# Supporting Information

Yao et al. 10.1073/pnas.1212790110

## SI Text

**Data Analysis.** The waveform data (downsampled to 10 Hz) for the four earthquakes recorded by the USArray or Hi-net stations (Fig. S1) are 0.05- to 4-Hz bandpass filtered, and the instrument responses are removed. We only use waveforms with P-wave signal-to-noise ratios above 20. Next, we align the waveforms using the first 8 s of the P waves using a cross-correlation method (1). Only waveforms with cross-correlation coefficients above 0.6 with respect to their reference stack are retained for further analysis (Fig. S2).

For the Tohoku, Maule, and 2005 Sumatra earthquakes, we divide the rupture region into $10 \times 10$ km source grids, but use $20 \times 20$ km grids for the 2004 Sumatra earthquake because this earthquake has a much larger rupture area (2). The equation systems (Eq. **1**) to be minimized are large and underdetermined, e.g., for the Tohoku earthquake we have 402 stations and $41 \times 41 =$ 1,681 potential source points, giving a complex-valued equation system with 1,681 unknowns and 402 observations. To solve Eq. **1**, we use the CVX package (4). A sliding window approach (5) is used to obtain the spectral data (complex variable) in each time window and at each frequency. Within the frequency band of 0.2–0.6 Hz, we use a 10-s time window, and for 0.06–0.2 Hz, a 20-s window, both sliding in steps of 2 s for the Tohoku, Maule, and 2005 Sumatra earthquakes. However, for the 2004 Sumatra earthquake, a 20-s time window sliding in 5-s steps is used for the entire 0.06- to 0.6-Hz band, considering its ~600-s-long rupture time (2, 6). We Fourier transform the waveform data to obtain spectra in each time window, using 128 data points for the 10-s window giving a frequency spacing of 0.08 Hz. For the 20-s window, we use 256 data points giving a frequency spacing of 0.04 Hz. The damping parameter $\lambda$ in Eq. **1** is set to $0.25\sqrt{N}$, where $N$ is the number of stations (5) for all earthquakes after careful resolution tests using synthetic data. For the detailed setup of CS, we refer to ref. 5.

**Synthetic Tests.** We perform synthetic tests to assess the resolution and reliability of the compressive-sensing (CS) method at various frequencies and compare it with results from the conventional beamforming method (7). For single-source problems, CS and conventional beamforming generally can identify the model source location equally well. However, CS has much better resolution than conventional beamforming in synthetic tests with two (Figs. S3 and S4) or three sources (Fig. S5).

In the two-source synthetic tests (Fig. S3), we generate synthetic spectral data (with 5% random Gaussian noise added) for the stations in the United States (Fig. S1, *Upper Left*) due to two synthetic sources in the region of the Tohoku earthquake. For all of the synthetic tests, the input sources are exactly located on the grid points where we invert for source spectra from CS. At 0.55 Hz, CS recovers the input two sources very well and there are two nearby sources (with amplitudes of 0.9 and 0.05) together representing the input source 1. Beamforming resolves these two sources quite well at 0.55 Hz for this source geometry. However, when two sources are set along the direction to the seismic array (Fig. S1, *Upper Left*), beamforming only shows one peak due to interference of the two sources but CS resolves the two input sources very well (Fig. S4). At 0.20 Hz, CS gives two source patches, each with two nearby sources (Fig. S3). The two largest sources are at the correct input source locations. The recovered source patches closely match the input two single sources. At 0.08 Hz, CS also recovers two patches of sources with locations only slightly off (about 10–30 km) the true locations (Fig. S3). The recovered source patch near the input source location 2 consists of three nearby sources. However, at lower frequencies (e.g., 0.20 Hz and 0.08 in Fig. S3), beamforming clearly fails to resolve the input sources, and only shows one peak between the two sources due to severe source spectral interference. These synthetic tests show that CS resolves two sources better than conventional beamforming in the 0.06- to 0.6-Hz frequency band of our analysis (Figs. S3 and S4).

In the three-source synthetic tests (Fig. S5), beamforming clearly fails to recover the source locations at both 0.55 and 0.2 Hz. However, CS recovers three source patches at or near the input source locations at 0.55 and 0.2 Hz. Similar to beamforming, CS cannot resolve the three sources at low frequencies, e.g., at 0.08 Hz, which reflects the resolution limit of CS at low frequencies for multiple sources. Nevertheless, CS is superior to beamforming in resolving multiple sources in the frequency band studied (0.06–0.6 Hz).

**Real Data Examples.** In our analyses of real data, we have found cases in which the CS and beamforming results largely agree, but also instances where they give different results. In Fig. S6, we show comparisons between CS and beamforming results at 0.2 Hz using the real data for two time windows of the 2010 Maule earthquake. Using the time window of 46–56 s, the CS and beamforming results (Fig. S6, lower row) seem to find very similar source locations. However, using the time window of 39–49 s, the CS and beamforming results (Fig. S6, upper row) are different: beamforming merges the two source patches as seen in the CS results (between latitude 34°S and 36°S) into one patch. Because the true source locations for the real data are unknown, it is not easy to assess differences between CS and beamforming, but our synthetic tests show that the CS results are generally more reliable.

1. Yao H, Shearer PM, Gerstoft P (2012) Subevents location and rupture imaging using iterative backprojection for the 2011 Tohoku Mw 9.0 earthquake. *Geophys J Int* 190(2): 1152–1168.
2. Ammon CJ, et al. (2005) Rupture process of the 2004 Sumatra-Andaman earthquake. *Science* 308(5725):1133–1139, 10.1126/science.1112260.
3. Chu RS, et al. (2011) Initiation of the great Mw=9.0 Tohoku-Oki earthquake. *Earth Planet Sci Lett* 308(3-4):277–283.
4. Grant M, Boyd SP (2011) CVX: Matlab software for disciplined convex programming, version 1.21. Available at http://cvxr.com/cvx. Accessed January 25, 2011.
5. Yao H, Gerstoft P, Shearer PM, Mecklenbräuker C (2011) Compressive sensing of the Tohuku-Oki Mw 9.0 earthquake: Frequency-dependent rupture modes. *Geophys Res Lett* 38(20):L20310, 10.1029/2011GL049223.
6. Ishii M, Shearer PM, Houston H, Vidale JE (2005) Extent, duration and speed of the 2004 Sumatra-Andaman earthquake imaged by the Hi-Net array. *Nature* 435(7044): 933–936.
7. Gerstoft P, Fehler MC, Sabra KG (2006) When Katrina hit California. *Geophys Res Lett* 33 (17):L17308.

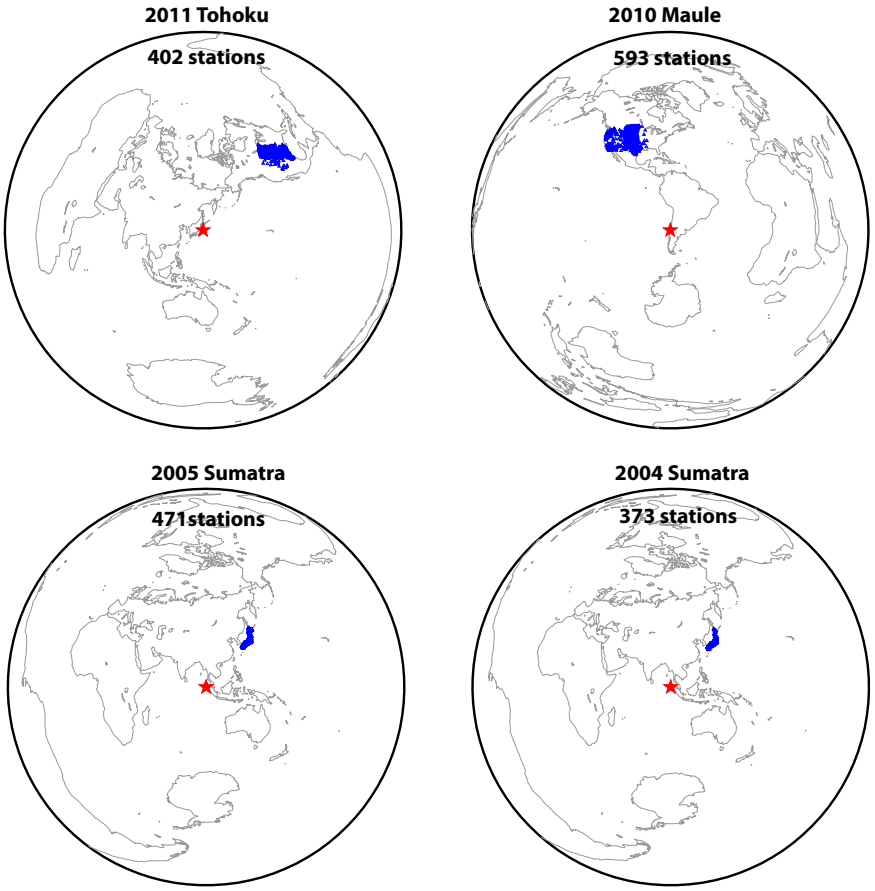


**Fig. S1.** Spatial distribution of the hypocenters (red stars) and stations (blue triangles) used for each earthquake. We used stations in the United States for the analysis of the Tohoku and Maule earthquakes, including nearly 400 USArray stations. For the 2004 and 2005 Sumatra earthquakes, we used the Japanese Hi-net array stations. The number of stations with good-quality waveform data (*Materials and Methods*) is indicated in each plot. For the Tohoku earthquake, the hypocenter is from ref. 3. For the other three earthquakes, the hypocenter is from the catalog of National Earthquake Information Center of the US Geological Survey.

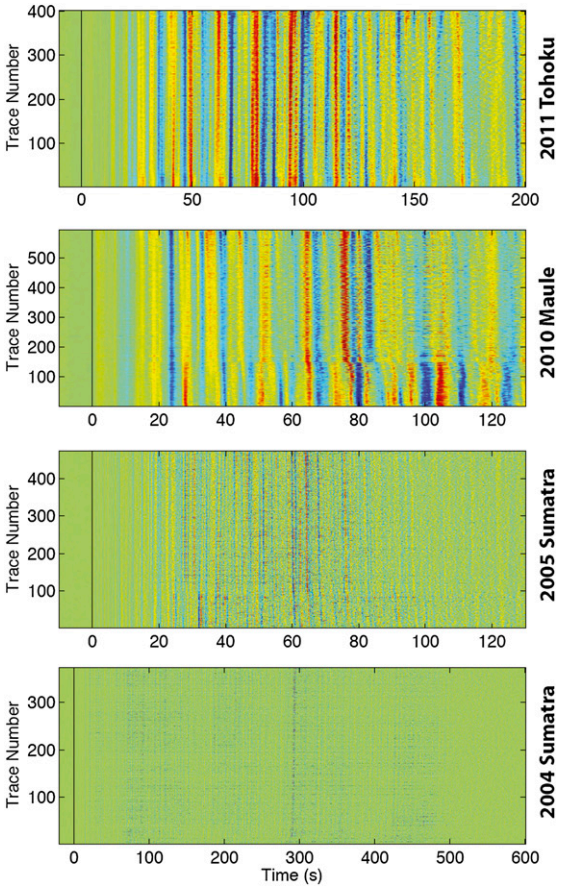


**Fig. S2.** The aligned waveforms (in the frequency band of 0.05–4 Hz) with good quality used for the CS analysis of each earthquake. The waveforms recorded by the Hi-net stations for the 2004 and 2005 Sumatra earthquakes appear to be dominated by higher frequency energy than the waveforms recorded by the stations in the United States for the Tohoku and Maule earthquakes. This is mainly because the Hi-net stations use short-period high-sensitivity seismographs that suppress energy less than 1 Hz. However, the stations used in the US are all equipped with broadband seismographs.

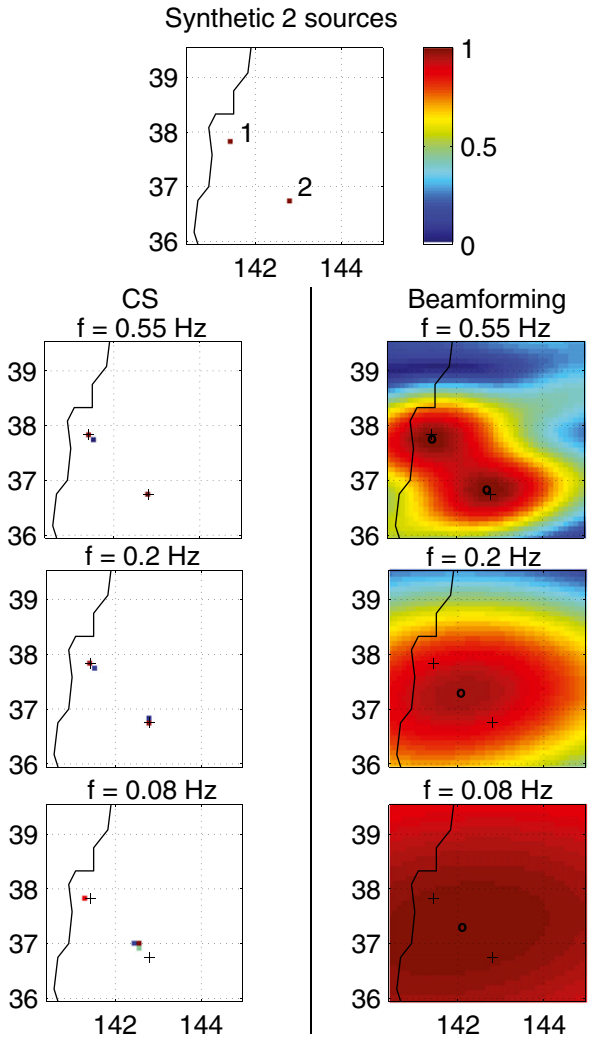


**Fig. S3.** Synthetic tests for a two-source problem (*Top*) using the compressive sensing method (left column) and the conventional beamforming method (right column) at frequencies 0.55 Hz (second row), 0.20 Hz (third row), and 0.08 Hz (last row). The input sources have unit amplitude. The crosses show the input source locations in the synthetic test (*Top*). The circles show beamforming peaks. The color bar shows the source amplitude or beamforming amplitude.

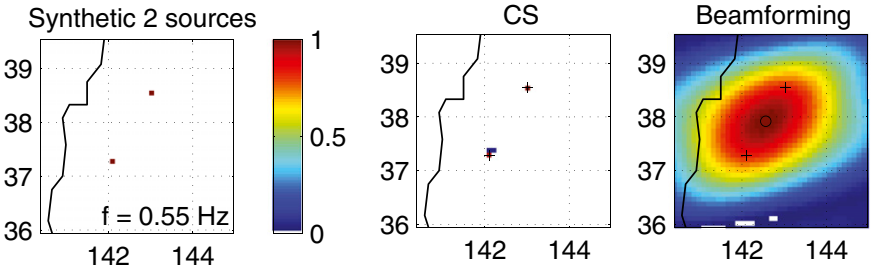


**Fig. S4.** Another synthetic test at $f = 0.55$ Hz for a two-source problem similar to Fig. S3.

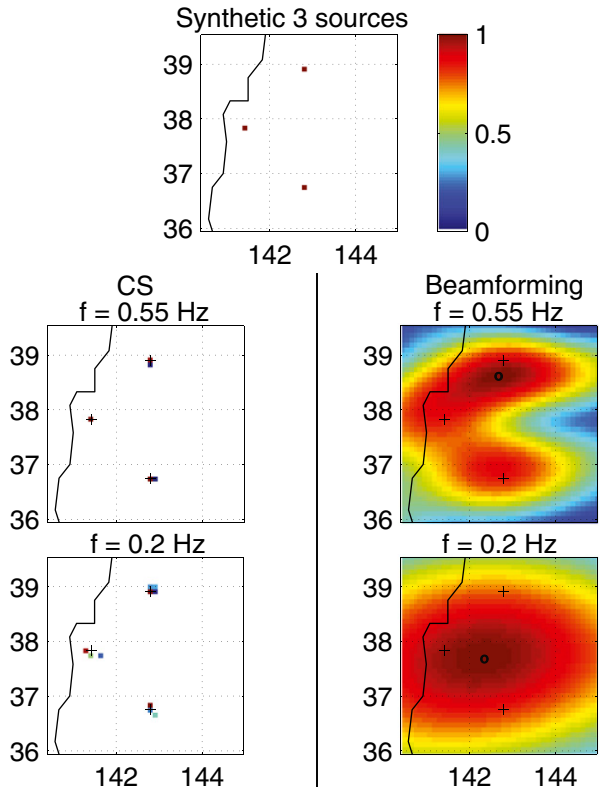


**Fig. S5.** Similar to Fig. S3 but for three-source synthetic tests at 0.55 and 0.2 Hz.

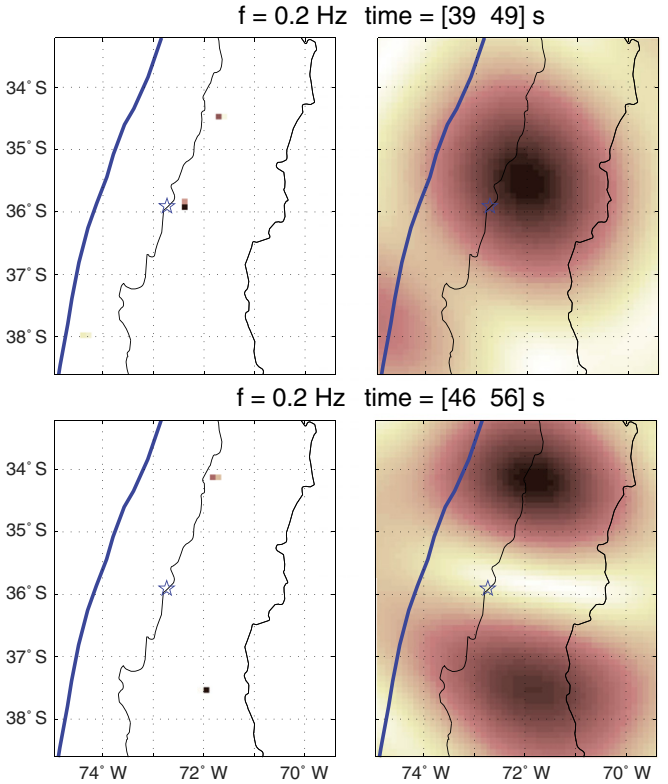


**Fig. S6.** Maule earthquake: comparison between the CS (*Left*) and beamforming (*Right*) results at 0.2 Hz for the two time windows: (*Upper*) between 39 and 49 s; (*Lower*) between 46 and 56 s. See Fig. S2 for the waveforms of the Maule earthquake.