire uncleaned* LIGO S4 science-mode data set with a frequency resolution of 0.125 Hz. The black curves show the *cleaned* S4 data used in this analysis with a frequency resolution of 0.5 Hz. The top (bottom) plot is the mean of the 10 H1 (7 L1) 30-hour data segments used in this Einstein@Home analysis.

taining 600 SFTs from 10 segments) and 5 157 600 bytes for L1 (containing 420 SFTs from 7 segments).

The detector data contains dozens of narrow-band spectral lines whose origin is instrumental, for example, the harmonics of the 60 Hz mains frequency, and violin modes of the mirror suspensions in the range from 342–350 Hz (H1) and 335–355 Hz (L1). To simplify later analysis, line features that are known to be instrumental artifacts are removed ("cleaned") from the data by replacing the frequency-domain data bins with computer-generated random Gaussian values. The frequencies of these lines are shown in Table II. The cleaning algorithm uses a moving-in-frequency median of the power in individual frequency bins to determine the instrumental noise floor. To prevent bias at the boundaries of the cleaned regions, the mean of the random values to replace the line features interpolates linearly between the noise floor at either side of the line feature. The median noise strain amplitude spectra of the final cleaned H1 and L1 data sets are shown in Fig. 1.

## III. DATA PROCESSING

Figure 2 is a schematic flow-diagram of the Einstein@Home data processing which is described in this section and in the following section on post-processing. It shows what parts of the analysis were done by project participants, what parts were done on project servers, and the relationships between these.

### A. BOINC workunit distribution and validation

The computational work of the search is partitioned into 6 731 410 workunits (separate computing tasks) and processed using the Berkeley Open Infrastructure for Network Computing (BOINC) [43–45]. Because the work is done on computers that are not owned or controlled by our scientific collaboration or institutions, any individual result could be wrong. Error sources include defective hardware (such as over-clocked memory), defective software (erroneous system libraries), and malicious users (faking correct results). To identify and eliminate such errors, BOINC was configured so that each workunit is done independently by computers owned by at least three different volunteers.

The most common types of errors (lack of disk space, corrupted or missing input files, inconsistent internal state, etc.) are detected during program execution. If an error is detected during run-time, the program reports to the Einstein@Home server that the workunit was *unsuccessful*, and BOINC generates another instance of the workunit, to be sent to another volunteer's computer. This behavior is repeated as necessary until three successful results have been obtained.

The three successful results obtained for each workunit are then compared by an automatic validator, which rejects results that do not agree closely. The validation process is more complicated than simple byte-by-byte comparison of output files, because Einstein@Home supports multiple
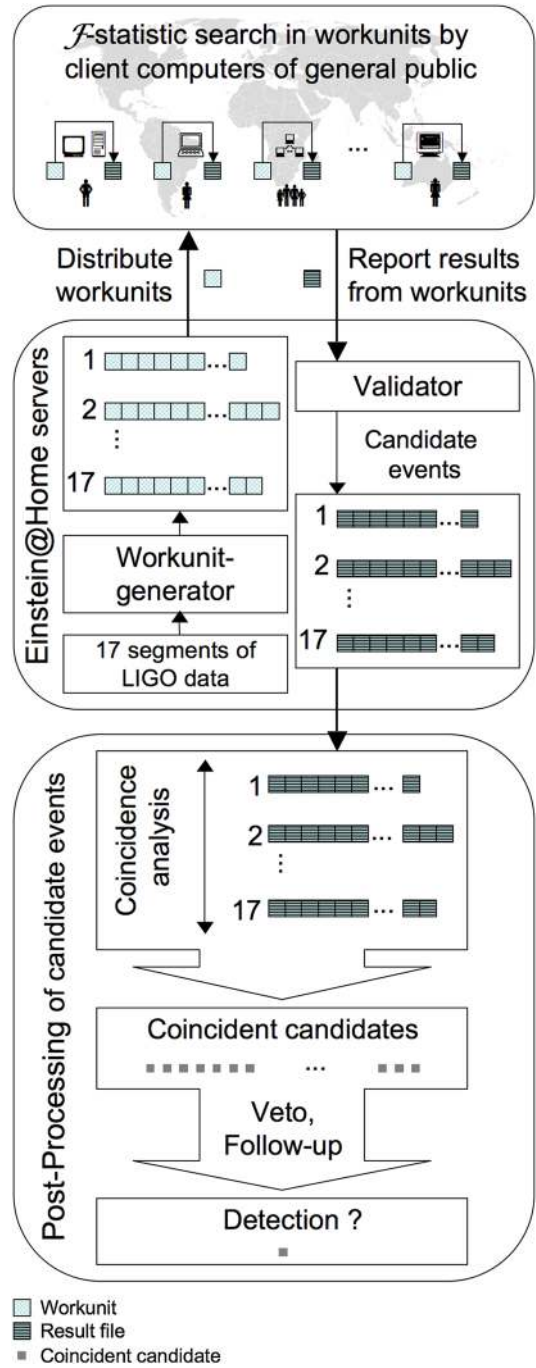


FIG. 2 (color online). Schematic overview of the Einstein@Home data processing and subsequent post-processing. Instrumental data enters at the start of the flow diagram on the bottom left-hand side of the middle "box."

computing platforms (Windows, GNU/Linux, Mac OS X on Intel and PPC, FreeBSD, and Solaris) and differences in CPU hardware, compiler instruction ordering, and floating-point libraries mean that correct and valid result files may exhibit small numerical differences. The automatic validation takes place in two steps.

The output files have a fixed five-column format and contain 13 000 candidate events, with one line per candi-

date event, as described in Sec. III C. The first validation step checks that the file syntax is correct and that each value is within the allowed range for that column. This detects most file corruption.

Then the validator does comparison of all possible pairs of result files. For a given pair of result files, the validator checks that corresponding candidate events lie on the same template grid point and have $\mathcal{F}$-statistic values that agree to within 1%. Since each file contains the 13 000 events with the largest values of the $\mathcal{F}$-statistic, numerical fluctuations in determining the value of $\mathcal{F}$ can lead to slightly different lists being returned on different platforms. Hence the validator tolerates unpaired candidate events if they have $\mathcal{F}$ values within 1% of the smallest value on the list.

A workunit is validated once it has three results that agree with one another to within these tolerances. If the three results do not pass this validation step, the Einstein@Home server generates more instances of this workunit until three *valid* results have been obtained. For the search described in this paper, a "post-mortem" analysis of the computation shows that the probability of a successful but invalid result is small (0.36%), and the errors which make a successful result invalid are typically unique and irreproducible. Hence we estimate that it is highly improbable that even a single incorrect result has been marked as "valid" by the automatic validator.

### B. Workunit design

The different workunits cover (search) different parts of parameter space. A key design goal of these workunits is that they should have roughly equal computational run times, of the order of ~8 CPU hours, and that the computational effort to carry out the entire search should last about 0.5–1 years. Another key design goal is that each workunit uses only a small reusable subset of the total data set. These allow Einstein@Home volunteers to do useful computations on the one-day time scale, and minimizes the download burden on their internet connections and on the Einstein@Home data servers.

Each workunit uses only one segment of data over a narrow frequency range, but searches the whole sky and the full range of frequency-derivatives $\dot{f}$ at that frequency. Therefore, the entire search is divided into computational units over different data segments and frequency bands. In the following it will be useful to label the workunits by three indices $(j, k, \ell)$, where $j = 1, \ldots, 17$ denotes the data segment, $k = 1, \ldots, 2900$ labels the 0.5 Hz band covered by the input data file, and $\ell = 1, \ldots, M(j, k)$ enumerates the individual workunits associated with data segment $j$ and frequency band $k$. Note that each workunit uses a frequency band that is *smaller* than the 0.5 Hz covered by the input data files, i.e. $M(j, k) \geq 1$.

#### 1. Grid in search parameter space

The parameter space is gridded in such a way that no point has a "squared-distance" from its nearest grid point

that exceeds a certain "maximal mismatch." The distance is defined by a metric on parameter space, first introduced in [21,22]. The squared distance is the fractional loss of squared signal-to-noise ratio (SNR$^2$, proportional to $\mathcal{F}$) due to waveform mismatch between the putative signal and the template. The search grid was constructed based on the projected metric on the subspace orthogonal to the frequency direction $\partial_f$ with $\dot{f} = 0$. For any given workunit, the parameter-space grid is a Cartesian product of uniformly spaced steps $df$ in frequency, uniformly spaced steps $d\dot{f}$ in frequency derivative, and a two-dimensional sky-grid, which has nonuniform spacings determined by the projected metric. For frequencies in the range $[50, 290)$ Hz, the maximal mismatch was chosen as $m = 0.2$ (corresponding to a maximal loss in SNR$^2$ of 20%), while in the range $[300, 1500)$ Hz, the maximal mismatch was $m = 0.5$. Because of a bug in the script generating the sky-grids, the range $[290, 300)$ Hz was covered by frequency and spin-down steps corresponding to $m = 0.2$, whereas the sky-grids were constructed for $m = 0.5$. The distribution of actual mismatches in this frequency range will therefore be somewhat in between those of the low-frequency and high-frequency workunits.

It can be shown [46] that these relatively large mismatches give near-optimal sensitivity for a coherent search at fixed CPU power. Choosing finer grid spacings (i.e. a smaller mismatch) would require searching more grid points, thus reducing the maximal possible coherent integration time. A coarser search grid would allow longer integrations but at a larger average loss in SNR. Because of these two competing tendencies, the sensitivity as a function of mismatch $m$ has a maximum in the range $m \sim 0.25$–0.7, depending on the choice of false-dismissal rate from the grid mismatch. Full details of the parameter-space grid and workunit construction are given in [46]; a short summary follows.

#### 2. Search grid in frequency and frequency derivative

The step size in frequency was determined using the metric-based expression

$$df_j = \frac{2\sqrt{3m}}{\pi T_{\mathrm{span},j}}, \qquad (1)$$

so the frequency spacing depends on $T_{\mathrm{span},j}$ of the data segment $j$. For the low-frequency workunits ($f < 300$ Hz), this results in frequency steps in the range $df_j \in [3.43, 4.06]$ $\mu$Hz, while for high-frequency workunits $df_j \in [5.42, 6.41]$ $\mu$Hz.

The range of frequency derivatives $\dot{f}$ searched is defined in terms of the "spin-down age" $\tau \equiv -f/\dot{f}$, namely, $\tau \geq 1000$ years for low-frequency and $\tau \geq 10\,000$ years for high-frequency workunits. Neutron stars younger than the limited range of this search probably would have left a highly visible (Sedov phase) supernova remnant or a pulsar

wind nebula. Thus our search for unknown neutron stars targeted older objects, which also resulted in less computational cost. The search also covers a small "spin-up" corresponding to $\tau \leq -10\,000$ years for low-frequency and $\tau \leq -100\,000$ years for high-frequency workunits. So the actual ranges searched are $\dot{f} \in [-f/\tau, 0.1f/\tau]$. In $\dot{f}$ the grid points are spaced according to

$$d\dot{f}_j = \frac{12\sqrt{5m}}{\pi T_{\mathrm{span},j}^2},\tag{2}$$

resulting in resolutions $d\dot{f}_j \in [1.84, 2.59] \times 10^{-10}$ Hz/s for low-frequency workunits, and $d\dot{f}_j \in [2.91, 4.09] \times 10^{-10}$ Hz/s for high-frequency workunits, depending on the duration $T_{\mathrm{span},j}$ of different segments $j$.

### 3. Search grid in the sky parameters

The resolution of the search grid in the sky-directions depends both on the start time $t_j$ and duration $T_{\mathrm{span},j}$ of the segment, as well as on the frequency. The number of grid points on the sky scales as $\propto f^2$, and approximately as $\propto T_{\mathrm{span},j}^{2.4}$ for the range of $T_{\mathrm{span},j} \sim 30$–40 h used in this search. Contrary to the simple uniform spacings in $f$ and $\dot{f}$, the sky-grids are computed beforehand and shipped together with the workunits. In order to simplify the construction of the workunits and reduce the number of different input files to be sent, the sky-grids are fixed over a frequency range of 10 Hz, but differ for each data segment $j$. The sky-grids are computed at the higher end of each 10 Hz band, so they slightly "over-cover" the sky instead of being too coarse. The search covers a frequency range of 1450 Hz, and so there are 145 different sky-grids for each segment. To illustrate this, four of these sky-grids are shown in Fig. 3 corresponding to two different data segments at two distinct frequency bands.

To ensure that each workunit takes a similar amount of CPU time, the total number of template grid points of each workunit is chosen to be *approximately* constant for all workunits. However, in practice, this number (and hence the average CPU time) can vary by up to a factor of 2 due to discretization effects. The number of points in the sky-grids grows with frequency as $f^2$ and the number of points in the spin-down grid grows linearly with $f$. Thus, to keep the number of templates (and therefore the CPU time) approximately constant, the frequency range covered by each workunit decreases as $f^{-3}$. Hence for fixed $j$, $M(j, k)$ is roughly proportional to $k^3$.

### C. The output of a workunit

The result from completing one workunit on an Einstein@Home host computer is a ZIP-compressed ASCII text file containing the 13 000 candidate events with the largest values of the $\mathcal{F}$-statistic found over the parameter-space grid points analyzed by that workunit. Each line of the output file contains five columns: frequency (Hz), right ascension angle (radians), declination angle (radians), spin-down-rate (Hz/s), and $2\mathcal{F}$ (dimensionless). The frequency is the frequency at the solar system barycenter (SSB) at the instant of the first data point in the corresponding data segment.

The number 13 000 was decided in advance, when the workunits were first launched on the Einstein@Home project, which was about 1 yr before the post-processing pipeline was developed. The network bandwidth required to retain more than 13 000 candidates per workunit, and the storage space required to preserve them, would have exceeded the capacity of the Einstein@Home server and its
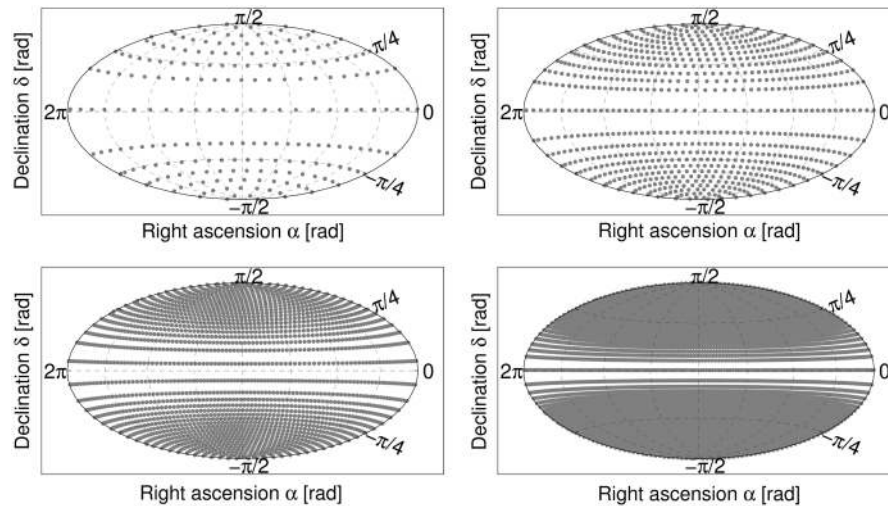


FIG. 3. Four different sky-grids in Hammer-Aitoff [50] projection. The top row is for frequency $f = 60$ Hz and the bottom row is for $f = 310$ Hz. The left column shows data segment $j = 1$ (from H1) with a spanned time of $T_{\mathrm{span},1} = 33.8$ h, while the right column shows data segment $j = 15$ (from L1) with a spanned time of $T_{\mathrm{span},15} = 39.8$ h. The grid points are spaced more closely for a longer spanned time, and for a higher frequency.
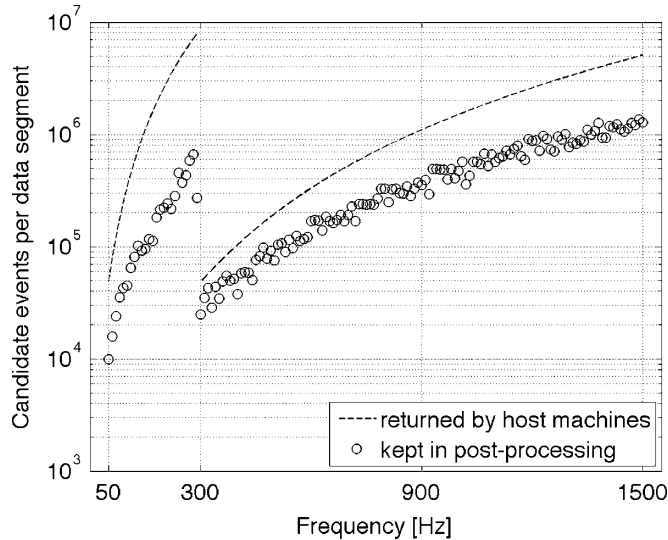
FIG. 4. The circles show the number of candidate events $\mathcal{E}_{\text{seg}}(k)$ retained per data segment and per 0.5 Hz frequency band in the post-processing in each 10 Hz band. The dashed curve represents the number of candidate events which are returned from volunteering hosts participating in Einstein@Home. The strange location of the point at 290 Hz is explained in Sec. III B 1.

internet connection. For frequency-band and data segment combinations with small numbers of workunits, for example, the $j = 1$ data set from 301.0 Hz to 301.5 Hz, almost all of the 13 000 candidate events are later used in the post-processing pipeline. However (as can be seen later in Fig. 4) for most frequency bands the post-processing pipeline only needed and used a fraction of the events that were returned.

Returning the "loudest" 13 000 candidate events effectively corresponds to a floating threshold on the value of the $\mathcal{F}$-statistic. This avoids large lists of candidate events being produced in regions of parameter space containing non-Gaussian noise, such as instrumental artifacts that were not removed *a priori* from the input data.

### D. Total computation time

The analysis of the entire 1450 Hz band computed $2\mathcal{F}$ values for 63 627 287 767 483 distinct points in the four-dimensional parameter space. The bulk of this processing was carried out on the Einstein@Home project between December 24, 2005 and June 30, 2006, and used approximately $4 \times 10^{11}$ CPU-seconds, or 13 000 CPU-years. In any given week, about 60 000 host machines were active. Thus, a typical host machine contributed about 40% of its potential CPU cycles to the project.

### IV. POST-PROCESSING

As shown previously in Fig. 2, after result files are returned to the Einstein@Home servers by project participants, further post-processing is carried out on those serv-

ers and on dedicated computing clusters. The goal of this post-processing pipeline is to identify consistent candidate events that appear in many of the 17 different data segments.

In this paper, a consistent (coincident) set of "candidate events" is called a "candidate." Candidate events from different data segments are considered coincident if they cluster closely together in the four-dimensional parameter space. A clustering method using a grid of "coincidence cells" will reliably detect strong CW signals, which would produce candidate events with closely matched parameters.

The post-processing pipeline operates in 0.5 Hz-wide frequency bands, and can be summarized in three steps. In *step one*, a subset of candidate events is chosen from those returned to the project server. This fixes the coincident false alarm probability. In *step two*, the frequency values of candidate events are shifted to the same fiducial time. In *step three*, a grid of cells is constructed in the four-dimensional parameter space, and each candidate event is assigned to a particular cell. Candidates (as defined above) are cells containing unusually large numbers of candidate events. In the following, the details involved in each step are described.

### A. Preparation and selection of candidate events

In the *first step* the individual result files are prepared for the later analysis by uncompressing them and keeping only a subset of the candidate events: from the $(j, k, \ell)$'th workunit only the $\mathcal{E}(j, k, \ell)$ candidate events with the largest values of $2\mathcal{F}$ are retained.

The number of these candidate events is chosen *a priori* to obtain a predetermined fixed false alarm probability in the later coincidence step. The false alarms should be approximately uniformly distributed among the workunits, since each workunit examines a similar number of independent grid points in parameter space. The number of candidate events is chosen so that, after doing the clustering, in a 0.5 Hz-wide frequency band the probability that one or more coincidence cells (in step three) has $\mathcal{C}_{\text{max}} = 7$ or more coincidences is $P_F = 0.001$. Thus, in the analysis of 2900 such frequency bands, in random noise one would expect to find only about three candidates with seven or more coincidences. (As explained later in Sec. IV F 1, this overall probability for the entire search is somewhat increased because the coincidence-cell grids are also shifted by half their width in 16 possible combinations).

In the notation introduced in the previous section, the number of candidate events kept from the $(j, k, \ell)$'th workunit takes the form

$$\mathcal{E}(j, k, \ell) = \frac{\mathcal{E}_{\text{seg}}(k)}{M(j, k)}, \tag{3}$$

where $\mathcal{E}_{\text{seg}}(k)$ is shown in Fig. 4. Because the individual workunits are constructed to use approximately the same

amount of CPU time, each workunit examines approximately the same number of templates in parameter space, so the same number of candidate events are retained from all workunits which have the same input data file and data segment. This implies that the number of candidate events that are kept per data segment $j$ and per frequency band is independent of the data segment $j$:

$$\sum_{\ell=1}^{M(j,k)} \mathcal{E}(j, k, \ell) = \mathcal{E}_{\text{seg}}(k). \quad (4)$$

Since the sky-grids are fixed in 10 Hz intervals, $\mathcal{E}_{\text{seg}}(k)$ takes the same value for all values of $k$ in the range of $20p + 1, \ldots, 20(p + 1)$ where $p$ labels the sky-grids by an integer in the range $p \in 0, \ldots, 144$.

It is illustrative to look at a specific case. For example consider the 0.5 Hz band covering [301.0, 301.5] Hz; this band is labeled by $k = 503$. As is shown in Fig. 4, in this band the post-processing pipeline retains $\mathcal{E}_{\text{seg}}(k = 503) = 24\,960$ candidate events from each of the 17 different 30-h data segments. The 30-h data segment from H1 with the *shortest* time span ($j = 1$) has approximately $4.3 \times 10^8$ templates divided among just $M(j = 1, k = 503) = 2$ workunits, so $12\,480$ candidate events are retained from each of these workunits. The 30-h data segment from L1 with the *longest* time span ($j = 15$) has approximately $1.7 \times 10^9$ templates divided among $M(j = 15, k = 503) = 7$ workunits, so 3565 candidate events are retained from each of these workunits. In the later stage of the post-processing, this ensures that each of the different data segments contributes equally to the probability of generating false alarms in the coincidence step.

### B. Number of cells in the post-processing coincidence grid

It is important to calculate the number of coincidence cells in the coincidence grid. Together with the desired false alarm probability, this determines the number of candidate events to retain in the post-processing pipeline.

The number of coincidence cells $N_{\text{cell}}(k)$ contained in each 0.5 Hz frequency band $k$ (used for doing the clustering in step three) is determined by the sizes of the cells. This is given by

$$N_{\text{cell}}(k) = \left(\frac{0.5\ \text{Hz}}{\Delta f}\right)\left(\frac{1.1f}{\tau \Delta \dot{f}}\right) \int_{-\pi/2}^{\pi/2} \frac{d\delta}{\Delta \delta(\delta)} \int_0^{2\pi} \frac{d\alpha}{\Delta \alpha(\delta)}, \quad (5)$$

where $\Delta f$ denotes the coincidence cell width in frequency, $\Delta \dot{f}$ denotes the width in spin-down, and $\Delta \alpha(\delta)$ and $\Delta \delta(\delta)$ denote the coincidence-cell widths in right ascension and declination (both of which vary with declination $\delta$). The choice of the coincidence-cell sizes will be explained in detail later when step three will be described.

### C. False alarm rate and the number of candidate events retained

The number of candidate events that must be retained is determined by the number of cells in the coincidence grid $N_{\text{cell}}(k)$ and by the desired probability of false alarm $P_{\text{F}}$ for false coincidence of candidate events from $\mathcal{C}_{\text{max}}$ or more data segments in each 0.5 Hz band. To relate these quantities, consider the case of random instrumental noise, in which the candidate events are distributed uniformly about the coincidence grid. Concentrate on a single 0.5 Hz band $k$, and consider the first of the $N_{\text{seg}} = 17$ data segments. A total of $\mathcal{E}_{\text{seg}}(k)$ candidate events must be distributed uniformly among $N_{\text{cell}}(k)$ coincidence cells. Each candidate event falls in a random coincidence cell, independent of the locations of the previous candidate events. The probability that the first candidate event falls in the first coincidence cell is $1/N_{\text{cell}}(k)$, and hence the probability that the first coincidence cell remains empty is $1 - 1/N_{\text{cell}}(k)$. If the remaining $\mathcal{E}_{\text{seg}}(k) - 1$ candidate events fall independently and at random into the coincidence cells, then this generates a binomial distribution, and the probability that the first coincidence cell contains no candidate events is

$$p_k(0) = \left(1 - \frac{1}{N_{\text{cell}}(k)}\right)^{\mathcal{E}_{\text{seg}}(k)}. \quad (6)$$

Since the first coincidence cell is equivalent to any other, the probability that the candidate events from the first data segment populate any given coincidence cell with one or more candidate events is thus given by

$$\epsilon(k) = 1 - p_k(0) = 1 - \left(1 - \frac{1}{N_{\text{cell}}(k)}\right)^{\mathcal{E}_{\text{seg}}(k)}. \quad (7)$$

In random noise, the candidate events produced by each different data segment are independent, so that the coincidence cells that are "marked" by one or more candidate events are also described by a (different) binomial distribution. Without loss of generality, again consider the first coincidence cell. The probability that it contains candidate events from $n$ distinct data segments is then given by

$$\binom{N_{\text{seg}}}{n} [\epsilon(k)]^n [1 - \epsilon(k)]^{N_{\text{seg}} - n}, \quad (8)$$

where $\binom{a}{b} = \frac{a!}{b!(a-b)!}$ is the binomial coefficient. Thus the probability per coincidence cell of finding $\mathcal{C}_{\text{max}}$ or more coincidences is given by

$$p_{\text{F}}(k) = \sum_{n=\mathcal{C}_{\text{max}}}^{N_{\text{seg}}} \binom{N_{\text{seg}}}{n} [\epsilon(k)]^n [1 - \epsilon(k)]^{N_{\text{seg}} - n}. \quad (9)$$

The probability $P_{\text{F}}$ that there are $\mathcal{C}_{\text{max}}$ or more coincidences in one or more of the $N_{\text{cell}}$ cells per 0.5 Hz band $k$ is

$$P_{\text{F}}(k) = 1 - [1 - p_{\text{F}}(k)]^{N_{\text{cell}}}. \quad (10)$$

If $p_{\text{F}}(k) \ll 1$ and $N_{\text{cell}} p_{\text{F}}(k) \ll 1$ then the false alarm

probability $P_F$ is approximately

$$P_F(k) \approx N_{\text{cell}} p_F(k). \qquad (11)$$

For the desired $P_F = 0.1\% = 10^{-3}$ with $\mathcal{C}_{\text{max}} = 7$ per 0.5 Hz band $k$, this equation is solved numerically to find $\mathcal{E}_{\text{seg}}(k)$. The results for $\mathcal{E}_{\text{seg}}(k)$ are shown in Fig. 4.

### D. Choice of false alarm probability and detection threshold

The goal of this work is to make a *confident detection*, not to set upper limits with the broadest possible coverage band. This is reflected in the choice of the *expected false alarm probability* and the choice of a *detection threshold*.

The detection threshold of 12 events was chosen because, as described in Sec. VII, the hardware injections are only "turned on" for 12 of the 17 data segments. A detection threshold of 12 events ensures that these simulated signals are properly detected by the post-processing pipeline.

The choice of false alarm probability ($P_F = 0.1\% = 10^{-3}$ per 0.5 Hz band to have coincidences in $\mathcal{C}_{\text{max}} = 7$ or more data segments) is a pragmatic choice, which leads to an extremely small false alarm rate at the detection threshold. For actual data, the probability of finding 7 or more coincidences in a given 0.5 Hz band can be somewhat larger than the target value of 0.1% because the candidate events are not uniformly distributed over the grid of coincidence cells and because (as described in Sec. IV F 1) 16 sets of coincidence cells are used for each 0.5 Hz band.

In random noise, the probability of reaching the detection threshold of 12 coincidences depends on the number of cells in the coincidence grid, which is a function of frequency. Some representative numbers are given in Table III; they vary from about $10^{-15}$ to $10^{-13}$ depending upon the 0.5 Hz band. The false alarm probabilities decrease very rapidly with increasing coincidence number. For example the probability of finding 14 or more coincidences in random noise varies from about $10^{-18}$ to about $10^{-20}$.

One might ask why we chose to specify a uniform false alarm probability, across all frequencies, of 0.1% for $\mathcal{C}_{\text{max}} = 7$, rather than directly specify a much lower false alarm probability at the detection threshold $\mathcal{C} = 12$. This was because we wanted the most significant coincident events due to noise alone to have $\mathcal{C}$ values a few less than our detection threshold, and we wanted *these* candidates to be uniformly distributed over frequency bands. Any detected signals could then be compared against fairly uniform fields of noise candidates in other frequency bands. If a uniform false alarm probability had been specified at the $\mathcal{C} = 12$ level, then the expected noise candidates with $\mathcal{C} \sim 7$ would *not* have been uniformly distributed over frequency, due to the differing numbers of coincidence cells in each frequency band.

The choice of detection threshold (high) and false alarm probability (low) sacrifices a small amount of sensitivity, but ensure that high numbers of coincidences are extremely improbable in random noise. A strong signal (say a factor of 1.5 above the upper curve in Fig. 9) would be expected to produce 15 or more coincidences in this detection pipeline. With the thresholds that we have adopted, this would stand out very strongly: the probability of having even one such event appear in coincidence in random noise is about $10^{-22}$ per 0.5 Hz band.

### E. Shifting candidate event frequencies to a fixed fiducial time

In the *second step* of the post-processing, the frequency value of each retained candidate event is shifted to the same fiducial time: the GPS start time of the earliest ($j = 4$) data segment, $t_{\text{fiducial}} = t_4 = 793\,555\,944$ s. This shifting is needed because a CW source with nonzero spin-down would produce candidate events with different apparent frequency values in each data segment. This step would shift these candidate events back to the same frequency value:

$$f(t_{\text{fiducial}}) = f(t_j) + [t_{\text{fiducial}} - t_j]\dot{f}, \qquad (12)$$

where $\dot{f}$ and $f(t_j)$ are the spin-down rate and frequency of a

TABLE III. False alarm probabilities $P_F$ in four different 0.5 Hz frequency bands labeled by the integer $k$. The frequency at the lower boundary of the 0.5 Hz band $k$ is denoted by $f_k$. The number of coincidence cells in the $k$'th half-Hz frequency band is denoted by $N_{\text{cell}}(k)$. The probability of finding 7 or more coincidences ($\mathcal{C} \geq 7$) in randomly distributed noise is fixed to be 0.1%. The probability of finding 12 or more coincidences (the detection threshold, $\mathcal{C} \geq 12$) in random noise varies over two orders of magnitude, from about $10^{-15}$ to $10^{-13}$. The probability of finding 14 or more coincidences ($\mathcal{C} \geq 14$) in random noise varies from about $10^{-18}$ to about $10^{-20}$.

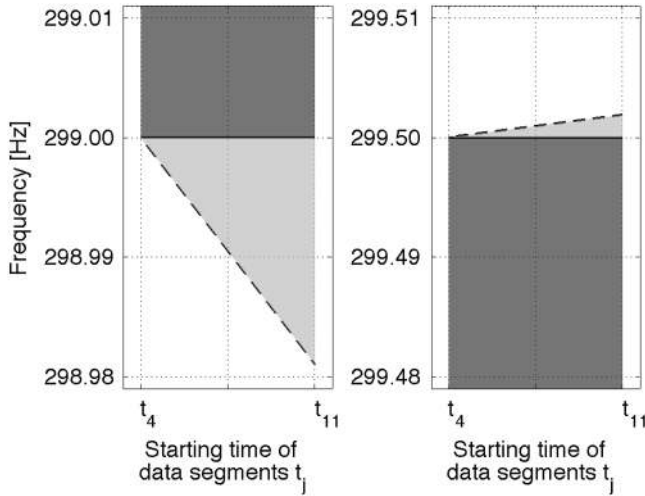| $f_k$ [Hz] | $k$ | $N_{\text{cell}}(k)$ | $P_F(\mathcal{C} \geq 7)$ | $P_F(\mathcal{C} \geq 12)$ | $P_F(\mathcal{C} \geq 14)$ |
|---|---|---|---|---|---|
| 50.0 | 1 | 734 500 | $10^{-3}$ | $1.5 \times 10^{-13}$ | $3.0 \times 10^{-18}$ |
| 290.0 | 481 | 35 198 800 | $10^{-3}$ | $8.7 \times 10^{-15}$ | $5.7 \times 10^{-20}$ |
| 301.0 | 503 | 2 161 284 | $10^{-3}$ | $6.7 \times 10^{-14}$ | $9.9 \times 10^{-19}$ |
| 1499.5 | 2900 | 233 345 981 | $10^{-3}$ | $2.2 \times 10^{-15}$ | $8.4 \times 10^{-21}$ |

FIG. 5. Additional "wings" of candidate events at the boundaries of each 0.5 Hz frequency band must be included in the coincidence analysis stage of the post-processing. This is because spin-down can carry a source below this half-Hz band, and spin-up can carry it above the band. To illustrate this, the frequency band $k = 498$ (covering $[299, 299.5)$ Hz) is (partly) shown by the dark-gray shaded area. The dashed sloped lines show the boundaries of the small additional regions (light gray) in frequency space whose candidate events must also be considered in the post-processing. Because the allowed spin-up range is 10 times smaller than the allowed spin-down range, the upper boundary has a slope 10 times smaller than the lower boundary.

candidate event reported by the search code in the result file, and $t_j$ is the time stamp of the first datum in the data segment. At the end of the second step, all candidate events for the 0.5 Hz band are merged into one file.

These candidate events are collected from a frequency interval that is slightly wider than 0.5 Hz. To see why this is necessary, consider a potential source whose frequency in the first data segment ($j = 4$) is at the lower (or upper) boundary of the 0.5 Hz interval. If the source has the minimum (or maximum) allowed value of $\dot{f}$, then in the *later* data segments it moves into, or is recorded in, the previous (or next) 0.5 Hz band. This effect is most apparent for the last $j = 11$ data segment, as illustrated in Fig. 5. So in collecting the candidate events for analysis of a given 0.5 Hz band, the frequency range is enlarged slightly for events coming from later and later data segments, as shown in Fig. 5.

### F. Search for coincident candidate events

The *third step* and final stage of the post-processing is to search for parameter-space coincidence among the candidate events. If a CW source is present that is strong enough to be confidently detected, then it would produce large $\mathcal{F}$-statistic values (i.e. candidate events) in many or all of the 17 data segments. In addition, the values of the frequency at the fiducial time $f(t_{\text{fiducial}})$, sky position (given by right ascension $\alpha$ and declination $\delta$), and spin-down $\dot{f}$

for these candidate events would agree among all data segments (within some coincidence "window" or "cell").

#### *1. Coincidence search algorithm*

To find coincidences, a grid of cells is constructed in four-dimensional parameter space, as described previously. This analysis uses rectangular cells in the coordinates $(f, \dot{f}, \alpha \cos\delta, \delta)$. The dimensions of the cells are adapted to the parameter-space search. Each candidate event is assigned to a particular cell. In cases where two or more candidate events from the same data segment $j$ fall into the same cell, only the candidate event having the largest value of $2\mathcal{F}$ is retained in the cell. Then the number of candidate events per cell coming from distinct data segments is counted, to identify cells with more coincidences than would be expected by random chance.

The search for coincident cells containing large numbers of candidate events is done with an efficient code that uses linked-list data structures, $O(N \log N)$ sort algorithms, and $O(\log N)$ bisection search algorithms. To ensure that candidate events located on opposite sides of a cell border are not missed, the entire cell coincidence grid is shifted by half a cell width in all possible $2^4 = 16$ combinations of the four parameter-space dimensions. Hence 16 different coincidence-cell grids are used in the analysis.

The cells in the coincidence grid are constructed to be as small as possible to reduce the probability of coincidences due to false alarms. However, since each of the 17 different data segments uses a different parameter-space grid, the coincidence cells must be chosen to be large enough that the candidate events from a CW source (which would appear at slightly different points in parameter space in each of the 17 data segments) would still fall in the same coincidence cell.

#### *2. Frequency and spin-down coincidence windows*

In frequency, the spacing of the parameter-space grid is largest for the data segment with the smallest value of $T_{\text{span},j}$, which is the first data segment $j = 1$. At first, this would appear to be the correct size $\Delta f$ for the coincidence cell in the frequency direction. However since the frequency of a candidate event must be shifted to a fixed fiducial time according to its spin-down value, and since that reported spin-down value might not be more accurate than the $\dot{f}$ spacing, the size of the coincidence cell must be somewhat larger to accommodate the effects of this discretization error in $\dot{f}$. The coincidence window in the frequency direction is thus determined by

$$\Delta f = \max_j (df_j + \Delta t d\dot{f}_j), \tag{13}$$

where the maximization over $j$ selects the data segment with the smallest $T_{\text{span},j}$ (which is $j = 1$) and

$$\Delta t = |\max_j t_j - \min_j t_j| = t_{11} - t_4 = 1\,997\,256 \text{ s} \quad (14)$$

is the total time span between the latest and earliest data segments. For safety, e.g. against noise fluctuations that could shift a candidate peak, $\Delta f$ has been increased by a further 40% below 300 Hz, so that the width of the coincidence cell in frequency is $\Delta f = 0.77 \times 10^{-3}$ Hz, and by 30% above 300 Hz, so that $\Delta f = 1.08 \times 10^{-3}$ Hz.

For the spin-down parameter, the size of the coincidence cell is given by the largest $d\dot{f}_j$ spacing in the parameter-space grid, which is also determined by the smallest value of $T_{\text{span},j}$. For safety this is also increased by 40% below 300 Hz giving $\Delta \dot{f} = 3.7 \times 10^{-10}$ Hz s$^{-1}$, and by 30% above 300 Hz giving $\Delta \dot{f} = 5.18 \times 10^{-10}$ Hz s$^{-1}$.

### 3. Coincidence windows in apparent sky position

Determining the optimal size for the coincidence cells in the sky coordinate directions is more difficult. Each of the 17 different data segments uses a different sky-grid, as illustrated in Fig. 3. Ideally the size of the coincidence cells in these sky directions must be just large enough to enclose one parameter-space grid point from each of the 17 different sky-grids. A practical solution to determine the coincidence cells, which is close to optimal, makes use of an observation concerning the parameter-space metric that first appears in [25].

To understand the properties of the parameter-space metric, first consider the relative orders-of-magnitude of the different frequency modulation effects. The fractional Doppler shift due to the Earth's annual orbital motion about the Sun has magnitude $|\mathbf{v}_{\text{orbital}}|/c = 10^{-4}$ and the fractional Doppler shift due to the detector's daily motion about the Earth rotation axis has magnitude $|\mathbf{v}_{\text{rotational}}|/c = 10^{-6}$. For the $T_{\text{span}} \approx 40$ h period of a single coherent integration, one can approximate the motion of the Earth's center of mass as motion with constant acceleration (along a parabolic trajectory) rather than as circular motion. The neglected term in the fractional Doppler shift has magnitude $|\ddot{\mathbf{v}}_{\text{orbital}}|T_{\text{span}}^2/2c \approx |\mathbf{v}_{\text{orbital}}||\boldsymbol{\omega}|^2 T_{\text{span}}^2/2c \approx 4 \times 10^{-8}$, where $|\boldsymbol{\omega}| = 2\pi/\text{year}$ is the magnitude of the Earth's orbital angular velocity about the sun. This term is a factor of 25 smaller than $|\mathbf{v}_{\text{rotational}}|/c$ and hence can be neglected. With this approximation, the orbital motion of the Earth is simply responsible for an apparent shift in the frequency $f$ and spin-down rate $\dot{f}$ of a source: the effects of the Earth's center of mass motion are degenerate with a shift in frequency and spin-down. So the Earth's orbital motion causes a signal only to shift to a different template in $f$ and $\dot{f}$; the Earth's rotation has a period of one sidereal day and its effects cannot be modeled by a shift in $f$ or $\dot{f}$. Note that terms are neglected *only* in determining where to place search-grid points in parameter space (because the neglected terms have an insignificant effect on where the grid points are placed). The actual filtering of the data uses

"exact" barycentering routines (total timing errors below 3 $\mu$s).

The search grid in parameter space is a Cartesian product of a frequency grid, a spin-down grid, and a two-dimensional sky-grid. Since the search maximizes the detection statistic over frequency and spin-down, the metric used to place grid points on the sky [46] may be obtained by minimizing the four-dimensional metric over frequency and spin-down and projecting it into the sky directions. As shown in the previous paragraph, this two-dimensional projected sky metric is well-approximated by assuming that the Earth is spinning about its axis but has its center of mass at rest. If the coherent integration period is an integer number of days, then by symmetry the two-dimensional metric on the sky is invariant under rotation about Earth's axis ($\partial_\alpha$ is a Killing vector). This is still an approximate symmetry for the search described here, since the coherent integration period and $T_{\text{span}}$ are longer than the rotation period (one day).

One can easily find coordinates in which this approximate sky metric (the four-dimensional metric, minimized over frequency and spin-down and projected onto the sky directions) is proportional to diag(1, 1). These new sky-coordinates are obtained by perpendicular projection of a point on the two-sphere (celestial sphere) vertically down into the unit radius disk that lies in the equatorial plane. If $\hat{\mathbf{n}}$ denotes a unit vector pointing from the SSB to the source the new coordinates are the components of $\hat{\mathbf{n}}$ in the equatorial plane: $n_x = \cos\delta \cos\alpha$, $n_y = \cos\delta \sin\alpha$. Points which are equally spaced in these coordinates correspond to equal spacing in Doppler shift, since source Doppler shift due to the Earth's rotation is just proportional to the component of the source direction vector in the equatorial plane. It then follows from rotational invariance that (with these approximations) the projected sky metric in these coordinates is proportional to diag(1, 1) [47]. The effect may be immediately seen in Fig. 6: the grid of "equally spaced" points forms a (roughly) uniform square grid on the unit radius disk in the equatorial plane. Computing the Jacobian of the coordinate transformation shows that in the original coordinates $(\alpha, \delta)$ the coordinate-space density of grid points should be proportional to $|\cos\delta \sin\delta| = |\sin(2\delta)|$.

This simple behavior of the projected sky metric guides the construction of the coincidence windows in the sky directions. Define polar coordinates $(r, \alpha)$ on the unit radius disk in the equatorial plane by $r = \cos\delta$. The coordinate boundaries of uniformly distributed coincidence cells containing a single parameter grid point are then given by $r\,d\alpha = dr = \text{const}$. When written in terms of the original coordinates this becomes

$$\cos(\delta)d\alpha = |\sin(\delta)|d\delta = \text{const}. \quad (15)$$

This is not directly useful, because it is singular as $\delta \to 0$, but suggests a coincidence-window size which varies with
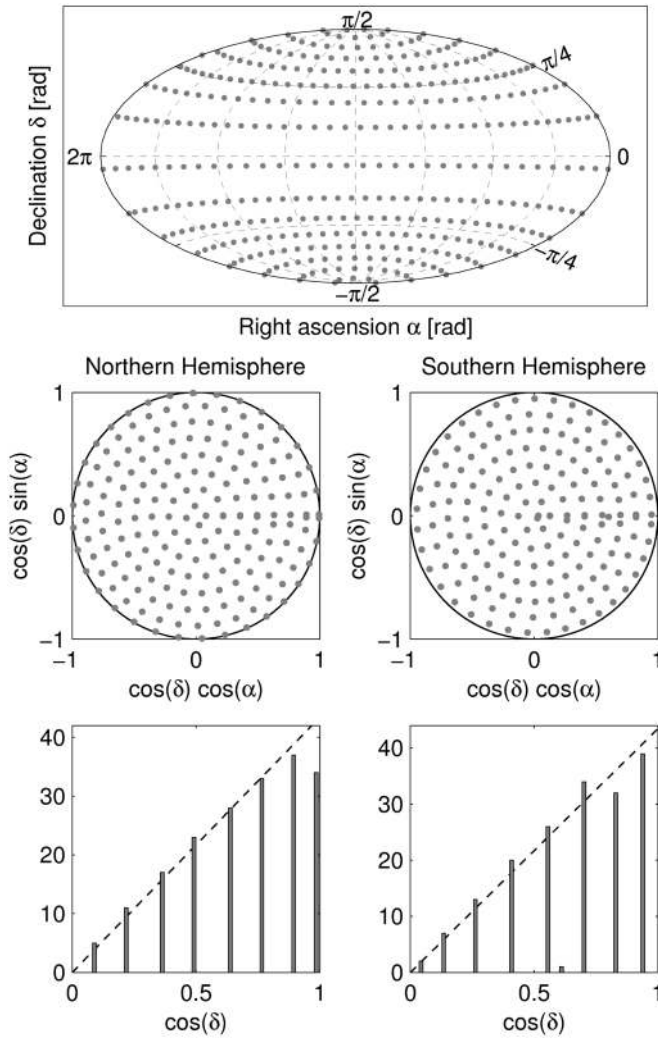
FIG. 6. Example sky-grid and its projection onto the equatorial plane. This sky-grid corresponds to the data segment $j = 7$ used in the frequency range from 60 to 70 Hz. The top plot shows a Hammer-Aitoff projection of the sky-grid. The middle plots show the projection of the sky-grid points in the northern hemisphere (left column) and in the southern hemisphere (right column) onto the equatorial plane. The bottom plots show histograms of $\cos(\delta)$ and the dashed line represents a linear fit to the distribution showing its uniformity.

declination according to the model

$$\Delta\alpha(\delta) = \Delta\alpha(0)/\cos(\delta),$$

$$\Delta\delta(\delta) = \begin{cases} \Delta\delta(0) & \text{if } |\delta| < \delta_c, \\ \Delta\alpha(0)/|\sin(|\delta| - \kappa\Delta\alpha(0))| & \text{if } |\delta| \geq \delta_c. \end{cases}$$
$$(16)$$

To ensure continuity at $\delta = \delta_c$, the transition point $\delta_c$ is defined by the condition $\Delta\alpha(0)/|\sin(|\delta_c| - \kappa\Delta\alpha(0))| = \Delta\delta(0)$. $\kappa$ is a tuning parameter of order unity, described below. An example of this coincidence-window model is shown in Fig. 7.
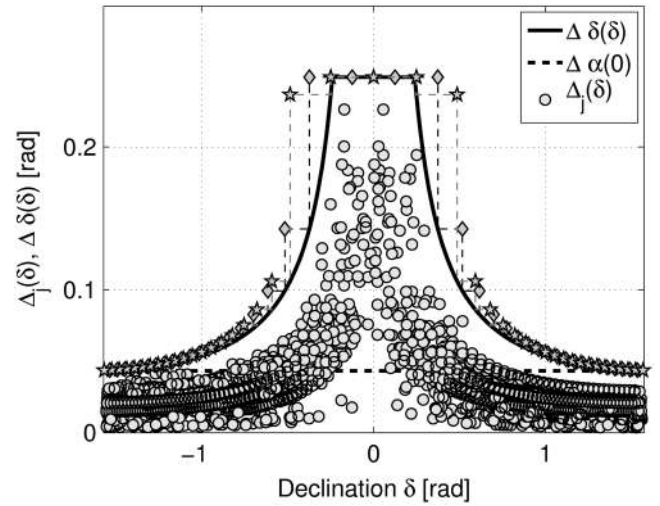


FIG. 7. The sky coincidence-window model for the frequency band from 510–520 Hz. The horizontal axis shows the declination $\delta$ in radians. On the vertical axis, the circles labeled $\Delta_j(\delta)$ correspond to the maximum distance in radians to neighboring $\delta$-points on either side. The solid curve shows the declination coincidence-window model $\Delta\delta(\delta)$ with the parameters $\Delta\delta(0) = 0.2489$, $\Delta\alpha(0) = 0.0433$, and $\kappa = 1.5$ used in this frequency band. It lies just above the largest declination separations shown. The stars denote the borders of the declination coincidence cell-grid and the diamonds represent the borders of the shifted declination coincidence cell-grid.

For each of the 145 different 10 Hz bands, the window size is determined by the three constants $\Delta\alpha(0)$, $\Delta\delta(0)$, and $\kappa$. For each sky-grid $p$ these values are directly determined from the sky-grids used in the search as follows. For each 10 Hz frequency band the maximum distances between adjacent declination points to either side are calculated for each of the 17 sky-grids as a function of declination $\delta$. In this way, 17 different overlaying curves $\Delta_j(\delta)$ (one per data segment) are obtained. These are indicated by the circles in Fig. 7 for a representative 510–520 Hz frequency band as illustration. Then the parameter $\Delta\delta(0)$ is obtained by considering the maximum separation to either side between all neighboring declination grid points *and* between the 17 different sky-grids, increased by a 10% safety factor as

$$\Delta\delta(0) = 1.1 \max_{j,\delta}\{\Delta_j(\delta)\}. \qquad (17)$$

The largest separations near the poles ($1.4 < |\delta| \leq \pi/2$) are then found and increased by a safety factor of 20% to determine the parameter $\Delta\alpha(0)$ via

$$\Delta\alpha(0) = 1.2 \max_{j,\delta}\{\Delta_j(|\delta| > 1.4)\}. \qquad (18)$$

Finally, the parameter $\kappa$ was chosen by visually examining diagrams similar to Fig. 7 for all 145 of the 10 Hz bands. A $\kappa$ value of 1.5 was found to be sufficient in most cases, while some bands required somewhat higher or lower
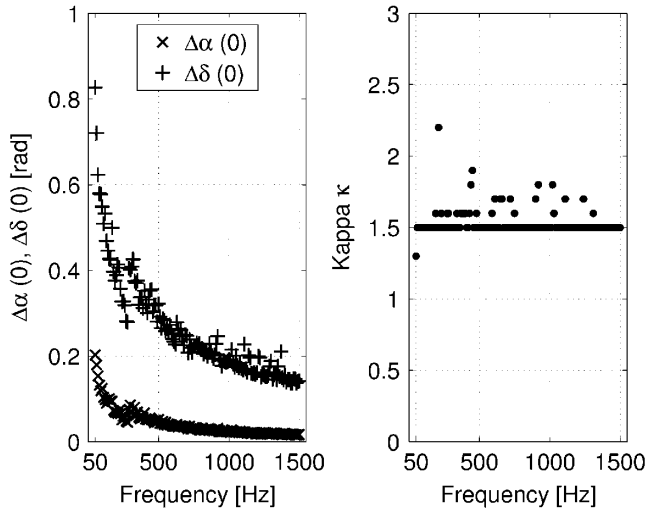
FIG. 8.   The parameters $\Delta\alpha(0)$, $\Delta\delta(0)$, and $\kappa$ of the sky coincidence-window model as a function of the 10 Hz frequency band.

values. For each triple of sky-coincidence parameters, tests were then performed to check that each sky-cell contained at least one sky-point from each data segment. In Fig. 7 the complete declination coincidence-window model given by Eq. (16) is represented by the solid black curve.

The three parameters for all sky-grids as a function of frequency are shown in Fig. 8. As stated above, the sky-grids are constant for 10 Hz-wide steps in frequency, and so these parameters vary with the same step size.

## G. Output of the coincidence search and significance of a candidate

The output of the post-processing pipeline is a list of the most coincident candidates. In each frequency band of coincidence-window width $\Delta f$, the coincidence cell containing the largest number of candidate events is found. Thus for each frequency band the pipeline finds the most coincident candidate maximized over the sky and over the spin-down parameter range. The pipeline outputs the average frequency of the coincidence cell, the average sky position and spin-down of the candidate events, the number of candidate events in the coincidence cell, and the significance of the candidate.

The "significance" of a candidate was first introduced in [27]. A candidate, consisting of the candidate events $1, \ldots, Q$, has significance

$$\mathcal{S}(1, \ldots, Q) = \sum_{q=1}^{Q} (\mathcal{F}_q - \ln(1 + \mathcal{F}_q)), \qquad (19)$$

where $Q \leq 17$ is the number of candidate events in the same coincidence cell. To understand the meaning of the significance, consider the case of pure Gaussian noise with no signal present. In this case the values of $2\mathcal{F}$ have a central $\chi^2$ distribution with 4 degrees of freedom. The

corresponding probability density function $p_0$ of $2\mathcal{F}$ is given by

$$p_0(2\mathcal{F}) = \frac{\mathcal{F}}{2} e^{-\mathcal{F}}. \qquad (20)$$

The false alarm probability $P_0$ that $2\mathcal{F}$ exceeds a certain threshold $2\mathcal{F}_0$ when there is no signal present has the form

$$P_0(2\mathcal{F}_0) = (1 + \mathcal{F}_0)e^{-\mathcal{F}_0}. \qquad (21)$$

The joint false alarm probability of candidate events $1, \ldots, Q$ can be written as $\prod_{q=1}^{Q} P_0(2\mathcal{F}_q)$. Therefore, in this analysis candidates are ranked according to

$$1 - \prod_{q=1}^{Q} P_0(2\mathcal{F}_q) = 1 - e^{-\mathcal{S}}, \qquad (22)$$

where $\mathcal{S} = \sum_{q=1}^{Q} -\ln P_0(2\mathcal{F}_q)$ is exactly the significance defined in Eq. (19). Thus ranking candidates by $\mathcal{S}$ is equivalent to ranking them by false alarm probability: candidates with large positive significance would not be expected to occur in Gaussian random noise. As will be described later in Sec. VIII the significance is used to rank equally coincident candidates within the same narrow frequency band. In such cases only the candidate with the largest significance is retained.

The post-processing pipeline has been validated by internal testing, and also using simulated CW signals created via so-called "software injections" [48]. In addition, Sec. VII presents realistic end-to-end testing of the analysis pipeline using "hardware injections," where simulated isolated-pulsar signals are physically added into the interferometer control systems to produce instrumental signals that are indistinguishable from those that would be produced by physical sources of gravitational waves.

## V. ESTIMATED SENSITIVITY

The sensitivity of the search is estimated using Monte Carlo methods for a population of simulated sources. The goal is to find the strain amplitude $h_0$ at which 10%, 50%, or 90% of sources uniformly populated over the sky and in their "nuisance parameters" (described below) would be detected. As previously discussed, the false alarm probability (of getting 7 or more coincidences in a 0.5 Hz frequency band) is of order $10^{-3}$. In this analysis, "detectable" means "produces coincident events in 12 or more distinct data segments." The false alarm probability for obtaining 12 or more coincidences in a 0.5 Hz band is of order $10^{-14}$, making it extremely unlikely for candidate events from random noise to show up consistently in 12 or more segments of data. This is therefore an estimate of the signal strength required for high-confidence detection. The pipeline developed for this purpose operates in 0.5 Hz frequency bands and consists of testing a large number of

distinct simulated sources (trials) to see if they are detectable. A "trial" denotes a single simulated source which is, or is not, detected.

### A. Iteration method

For every trial, source parameters are randomly chosen independent of the previous trial, except for the intrinsic amplitude $h_0$. For the first trial $h_0$ is set to a starting value $30\sqrt{S_h/30 \text{ hours}}$. The rule for varying $h_0$ depends upon the last $N_{\text{last}}$ trials, where $N_{\text{last}}^{10\%} = 100$, $N_{\text{last}}^{50\%} = 20$, and $N_{\text{last}}^{90\%} = 100$. In the past $N_{\text{last}}$ trials, if more than 10%, 50%, or 90% of simulated sources have been detected then $h_0$ is *decreased* by $0.25h_0/n_{\text{trial}}$ for the following trial, where $n_{\text{trial}}$ is an integer in the range $0 \leq n_{\text{trial}} \leq 1000$ that is incremented with each additional trial. On the other hand, if less than 10%, 50%, 90% of simulated sources have been detected then $h_0$ is *increased* by $0.25h_0/n_{\text{trial}}$ for the next trial. This process is followed until $h_0$ has converged to a stationary range after 1000 trials. Then the median of $h_0$ is found using the $h_0$-values starting from the trial in which the desired detection level has been reached the first time during the $N_{\text{last}}$ trials. The following describes the pipeline for a single trial.

### B. Population of simulated sources

For each trial, a random source frequency is chosen from a uniform distribution in a given 0.5 Hz frequency band and a frequency derivative is drawn from a uniform distribution in the range covered by the Einstein@Home search. A sky position is chosen from a uniform distribution on the celestial sphere, and random values are chosen for the pulsar "nuisance parameters." These are the inclination parameter $\cos(\iota)$, initial phase $\Phi_0$, and polarization angle $\psi$ as defined in [19], and are all drawn from the appropriate uniform distributions.

### C. Determination of $2\mathcal{F}$ values for a single simulated source

The noise floors of the different SFTs are estimated at the source's frequency intervals using a running median with a window of $\pm 25$ frequency bins. Figure 1 showed the average of these for the data segments used from each instrument.

Then for each set of parameters the detection statistic $2\mathcal{F}$ is estimated using a semianalytic method. From the estimated noise floor at the simulated source's frequency and given the other source parameters, the expectation value of the $\mathcal{F}$-statistic is calculated analytically as given in [19]. A random number is then drawn from a noncentral $\chi^2$ distribution with 4 degrees of freedom and with the corresponding mean value.

These random numbers, drawn from the appropriate distribution of $2\mathcal{F}$ values, would be sufficient to determine

the sensitivity of the search, if the template grid in parameter space were very closely spaced, so that the template bank always contained at least one waveform that was a very close match to the putative signal. However the grid in parameter space used in this search is quite "coarse," corresponding to a mismatch of 20% below 290 Hz and 50% above 300 Hz, so that the $2\mathcal{F}$ value that would be returned by the search might be significantly lower than the value drawn from the distribution above. To account for this effect, the sensitivity prediction considers the mismatch between the parameters of the simulated signals (determined by a random number generator) and the template grid points of the search (fixed as described earlier). For each simulated source, the search grid point that is nearest in the sense of the metric is located. Then, using the parameter-space metric, the mismatch between the simulated signal and the closest search template is computed. This gives the fractional amount by which the $2\mathcal{F}$ value is reduced.

From this ensemble of $2\mathcal{F}$ values, one can determine the number of coincidences that would be produced by each simulated source. As previously described, the postprocessing sets an effective lower threshold on the $\mathcal{F}$-statistic of the retained candidate events. For each simulated source, these thresholds are determined by examining the exact workunits that would have contained the corresponding signal. Then the number of data segments for which the estimated $2\mathcal{F}$ values are above threshold is counted. If the $2\mathcal{F}$ values are above threshold in 12 or more of the 17 data segments, the simulated source is labeled as "detected," else it is labeled as "undetected."

### D. Search sensitivity, estimated errors, and comparison with the expected sensitivity

Shown in Fig. 9 are the resulting search sensitivity curves as functions of frequency. Each data point on the plot denotes the results of 1000 independent trials. These show the values of $h_0$ as defined in [19] such that 10%, 50%, and 90% of simulated sources produce 12 or more coincidences in the post-processing pipeline. The dominant sources of error in this sensitivity curve are uncertainties in the magnitudes of the LIGO detector response functions (calibration errors). Details of these frequency-dependent uncertainties may be found in Ref. [41]. The uncertainties are typically of order 5% (L1) and 8% (H1) in the frequency band from 50–1500 Hz, and are always less than 10%. Systematic errors, which arise because of the finite number of Monte Carlo trials and similar effects, are less than $\pm 2\%$. These can be added in quadrature to the uncertainties given in [41] to obtain frequency-dependent error bounds in the sensitivity curve. The resulting error in this sensitivity plot is below 10% at all frequencies.

The behavior of the curves shown in Fig. 9 essentially reflect the instrument noise given in Fig. 1. One may fit the curves obtained in Fig. 9 to the strain noise power spectral
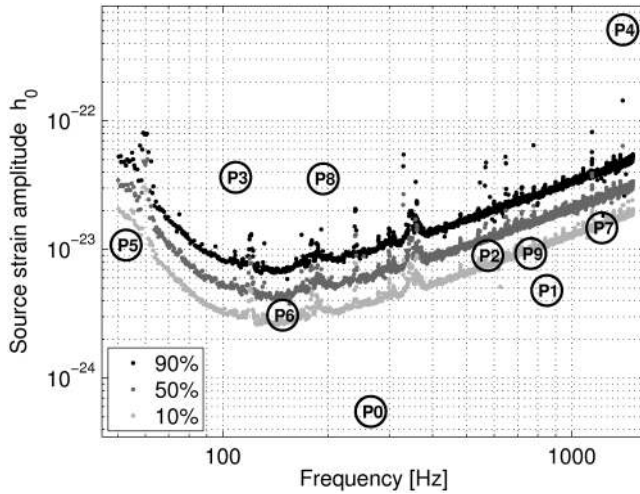
FIG. 9. Estimated sensitivity of the Einstein@Home search for isolated CW sources in the LIGO S4 data set. The set of three curves shows the source strain amplitudes $h_0$ at which 10% (bottom), 50% (middle), and 90% (top) of simulated sources would be confidently detected by the Einstein@Home pipeline (appearing in coincidence in 12 or more of the 17 data segments). The centers of the circles labeled P0 to P9 give the strain amplitudes of the S4 hardware injections as listed in Table IV. Based on this curve, one would expect that the simulated signals P3, P4, and P8 could be confidently detected, and that P0, P1, and P5 would not be detected.

density $S_h(f)$ and then describe the three sensitivity curves in Fig. 9 by

$$h_0^{\mathcal{D}}(f) \approx R_{\mathcal{D}} \sqrt{\frac{S_h(f)}{30 \text{ hours}}}, \qquad (23)$$

where the prefactors $R_{\mathcal{D}}$ for different detection probabilities $\mathcal{D} = 90\%$, 50%, and 10% are well fit below 300 Hz by $R_{90\%} = 31.8$, $R_{50\%} = 20.1$, and $R_{10\%} = 12.6$, above 300 Hz by $R_{90\%} = 33.2$, $R_{50\%} = 21.0$, and $R_{10\%} = 12.9$.

Some other published CW searches were done at 95% detection confidence. For comparison in the next section, the sensitivity of *this* search at that confidence is $R_{95\%} = 36.2$ below 300 Hz and $R_{95\%} = 37.9$ above 300 Hz. The iteration method previously described used $N_{\text{last}}^{95\%} = 200$.

### E. Comparison with sensitivity of other search and upper limit methods

The methods used here would be expected to yield very high confidence if a strong signal were present. It is interesting to compare the sensitivity of this detection method with the sensitivity of CW upper limits such as that of Ref. [27]. The sensitivity of the high-confidence detection method used here is well-described by Eq. (23). The same equation describes the results of the S2 $\mathcal{F}$-statistic loudest-event upper limit analysis [27], but in that work the 95% detection confidence curve has a prefactor $R_{95\%} = 29.5$. It is useful to understand the source of this apparent differ-

ence in strain sensitivity (a factor of $37.9/29.5 = 1.28$). There are three main contributors to this difference.

The most significant difference between the two analyses is the spacing of the search-grid templates. In this search, the templates are significantly farther apart (worst-case 50% loss of signal-to-noise ratio, or expected $2\mathcal{F}$) than in [27], where the worst-case mismatch was negligible. The effect of employing different mismatches has been studied by running the sensitivity-estimation pipeline using simulated sources only at the template grid points, and reduces $R_{95\%}$ in Eq. (23) by a factor of 1.17.

Another difference between the two analyses is the detection criteria. In this work, detection requires a signal to produce 12 or more coincidences between the 17 different data segments. This corresponds to a false alarm probability (in Gaussian noise) of order $10^{-14}$ per 0.5 Hz frequency band. This is different from [27], where simulated signals are compared against the loudest candidate found (largest $2\mathcal{F}$). An equivalent detection criterion for *this* work would be to compare the simulated signals against the loudest candidates (per 0.5 Hz band). These typically had 7 or 8 coincidences, corresponding to a Gaussian noise false alarm probability of order $10^{-3}$ and $10^{-5}$, respectively. To estimate the effect on the sensitivity, the sensitivity-estimation pipeline was rerun, but now requiring the signal to exceed the $2\mathcal{F}$-thresholds in only 7 of the 17 data segments. This reduced $R_{95\%}$ in Eq. (23) by an additional factor of 1.14.

The least important difference between the two analyses is the effective threshold on the $\mathcal{F}$-statistic. As explained in Secs. IV A and IV C, only a subset of candidate events with the largest values of $2\mathcal{F}$ are retained in the post-processing, fixing the false alarm probability. The smallest $2\mathcal{F}$-value on this list is typically around 28 or slightly higher. In [27] a fixed threshold of $2\mathcal{F} = 20$ has been used. Then, events with a combined significance below $\mathcal{S} = 64.4$ [see Eq. (19)] were also dropped. While it is difficult to compare these two criteria, they seem to be fairly close.

Taken together, the differences in grid spacing and detection thresholds are responsible for, and explain, the sensitivity difference in the two analyses (a factor of $1.17 \times 1.14 = 1.33 \approx 1.28$).

### VI. VETOING OF INSTRUMENTAL LINE ARTIFACTS

When the instrument data was prepared and cleaned, narrow-band instrumental line features of known origin were removed, as previously described in Sec. II. However, the data also contained stationary instrumental line features that were not understood, or were poorly understood. These features were *not* removed from the input data for the search. As a consequence, the output from the post-processing pipeline contains instrumental artifacts that in some respects mimic CW signals. But these

artifacts tend to cluster in certain regions of parameter space, and in many cases they can be automatically identified and vetoed. In previous incoherent searches for CW sources in LIGO data [32] the S-veto method has been employed, which excludes the regions of parameter space where there is little or no frequency modulation from the Earth's motion, leading to a relatively stationary detected frequency. This cannot directly be applied to a coherent matched-filtering search using the $\mathcal{F}$-statistic. Thus the method used here will be similar, but arises from a conceptually different approach that is appropriate for an $\mathcal{F}$-statistic search.

### A. Parameter-space locations of instrumental lines

For a coherent observation of 30 h the parameter-space regions where instrumental lines tend to appear are determined by global-correlation hypersurfaces of the $\mathcal{F}$-statistic [49] in parameter space. The locations of these instrumental line candidate events are described by

$$\dot{f} + \frac{(\boldsymbol{\omega} \times \mathbf{v}_{av}) \cdot \hat{\mathbf{n}}}{c} f(t_{\text{fiducial}}) = 0, \qquad (24)$$

where $c$ denotes the speed of light, $\hat{\mathbf{n}}$ is a unit vector pointing to the source's sky-location in the SSB frame and relates to the equatorial coordinates $\alpha$ and $\delta$ by $\hat{\mathbf{n}} = (\cos\delta \cos\alpha, \cos\delta \sin\alpha, \sin\delta)$, $\boldsymbol{\omega}$ is the angular velocity vector of the Earth as it orbits around the Sun ($|\boldsymbol{\omega}| \approx 2\pi/\text{year}$), and $\mathbf{v}_{av}$ is the average velocity of the Earth ($|\mathbf{v}_{av}| \approx 9.9 \times 10^{-5}c$). This equation can also be understood on simple physical grounds. The l.h.s. of Eq. (24) is the rate of change of detector output frequency, for a source whose SSB frequency and spin-down are $f$ and $\dot{f}$. An instrumental line, which has fixed detector frequency, mimics such a source when the l.h.s. vanishes.

The potential CW sources whose locations in parameter space are consistent with Eq. (24) will not produce a modulation pattern that would distinguish them from an instrumental line. As the resolution in parameter space is finite, the post-processing analysis eliminates (vetoes) candidates that satisfy the condition

$$\left| \dot{f} + \frac{(\boldsymbol{\omega} \times \mathbf{v}_{av}) \cdot \hat{\mathbf{n}}}{c} f(t_{\text{fiducial}}) \right| < \epsilon, \qquad (25)$$

where the parameter $\epsilon > 0$ provides the tolerance needed due to the parameter-space gridding. This tolerance parameter can be understood as

$$\epsilon = \frac{\Delta f}{\Delta T} N_{\text{cell}}, \qquad (26)$$

where $\Delta f$ denotes width in frequency (corresponding to the coincidence-cell width in the post-processing) up to which candidate events can be resolved during the characteristic length of time $\Delta T$, and $N_{\text{cell}}$ represents the size of the vetoed or rejected region, measured in coincidence cells. In this analysis $\Delta T = 2\,122\,735$ s ($\approx 24$ days) is

the total time interval spanned by the data

$$\Delta T = |\max_j t_j^{\text{end}} - \min_j t_j| = t_{11}^{\text{end}} - t_4. \qquad (27)$$

For potential sources that satisfy (25), the modulation due to the Earth's motion does not spread the signal across more than $N_{\text{cell}}$ coincidence cells during $\Delta T$.

### B. Fraction of parameter space excluded by the veto

One can visualize and calculate the volume of the region in four-dimensional parameter space that is excluded by this veto. For a given source sky position, Eq. (25) is linear in $f$ and $\dot{f}$. Thus, for fixed sky position $\hat{\mathbf{n}}$, the veto condition (25) defines two parallel lines in the $(f, \dot{f})$-plane. Candidate events that lie in the region between the lines are discarded (vetoed). Candidates that lie outside this region are retained (not vetoed). The locations of these two lines in the $(f, \dot{f})$ plane depend upon the sky position. The fractional volume excluded by the veto depends upon whether or not (as the source position varies over the sky) the excluded region between the lines lies inside or outside of the boundaries of the search, or intersects it. In this work, for the search region $-f/\tau < \dot{f} < 0.1f/\tau$ described in the abstract, the excluded region lies entirely within the parameter space above 300 Hz, but crosses the boundaries below 300 Hz. This is because a wider range of spin-down ages is searched below 300 Hz.

The fractional volume of the region in parameter space excluded by the veto may be easily calculated. The details of the calculation are in Appendix A. The resulting fraction of sky excluded by the veto (uniformly averaged over spin-down) as a function of frequency is shown in Fig. 10. In this search, the fraction of the sky excluded for frequencies $f \in [300, 1500)$ Hz has been fixed at the constant fraction 30%. In this search, the fraction of the sky excluded for frequen-
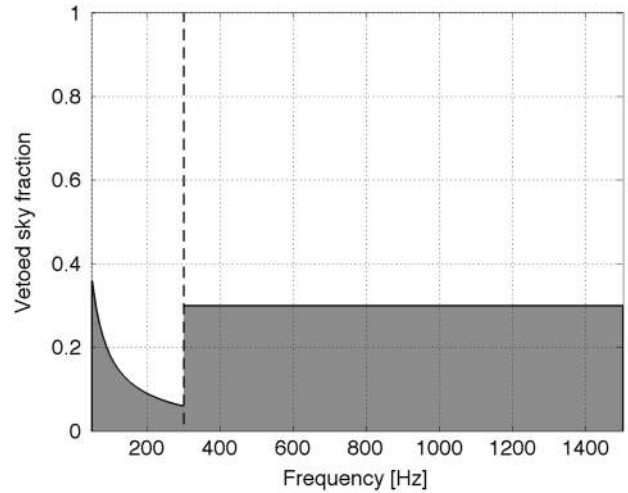


FIG. 10. The average fraction of sky excluded by the veto method as a function of frequency, uniformly averaged over the searched spin-down range.

TABLE IV. Parameters for hardware-injected CW signals during the S4 run, labeled Pulsar0 to Pulsar9. The parameters are defined at the GPS reference time $t_{ref} = 793\,130\,413$ s in the Solar System barycenter. These are the frequency $f(t_{ref})$, the spin-down $\dot{f}$, the sky position right ascension $\alpha$ and declination $\delta$, the polarization angle $\psi$, the initial phase $\Phi_0$, the inclination parameter $\cos\iota$, and the dimensionless strain amplitude $h_0$. Because the calibration was only accurately determined after S4 was finished, the H1 strain amplitudes should be multiplied by the correction factor 1.12. The L1 amplitudes should be multiplied by 1.15 for Pulsar1, 1.18 for Pulsar9, and 1.11 for the others.

| Name | $f(t_{ref})$ [Hz] | $\dot{f}$ [Hz s$^{-1}$] | $\alpha$ [rad] | $\delta$ [rad] | $\psi$ [rad] | $\Phi_0$ [rad] | $\cos\iota$ [rad] | $h_0$ |
|------|------|------|------|------|------|------|------|------|
| Pulsar0 | 265.576 933 18 | $-4.15 \times 10^{-12}$ | 1.248 817 | $-0.981\,180$ | 0.770 087 | 2.66 | 0.794 905 | $4.93 \times 10^{-25}$ |
| Pulsar1 | 849.070 861 08 | $-3.00 \times 10^{-10}$ | 0.652 646 | $-0.514\,042$ | 0.356 036 | 1.28 | 0.463 799 | $4.24 \times 10^{-24}$ |
| Pulsar2 | 575.163 567 32 | $-1.37 \times 10^{-13}$ | 3.756 929 | 0.060 109 | $-0.221\,788$ | 4.03 | $-0.928\,575$ | $8.04 \times 10^{-24}$ |
| Pulsar3 | 108.857 159 40 | $-1.46 \times 10^{-17}$ | 3.113 189 | $-0.583\,579$ | 0.444 280 | 5.53 | $-0.080\,666$ | $3.26 \times 10^{-23}$ |
| Pulsar4 | 1402.110 490 84 | $-2.54 \times 10^{-08}$ | 4.886 707 | $-0.217\,584$ | $-0.647\,939$ | 4.83 | 0.277 321 | $4.56 \times 10^{-22}$ |
| Pulsar5 | 52.808 324 36 | $-4.03 \times 10^{-18}$ | 5.281 831 | $-1.463\,269$ | $-0.363\,953$ | 2.23 | 0.462 937 | $9.70 \times 10^{-24}$ |
| Pulsar6 | 148.440 064 51 | $-6.73 \times 10^{-09}$ | 6.261 385 | $-1.141\,840$ | 0.470 985 | 0.97 | $-0.153\,727$ | $2.77 \times 10^{-24}$ |
| Pulsar7 | 1220.933 156 55 | $-1.12 \times 10^{-09}$ | 3.899 513 | $-0.356\,931$ | 0.512 323 | 5.25 | 0.756 814 | $1.32 \times 10^{-23}$ |
| Pulsar8 | 193.949 772 54 | $-8.65 \times 10^{-09}$ | 6.132 905 | $-0.583\,263$ | 0.170 471 | 5.89 | 0.073 904 | $3.18 \times 10^{-23}$ |
| Pulsar9 | 763.847 321 649 9 | $-1.45 \times 10^{-17}$ | 3.471 208 | 1.321 033 | $-0.008\,560$ | 1.01 | $-0.619\,187$ | $8.13 \times 10^{-24}$ |

cies $f \in [50, 300)$ Hz has been chosen to depend upon the values of $f$ and $\dot{f}$, where the uniform average of the excluded sky fraction over the spin-down range considered in this analysis is 36% at 50 Hz and 6% just below 300 Hz. Finally, Fig. 15 shows a summary diagram illustrating which candidates have been vetoed by this method.

## VII. HARDWARE-INJECTED SIGNALS

A good way to test and validate search algorithms and code is to add simulated signals into the detector's data

TABLE V. The time overlap between the Einstein@Home data segments and the hardware injections. The hardware injections were only turned on about 2/3 of the time. The columns are data segment index $j$, detector, the duration of the overlap, and the fractional overlap (obtained by dividing the third column by 30 hours $= 108\,000$ s).

| $j$ | Detector | Overlapping Duration [s] | Fractional Overlap |
|-----|----------|--------------------------|---------------------|
| 1 | H1 | 107 201 | 99.3% |
| 2 | H1 | 107 554 | 99.6% |
| 3 | H1 | 107 272 | 99.3% |
| 4 | H1 | 0 | 0 |
| 5 | H1 | 99 799 | 92.4% |
| 6 | H1 | 0 | 0 |
| 7 | H1 | 101 991 | 94.4% |
| 8 | H1 | 21 268 | 19.7% |
| 9 | H1 | 100 773 | 93.3% |
| 10 | H1 | 0 | 0 |
| 11 | L1 | 23 164 | 21.5% |
| 12 | L1 | 106 760 | 98.9% |
| 13 | L1 | 107 294 | 99.4% |
| 14 | L1 | 102 711 | 95.1% |
| 15 | L1 | 0 | 0 |
| 16 | L1 | 0 | 0 |
| 17 | L1 | 98 696 | 91.4% |

stream. This can either be done while the experiment is in progress (real-time injections) or after the data has been collected (software injections). If it is done while the experiment is in progress, the simulated signals can either be added into the hardware (into feedback and error-point control signals) or after data acquisition.

At the time that the S4 run was carried out, ten simulated CW signals were injected at the hardware level: using magnetic coil actuators, the interferometer mirrors were physically made to move as if a gravitational wave was present.

### A. Parameters of hardware injections

Table IV shows the parameters of the hardware injections that were carried out at the LIGO detectors during the

TABLE VI. The estimated (predicted) and obtained (measured) results for the hardware-injected pulsar signals. For each simulated signal the predicted number of coincidences $\mathcal{C}$ and a predicted value for the significance $\mathcal{S}$ is given, as well as the measured number of coincidences and measured value for the significance from the Einstein@Home search. The measured values are obtained by maximizing over a narrow band of $2 \times 10^{-4}f$ on either side of the injection frequency, the whole sky, and the entire spin-down range. As explained in the text, Pulsar4 and Pulsar8 are not expected to have the correct significance. Pulsar0, Pulsar1, Pulsar5, and Pulsar6 are not listed. They are so weak that they produce less than 7 coincidences, consistent with random noise containing no signal.

| Name | Predicted $\mathcal{C}$ | Measured $\mathcal{C}$ | Predicted $\mathcal{S}$ | Measured $\mathcal{S}$ |
|------|-------------------------|------------------------|-------------------------|------------------------|
| Pulsar2 | 12 | 13 | 263.1 | 249.3 |
| Pulsar3 | 12 | 12 | 3160.9 | 2397.5 |
| Pulsar4 | 12 | 12 | 35 108.2 | 1749.6 |
| Pulsar7 | 6 | 7 | 93.2 | 100.0 |
| Pulsar8 | 12 | 13 | 3692.6 | 2263.6 |
| Pulsar9 | 7 | 7 | 131.2 | 98.9 |

S4 run, mimicking gravitational-wave signals from ten different isolated pulsars with different frequencies, sky locations, and frequency derivatives. The ten artificial pulsars are denoted Pulsar0 to Pulsar9. At the time of the injections, lack of complete knowledge of the instrument's response function (calibration) meant that the hardware injections did not actually have the intended strain amplitudes as given in the table. The effective strain amplitudes may be computed from correction factors provided in Ref. [32]. These factors are 1.12 for all simulated pulsars in the H1 detector. In the L1 detector, the correction factor

is 1.11 for all simulated pulsars, except for Pulsar1 (1.15) and Pulsar9 (1.18).

### B. Duty cycle of hardware injections

During S4 the hardware injections were not active all of the time. Table V shows the fractional overlap between the times when the hardware injections were active and the times of the S4 Einstein@Home data segments. As can be seen from the table, the hardware injections were only turned on during 12 of the data segments analyzed in this paper, and for two of those 12 data data segments, the
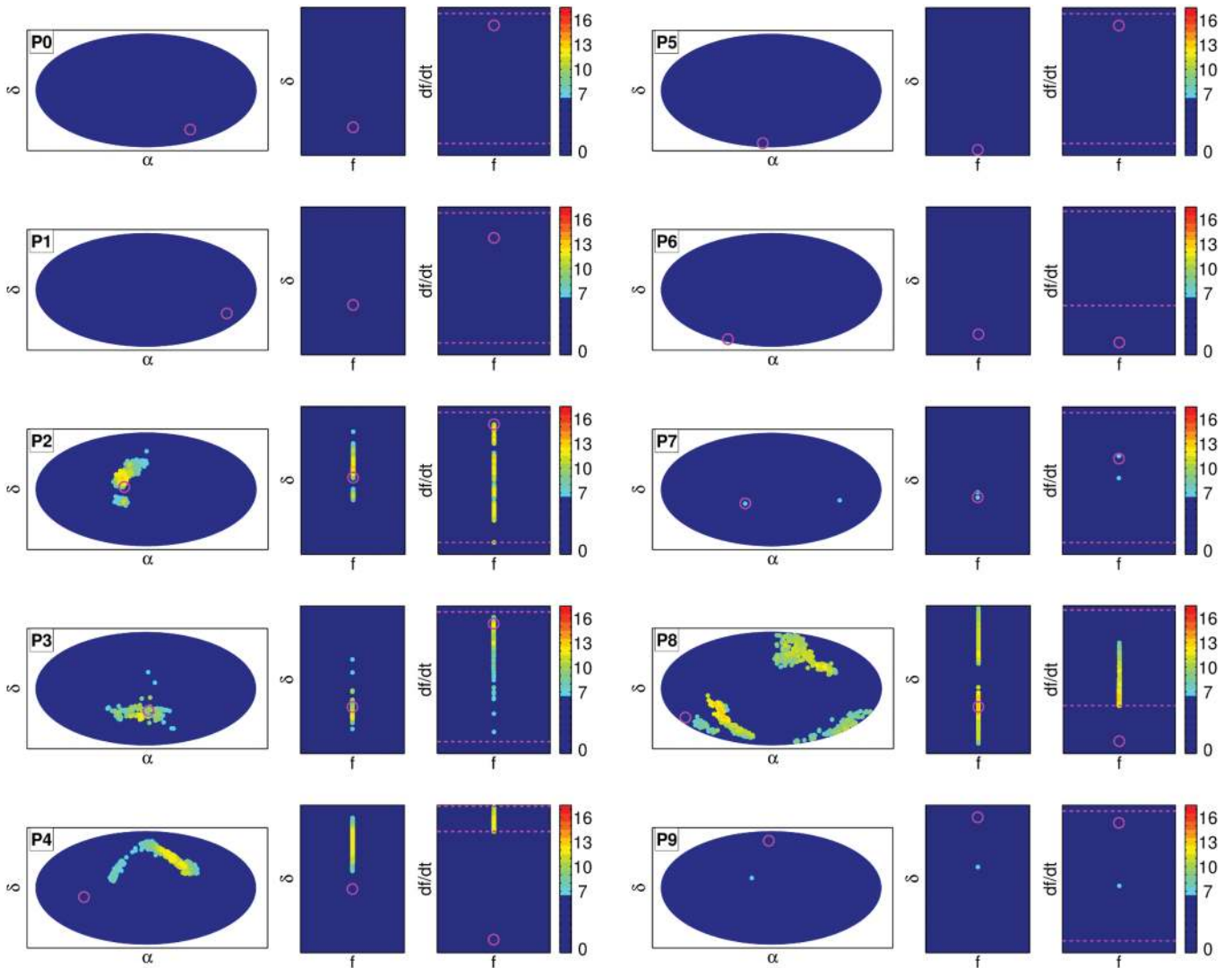


FIG. 11 (color). Einstein@Home results showing the 10 hardware-injected pulsar signals labeled P0 to P9. Here, a narrow band of width $2 \times 10^{-4} f$ to either side of each injection's frequency $f$ is considered. The color-bar in each plot indicates the number of coincidences. As shown in the color scale, only candidates having 7 or more coincidences appear. For each hardware injection a group of three different subplots are given representing different projections of the parameter space. The left subplot is a Hammer-Aitoff projection of the entire sky. The middle subplot shows declination $\delta$ versus frequency $f$. The right subplot shows spin-down $\dot{f}$ versus frequency $f$, where the region between the two horizontal magenta dashed lines refers to searched range of spin-downs. The center of a magenta circle represents the location of the injection. P4 and P8 appear at the wrong sky position because their intrinsic spin-downs lie outside the searched range. Table VI shows a comparison with the expectations for these simulated signals.

injections were only turned on for about 20% of the data taking time. In the remaining ten data segments, the hardware injections were turned on for almost the entire segment. This needs to be taken into account when analyzing the Einstein@Home search results for these injections. Because of this, the maximum possible number of coincidences expected from these simulated signals is 12, even though 17 data segments are analyzed.

### C. Results from the hardware injections

For each hardware-injected pulsar signal Table VI compares a prediction for the outcome of the Einstein@Home search to the actual results found through the Einstein@Home analysis pipeline. The predicted values given in Table VI are obtained by feeding the sensitivity-estimation pipeline, which was described in Sec. V, with the parameters of the simulated pulsars and only considering data segments where the hardware injections were active.

As shown in Table VI and consistent with Fig. 9, the hardware-injected signals Pulsar0, Pulsar1, Pulsar5, and Pulsar6 are too weak to be confidently detected by the search. In contrast, Pulsar2, Pulsar3, Pulsar4, and Pulsar8 are clearly detected. The parameters of Pulsar7 and Pulsar9 are such that in both cases the search pipeline found 7 coincidences, but this is consistent with the level of coincidences that would result from Gaussian noise with no signal present, and so these are not confidently detected.

Figure 11 presents the results of the search for all hardware injections. Small subspaces of the search parameter space around the hardware injections are shown, as well as the locations of the artificial signal parameters. The subspaces considered in Fig. 11 and also for the (measured) results presented in Table VI are constrained to a band of $2 \times 10^{-4} f$ to either side of the injected frequency. This choice of frequency bandwidth is motivated by the maximum Doppler shift due to the Earth orbital motion.

The significant sky position offset between the *a priori* location of the simulated source and the location where the search located Pulsar4 and Pulsar8 is explained by the global correlations [49] in parameter space. This arises because for Pulsar4 and Pulsar8, the spin-down range that is searched (region between dashed lines in the far right column of Fig. 11) is too small to include the actual spin-down value used in creating the simulated signals. Because of the global parameter-space correlations the offset between the actual and detected spin-down value gives rise to the offset in the sky position. The observed structure of large-coincident events in the sky is consistent with the global-correlation hypersurface description in [49]. This is also why Pulsar4 shows a considerable discrepancy between the significance that would have been expected if the search grid had covered the *a priori* parameters, and the significance that was actually observed, as shown in Table VI.
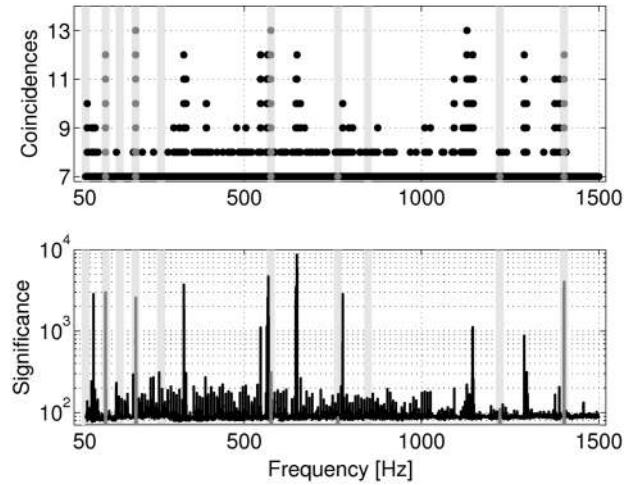


FIG. 12. Numbers of coincidences of 7 or more (top) and the significance (bottom) of all candidates found in the Einstein@Home post-processing, shown as functions of frequency. The light-gray shaded rectangular regions highlight the S4 hardware injections, listed in Table IV. The data points colored in dark-gray show the candidates resulting from the hardware-injected CW signals.
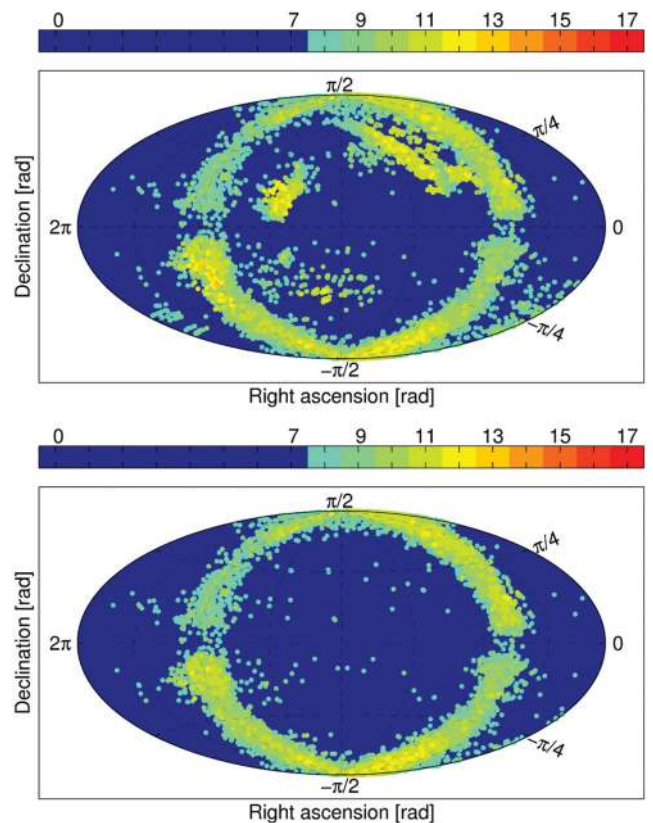


FIG. 13 (color). All candidates obtained from the post-processing that have more than 7 coincidences, shown in Hammer-Aitoff projections of the sky. The color bar indicates the number of coincidences of a particular candidate (cell). The upper plot includes the S4 hardware-injected pulsars. In the lower plot, bands of $2 \times 10^{-4} f$ width to either side of the hardware injections' frequencies $f$ have been removed.

## VIII. RESULTS

This section presents the results of the Einstein@Home S4 CW search. Figs. 12 and 13 give a summary of all post-processing results, from 50 to 1500 Hz. In Fig. 12 the coincidences and significance of all candidates that have 7 or more coincidences are shown as functions of frequency. Figure 13 presents the same information as given in Fig. 12, but projected on the sky, and showing all cells that have more than 7 candidate events.

In Fig. 13 the number of coincidences is maximized over the entire sky and full spin-down range. The color indicates the numbers of coincidences, where the same color scale has been used in each plot. The maximum possible number of coincidences ranges from a minimum of 0 to a maximum of 17 (the number of data segments analyzed). The meaning of 0 coincidences is that there is no candidate event found, 1 coincidence means a single candidate events is found (which is always coincident with itself).

TABLE VII. The post-processing candidates that have 10 or more coincidences. The frequency $f_{\rm cand}$ corresponds to the most coincident candidate in the band. The lowest frequency of a candidate in the band is labeled by $f_{\rm start}$. The difference from the highest frequency is given by $\Delta f_{\rm cand}$. The parameters $\delta_{\rm cand}$, $\alpha_{\rm cand}$, $f_{\rm cand}$, $\mathcal{C}_{\rm cand} = \mathcal{C}^{\rm H1}_{\rm cand} + \mathcal{C}^{\rm L1}_{\rm cand}$, and $\mathcal{S}_{\rm cand}$ are for the most significant most coincident candidate within the frequency band, where $\mathcal{C}^{\rm H1}_{\rm cand}$ and $\mathcal{C}^{\rm L1}_{\rm cand}$ denote the number of coincidences contributing to $\mathcal{C}_{\rm cand}$ from detector H1 and L1, respectively. The column "Information" lists information about the source. The following are understood sources of narrow-band line noise in the instrument: "Demod" are the electronics boards that demodulate the signal at the antisymmetric port of the interferometer, "H1 (or L1) MC 1" is a violin mode resonance of the first mode cleaner mirror, "H1 MC 2/3" are violin mode resonances of the second and third mode cleaner mirrors, "TM violin" are harmonics of the test mass violin modes, "EX + 15v" is a 15 volt power supply at the end station of the X arm, "EM Interference" is electromagnetic interference, "H1 Cal" are side-bands of calibration lines at 393.1 Hz and 1144.3 Hz.

| $f_{\rm cand}$ [Hz] | $f_{\rm start}$ [Hz] | $\Delta f_{\rm cand}$ [Hz] | $\delta_{\rm cand}$ [rad] | $\alpha_{\rm cand}$ [rad] | $\dot{f}_{\rm cand}$ [Hz s$^{-1}$] | $\mathcal{C}_{\rm cand}$ | $\mathcal{C}^{\rm H1}_{\rm cand}$ | $\mathcal{C}^{\rm L1}_{\rm cand}$ | $\mathcal{S}_{\rm cand}$ | Information |
|---|---|---|---|---|---|---|---|---|---|---|
| 193.9276 | 193.9263 | 0.040 112 | $-0.583\,514$ | 4.723 595 | $-5.6001 \times 10^{-09}$ | 13 | 7 | 6 | 2263.6 | Pulsar 8 |
| 575.1681 | 575.1562 | 0.030 612 | 0.285 505 | 3.834 511 | $-5.0913 \times 10^{-10}$ | 13 | 7 | 6 | 249.3 | Pulsar 2 |
| 1128.1147 | 1128.0336 | 0.220 321 | $-1.395\,918$ | 0.744 273 | $-3.4249 \times 10^{-09}$ | 13 | 10 | 3 | 219.3 | H1 MC 2/3 |
| 108.8549 | 108.8522 | 0.008 158 | $-0.705\,729$ | 3.361 465 | $-4.4362 \times 10^{-11}$ | 12 | 9 | 3 | 2397.5 | Pulsar 3 |
| 329.6107 | 329.5507 | 0.066 447 | 1.027 320 | 1.336 051 | $-5.7799 \times 10^{-10}$ | 12 | 10 | 2 | 3127.1 | Demod |
| 545.9973 | 545.9929 | 0.109 58 | $-0.293\,877$ | 4.849 960 | $-1.5782 \times 10^{-09}$ | 12 | 10 | 2 | 893.3 | H1 MC 2/3 |
| 566.0868 | 566.0490 | 0.105 853 | $-1.367\,663$ | 0.665 233 | $-1.626 \times 10^{-09}$ | 12 | 10 | 2 | 2340.8 | H1 MC 2/3 |
| 568.0886 | 567.9893 | 0.165 769 | $-1.323\,532$ | 0.726 729 | $-1.7149 \times 10^{-09}$ | 12 | 10 | 2 | 4137.7 | H1 MC 2/3 |
| 648.8288 | 648.6930 | 0.206 223 | $-1.232\,868$ | 1.005 733 | $-1.0298 \times 10^{-09}$ | 12 | 10 | 2 | 1870.8 | H1 MC 1 |
| 1143.9976 | 1143.9182 | 0.232 221 | $-1.491\,264$ | 1.314 456 | $-7.7434 \times 10^{-10}$ | 12 | 10 | 2 | 1028.8 | H1 Cal |
| 1144.5198 | 1144.4533 | 0.228 407 | $-1.535\,248$ | 4.497 733 | $-2.5257 \times 10^{-11}$ | 12 | 10 | 2 | 989.8 | H1 Cal |
| 1289.6769 | 1289.5081 | 0.242 915 | 1.461 093 | 0.266 878 | $-2.0949 \times 10^{-09}$ | 12 | 10 | 2 | 493.7 | H1 MC 1 |
| 1402.2838 | 1402.2677 | 0.063 117 | 1.025 583 | 2.50 2838 | $-3.8482 \times 10^{-09}$ | 12 | 6 | 6 | 1749.6 | Pulsar 4 |
| 329.7593 | 329.7396 | 0.066 078 | $-1.536\,179$ | 4.887 048 | $-5.5375 \times 10^{-10}$ | 11 | 10 | 1 | 3038.3 | Demod |
| 335.7735 | 335.7100 | 0.065 415 | 0.469 606 | 0.955 884 | $-1.0646 \times 10^{-09}$ | 11 | 10 | 1 | 298.5 | EM Interference |
| 545.9232 | 545.8662 | 0.063 608 | $-1.060\,735$ | 1.078 303 | $-8.032 \times 10^{-10}$ | 11 | 10 | 1 | 196.8 | H1 MC 2/3 |
| 564.1219 | 564.0096 | 0.113 783 | 0.386 877 | 1.111 355 | $-1.6868 \times 10^{-09}$ | 11 | 10 | 0 | 1069.3 | H1 MC 2/3 |
| 646.3758 | 646.3206 | 0.127 884 | $-1.281\,366$ | 0.897 933 | $-1.8931 \times 10^{-09}$ | 11 | 10 | 1 | 3202.7 | H1 MC 1 |
| 1092.1387 | 1091.9671 | 0.217 482 | $-0.523\,866$ | 1.302 500 | $-6.4347 \times 10^{-11}$ | 11 | 10 | 1 | 196.7 | H1 MC 2/3 |
| 1136.2217 | 1136.1460 | 0.168 345 | $-1.216\,945$ | 0.935 876 | $-3.4811 \times 10^{-09}$ | 11 | 10 | 1 | 165.6 | H1 MC 2/3 |
| 1142.8210 | 1142.7200 | 0.231 73 | $-1.310\,037$ | 1.114 563 | $-3.5022 \times 10^{-09}$ | 11 | 10 | 1 | 250.7 | H1 Cal |
| 1145.8318 | 1145.6515 | 0.231 067 | 1.330 065 | 0.976 422 | $-2.0297 \times 10^{-10}$ | 11 | 10 | 1 | 256.4 | H1 Cal |
| 1376.7370 | 1376.4697 | 0.271 536 | 0.201 677 | 1.282 354 | $-2.3875 \times 10^{-09}$ | 11 | 9 | 2 | 165.0 | TM violin |
| 1388.6402 | 1388.4070 | 0.279 967 | 1.176 082 | 0.850 794 | $-2.8907 \times 10^{-09}$ | 11 | 10 | 1 | 200.0 | TM violin |
| 56.9966 | 56.9966 | | $-0.935\,903$ | 0.150 238 | $-1.5029 \times 10^{-09}$ | 10 | 8 | 2 | 136.7 | EM Interference |
| 329.4918 | 329.4784 | 0.021 358 | $-1.307\,440$ | 4.692 056 | $-5.2405 \times 10^{-10}$ | 10 | 10 | 0 | 1137.8 | Demod |
| 392.8322 | 392.8322 | | $-1.210\,088$ | 1.268 596 | $-1.069 \times 10^{-09}$ | 10 | 9 | 1 | 150.9 | H1 Cal |
| 393.4060 | 393.4057 | 0.000 342 | 0.632 053 | 1.270 922 | $-1.1043 \times 10^{-09}$ | 10 | 9 | 1 | 154.7 | H1 Cal |
| 646.7224 | 646.7224 | 0.002 174 | $-1.446\,520$ | 0.825 633 | $-1.8813 \times 10^{-09}$ | 10 | 3 | 7 | 2774.4 | L1 MC 1 |
| 648.4132 | 648.4132 | 0.024 291 | 1.319 729 | 1.033 730 | $-1.8479 \times 10^{-09}$ | 10 | 3 | 7 | 5067.5 | L1 MC 1 |
| 658.6353 | 658.6353 | 0.000 055 | $-0.470\,832$ | 4.762 475 | $-1.6992 \times 10^{-09}$ | 10 | 3 | 7 | 261.4 | EX + 15v |
| 777.9202 | 777.8377 | 0.117 087 | 1.511 859 | 4.010 213 | $-5.6101 \times 10^{-10}$ | 10 | 3 | 7 | 1951.7 | EM Interference |
| 1296.4962 | 1296.4962 | | $-0.993\,190$ | 4.557 370 | $-1.0022 \times 10^{-09}$ | 10 | 3 | 7 | 247.1 | L1 MC 1 |
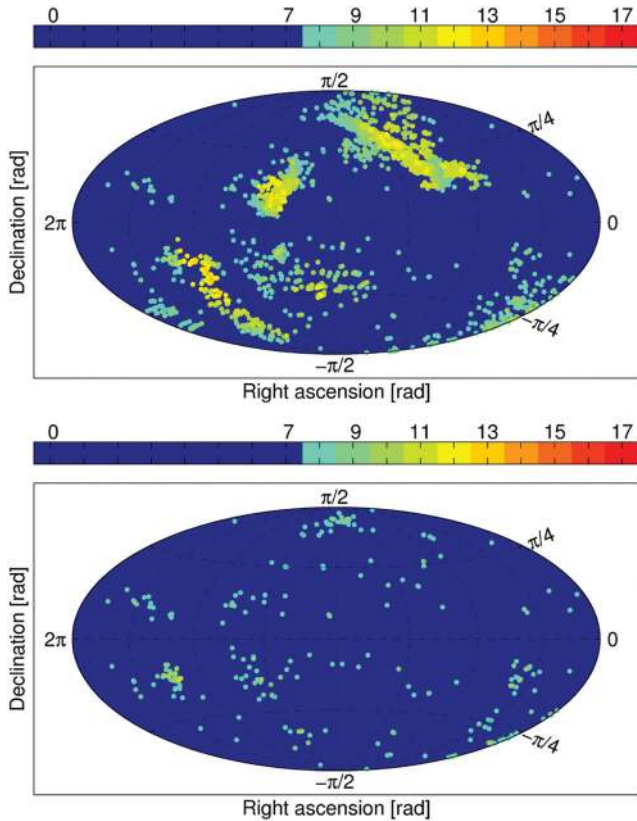
FIG. 14 (color). Candidates not eliminated by the veto. This shows Hammer-Aitoff sky projections of all candidates obtained from post-processing that had more than 7 coincidences and that passed the veto. The upper plot includes the S4 hardware injections. The lower plot removes bands of $2 \times 10^{-4} f$ width to either side of the hardware injections' frequencies $f$. In comparison to Fig. 13, after excluding the hardware injections, the veto rejects 99.5% of all candidates.



FIG. 15 (color). Candidates discriminated by the veto method. All candidate cells obtained from post-processing that have more than 7 coincidences are shown, where the color bar indicates the number of coincidences of a particular cell. The vertical axis represents the veto quantity on the left-hand side of (25), as a function of frequency. Candidates located below the magenta line are eliminated by the veto. The four accumulations of highly coincident cells above the magenta line are the hardware-injected pulsars, which are not eliminated by the veto.

Four typical examples of different types of post-processing results, in different 10 Hz bands, are shown in Figs. 16–19 of Appendix B.

Table VII shows all candidates (cells) which have 10 or more coincidences. In cases where a set of candidates is clustered together at slightly different frequencies, Table VII lists the bandwidth in frequency covered by these candidates and shows the parameters of the most coincident candidate. If candidates within these narrow frequency bands have the same number of coincidences, then the candidate with the largest significance is shown.

Table VIII shows the same information *after* the veto method described in Sec. VI has been applied, for candidates with 9 or more coincidences. There are no candidates that exceed the detection threshold, appearing in 12 or more data segments. (Note that this would be a threshold for initiating a more extensive investigation of the candidate event, not a threshold for announcing a discovery!)

Figure 14 shows all candidates from the post-processing results that have not been discriminated by the veto intro-

duced in Sec. VI. Figure 15 illustrates the fraction of candidates that has been excluded by the veto. After removing fractional bands of $2 \times 10^{-4} f$ around the frequencies $f$ of the S4 hardware injections, the veto eliminates 99.5% of all candidates that have more than 7 coincidences.

## IX. CONCLUSION

These are the first published results from the Einstein@Home project, which was launched in February 2005. While no credible CW sources were found in this search of LIGO S4 data, the results clearly establish that this type of distributed computing project can carry out a credible and sensitive search for such signals.

In retrospect, it probably would have been a good idea to employ identical grids on the four-dimensional parameter space for all 17 data segments. This would have required more CPU time on the part of participants, but would have greatly simplified and sped up the development of the post-processing pipeline and would also have greatly simplified the interpretation of the results.

A similar search (also with a 30-h time baseline) has already been completed using 660 h of data from the beginning of the S5 science run. The post-processing of that data set is currently underway, using methods identical to those employed here.

TABLE VIII. Post-processing candidates that have 9 or more coincidences and that are not excluded by the veto. The frequency $f_{cand}$ corresponds to the most coincident candidate in the band. The lowest frequency of a candidate in the band is labeled by $f_{start}$. The difference from the highest frequency is given by $\Delta f_{cand}$. The parameters $\delta_{cand}$, $\alpha_{cand}$, $\dot{f}_{cand}$, $\mathcal{C}_{cand} = \mathcal{C}_{cand}^{H1} + \mathcal{C}_{cand}^{L1}$, and $\mathcal{S}_{cand}$ are for the most significant most coincident candidate within the frequency band, where $\mathcal{C}_{cand}^{H1}$ and $\mathcal{C}_{cand}^{L1}$ denote the number of coincidences contributing to $\mathcal{C}_{cand}$ from detectors H1 and L1, respectively. The column "Information" lists information about the source. The following are understood sources of narrow-band line noise in the instrument: "Demod" are the electronics boards which demodulate the signal at the antisymmetric port of the interferometer, "H1 MC 2/3" are violin mode resonances of the second and third mode cleaner mirrors, "EM Interference" is electromagnetic interference, "H1 Cal" are side-bands of a 1144.3 Hz calibration line. For the single candidate labeled "Unknown" in the last column no instrumental source could be confidently identified, however the 9 coincidences are far below the confident-detection threshold.

| $f_{cand}$ [Hz] | $f_{start}$ [Hz] | $\Delta f_{cand}$ [Hz] | $\delta_{cand}$ [rad] | $\alpha_{cand}$ [rad] | $\dot{f}_{cand}$ [Hz s$^{-1}$] | $\mathcal{C}_{cand}$ | $\mathcal{C}_{cand}^{H1}$ | $\mathcal{C}_{cand}^{L1}$ | $\mathcal{S}_{cand}$ | Information |
|---|---|---|---|---|---|---|---|---|---|---|
| 193.9276 | 193.9261 | 0.040 646 | −0.583 514 | 4.723 595 | $-5.6001 \times 10^{-09}$ | 13 | 7 | 6 | 2263.6 | Pulsar 8 |
| 575.1681 | 575.1562 | 0.039 394 | 0.285 505 | 3.834 511 | $-5.0913 \times 10^{-10}$ | 13 | 7 | 6 | 249.3 | Pulsar 2 |
| 108.8549 | 108.8518 | 0.008 506 | −0.705 729 | 3.361 465 | $-4.4362 \times 10^{-11}$ | 12 | 9 | 3 | 2397.5 | Pulsar 3 |
| 1402.2838 | 1402.2488 | 0.086 78 | 1.025 583 | 2.502 838 | $-3.8482 \times 10^{-09}$ | 12 | 6 | 6 | 1749.6 | Pulsar 4 |
| 545.9987 | 545.9568 | 0.141 563 | −0.398 855 | 5.013 332 | $-4.6693 \times 10^{-10}$ | 11 | 10 | 1 | 794.1 | H1 MC 2/3 |
| 56.9966 | 56.9963 | 0.000 933 | −0.935 903 | 0.150 238 | $-1.5029 \times 10^{-09}$ | 10 | 8 | 2 | 136.7 | EM Interference |
| 329.4849 | 329.4833 | 0.005 843 | −0.344 739 | 5.171 401 | $-5.6694 \times 10^{-10}$ | 10 | 10 | 0 | 1024.0 | EM Interference |
| 329.6040 | 329.6040 | | −0.439 100 | 1.006 331 | $-5.6923 \times 10^{-10}$ | 10 | 9 | 1 | 2625.6 | EM Interference |
| 329.7434 | 329.7413 | 0.032 463 | −0.338 712 | 5.025 108 | $-5.6923 \times 10^{-10}$ | 10 | 9 | 1 | 2490.4 | EM Interference |
| 567.9984 | 567.9984 | 0.051 768 | −0.353 846 | 5.116 972 | $-1.5532 \times 10^{-09}$ | 10 | 9 | 1 | 409.3 | EM Interference |
| 69.6964 | 69.6964 | | −1.223 613 | 4.232 687 | $-5.4823 \times 10^{-10}$ | 9 | 9 | 0 | 130.3 | EM Interference |
| 317.4207 | 317.4207 | | 1.389 330 | 2.663 214 | $-8.0338 \times 10^{-10}$ | 9 | 3 | 6 | 157.8 | EM Interference |
| 329.5615 | 329.5615 | | −1.027 976 | 3.822 726 | $-6.3014 \times 10^{-10}$ | 9 | 7 | 2 | 2176.0 | Demod |
| 335.7541 | 335.7141 | 0.056 927 | 1.395 059 | 3.271 989 | $-6.362 \times 10^{-10}$ | 9 | 9 | 0 | 259.3 | EM Interference |
| 795.4783 | 795.4783 | | 0.245 291 | 3.211 417 | $-1.4374 \times 10^{-09}$ | 9 | 7 | 2 | 110.7 | EM Interference |
| 1092.1564 | 1092.1564 | | −0.252 089 | 1.099 873 | $-2.1099 \times 10^{-11}$ | 9 | 8 | 1 | 147.1 | H1 MC 2/3 |
| 1117.3032 | 1117.3032 | | −0.207 300 | 4.051 169 | $-3.3192 \times 10^{-09}$ | 9 | 4 | 5 | 116.1 | Unknown |
| 1145.6678 | 1145.6678 | | −0.247 554 | 5.067 301 | $-3.1679 \times 10^{-09}$ | 9 | 6 | 3 | 168.7 | H1 Cal |

Future Einstein@Home searches overcome some of the sensitivity limitations discussed at the end of Sec. V by doing the incoherent step (called "post-processing" in this paper) on the host machines. This allows the use of the optimal threshold of $2\mathcal{F} \sim 5$, so those searches are expected to be the most sensitive blind CW searches that will be possible using LIGO data. Results from those searches should become available within the next one to two years, and are expected to offer more than 1 order of magnitude improvement in strain sensitivity compared with the work presented here.

In the longer term, further increases in sensitivity will probably result mostly from improvements in the detectors rather than from improvements in the data analysis methods. In 2009 LIGO is expected to begin its S6 run with an "enhanced" detector configuration that should improve on S5 sensitivity by at least a factor of 2. By 2014, an advanced LIGO detector configuration should give at least another factor of 5 improvement. By combining these data sets with those from LIGO's international partner projects Virgo and GEO, there is real hope that the first direct CW detection can be made using methods like the ones described here.

## ACKNOWLEDGMENTS

TABLE IX.   For given values of frequency $f$ and spin-down $\dot{f}$, this shows the fractional volume of the sky excluded by the veto (25). There are six possible cases, depending upon the values of $\cos\theta_-$ and $\cos\theta_+$. (There are six rather than nine cases because $\cos\theta_+$ is never greater than $\cos\theta_-$.) For the ranges of $f$ and $\dot{f}$ considered in this work, case 4 applies above 300 Hz. Between 50 Hz and 300 Hz, because of the wider range of $\dot{f}$ considered, the three cases 4, 5, and 6 are found. Values of $\cos\theta$ outside the range $[-1, 1]$ correspond to imaginary (unphysical) values of $\theta$. In such cases the upper and/or lower limits of integration are replaced by $\theta = \pi$ or $\theta = 0$, respectively, as can be seen from the final column of this table.

| Case | Range of $\cos\theta_+ = \frac{-\epsilon-\dot{f}}{\gamma f}$ | Range of $\cos\theta_- = \frac{\epsilon-\dot{f}}{\gamma f}$ | Excluded sky fraction $\Omega_{\text{excluded}}/4\pi$ |
|------|------|------|------|
| 1 | $(-\infty, -1)$ | $(-\infty, -1)$ | 0 |
| 2 | $(-\infty, -1)$ | $[-1, 1]$ | $(\cos\theta_- + 1)/2 = (1 + \frac{\epsilon-\dot{f}}{\gamma f})/2$ |
| 3 | $(-\infty, -1)$ | $(1, \infty)$ | 1 |
| 4 | $[-1, 1]$ | $[-1, 1]$ | $(\cos\theta_- - \cos\theta_+)/2 = \epsilon/\gamma f$ |
| 5 | $[-1, 1]$ | $(1, \infty)$ | $(1 - \cos\theta_+)/2 = (1 + \frac{\epsilon+\dot{f}}{\gamma f})/2$ |
| 6 | $(1, \infty)$ | $(1, \infty)$ | 0 |

## APPENDIX A: FRACTION OF PARAMETER SPACE EXCLUDED BY THE VETO METHOD

The fractional volume of the region in parameter space excluded by the veto method presented in [49] and used in Sec. VI may be easily calculated. Since the time $\Delta T$ is small compared to 1 yr, one may use the following approximation:

$$(\boldsymbol{\omega} \times \mathbf{v}_{\text{av}}) \cdot \hat{\mathbf{n}} \approx |\boldsymbol{\omega}||\mathbf{v}_{\text{av}}| \cos\theta, \qquad (A1)$$

where $\theta \in [0, \pi]$ is the angle between the SSB-to-Earth vector and the source sky position $\hat{\mathbf{n}}$. The veto condition (25) may then be rewritten as

$$|\dot{f} + \gamma f \cos\theta| < \epsilon, \qquad (A2)$$

where $\gamma$ is defined as $\gamma = |\boldsymbol{\omega}||\mathbf{v}_{\text{av}}|/c$. For fixed values of $f$ and $\dot{f}$ the situation is depicted in Table IX. Depending upon the values of $(\pm\epsilon - \dot{f})/\gamma f$, a part of the sky might be excluded by the veto. As shown in the Table, there are six possible cases, determined by the values of

$$\cos\theta_- = \frac{\epsilon - \dot{f}}{\gamma f} \quad \text{and} \qquad (A3)$$

$$\cos\theta_+ = \frac{-\epsilon - \dot{f}}{\gamma f}. \qquad (A4)$$

For example in the case (labeled case 4 in Table IX) where both $\cos\theta_-$ and $\cos\theta_+$ lie in the range $[-1, 1]$ then the excluded region of the sky is an annulus defined by $0 \le \theta_- < \theta < \theta_+ \le \pi$, and the excluded solid angle is

$$\Omega_{\text{excluded}} = \int_0^{2\pi} d\phi \int_{\theta_-}^{\theta_+} d\theta \sin\theta \qquad (A5)$$

$$= 2\pi(\cos\theta_- - \cos\theta_+). \qquad (A6)$$

The fraction of sky excluded in this case is then

$$\frac{\Omega_{\text{excluded}}}{4\pi} = (\cos\theta_- - \cos\theta_+)/2 \qquad (A7)$$

$$= \frac{\epsilon}{\gamma f} \qquad (A8)$$

$$= \epsilon \frac{c}{|\boldsymbol{\omega}||\mathbf{v}_{\text{av}}|} \frac{1}{f}. \qquad (A9)$$

In the other cases listed in Table IX the excluded region of the sky might be a cap about $\theta = 0$ or about $\theta = \pi$ or the null set, or the entire sky.

In this search, the fraction of the sky excluded for frequencies $f \in [300, 1500)$ Hz has been fixed at the constant fraction $\Omega_{\text{excluded}}/4\pi = 30\%$. This is equivalent to choosing $\epsilon$ to be a linear function of frequency

$$\epsilon = 0.3 \frac{|\boldsymbol{\omega}||\mathbf{v}_{\text{av}}|}{c} f. \qquad (A10)$$

In this search, the fraction of the sky excluded for frequencies $f \in [50, 300)$ Hz has been chosen to depend upon the value of $\dot{f}$. As can be seen from Fig. 15, below 300 Hz, the instrument noise permits the use of a frequency-independent value $\epsilon = 5.4 \times 10^{-10}$ Hz/s which corresponds to $N_{\text{cell}} = 1.5$. Within the region of parameter space which is searched ($-f/\tau < \dot{f} < 0.1f/\tau$ for $\tau = 1000$ years) cases 4, 5, or 6 from Table IX occur depending on the spin-down value $\dot{f}$. If

$$\dot{f} > \epsilon - \frac{|\boldsymbol{\omega}||\mathbf{v}_{\text{av}}|f}{c}, \qquad (A11)$$

then case 4 of Table IX applies, and the fraction of sky excluded is given by

$$\frac{\Omega_{\text{excluded}}}{4\pi} = \epsilon \frac{c}{|\boldsymbol{\omega}||\mathbf{v}_{\text{av}}|} \frac{1}{f}. \qquad (A12)$$

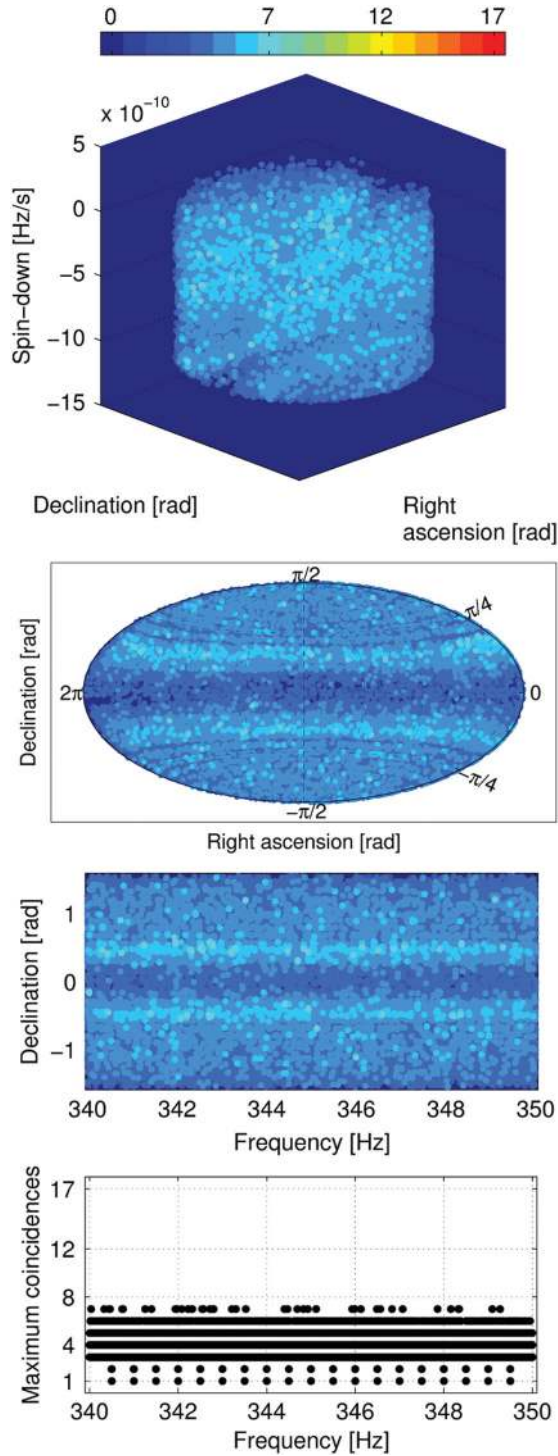This fraction ranges from 52% at 50 Hz to 8.7% at 300 Hz. If $\dot{f}$ is in the interval

FIG. 16 (color). Einstein@Home S4 post-processing results for the frequency band 340.0–350.0 Hz, which is pure Gaussian noise for L1 and mostly Gaussian noise for H1. This is because in this band the line-cleaning process has replaced all the L1 data and most of the H1 data with computer-generated random numbers (see Table II). From top to bottom the different plots show the numbers of coincidences in a 3D map of sky and spin-down, in a 2D Hammer-Aitoff projection of the sky, in a 2D plot of declination over frequency, and in a histogram as a function of frequency.
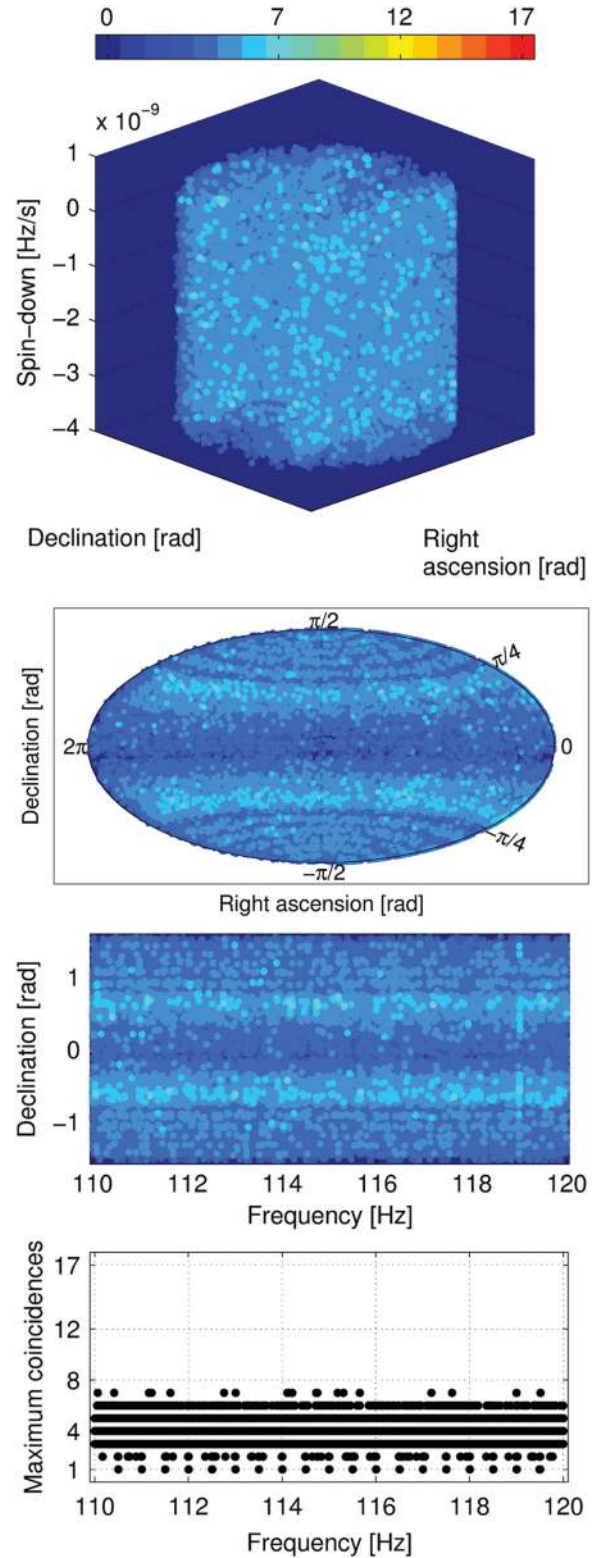


FIG. 17 (color). Einstein@Home S4 post-processing results for a quiet frequency band of real instrumental data from 110.0–120.0 Hz. From top to bottom the different plots show the numbers of coincidences in a 3D map of sky and spin-down, in a 2D Hammer-Aitoff projection of the sky, in a 2D plot of declination over frequency, and in a histogram as a function of frequency.
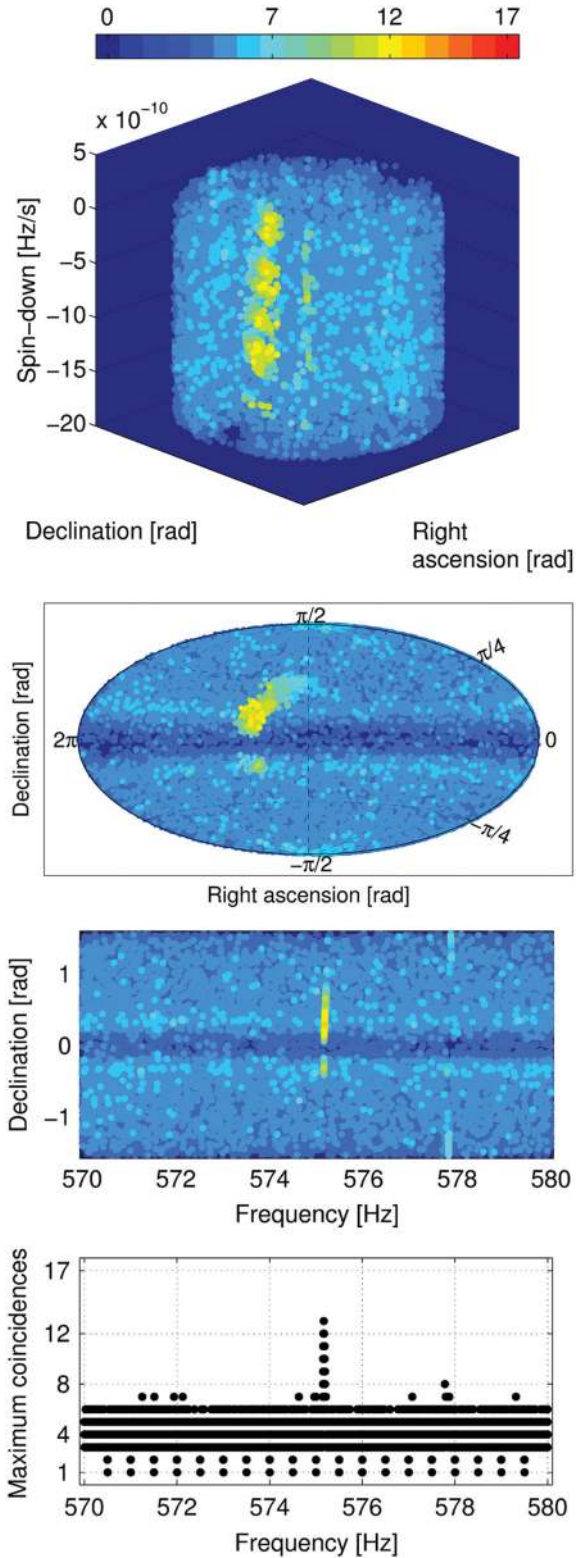
FIG. 18 (color). Einstein@Home S4 post-processing results for the frequency band 570.0–580.0 Hz including a hardware-injected CW signal (Pulsar2). From top to bottom the different plots show the numbers of coincidences in a 3D map of sky and spin-down, in a 2D Hammer-Aitoff projection of the sky, in a 2D plot of declination over frequency, and in a histogram as a function of frequency.
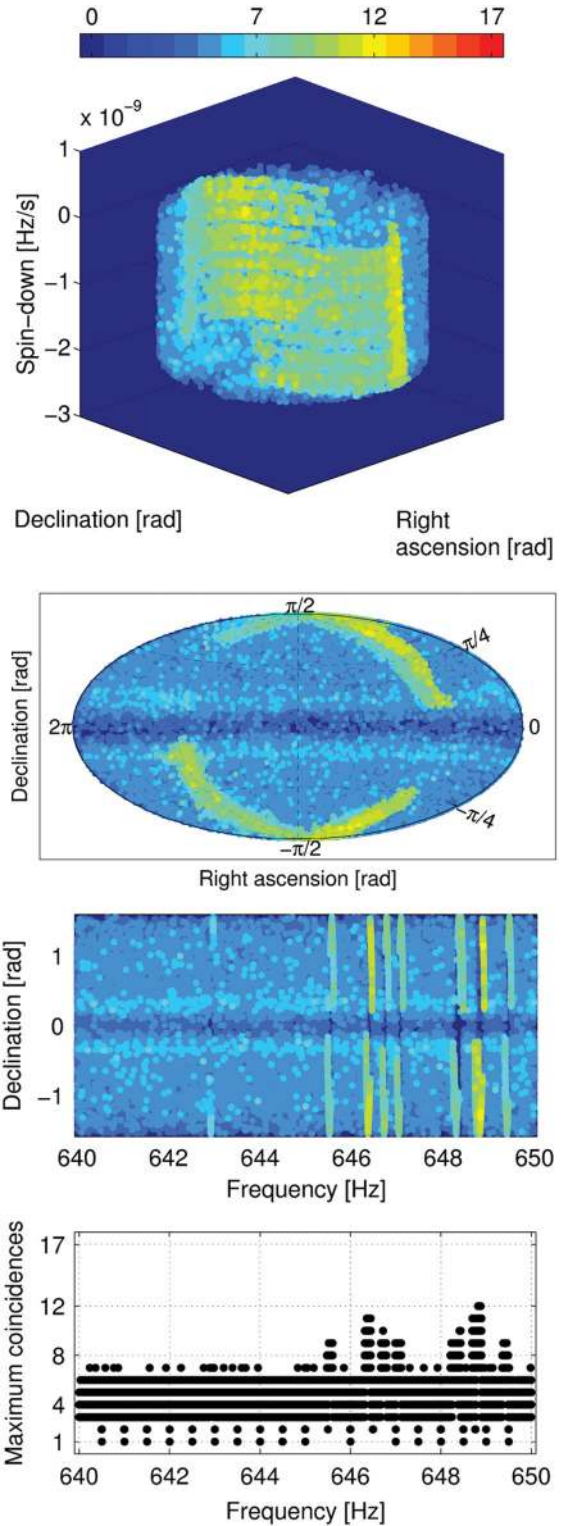
FIG. 19 (color). Einstein@Home S4 post-processing results for a noisy frequency band of data polluted by instrumental noise artifacts from 640.0–650.0 Hz. These spectral features are resonance modes of the mode cleaner optics suspensions. From top to bottom the different plots show the numbers of coincidences in a 3D map of sky and spin-down, in a 2D Hammer-Aitoff projection of the sky, in a 2D plot of declination over frequency, and in a histogram as a function of frequency.

$$\dot{f} \in \left[ -\epsilon - \frac{|\boldsymbol{\omega}||\mathbf{v}_{\mathrm{av}}|f}{c}, \epsilon - \frac{|\boldsymbol{\omega}||\mathbf{v}_{\mathrm{av}}|f}{c} \right], \qquad \text{(A13)}$$

then case 5 of Table IX applies, and the fraction of sky excluded is given by

$$\frac{\Omega_{\mathrm{excluded}}}{4\pi} = \frac{1}{2}\left( 1 + \frac{\epsilon + \dot{f}}{f|\boldsymbol{\omega}||\mathbf{v}_{\mathrm{av}}|/c} \right). \qquad \text{(A14)}$$

Finally, if

$$\dot{f} < -\epsilon - \frac{|\boldsymbol{\omega}||\mathbf{v}_{\mathrm{av}}|f}{c}, \qquad \text{(A15)}$$

then case 6 applies and none of the sky is excluded by the veto: $\Omega_{\mathrm{excluded}} = 0$. Below 300 Hz, one can compute a uniform average of the excluded sky fraction over the spin-down range considered in this analysis. As shown in Fig. 10 this gives an excluded sky fraction of 36% at 50 Hz and 6% just below 300 Hz.

## APPENDIX B: ILLUSTRATIVE EXAMPLES OF TYPICAL POST-PROCESSING RESULTS

This Appendix shows typical examples of different types of post-processing results, in four different 10 Hz bands. Figure 16 shows a 10 Hz frequency band containing pure Gaussian noise. Figure 18 shows the frequency band of the hardware-injected signal Pulsar2. Figure 17 shows a "quiet" 10 Hz band of real instrument data without any "noisy" lines. In contrast to this, Fig. 19 shows a noisy band which is polluted by instrumental noise artifacts.

As described in Sec. VIII, each of these plots shows the number of coincidences maximized over the entire sky and full spin-down range. The color indicates the numbers of coincidences, where the same color scale has been used in each figure.

[1] A. Einstein, Sitzungsber. K. Preuss. Akad. Wiss. 154 (1918).
[2] A. Einstein, Sitzungsber. K. Preuss. Akad. Wiss. 688 (1916).
[3] J. M. Weisberg and J. H. Taylor, in *Binary Radio Pulsars*, ASP Conference Series Vol. 328, edited by F. A. Rasio and I. H. Stairs (Astronomical Society of the Pacific, San Francisco, 2005), p. 25.
[4] J. Hough and S. Rowan, J. Opt. A Pure Appl. Opt. **7**, S257 (2005).
[5] A. Abramovici *et al.*, Science **256**, 325 (1992).
[6] B. Barish and R. Weiss, Phys. Today **52**, 44 (1999).
[7] D. Sigg (LSC), Classical Quantum Gravity **23**, S51 (2006).
[8] L. Bildsten, Astrophys. J. **501**, L89 (1998).
[9] G. Ushomirsky, C. Cutler, and L. Bildsten, Mon. Not. R. Astron. Soc. **319**, 902 (2000).
[10] C. Cutler, Phys. Rev. D **66**, 084025 (2002).
[11] A. Melatos and D. J. B. Payne, Astrophys. J. **623**, 1044 (2005).
[12] B. J. Owen, Phys. Rev. Lett. **95**, 211101 (2005).
[13] B. J. Owen, L. Lindblom, C. Cutler, B. F. Schutz, A. Vecchio, and N. Andersson, Phys. Rev. D **58**, 084020 (1998).
[14] N. Andersson, K. D. Kokkotas, and N. Stergioulas, Astrophys. J. **516**, 307 (1999).
[15] D. I. Jones and N. Andersson, Mon. Not. R. Astron. Soc. **331**, 203 (2002).
[16] C. Van Den Broeck, Classical Quantum Gravity **22**, 1825 (2005).
[17] J. W. Cooley and J. W. Tukey, Math. Comput. **19**, 297 (1965).
[18] N. Brenner and C. Rader, IEEE Acoustics, Speech and Signal Processing Mag. **24**, 264 (1976).
[19] P. Jaranowski, A. Królak, and B. F. Schutz, Phys. Rev. D **58**, 063001 (1998).
[20] P. Jaranowski and A. Królak, Phys. Rev. D **61**, 062001 (2000).
[21] R. Balasubramanian, B. S. Sathyaprakash, and S. V. Dhurandhar, Phys. Rev. D **53**, 3033 (1996).
[22] B. J. Owen, Phys. Rev. D **53**, 6749 (1996).
[23] T. A. Apostolatos, Phys. Rev. D **52**, 605 (1995).
[24] P. Brady, T. Creighton, C. Cutler, and B. F. Schutz, Phys. Rev. D **57**, 2101 (1998).
[25] P. Brady and T. Creighton, Phys. Rev. D **61**, 082001 (2000).
[26] B. Abbott *et al.* (LIGO Scientific Collaboration), Phys. Rev. D **69**, 082004 (2004).
[27] B. Abbott *et al.* (LIGO Scientific Collaboration), Phys. Rev. D **76**, 082001 (2007).
[28] The distributed computing project Einstein@Home can be found at http://einstein.phys.uwm.edu/.
[29] D. P. Anderson, J. Cobb, E. Korpela, M. Lebofsky, and D. Werthimer, Commun. ACM **45**, 56 (2002).
[30] Stefan M. Larson, Christopher D. Snow, Michael Shirts, and Vijay S. Pande, arXiv:0901.0866v1 (http://arxiv.org/abs/0901.0866).
[31] B. Abbott *et al.* (LIGO Scientific Collaboration), M. Kramer, and A. G. Lyne, Phys. Rev. D **76**, 042001 (2007).
[32] B. Abbott *et al.* (LIGO Scientific Collaboration), Phys. Rev. D **77**, 022001 (2008).
[33] B. Abbott *et al.* (LIGO Scientific Collaboration), Classical Quantum Gravity **24**, 5343 (2007).
[34] B. Abbott *et al.* (LIGO Scientific Collaboration), Phys. Rev. D **77**, 062004 (2008).
[35] B. Abbott *et al.* (LIGO Scientific Collaboration), Phys. Rev. D **77**, 062002 (2008).
[36] B. Abbott *et al.* (LIGO Scientific Collaboration), Astrophys. J. **659**, 918 (2007).
[37] B. Abbott *et al.* (LIGO Scientific Collaboration), Phys.

Rev. D **76**, 082003 (2007).

[38] B. Abbott *et al.* (LIGO Scientific Collaboration), Phys. Rev. D **76**, 022001 (2007).

[39] B. Allen and G. Mendell (LSC CW Working Group), LIGO Technical Document T040164-01-Z, 2004.

[40] X. Siemens *et al.*, Classical Quantum Gravity **21**, S1723 (2004).

[41] A. Dietz *et al.*, LIGO Technical Document T050262, 2005.

[42] Neil Ashby, Living Rev. Relativity **6**, 1 (2003), http://www.livingreviews.org/lrr-2003-1

[43] The Einstein@Home project is built upon the BOINC (Berkeley Open Infrastructure for Network Computing) architecture described at http://boinc.berkeley.edu/.

[44] D. P. Anderson, "A System for Public-Resource Computing and Storage," in GRID '04: Proceedings of the Fifth IEEE/ACM International Workshop on Grid Computing, 2004, pp. 4–10.

[45] D. P. Anderson, C. Christensen, and B. Allen, Proceedings of the 2006 ACM/IEEE conference on Supercomputing, 2006, p. 126.

[46] B. Krishnan, B. Owen, R. Prix, and A. Sintes, "Searching for Isolated Pulsars using Einstein@Home" (unpublished); LIGO Technical Document T080340-00-Z, 2008.

[47] P. Astone, K. M. Borkowski, P. Jaranowski, and A. Królak, Phys. Rev. D **65**, 042003 (2002).

[48] H. J. Pletsch, M.S. thesis, University of Wisconsin–Milwaukee, 2006; UWM Library QC 1000.P7266.

[49] H. J. Pletsch, Phys. Rev. D **78**, 102005 (2008).

[50] John P. Snyder, *Flattening the Earth: Two Thousand Years of Map Projections* (U. of Chicago Press, Chicago, 1993), pp. 130–133.