Digitization of the Carnegie Analog Broadband Instruments Tape Records (1965–1996)

Steven Golden*[1], Lara S. Wagner[1], Brian Schleigh[1], Daniela Power[1], Diana C. Roman[1], Selwyn I. Sacks[1], and Helen Janiszewski[1,2]

Abstract

Between 1965 and 2003, the Carnegie Institution of Washington's Department of Terrestrial Magnetism operated a continuous network of nine broadband seismographs with locations in South America, Japan, Iceland, Papua New Guinea, and Washington, D.C. The Carnegie seismographs designed in the 1960s by Selwyn Sacks were among the earliest broadband instruments, sensing between at least 30 s and ~30 Hz. Given the scarcity of historic seismic data of comparable bandwidth and dynamic range prior to the widespread shift to force-feedback instruments and digital recording around the mid-1980s, this dataset is still of high scientific value today.

The Carnegie seismographs recorded data to magnetic tapes meant to be read and analyzed using a custom playback system. Since 1989 these tapes have been stored in a climate-controlled, electromagnetically shielded room, which preserved them in reasonably good condition. However, some tapes now show signs of moisture damage, and reading them is difficult and time consuming by today's standards, creating a barrier to the use of this dataset. To overcome these issues, we have undertaken an ongoing effort to digitize this dataset with the goal of making it publicly available in Standard for the Exchange of Earthquake Data (SEED) format at the Incorporated Research Institutions for Seismology Data Management Center (IRIS DMC).

Cite this article as Golden, S., L. S. Wagner, B. Schleigh, D. Power, D. C. Roman, S. I. Sacks, and H. Janiszewski (2020). Digitization of the Carnegie Analog Broadband Instruments Tape Records (1965–1996), *Seismol. Res. Lett.* **XX**, 1–11, doi: 10.1785/0220190334.

Supplemental Material

Introduction

In the 1960s, scientists at the Carnegie Institution of Washington's Department of Terrestrial Magnetism (DTM) were interested in using earthquake seismology to study plate structure in the context of the then-new theory of plate tectonics. A major science objective was to compare the inelasticity structure of an island arc (Japan) with that of a continental arc (Peru and Chile in western South America) (Fig. 1). The plan was to compare shear-wave spectra from local earthquakes of variable depth with spectra from the *SKP* phase (~134°); for the *SKP* phase observations, vertical-component seismographs would be sufficient. The envisioned studies required the acquisition of new seismographs with a bandwidth of 30 s to 30 Hz and a dynamic range sufficient to record both weak teleseisms and strong local events. Most of the available seismographs at the time were considered to offer insufficient dynamic range or exhibit considerable distortion of large signals (Sacks, 1966), prompting in-house development of a new instrument.

Seven Carnegie seismographs were built and deployed at nine stations, with two clusters of three stations each in the Andes and in Japan and the remainder spread over other parts of the world (Table 1). The first station was installed in 1965 at Toconce (TCC) in northern Chile. This began the buildup of two triples of stations to study the Andes and Japan, respectively: the Andean continental arc set consisted of the three-component stations Cuzco (CUZ) and TCC for local observations and the single-component station Port Moresby (PMG) in Papua New Guinea for *SKP* observations. The Japan island arc set consisted of the three-component stations Matsushiro (MAT) and Kamikineusu (KMU) in Japan for local observations and the single-component station Trujillo (TRU) in Peru for *SKP* observations.

One seismograph was installed at the DTM on the Carnegie Institution's Broad Branch Road campus in Washington, D.C. The instrument was installed in a sturdy underground facility originally constructed to house a cyclotron. For decades, this

---

1. Department of Terrestrial Magnetism, Carnegie Institution for Science, Washington, D.C., U.S.A.; 2. Now at Department of Earth Sciences, University of Hawai'i at Manoa, Honolulu, Hawaii, U.S.A.

*Corresponding author: sgolden@carnegiescience.edu

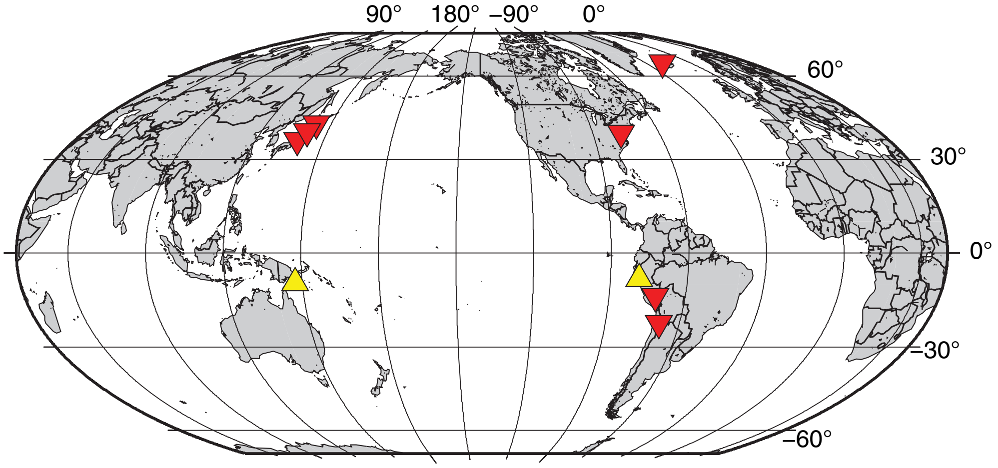

Figure 1. Map of seismic stations included in the Carnegie Analog Broadband Seismographs network. Red inverted triangles show the locations of three-component stations, and yellow triangles show the locations of single-component stations. The color version of this figure is available only in the electronic edition.

instrument streamed data to the Smithsonian National Museum of Natural History for live display on a drum recorder.

In 1971, station TCC was uninstalled, and the instrumentation was reused the following year to install station Akureyri (AKU) in Iceland. The purpose of AKU was to study the lithospheric growth rate in a sea-floor spreading region (Evans and Sacks, 1980). The instrumentation at station Matsushiro (MAT) was moved in 1984 to a site farther north near the village of Sawauchi (SWU) in northern Honshu. Unless stations were uninstalled for equipment relocation, they were kept in operational condition for as long as they were considered to be providing data of scientific value. For most stations, this was into the 1990s and in the case of Akureyri until 2003.

Our current efforts are to digitize as much of these data as possible using the original playback equipment coupled with modern digitization hardware. We plan to make all data publicly available through the Incorporated Research Institutions for Seismology Data Management Center (IRIS DMC) as soon as it has been fully quality controlled. What follows is a brief description of the instrumentation that was developed and used for this network, an overview of the digitization procedures, and a description of the work that has been accomplished so far.

Instrumentation and Station Operation

Detailed information on the design of the instrumentation used to collect these data, including the rationale behind building them and the operational procedures, can be found in Sacks (1966). Here, we provide a brief overview of this information for convenience. For further details, we refer readers to Sacks (1966).

The Carnegie analog broadband seismographs

The Carnegie seismographs were designed to record ground motion in a frequency band covering at least 30 s to 30 Hz with high dynamic range and low distortion. A series of lab experiments conducted at the DTM throughout the early

TABLE 1

Location, Years of Operation, and Number of Installed Components for All Nine Carnegie Seismograph Stations

Region	Location (Code)	Coordinates	Timespan	Components
Andes	Cusco, Peru (CUS)	13.563° S, 71.877° W	1966–1986	3
	Toconce, Chile (TCC)	22.275° S, 68.172° W	1965–1971	3
	Trujillo, Peru (TRU)	8.078° S, 78.861° W	1967–1986	1
Japan	Kamikineusu (KMU)	42.238° N, 142.967° E	1967–1996	3
	Matsushiro (MAT)	36.543° N, 138.207° E	1967–1984	3
	Sawauchi (SWU)	39.490° N, 140.790° E	1984–1996	3
Other	Akureyri, Iceland (AKU)	65.686° N, 18.099° W	1972–2003	3
	Port Moresby, Papua New Guinea (PMG)	9.406° S, 147.159° E	1966–1992	1
	Washington DC (DTM)	38.959° N, 77.063° W	1966–1994	3

1960s had found standard coil-spring-based seismograph designs unsatisfactory because of significant crosstalk and nonlinear distortion in the presence of large signals, such as those caused by large or local events. One underlying problem was identified to be the presence of higher mode oscillations in the coil springs, especially sideways violin-spring oscillations. To overcome this, a new design was conceived that would submerge a few coil spring windings in oil to dampen any higher mode oscillations without significantly affecting the main mode of the spring. The seismometers not only met the original design goal of a bandwidth from 30 s to 30 Hz but turned out to exceed it (see the Progress to Date section).

Tape recording

Another revolutionary decision was to have the seismographs record to magnetic tape. This would allow easier access to new analysis techniques, such as audio playback of the signals at advanced speed to quickly screen a recording for events, repeated playback of an event through different analog filters to analyze frequency content, and, eventually, relatively easy digitization of events.

The primary challenge with magnetic tape recordings was how to record at extremely low tape speeds to make one tape last for as long as possible. In the case of the Carnegie seismographs, a nominal tape speed of 0.006 inch/s allowed each tape to capture approximately three months of data. Signals were recorded unmodulated with AC bias; FM recording was considered for its superior low-frequency performance but rejected for insufficient dynamic range at higher frequencies when recording at such low tape speeds. With an unmodulated recording, the maximum dynamic range of the magnetic media was estimated to be ~40 dB, but by recording five copies of each signal using different gain and filter combinations, it became possible to extend the effective dynamic range to ~60 dB over a bandwidth of three octaves.

Each seismometer component was recorded in parallel onto five tape tracks using different gain and filter combinations within the electronics between sensor and tape head: filter settings were chosen such that each seismometer produced two channels of different gain proportional to velocity (SPM and SPH) and three channels of different gain proportional to displacement (LPL, LPM, and LPH), all with the gain increased by a factor of 20 between successive gain levels. This resulted in a total of 16 tracks at three-component stations and six tracks at single-component stations. Hence, the Carnegie seismographs were built in two different recording configurations: three-component stations recorded 16 tracks on 1-inch "wide" tape, whereas single-component stations recorded only six tracks on ½-inch "narrow" tape.

Fitting 16 tracks on 1-inch tape or six tracks on ½-inch tape required an unusually high track density not offered by tape head technology available at the time. To overcome this, the Carnegie recording and replay decks used pairs of heads arranged with a slight transverse offset to each other, such that their tracks were interleaved. Three-component stations used pairs of eight-track heads for a total of 16 tracks, whereas single-component stations used pairs of four-track heads for eight tracks, of which six were used for seismic applications.

The use of multiple recording and playback heads introduced an important potential error source: if the separation and alignment of the playback heads did not match those of the recording heads exactly, a small time offset between odd and even tracks could be introduced during playback. At the nominal tape recording speed of 0.006 inch/s, each millimeter of tape head misalignment translates to a time offset of 6.56 s. Examination of different recordings showed systematic timing errors between odd and even tracks on a typical order of 3 s, which corresponds to a head misalignment on the order of 0.5 mm. It is important to note that this error source does not affect any analysis that makes use of only odd or only even tracks.

As can be seen in Table 2, except for channel 12 (LPL-EW), all long-period (LP) channels are handled via head 1, whereas all short-period (SP) channels and the time code are handled via head 2. This particular layout was chosen to minimize misinterpretation potential caused by expected crosstalk between adjacent channels. Channel combinations for which exact interchannel timing matters most—that is mid and high gain channels that share the same band—always share a single head. This means that the timing for the horizontal components and vertical component for the SPM, SPH, LPM, and LPH recordings are identical, making these recordings ideally useful for single-station analyses such as receiver functions or shear-wave splitting that depend on the relative timing between orthogonal components but less so on the absolute timing of any given phase. Single-component ½-inch "narrow" tapes share the same track layout, except that they only make use of the first six tracks. (Half-inch tapes allow recording of up to eight tracks, but only six are needed for a single-component seismic station; the last two tracks remained unused.)

Station timekeeping and time track format

Station time was measured with high precision mechanical clocks of varying construction, which were fitted with an electromagnetic pickup to feed a "tick" per pendulum swing into the station's electronics. The latter then counted seconds, minutes, hours, and in later revisions also days, and from this generated a time code that was recorded on tape track number 6. The encoding of station time on the time track uses a custom analog digital hybrid format (Fig. 2), which consists primarily of a sequence of approximately rectangular second pulses, with the first pulse of each minute greatly enlarged in amplitude and duration (spanning two seconds).

Superimposed onto sections of the time track's continuous stream of second and minute pulses are high-amplitude sine signals marking the start of every hour and day: the first

TABLE 2

List of Tape Tracks, Their Assignment to Tape Head (Channel) within the Tape Head, the Component They Are Recording, and the Band (Gain) Code

Track	Original Channel Name (SEED name)	Head (Channel) within Head	Component	Recording Unit	Gain
1	LPL-V (LLV)	1 (1)	Vertical	Displacement	1 (low)
2	SPM-V (SMV)	2 (1)	Vertical	Velocity	20 (medium)
3	LPM-V (LMV)	1 (2)	Vertical	Displacement	1 (low)
4	SPH-V (SHV)	2 (2)	Vertical	Velocity	400 (high)
5	LPH-V (LHV)	1 (3)	Vertical	Displacement	400 (high)
6	TIME (TIM)	2 (3)	Time		
7	LPL-NS (LLN)	1 (4)	Horizontal-NS	Displacement	1 (low)
8	SPM-NS (SMN)	2 (4)	Horizontal-NS	Velocity	20 (medium)
9	LPM-NS (LMN)	1 (5)	Horizontal-NS	Displacement	20 (medium)
10	SPH-NS (SHN)	2 (5)	Horizontal-NS	Velocity	400 (high)
11	LPH-NS (LHN)	1 (6)	Horizontal-NS	Displacement	400 (high)
12	LPL-EW (LLE)	2 (6)	Horizontal-EW	Displacement	1 (low)
13	SPM-EW (SME)	1 (7)	Horizontal-EW	Velocity	20 (medium)
14	LPM-EW (LME)	2 (7)	Horizontal-EW	Displacement	20 (medium)
15	SPH-EW (SHE)	1 (8)	Horizontal-EW	Velocity	400 (high)
16	LPH-EW (LHE)	2 (8)	Horizontal-EW	Displacement	400 (high)

EW, east–west; L/M/H, low/medium/high gain respectively; LP/SP, long and short periods; NS, north–south; V, vertical.

minute of every hour has a very high amplitude 5 Hz sine added; the first 15 min of every hour, as well as every minute of the first hour in a new day, have a slightly smaller 0.25 Hz sine added. The rationale behind these signals was to aid manual playback system operation: a separate pickup head in the playback system made it possible to route the time track to a speaker during fast forwarding, which would turn the hourly sine signals into beeps to guide the operator toward a sought timespan on the tape. Unfortunately, these sine signals tend to be highly distorted and overlap in the frequency domain with the second and minute pulses, making automated signal separation difficult.

Starting with a clock upgrade performed at all stations in 1974, a binary coded decimal (BCD) encoded timestamp consisting of day-of-year, hour, and minute was added to each minute: for recordings where this code is present, at the start of each minute, every fifth second pulse is omitted, leaving groups of four pulses each standing between them. This is repeated seven times to form seven pulse groups. Within each pulse group, pulse-width modulation is used to encode the binary ones and zeros of a seven-digit BCD timestamp. Regardless of the varying pulse width of second and minute marks, it is always the starting, nominally rising, flank of the pulse that marks the corresponding second. Even if a recording contains timestamps, year numbers are not part of them and thus must be supplied by auxiliary metadata. Leap years are accounted for in the day numbering thanks to a switch on the station's day counter that let the operator select the number of days in the current year. As a small quirk, the last day of any year is numbered as day 000 of the following year.

Station operation procedures

Proper operating procedures of a Carnegie seismograph station would involve regular, ideally daily, visits by the station's assigned local operator to keep track of the tape counter and to confirm that all instrumentation was running properly. Operators were often not scientists but local laymen trained to use the instruments. Operators documented and corrected any anomalies that were found, and for a subset of station visits (originally about once per week), they performed a station calibration. As documented in the handwritten service sheets, early in the project, the maintenance schedule was well kept, but later on, visits to some stations became less rigorous.

During each station visit, the operator noted the value of a tape counter and the current station time. The operator compared the station clock with the signal of a radio clock and noted any observed discrepancies. If a small mismatch in time was found, the station operator temporarily slowed down or sped

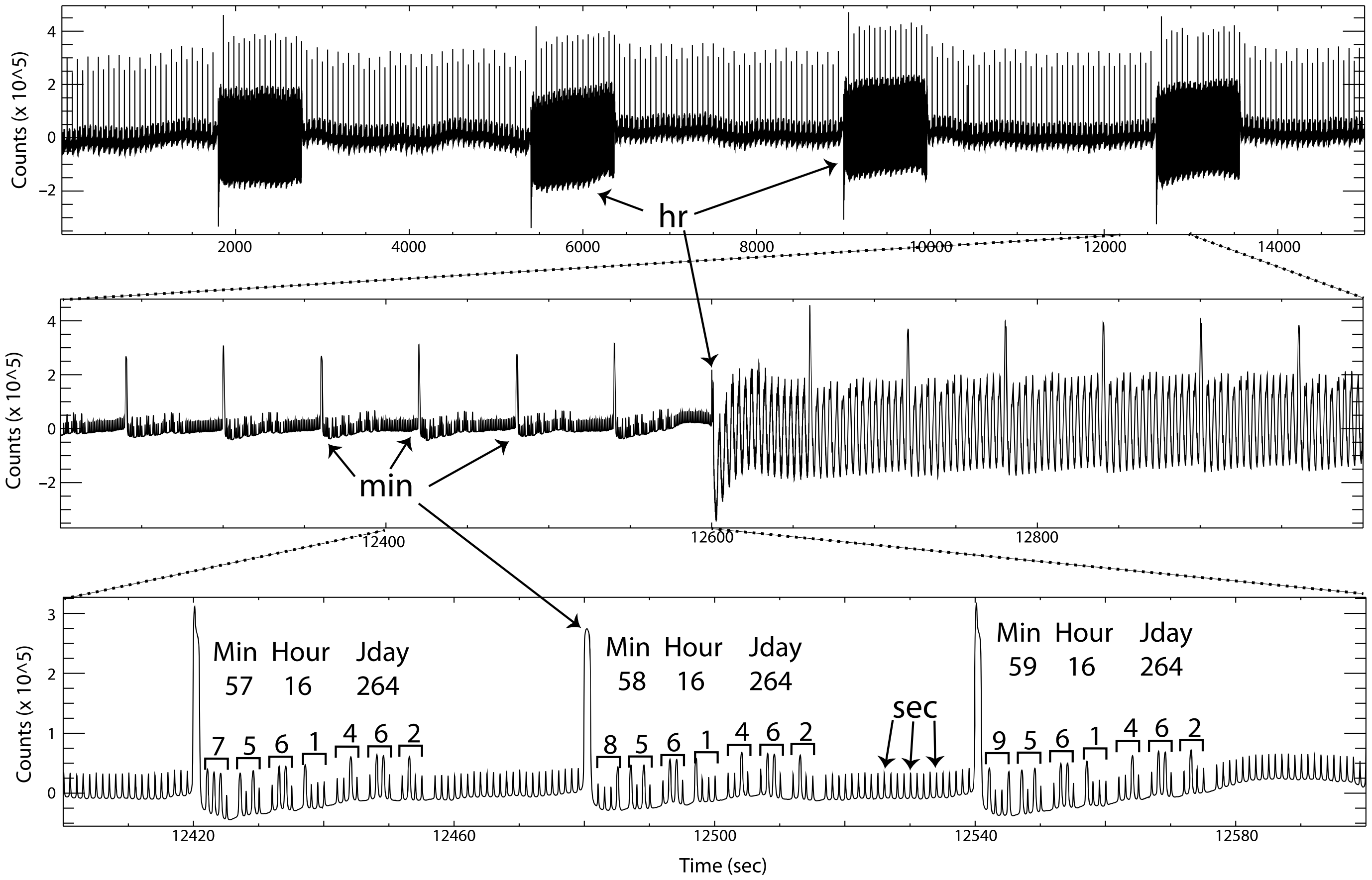

Figure 2. Example of the time track as recorded at station AKU on 21 September 1974 (day 264) covering the timespan 13:30 to 17:30. As seen in the top and center panels, the first 15 min of every hour has a large 0.25 Hz sinusoidal signal superimposed that acts as an hour mark. (Less visible in this representation is an additional 5 Hz signal during the first minute of every hour.) The bottom panel zooms in on three minutes just before the onset of an hour mark to show the minute and second pulses and the way binary coded decimal (BCD) time stamps are encoded into each minute's second pulses. For each four-second cluster of pulses, higher amplitude pulses are "1"s, and lower amplitude pulses are "0"s, allowing for binary decoding as shown.

up the station clock (e.g., by placing a small weight on its mechanical pendulum) and waited for its ticks to sync up with those of the radio clock. Typically, this procedure took only a few minutes and did not create any obvious time jump in the recording. Only very rarely did the operator perform a hard clock reset.

The nature of the timekeeping mechanism and procedure occasionally introduced errors and inconsistencies. For example, sometimes the hour and minute counters were correctly set but not the day counter. Occasionally, the station operator set the station clock to local time instead of universal time. All of these time-mislabeling errors are correctable but require raised awareness during time track decoding and subsequent quality control.

Calibration pulses were used at varying time intervals at all stations, initially weekly, but later on a more-or-less monthly schedule. For calibration purposes, each seismometer featured an extra coil, which was driven by a 4 s oscillator. The oscillator amplitude was then set to either 4, 80, or 1600 μm ground displacement equivalents. A full calibration sequence consisted of three individual calibrations of each component at increasing amplitude settings finished off by a final large pulse excitation of the undamped seismometer (button on the calibrator unit) to time its pendulum swings.

The calibration sequence manifests itself on the recordings as a sequence of sine signals with increasing amplitudes (Fig. 3): The first, 4 μm equivalent amplitude signal was intended to be fully recorded on all affected tape tracks without any saturation. The second, 80 μm equivalent amplitude calibration signal saturates the most sensitive displacement and velocity tracks (LPH and SPH) but is still well recorded on all other tracks (LPM, LPL, and SPM). The third, 1600 μm equivalent amplitude calibration signal was designed to saturate all channels; its purpose was to drive the mass far enough that its movement was easily observable through a window in the seismometer casing allowing the actual mass displacement to be measured via a reticule. Finally, the sequence concluded with a high amplitude free swing of the undamped mass.

When a tape was full or nearly full, it was shipped back to the DTM along with its corresponding service sheets. Each tape

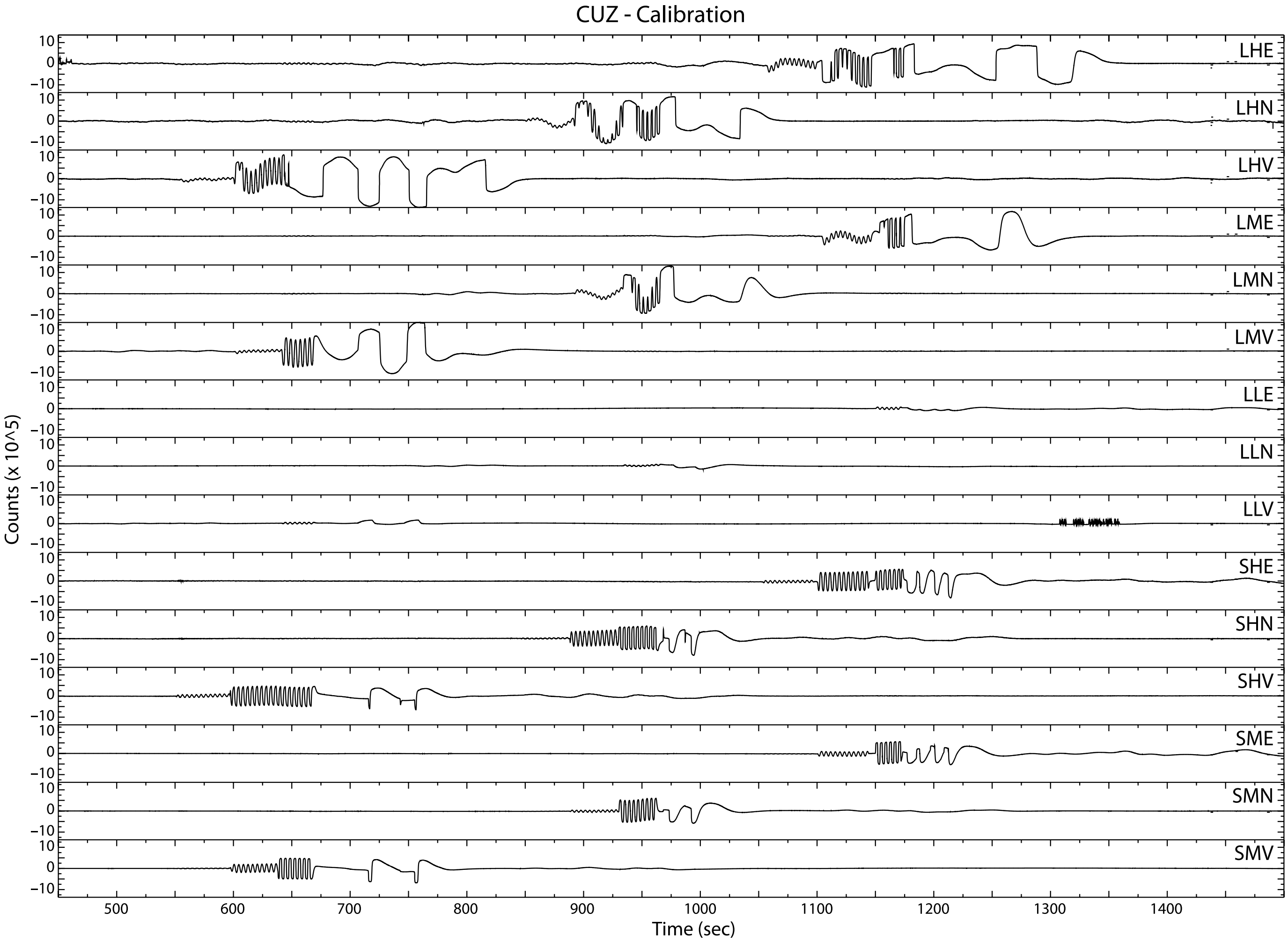

Figure 3. Example recording of a calibration sequence as performed at station MAT on 28 December 1974 on all recorded components. The sequence starts with the vertical component followed by north–south and finally east–west. For each component, the sequence starts with applying a 0.25 Hz signal with three successively increasing amplitude settings, where after the seismometer component was temporarily undampened and had a strong pulse applied to let the operator measures the free period of the mass through direct observation; the signal resulting from the large pulse is not usefully reproduced on tape because of saturation.

could hold about three months of data, but occasional prolonged system outages have caused some tapes' start and end dates to lie further apart than this.

Tape playback

Tapes were played back on a custom-built system consisting of two reel-to-reel tape decks sharing control, amplifier, and filter electronics, one for the wide 16-track tapes and one for the narrow 8-track tapes. Each tape deck contains one high-speed and four regular playback heads. The high-speed head could be used for listening in to the time track during fast-forward operation: by counting the beeps caused by the 4 s and 5 Hz hour and day marks and referencing a tape footage counter, it was relatively easy to navigate the tape to find events using faster (selectable) tape speeds.

Originally, the tape library and the playback systems were located in a space on the top floor of the Carnegie Institution's Abelson building. When the department moved into the campus' newly constructed Research Building in 1989, the tape library was relocated to a specially constructed separately climate controlled and electromagnetically shielded room. The new room was designed to offer ideal storage conditions; however, because of incorrect air ducting, room humidity frequently exceeded specification levels, especially during the first years after construction.

In the late 1980s, a plan was devised to mass-digitize event-based time series using a partially automated system built around a 16-bit digitizer feeding data into a MASSCOMP model computer, which featured a fast bus capable of handling the data rate. Unfortunately, a series of hardware failures and other issues with the chosen computer system eventually led to the abandonment of this first major digitization effort.

Digitization and Archiving Efforts

Digitization

The current digitization attempt began in 2016 using modern hardware. For our initial digitization efforts, we have focused on records after the 1974 station clock upgrades, which added BCD time stamping as described previously and shown in Figure 2. Future efforts will focus on decoding of the early clock time stamps to digitize earlier records. As a first step, we scanned in all surviving service sheets to use in assessing the quality of the data, to identify calibration pulses, and to look for possible timing issues. The data digitization itself is performed using a 16-channel, 16-bit Elsys TraNET 208S digitizer. Between the tape head and digitizer, the signals pass through a new custom-built preamplifier, which replaced the original preamplifier and filter units that had degraded. Through a series of tests with different tapes and preamplifier gain settings, we confirmed that 16 bits could provide sufficient resolution without clipping, although the gain had to be selected very carefully.

Digitization is performed at 100 kHz while the tape is played back at ~1000 times its recording speed, leading to an effective sample rate of ~100 Hz. The highest meaningful frequency content from the recording system is expected to be ~30 Hz, so this sample rate choice should preserve the full information from the original signal. Raw data from the Elsys digitizer are recorded in TPC5 format (see Data and Resources), an Elsys proprietary adaptation of the HDF5 format, which adds digitizer-specific metadata and stream naming conventions. All contained waveform data are readable with standard HDF5 tools and libraries.

The not always ideal storage conditions caused many tapes to absorb moisture into the binder responsible for holding the ferromagnetic oxide particles on the tape's plastic layer, leading to a condition known in the audio community as "sticky shed syndrome." Affected tapes exhibit heightened friction when gliding over the playback head, leading to audible noise inducing vibrations and rapid buildup of shed oxide on the head, the latter acting as a signal attenuating low-pass filter. After some trial and error, we found a mitigation procedure described by Ciletti (see Data and Resources) to work well with this dataset: tapes are "baked" before playback in a modified food dehydrator for about 48 hr at a temperature of 58°–62°C to drive out enough moisture to temporarily reduce binder stickiness to an acceptable level.

Postprocessing of digitized data

The digitized raw data are multichannel time series of the playback heads' output voltages, sampled at 100 kHz during tape playback with a slightly varying speed close to 1000 times the recording speed. These raw data are time stamped with the digitization time, not the original recording time. To make these useful for seismological analysis, the data are resampled to a constant rate in relation to original recording time and time stamped with absolute station time.

The data also need to be integrated to compensate for the time-differentiating effect of the playback head's mechanism: the head output voltage is proportional to the temporal magnetic flux change at the head's gap, which in turn depends on the speed at which the tape passed the head. This means that tape speed fluctuations introduce not only timing errors but also amplitude variations in the digitized signal. Integration before resampling can correct much of this error source.

These considerations led to the adoption of a two-pass postprocessing procedure (for details, see note S1 in the supplemental material to this article). First, the digitized record is decoded to estimate a mapping from raw data sample numbers to station time. A list of time spans with continuous coverage is built up during this process. Second, each digitized record is read again, numerically integrated, and then resampled to a constant rate in relation to station time. The result of this second pass is written to files in miniSEED format.

These postprocessing steps are performed by software we have created named "cabsproc" (for details, see note S2). The many substeps require additional metadata, such as the tape number and station code, the start year and month of a recording, and optional corrective information. To make postprocessing reproducible, all required metadata are provided in one configuration file per tape, which is written in a simple human-readable text format.

Quality control

For each processed tape, an automated script is used to extract sections covering historic large events that should have been captured by the tape run and write them to files in Seismic Analysis Code (SAC) format (Goldstein *et al.*, 2003). The script uses the USGS Earthquake Catalog (see Data and Resources), a selectable cutoff event magnitude, and the table of continuous time span coverage generated by the cabsproc postprocessing software to choose appropriate events. A cutoff magnitude of M6.0 was found to produce a manageable number of events to review. Labels for common phase arrivals are added to the SAC files based on travel-time predictions calculated with TauP (Crotwell *et al.*, 1999). The resulting SAC files are then visually reviewed to see whether these events were recorded at or near their predicted phase arrival times. This procedure offers a quick first check to determine whether the station clock was properly set during the event, whether the time track was properly decoded, and whether all seismograph components were operational.

We also tested the station clock time quality by comparing station calibration metadata and actual calibration pulses recorded at the station. This has, in a number of cases, enabled us to understand timing errors such as the undocumented use of local instead of universal time, or incorrect day numbers, whether caused by decoding issues or station programming errors. Where timing errors were obviously caused by mislabeling (e.g., incorrect hour or day number but correct minute

TABLE 3
Qualitative Data Quality Classification Categories Used for Data Summaries

Category	Numeric Code	Description
Good	1	No serious issues found; clock quality verified to be good within a few seconds by observing known earthquakes
Good?	2	No serious issues found, but timing quality not verified
Calibration	3	A calibration sequence was performed (documented in service sheets or found during data screening)
Conditional (description varies)	4 or 5	Serious data quality issues were found that make the dataset only conditionally usable, depending on individual data user requirements; examples are inaccurate timing, unusually high noise level, failure of a single component, and data fragmented into many short continuous sections
Bad? (description varies)	8	We were unable to recover any potentially useful data but are not certain whether future efforts might be able to retrieve more information
Bad	9	The data appear completely useless; examples: very high electronic or other noise, time code badly corrupted, frequent signal dropouts on tape
Gap	0	Data gaps caused by either station outage or inability to decode section of tape (short gaps suspected to be caused by decoding errors may or may not be closable with advanced reprocessing starting from the digitized raw data)
QC incomplete	−1	Tape has been digitized, but processing and QC have not been completed yet, or preliminary QC has led to ambiguous results
Not digitized	−2	Tape digitization has been skipped for now, typically because of suspicion that the tape may be bad (e.g., because it was preceded by multiple bad tapes in a row)

QC, quality control.

and second markers), we correct them for the archived record. Subtler errors on the order of a few seconds caused by drift of the station clock we simply document (if possible) but leave their correction to the discretion of the data user.

All quality control results are compiled into tables (see Tables S1–S3) for each station that subdivide the station's runtime into a sequence of time spans, each representing either a gap or a continuous section of data coverage, with qualitative ratings of timing and data quality (Table 3). The intent is to let data users quickly decide based on this table whether an event of interest is covered by the dataset and at what approximate quality.

More advanced or quantitative quality control metrics, such as spectral power density functions, can be envisioned but have not yet been systematically implemented for this dataset.

Progress to Date

As of the end of September 2019, 175 tapes (of 812 total) have been digitized (Table 4). As discussed earlier, we have focused our preliminary efforts on digitizing records from three-component stations recorded with the revised clock mechanism during or after 1974. From the quality control documented in the individual spreadsheets for each station (see Tables S1–S3), we find varying rates of good data (quality values 1 or 2) and potentially useable data (quality values 4 or 5) relative to the total amount of time recorded. For AKU, we find only 23.3% of the data are good, and another 32.4% are potentially useable. For CUZ, good data comprise 48.6% of the total data time recorded, but only an additional 6.1% are potentially usable.

TABLE 4
Overview of Data Volume and Digitization Status as of September 2019

Station	Tapes Recorded	Tapes Digitized	Years Digitized
AKU	143	59	1974–1988
CUZ	96	54	1974–1988
DTM	117	0	N/A
KMU	128	20	1974–1979
MAT	80	40	1974–1983
PMG*	103	0	N/A
SWU	54	0	N/A
TCC	18	0	N/A
TRU*	73	0	N/A

Stations in bold are prioritized over other stations. N/A, not applicable.
*Single-component stations.

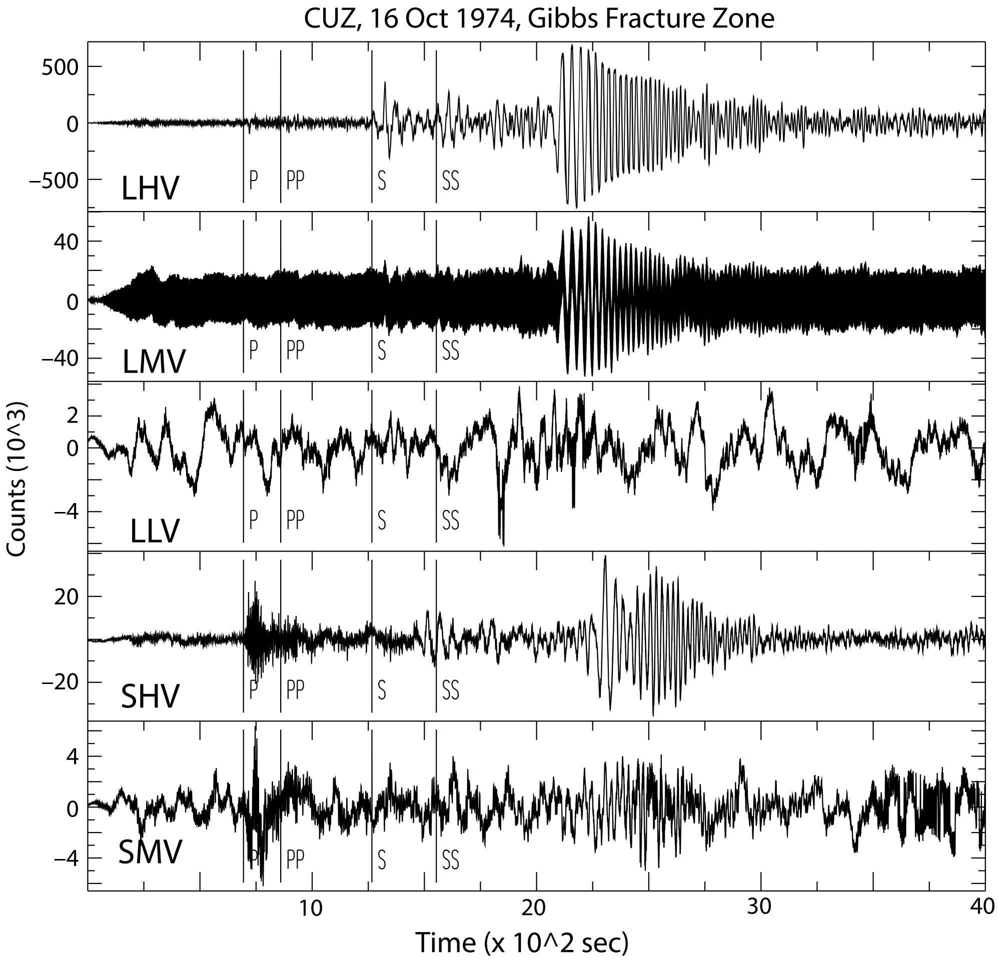

Figure 4. Plot of the five channels that recorded the vertical component of the 16 October 1975 Gibbs Fracture Zone earthquake, band-pass filtered between 200 s and 30 Hz. The top three traces record displacement with high, medium, and low gain. The bottom two traces record velocity with high and medium gain.

MAT has so far produced the most success, with 68.1% of the data qualifying as good and an additional 12.8% as potentially usable. Slightly more useful data (particularly for quality value 8) may be recovered using an improved time decoding algorithm; to allow others to attempt this in the future, we plan to archive digitized raw data along with the time-corrected miniSEED files.

Data quality varies greatly among tapes. In most cases, these differences can be traced back to more-or-less well-documented changes during the recording interval. Differences in how tapes from different manufacturers and batches held up to aging appear to have also played a role. Whereas early data recorded on Scotch brand tape held up comparatively well, data recorded on AMPEX tapes, especially during the late 1970s, were often challenging to recover. We also noticed that tapes recorded at stations located in dry climates were generally better preserved than tapes recorded at stations in humid climates despite identical long-term storage conditions.

Sample data and analyses

To assess the effective bandwidth and future utility of the digitized data set, we analyzed a series of teleseisms recorded at both MAT and CUZ stations. An example of a typical event is shown in Figures 4 and 5 for the CUZ recording of the 16 October 1974 Gibbs Fracture Zone event (M_w 7.0, distance 74.3°, back azimuth 23.9°) (Kanamori and Stewart, 1976; Aderhold and Abercrombie, 2016). Figure 4 shows the differences between the five different recordings of the vertical component for this event. The upper three are the LP traces that record ground displacement. The bottom two are the SP traces that record velocity. For the LP traces, the high gain (top) and middle gain (third trace) both show some evidence for clipping of the high-amplitude surface waves. The SP traces unsurprisingly do not record the high amplitudes of the LP Rayleigh waves but do show clearly incoming *P*-wave arrivals.

Figure 5 shows the high and middle gain LP records, band-pass filtered using a narrow (~100–140 s wide) filter around a central period. These show consistent recordings between the high gain and medium gain traces for periods shorter than or equal to 77 s. We have made similar observations for other events and at other stations, suggesting that for a sufficiently large event, these data faithfully recover periods much longer than the 30 s specifications for which these instruments were produced.

Although the absolute timing of the instruments is somewhat variable, the relative timing between traces is robust. This makes these records valuable for single-station analyses such as receiver functions and shear-wave splitting. To demonstrate this, we calculated *P*-to-*S* receiver functions for four teleseismic earthquakes recorded by the station MAT using the time iterative deconvolution method of Ligorria and Ammon (1999) (Fig. 6). Prior to deconvolution, the data are windowed around the direct *P* arrival, and a high-pass filter with a corner frequency of 0.07 Hz is applied. During deconvolution a

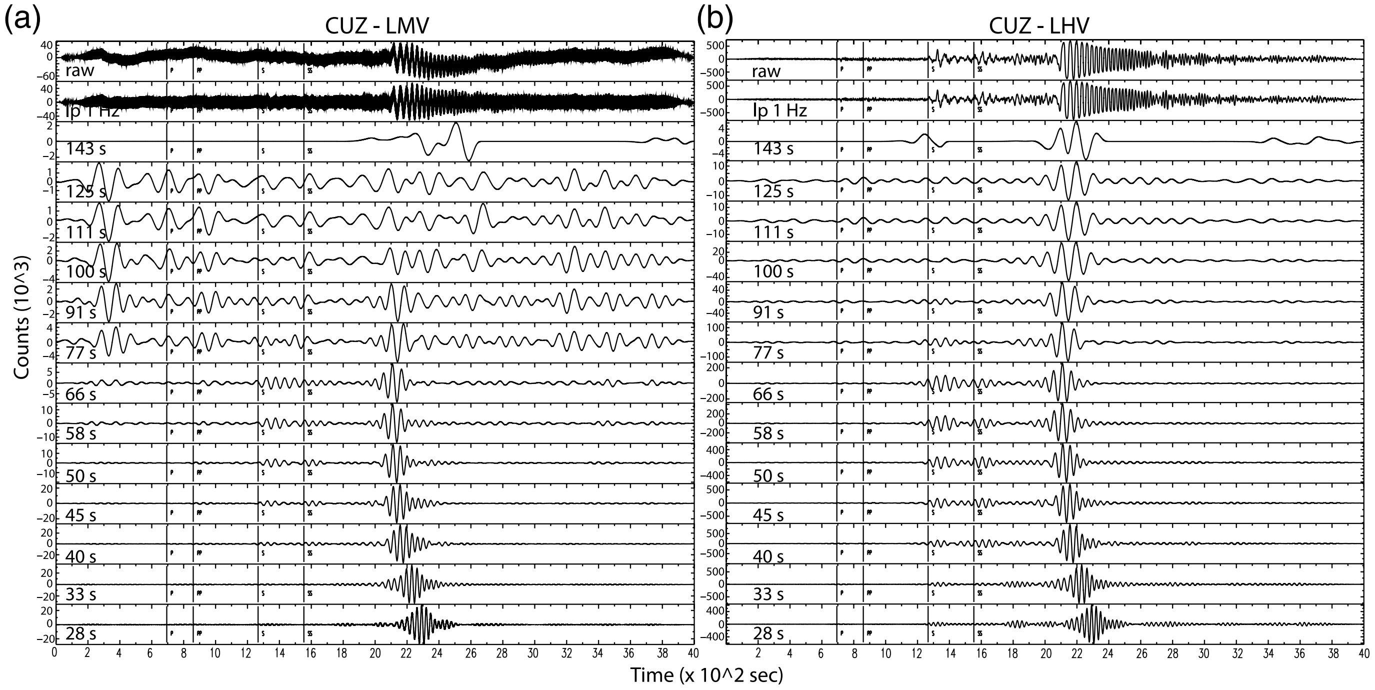

Figure 5. (a) Medium gain and (b) high gain displacement records of the 16 October 1975 Gibbs Fracture Zone earthquake, filtered over a range of different frequencies. Notice the similarities between the records of both tracks for periods shorter than 91 s.

low-pass 0.5 Hz Gaussian filter is applied. The receiver functions are primarily characterized by a large negative arrival at 2 s; a series of positive arrivals from 3 to 5 s, possibly indicating a complex Moho structure; and a positive arrival at 14 s for earthquakes with back azimuths >270 that is otherwise absent, possibly related to the subducting crust. To verify these results, we calculate receiver functions using the same procedure for earthquakes that occurred from 2017 to 2019 recorded at the Global Seismic Network, STS-1 seismic station MAJO, which is located <1 km from MAT. The same key features are observed: a negative arrival followed by a complex set of positive arrivals within the first 5 s and a positive arrival at

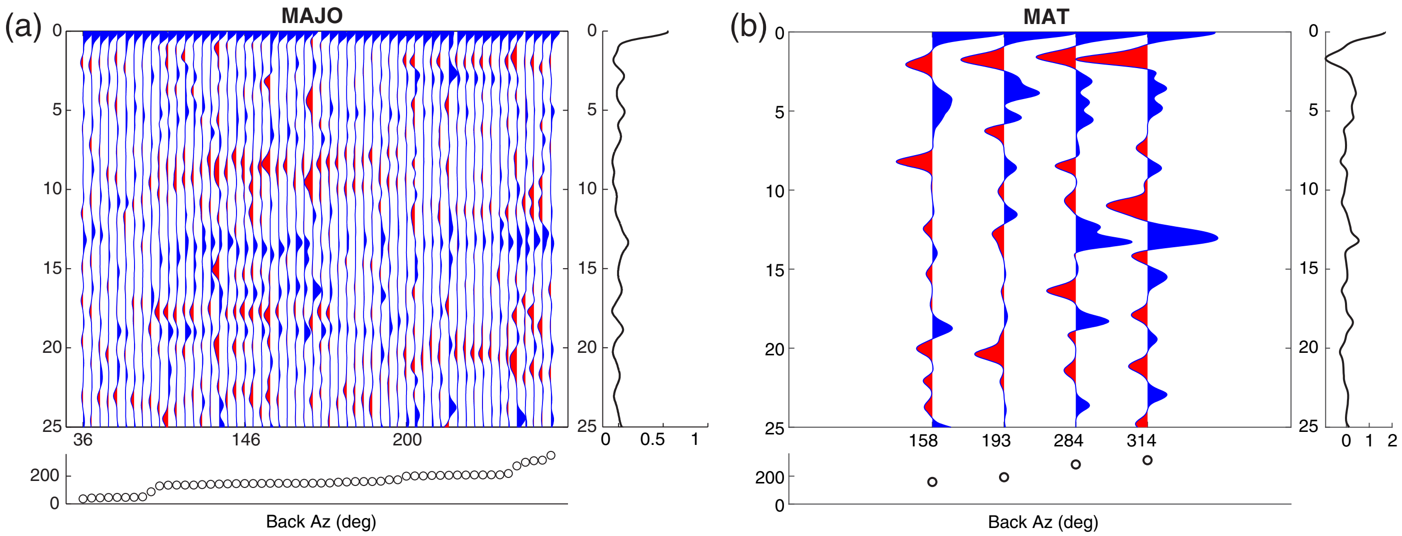

Figure 6. Receiver functions calculated for the Carnegie station MAT (b) and for the modern STS-1 Global Seismic Network station MAJO (a), located less than a mile from the MAT installation site. Blue pulses indicate positive arrivals (nominally increases in velocity with depth). Red signals indicate negative arrivals (nominally decreases in velocity with depth). Receiver functions are plotted from left to right as a function of the back azimuth of the source. The color version of this figure is available only in the electronic edition.

14 s with amplitude that varies with back azimuth, indicating that the results from MAT are robust.

Conclusion and Future Work

Digitization of this dataset is not complete. We plan to continue digitizing more tapes for the foreseeable future. We plan to archive and distribute what is already digitized via the IRIS DMC (network code DT) in the near future. Playback of a single tape takes about two hours followed by another two hours of forced downtime to compress and save the digitized data while waiting for machine parts that require cleaning after each run to thoroughly dry. Although this, in theory, may allow comfortable digitization of two tapes per workday, we found an overall best-case digitization rate of up to six tapes per week (~1.5 station-years) more realistic. We therefore encourage people interested in very specific stations and time windows to contact us for discussing the possibility to prioritize certain tapes in the processing queue.

Data and Resources

Quality control efforts described subsequently make use of the U.S. Geological Survey Earthquake Catalog at https://earthquake.usgs.gov/earthquakes/search (last accessed September 2019). Information on desiccation methods for tapes that had absorbed moisture comes from Ciletti, E. (1998): If I Knew You Were Coming I'd Have Baked a Tape! A Recipe for Tape Restoration (https://www.tangible-technology.com/tape/baking1.html, last accessed October 2019). Information on the Elsys TPC5 format comes from TPC5 and TPS5 Format Specification v.1.5 (2012), available at https://www.elsys-instruments.com/en/support/TPC5-Filespecification_1.5.pdf (last accessed October 2019). The supplemental material consists of two notes and three supplementary tables.

Acknowledgments

The authors thank the Carnegie Institution for Science for their support of this digitization effort. The authors also thank the many station operators who served over the years at each of these stations who made these archival efforts possible.

References

Aderhold, K., and R. E. Abercrombie (2016). The 2015 Mw 7.1 earthquake on the Charlie-Gibbs transform fault: Repeating earthquakes and multimodal slip on a slow oceanic transform, *Geophys. Res. Lett.* **43,** 6119–6128, doi: 10.1002/2016GL068802.

Crotwell, H. P., T. J. Owens, and J. Ritsema (1999). The TauP Toolkit: Flexible seismic travel-time and ray-path utilities, *Seismol. Res. Lett.* **70,** 154–160.

Evans, J. R., and I. S. Sacks (1980). Lithospheric structure in the North Atlantic from Observations of Love and Rayleigh Waves, *J. Geophys. Res.* **85,** no. BI2, 7175–7182.

Goldstein, P., D. Dodge, M. Firpo, and L. Minner (2003). SAC2000: Signal processing and analysis tools for seismologists and engineers, Invited contribution to *The IASPEI International Handbook of Earthquake and Engineering Seismology*, W. H. K. Lee, H. Kanamori, P. C. Jennings, and C. Kisslinger (Editors), Academic Press, London, United Kingdom.

Kanamori, H., and G. S. Stewart (1976). Mode of the strain release along the Gibbs Fracture Zone, Mid-Atlantic Ridge, *Phys. Earth Planet. In.* **11,** 312–332.

Ligorria, J. P., and C. J. Ammon (1999). Iterative deconvolution and receiver-function estimation, *Bull. Seismol. Soc. Am.* **89,** no. 5, 1395–1400.

Sacks, I. S. (1966). A Broad-band large dynamic range seismograph, in *The Earth beneath the Continents*, J. S. Steinhart and T. J. Smith (Editors), Geophysical Monograph, Vol. 10, American Geophysical Union, Washington, D.C., 543–553, doi: 10.1029/GM010p0543.

Manuscript received 31 October 2019