Physics of the Earth and Planetary Interiors, 31 (1983) 202–230
Elsevier Science Publishers B.V., Amsterdam – Printed in The Netherlands

The rupture process and asperity distribution of three great earthquakes from long-period diffracted P-waves

Larry Ruff * and Hiroo Kanamori

Seismological Laboratory, California Institute of Technology, Pasadena, CA 91125 (U.S.A.)

(Received May 13, 1982; revision accepted August 25, 1982)

Ruff, L. and Kanamori, H., 1983. The rupture process and asperity distribution of three great earthquakes from long-period diffracted P-waves. Phys. Earth Planet. Inter., 31: 202–230.

Maximum earthquake size varies considerably amongst the subduction zones. This has been interpreted as a variation in the seismic coupling, which is presumably related to the mechanical conditions of the fault zone. The rupture process of a great earthquake indicates the distribution of strong (asperities) and weak regions of the fault. The rupture process of three great earthquakes (1963 Kurile Islands, $M_W = 8.5$; 1965 Rat Islands, $M_W = 8.7$; 1964 Alaska, $M_W = 9.2$) are studied by using WWSSN stations in the core shadow zone. Diffraction around the core attenuates the P-wave amplitudes such that on-scale long-period P-waves are recorded. There are striking differences between the seismograms of the great earthquakes; the Alaskan earthquake has the largest amplitude and a very long-period nature, while the Kurile Islands earthquake appears to be a sequence of magnitude 7.5 events.

The source time functions are deconvolved from the observed records. The Kurile Islands rupture process is characterized by the breaking of asperities with a length scale of 40–60 km, and for the Alaskan earthquake the dominant length scale in the epicentral region is 140–200 km. The variation of length scale and M_W suggests that larger asperities cause larger earthquakes. The source time function of the 1979 Colombia earthquake ($M_W = 8.3$) is also deconvolved. This earthquake is characterized by a single asperity of length scale 100–120 km, which is consistent with the above pattern, as the Colombia subduction zone was previously ruptured by a great ($M_W = 8.8$) earthquake in 1906.

The main result is that maximum earthquake size is related to the asperity distribution on the fault. The subduction zones with the largest earthquakes have very large asperities (e.g. the Alaskan earthquake), while the zones with the smaller great earthquakes (e.g. Kurile Islands) have smaller scattered asperities.

1. Introduction

The distribution and description of the largest earthquakes are of fundamental interest in seismology. With the improvement of measuring earthquake size by use of the seismic moment, the great earthquakes of this century are now identified (see Table I). The more recent events have been studied using very long-period surface waves to determine the seismic moment and fault orientation. Given the interest in these large events, it is somewhat surprising that there has been no systematic study of the detailed rupture processes. The primary reason for the lack of source time function investigations is that the P and S waves are off-scale for the largest earthquakes. In this paper, diffracted P-waves are used to recover the source time functions of the three most recent great earthquakes (Kurile Islands, 1963, Alaska, 1964, Rat Islands, 1965).

The source time functions have been de-

* Present address: Seismological Observatory, Department of Geological Sciences, University of Michigan, Ann Arbor, MI 48109, U.S.A.

TABLE I

The ten largest earthquakes in this century (moments from Kanamori, 1977a, also see Kanamori and Anderson, 1975)

Earthquake (year, location)	M_0 ($\times 10^{27}$ dyne·cm)	M_W
1960 Chile	2000	9.5
1964 Alaska	820	9.2
1957 Aleutian Islands	585	9.1
1952 Kamchatka	350	9.0
1906 Colombia–Ecuador	204	8.8
1965 Aleutian Islands	125	8.7
1950 Assam	100	8.6
1938 Banda Sea	70	8.5
1922 Chile	69	8.5
1963 Kurile Islands	67	8.5

termined for many smaller earthquakes (i.e., $M_W < 8$), and some of these earthquakes consist of distinct multiple events, which are usually interpreted as the sequential breaking of asperities (regions of higher strength on the fault plane). It has commonly been assumed that great earthquakes consist of a sequence of smaller (i.e., "magnitude 8") events, and the characteristic asperity size is the same as that for an individual magnitude 8 event. In this view, a great earthquake occurs when adjacent asperities along the subduction zone are all stressed close to the failure stress, such that the failure of one asperity can trigger the adjacent asperities.

An alternative view is that great earthquakes occur where the characteristic asperity size is large and results in a substantially increased loading of the surrounding asperities, large or small, continuing the rupture over a larger area. Consequently, a great earthquake is directly associated with a larger asperity size. Hence, great earthquakes occur in places where the mechanical coupling between the two plates is stronger. Various approaches to seismic coupling and its relation to large earthquakes have been discussed and developed in Kanamori (1971), Kelleher et al. (1974), Kelleher and McCann (1976), Kanamori (1977b), Toksöz and Hsui (1978), Uyeda and Kanamori (1979), Ruff and Kanamori (1980), Kanamori (1981), Sykes and Quittmeyer (1981), Lay and Kanamori (1981) and Ruff and Kanamori (1982).

Thus, in addition to the intrinsic interest in the rupture history of great earthquakes, the interpretation of the source time functions will test the asperity model of seismic coupling at subduction zones. In this paper, we present a detailed discussion of the data, the deconvolution method and then interpret the results in terms of asperity length scale.

2. Diffracted P-waves

The rupture duration of large earthquakes ranges from ~ 10 s (magnitude 7.5 events) to possibly 300 s for the 1964 Alaska and 1960 Chile events. Long-period body waves are best suited for studying the rupture history as they are recorded in the period range from 10 to ~ 200 s. The WWSSN network was established in the early 1960's to provide world-wide coverage with standard instruments. This data base has allowed extensive use of synthetic seismograms to model earthquake sources. The earthquakes are usually modeled by using seismograms recorded in the distance range $\Delta = 30–90°$, where there are no complications in the P and S waves due to Earth structure. In this distance range the elastic Earth response can be adequately modeled by using geometric ray theory, i.e. there is no significant distortion of the waveshape. At distances less than 30°, the upper mantle triplications cause multiple arrivals in the seismograms. Upper mantle structure must be properly characterized before these recordings can be used for source studies. The core shadow zone starts at 85° for long period P-waves, and the P-waves recorded beyond this distance have been diffracted around the core, which causes a frequency dependent amplitude decay. The PKP and PKIKP phases, which are recorded from $\Delta = 120–180°$, are not used for source studies because there are two or more arrivals over much of the distance range. The amplitudes of these phases are fairly sensitive to structure at the core–mantle and inner core–outer core boundaries, and there is a rapid spatial variation in the PKP amplitude away from the PKP caustic at $\Delta \sim 140°$.

The long period P-waves of the three great earthquakes are off-scale at virtually all of the

WWSSN stations at distances less than 90°. All PKP and PKIKP recordings are also off-scale. The source history might be obtained from special well-calibrated low-gain long period instruments. However, for these three events we have been able to collect only one good record; a low-gain, high paper-speed strain record of the 1964 Alaska event recorded at Pasadena. Given this lack of data, it is understandable why there have been no long-period body wave source studies of these events.

We can exploit the fact that diffraction around the core attenuates the P-waves, thereby providing on-scale long period recordings at WWSSN stations sufficiently far into the shadow zone. There are in fact two advantages in using diffracted P-waves instead of the mid-mantle P-waves. As seen in Fig. 1, diffracted P-waves have the longest time "window" in which to view the source time function. That is, at $\Delta = 110°$, there are four minutes in which no other major seismic phase arrives. As we expect source durations to be greater than a minute, it is convenient not to have later arrivals interfering with the source time function. The other advantage of diffracted P-waves is that, since the higher frequencies are preferentially attenuated, diffraction acts as a natural smoothing filter. This is useful as we are interested in periods greater than 5 s.

There are recordings of long period diffracted P-waves for the three great earthquakes, and considering the availability and advantages of using diffracted P-waves, an obvious question is why diffracted P-waves have not been used previously. The answer appears to be the "fear of the shadow zone". Just as seismograms at ranges $< 30°$ have been avoided due to structural complexities, diffracted P-waves have been avoided on the assumption that reliable estimates of the source could not be obtained due to the complexity of the diffracted response. While the upper mantle response is unquestionably complicated, the diffracted response is not complex for long periods. In fact, the response is similar to an attenuation filter. Though calculating diffracted responses is non-trivial, synthetic seismogram methods have been developed which can generate the diffracted response for various structures at the core–mantle boundary. The structure in the lowermost 200 km of the mantle is uncertain and may vary laterally. The current allowable structure variations can affect the short period P-waves in the shadow (e.g., Ruff

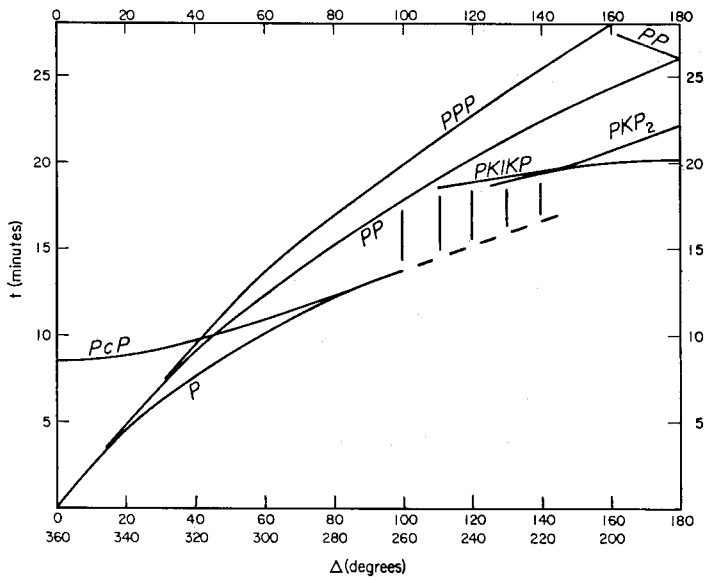

Fig. 1. Simplified travel-time curves for P phases. Diffracted P is indicated by the dashed extension of P and PcP. There is a time window of 3 to 4 minutes after diffracted P (vertical lines) in which no major seismic phase arrives (modified from Herrin, 1968).

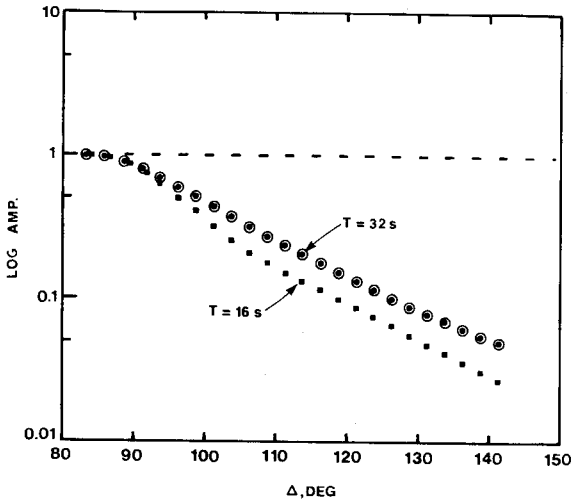

Fig. 2. The decay of P wave amplitude into the shadow zone. The amplitude decay is frequency dependent, and curves for two representative periods are shown.

and Helmberger, 1982), but fortunately long period P-waves lack the resolution to be sensitive to these variations (e.g., Doornbos and Mondt, 1979; Mula, 1981). The long period diffracted responses are invariably smooth and do not have later arrivals, such as the upper mantle response. Examples of the amplitude decay into the shadow for two periods are shown in Fig. 2. Observations of the long period amplitudes are quite scattered. Although we can make reasonable amplitude corrections based on theoretical results, the intrinsic scatter in observed amplitudes would always cause a factor of 2 scatter in the earthquake moments. We have not included the diffracted response in the modeling of the data, as the differences in the time functions between the events are sufficiently large not to demand a formal correction of the diffraction effect.

3. Data

The main conclusion of this paper is readily apparent in the seismograms. In addition to comparing the seismograms of the great earthquakes, we include a "typical magnitude 8" event. The WWSSN long period instruments were changed from a 30–100 to 15–100 (seismometer period–galvanometer period) configuration in mid 1965. Therefore, to directly compare the seismograms, the reference magnitude 8 event must have occurred between 1962 and 1965. The only event that qualifies is the 1964 Niigata, Japan event, with $M_W = 7.6$. This event had a thrust mechanism (Fig. 3), but was not a subduction zone earthquake. Figure 4 shows some of the records for the Niigata event. Note that the dominant effect of diffraction is to decrease the amplitude, with the waveform changing slowly to a smoothed version of the pre-shadow waveshape. Figure 5 shows the seismograms for the Kurile Islands event. The amplitudes of the Kurile Islands records are larger than those of the Niigata event. Also, the Kurile Islands records are composed of many pulses, as compared to the Niigata records which we consider a single pulse, or event. Clearly the Kurile Islands earthquake is composed of many events of at least the same size of the Niigata event. The records of the Rat Islands event (Fig. 6) appear similar to those of the Kurile Islands event, though of a larger amplitude.

The Alaskan records in Fig. 7 are impressive, and two features of importance are easily seen: (i) the amplitude is substantially larger than that of the other events, (ii) the dominant period of the waveform is noticeably longer than the dominant period of the other events. (Representative records of the four earthquakes can be directly compared in Fig. 8.) In particular, the first pulse of the Alaska records remains positive for approximately 30 s. This straightforward observation has immediate and strong implications for the nature of the source time function and rupture process. Regardless of the source mechanism, it is quite difficult to force the 30–100 long period response to remain at one polarity for 30 s. A ramp time function with a duration > 30 s is not adequate. This is illustrated in Fig. 9 where the BUL seismogram is modeled using two different time functions, a ramp and a half-cosine. Note that the fault orientation and source depth used in Fig. 9(a) and (b) are quite different. These two focal mechanisms are both acceptable in terms of the long period surface waves (Kanamori, 1970b), although the deeper mechanism is not consistent with the first motions (Fig. 3). Despite the uncertainty in

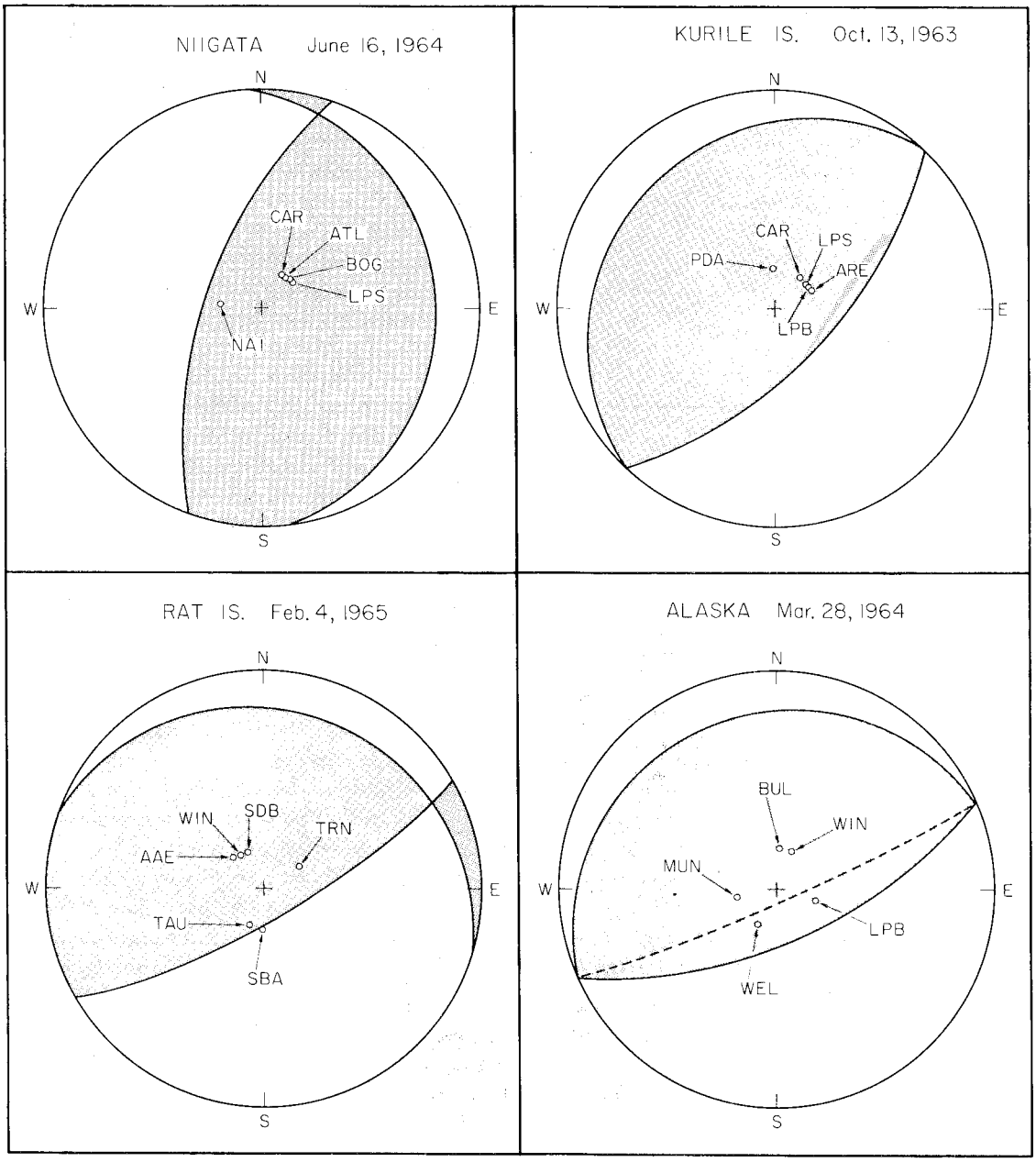

Fig. 3. Focal mechanisms and locations of stations on the focal sphere. The shaded regions depict the compressional quadrants in the lower hemisphere projections. The mechanisms are from studies which used long period surface waves. The mechanism for Alaska is the "deep" ($h = 70$ km) solution of Kanamori (1970b), which is not consistent with first motions. The dotted plane is consistent with first motions.

the focal mechanism, the broad first pulse of the Alaskan records apparently requires a time function that curves upward for at least 30 s. The time function must also continue as a smooth function, as any abrupt jump would cause a very large, sharp pulse that is not observed (for the first 3

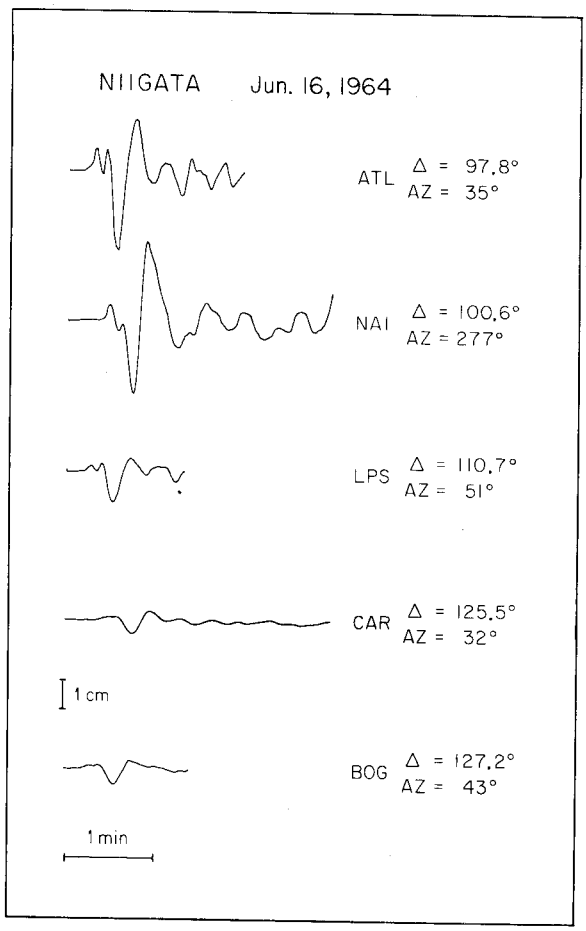

Fig. 4. Representative seismograms of the Niigata earthquake. Notice the smooth amplitude decay into the shadow zone. This earthquake is characterized by a single pulse The amplitude scale gives the trace amplitude for a magnification of 1500.

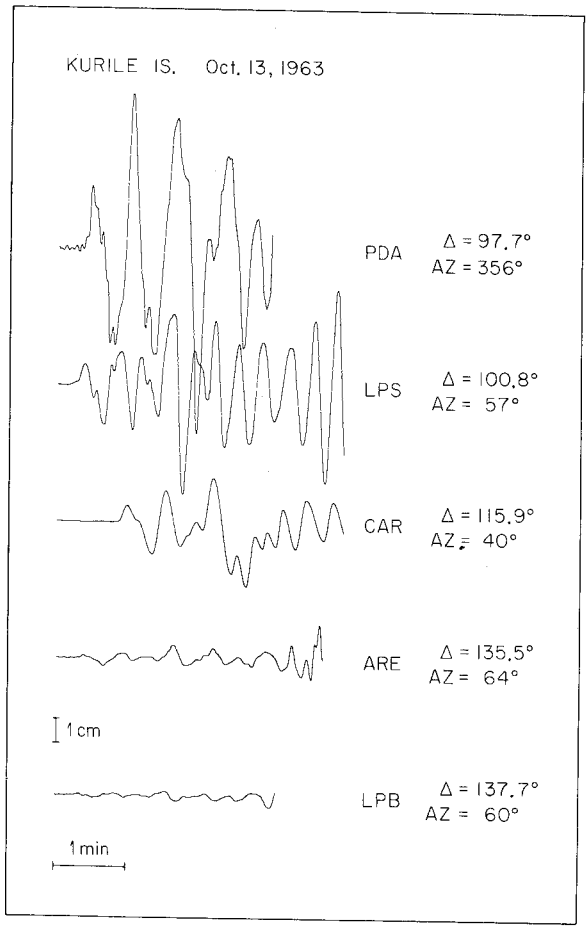

Fig. 5. Seismograms of the Kurile Islands earthquake. There is a sequence of pulses and the overall amplitude is larger than the amplitude of the Niigata seismograms.

minutes of the rupture). It is important to notice that this long period character exhibited by BUL is not just a consequence of diffraction. All of the Alaskan records show the same initial pulse width, including the LPB(NS) record at $\Delta = 97.8°$, where the diffraction effect is minimal (Fig. 10). The time function required at LPB is similar to that at BUL. Based on the time functions in Fig. 10, it seems that the moment released in the first 90 s of the Alaskan earthquake is greater than the total moment of the Kurile Islands earthquake.

The main conclusion is already apparent. Based on a direct comparison of the seismograms combined with some simple forward modeling, the initial rupture of the Alaskan earthquake was a slow, smooth rupture with an effective rise time of greater than 40 s, compared to the Niigata event with a total rupture time of 20 s. As will be discussed later, this very long rise time translates into a source dimension of 140–200 km for the Alaskan asperity that initiated the Alaskan earthquake.

To quantify the above conclusion, we will need to determine the source time functions for each of the events. A method for determining source time functions is presented in the next two sections.

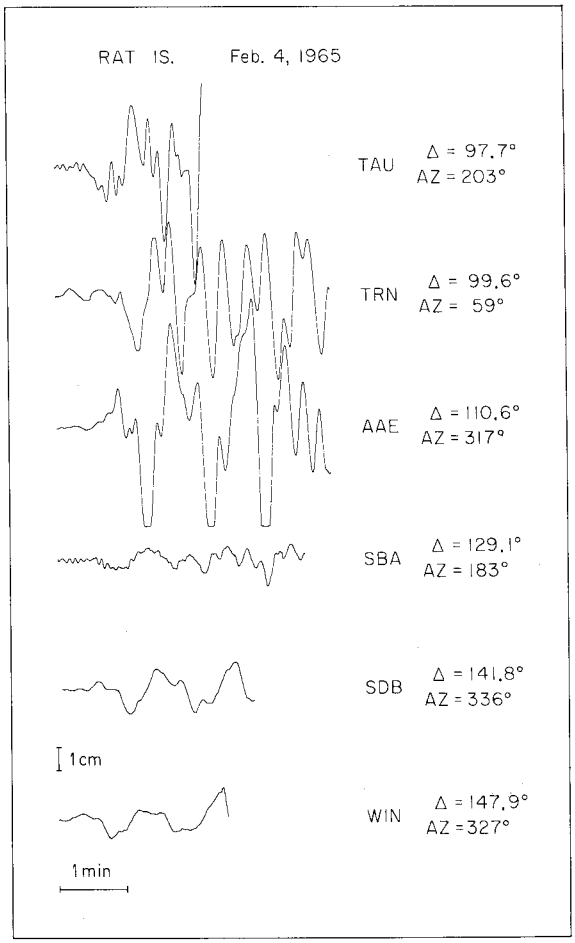

Fig. 6. Seismograms of the Rat Islands earthquake. The amplitude is larger than the Kurile Islands amplitude (note the diminished amplitude scale), and the pulses appear to have a longer dominant period.

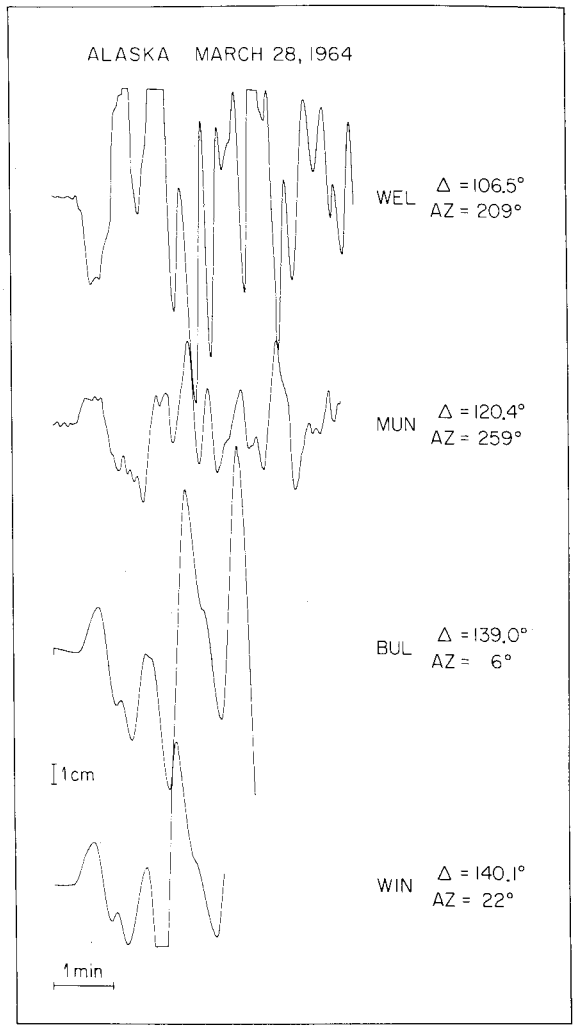

Fig. 7. Seismograms of the Alaskan earthquake. The amplitudes are very large and the dominant period is noticeably longer than for the other earthquakes.

4. Body wave seismograms of earthquake sources

Synthetic seismograms for body waves at teleseismic distances can be constructed as $s(t) = \mathrm{I}(t)*g(t)*m(t)$, where * denotes convolution, $s(t)$ is the seismogram, $\mathrm{I}(t)$ is the instrument impulse response, $g(t)$ contains the source geometry and the Earth's impulse response and $m(t)$ is the source time function, i.e. the faulting behavior as a function of time. Assuming that the average fault geometry is known, $m(t)$ is then obtained by inverting the convolution operation. This can be done in either the frequency domain or directly in the time domain. We will use a time domain method.

4.1. Point source representation

The earthquake source is usually considered as a shear dislocation across the fault surface. If the region of displacement discontinuity is small (i.e. source dimension is smaller than the wavelength of interest), then the earthquake source can be con-

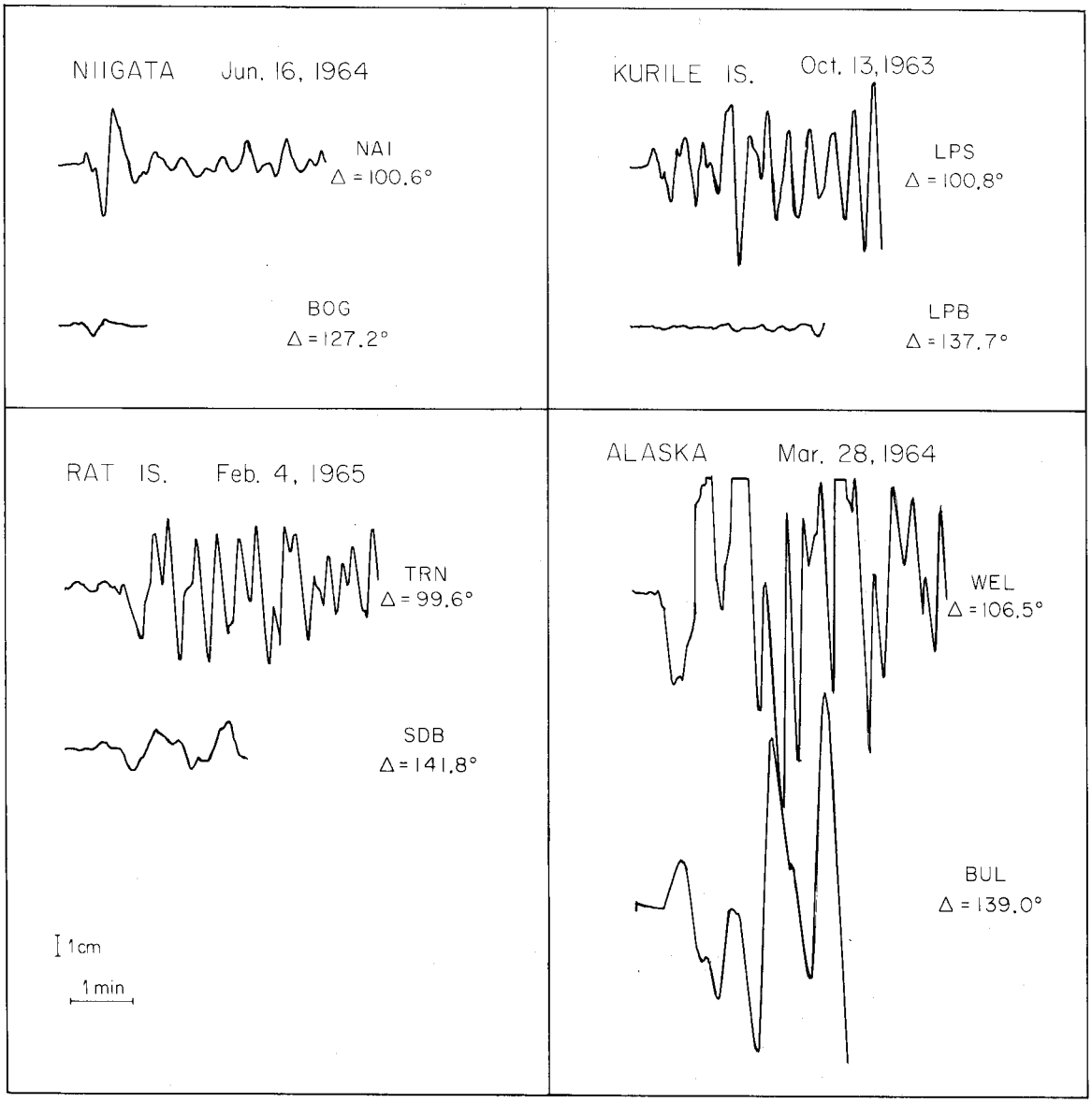

Fig. 8. Two representative seismograms from each earthquake plotted at the same amplitude and time scales. The large size and long period nature of the Alaska seismograms is clearly evident.

sidered as a point source. The source displacement can be characterized by a system of force couples, conveniently grouped into the moment tensor (see Aki and Richards, 1980). If we restrict our attention to a shear dislocation across a planar surface, then the moment tensor can be reduced to a double couple of some orientation. For the teleseismic (station distance $\gg$ source dimension) body wave problem, the double couple source is represented by the radiation pattern which gives the displacement amplitudes of P or S waves leaving the source (for details, see Langston and Helmberger, 1975; Kanamori and Stewart, 1976). For a double couple point source, the source time

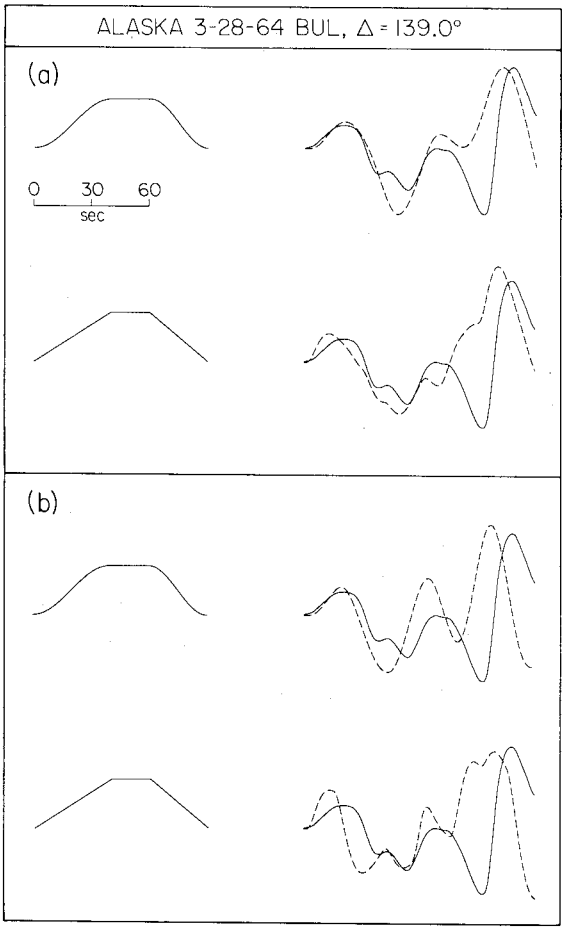

Fig. 9. Forward modeling of the Alaskan record at BUL. The time functions are on the left, the observed seismogram is the solid trace, and the synthetic seismograms are the dashed traces. The width of the first peak of the seismogram cannot be matched by a time function with a simple ramp. Some type of "rising-ramp", such as a half-cosine, is required by the first part of the seismogram regardless of the mechanism used ((a) and (b) show the synthetics for the deep and shallow mechanisms of Kanamori (1970b)).

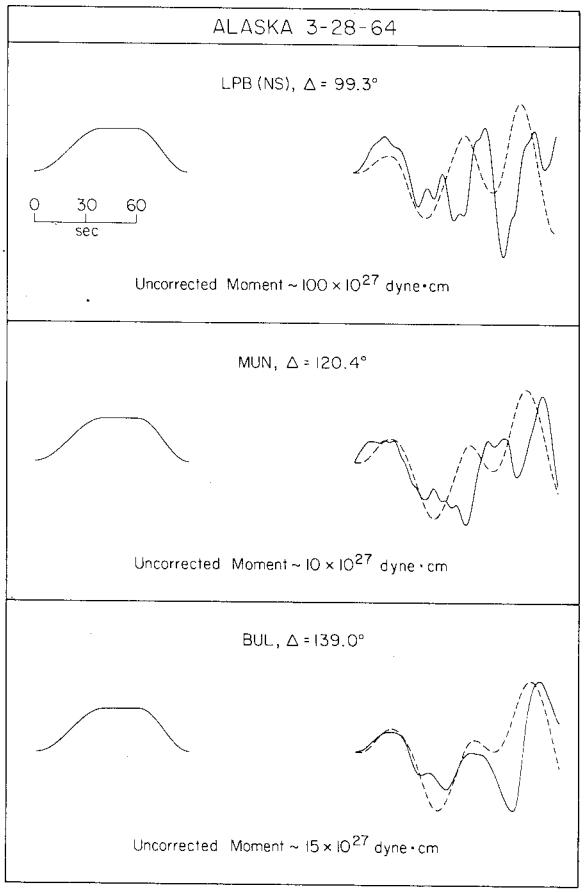

Fig. 10. The Alaskan time function determined by forward modeling. The half-cosine time function used in Fig. 9(a) is applied to the seismograms recorded at MUN and LPB (NS component). The width of the first peak is not an effect of diffraction and a rising-ramp time function is required at all stations.

function is the rate of moment release and has the same time dependence at all azimuths. The effect of propagation through the earth is then given by the Green's function for a point double couple source. Thus, the direct P-wave recorded at distance Δ and azimuth ϕ is given by

$$s(t, \Delta, \phi) = I(t)^* g_p(t; \Delta, h)^* R_p(\Delta, \phi, h)\, \dot{M}(t) \quad (1)$$

where $I(t)$ is the instrument response (includes the free surface effect), $g_p(t; \Delta, h)$ is the Green's function, $R_p(\Delta, \phi, \mathrm{h})$ is the radiation amplitude, and $\dot{M}(t)$ is the moment rate. When the station distance is in the range where geometric ray theory is applicable, 30–90°, the Green's function is simply the attenuation operator ($Q(t)$) multiplied by the geometric spreading factor and source constants, and delayed by the travel time: $g_p(t; \Delta, h) = \xi(\Delta, h)\, Q(t)^* \delta(t - T_p)$, where $\delta(t)$ is the Dirac delta function). For shallow earthquakes, the reflections from the surface, pP and sP, arrive soon after the direct P-wave and need to be incorpo-

rated into the synthetic. The seismogram is then

$$s(t, \Delta, \phi) = I(t)^* \left[R_P(\Delta, \phi, h)\, \delta(t - T_P) + r_{pP} R_{pP}(\Delta, \phi, h)\, \delta(t - T_{pP}) + r_{sP} R_{sP}(\Delta, \phi, h)\, \delta(t - T_{sP}) \right] {}^*\xi(\Delta, h) Q_P(t)^* \dot{M}(t) \quad (2)$$

where R_{pP} gives the amplitude for the upgoing P wave, r_{pP} is the surface reflection coefficient, T_{pP} is the total pP travel time, and similarly for sP. Note that the above expression can be shortened to $s(t) = T(t)^* g(t;\ \Delta,\ \phi,\ h)^* m(t)$ where $m(t) = \dot{M}(t)$, the source time function, and $g(t;\ \Delta,\ \phi,\ h)$ is referred to as the "half space" Green's function for a particular mechanism. With the above expression, it is quite easy to construct synthetic seismograms for a given focal mechanism and source time function.

4.2. *Finite source dimension*

As we model earthquakes with a fault length of 100 km or more, the point source approximation is not valid. For a finite source, we need to integrate the contributions from point sources distributed over the fault plane

$$s(t, \Delta, \phi) = I(t)^* \iint_A g(t;\ \Delta, \phi, h)^* \mu \dot{D}(t)\, dA \quad (3)$$

where $\dot{D}$ is the time derivative of the particle displacement and μ is the shear modulus in the source region. $g(t;\ \Delta, \phi, h)$ will vary over the fault plane, and yields the directivity effect due to the variation in Δ, ϕ, and h across the fault plane. Additionally, $g(t;\ \Delta, \phi, h)$ could vary if the focal mechanism changed over the fault surface. Usually the variation in $g(t)$ over the fault surface is slight, so that an average $g(t)$ can be taken outside of the integral

$$s(t, \Delta, \phi) = I(t)^* \bar{g}(t;\ \Delta, \phi, h) * \mu \iint_A \dot{D}(t)\, dA \quad (4)$$

which has the same form as the point source expression except that the source time function is now $\mu \iint \dot{D} dA$. To interpret this, it is convenient to use a rupture front sweeping over the fault area. This allows a change of variable from dA to $d\tau$, where $A(\tau)$ gives the location of the rupture front as parameterized by the rupture time, τ. Thus

$$s(t, \Delta, \phi) = I(t) * \bar{g}(t;\ \Delta, \phi, h) * \mu \int_0^\infty \dot{D}(t, \tau)\, (dA/d\tau)\, d\tau \quad (5)$$

If the particle displacement history is the same everywhere on the fault, simply delayed by the rupture time, then $\dot{D}(t, \tau) \rightarrow \dot{D}(t - \tau)$, and

$$s(t, \Delta, \phi) = I(t) * \bar{g}(t;\ \Delta, \phi, h) * \mu \int_0^\infty \dot{D}(t - \tau) \dot{A}(\tau)\, d\tau \quad (6)$$

The integral is now in the form of the convolution integral. Thus, with the above assumptions, the time function $m(t)$ is the convolution of the particle velocity and the rate of increase of the faulted area. Two simple examples of unilateral and circular rupture are shown in Fig. 11. For large earthquakes, the duration of rupture across the fault plane can be longer than the characteristic time of particle displacement, hence the duration of the time function indicates the rupture time

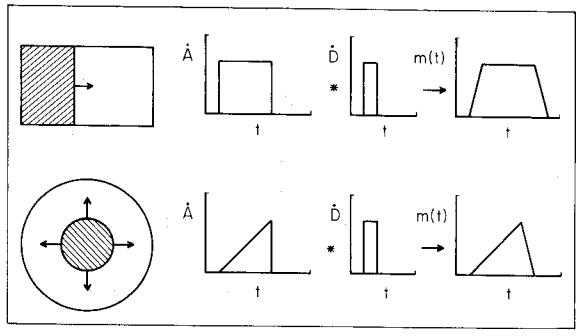

Fig. 11. Two different kinematic rupture modes, unilateral and circular. The time function for unilateral rupture is depicted in the top row. At the far left, a rectangular fault is shown with the rupture front sweeping to the right. When the particle velocity is the same everywhere on the fault (this is a crude approximation), the time function, $m(t)$, results from the convolution of area rate with particle velocity. As area rate is characterized by a boxcar function, the time function will have the shape of a trapezoid. For the case of circular rupture (shown on bottom row), a ramp function characterizes the area rate, leading to a sawtooth time function for an abrupt termination of rupture.

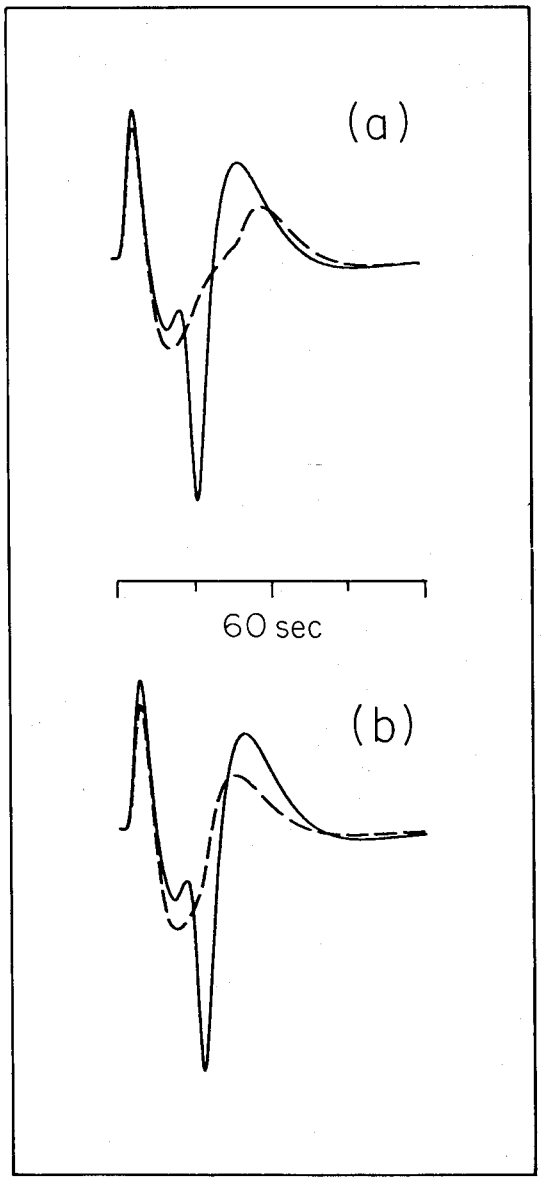

Fig. 12. Half-space Green's functions for a point source and distributed sources. The solid trace in (a) and (b) is the point source Green's function at BUL for the Alaskan earthquake with $h = 30$ km. The dashed trace in (a) is the Green's function for a source distributed from the surface to a depth of 50 km, while the dashed trace in (b) is for a source distributed between the surface and 30 km. The WWSSN 30–100 instrument is used. The distributed source Green's functions appear as filtered versions of the point source Green's function.

which is the characteristic fault length divided by the rupture velocity. In some cases, the observed shape of the time function indicates the mode of rupture, e.g. circular or unilateral.

Returning to eq. 6, the directivity effect can be included by allowing $dA/d\tau$ to vary with azimuth to station. Directivity is usually ignored for P-waves, as the total effect is only a few seconds for fault lengths less than 100 km.

To determine the source time history by using eq. 5, it is necessary to choose an average Green's function, with the simplest choice being the half-space Green's function, selecting an average source depth. As the half-space Green's functions vary most rapidly with depth, perhaps a finite source distributed over a depth interval is not well represented by a fixed depth Green's function. Figure 12 compares a fixed depth Green's function and Green's functions for finite sources. The finite Green's function appears to be a filtered version of the point source Green's function. Hence, the synthetic seismograms for the point source and distributed source could be quite similar after convolution with a long period time function. We have used fixed depth Green's functions in the following analysis, and we will later show a comparison of time functions obtained from fixed depth and distributed source Green's functions. We do not imply that the Great Alaskan earthquake is a point source, as the source finiteness will be exhibited in the time function through the $dA/d\tau$ contribution in the integral. Choosing an average Green's function simply means that we are not attempting to model the details of the time function; i.e. we are mainly interested in the gross features of the rupture histories.

5. Time function deconvolution

This section, presents the methodology of deconvolution. If one is not concerned with the details of the analysis, the time functions can be regarded as lucky guesses, and their worth evaluated by, (i) comparing the associated synthetic seismograms to the data, and (ii) checking if the time functions determined for individual stations are similar.

The seismogram is modeled as $s(t) = g(t)*m(t)$, where $g(t)$ now includes the instrument, propagation effects, and the assumed source mechanism, and $m(t)$ is the effective far-field time function. Determining a satisfactory $m(t)$ by trial-and-error modeling for a large number of complicated seismograms. It is preferable to deconvolve $m(t)$ directly from the observed seismogram. Deconvolution is a subject of general interest in signal analysis and is routinely used in seismic reflection data processing. However, no generally useful computational technique is available. Deconvolution is unstable, with respect to the reliability of $m(t)$, for most applications and this problem is handled in different ways. We review the known properties of deconvolution, then discuss our particular application, including the reliability and resolution of the solutions.

5.1. *Frequency domain deconvolution*

Though $m(t)$ will be deconvolved in the time domain, the basics of frequency domain deconvolution are presented due to the conceptual simplicity of this approach. Note that the effects of discrete sampling will not be treated (which occur when using the Fast Fourier Transform algorithm).

Frequency domain deconvolution is simple because convolution in the time domain is equivalent to multiplication in the frequency domain. Let $S(f)$, $G(f)$, and $M(f)$ be the Fourier transforms of $s(t)$, $g(t)$, and $m(t)$ respectively. Then, the Fourier transform of the equation $s(t) = g(t)*m(t)$ is $S(f) = G(f)M(f)$. Convolutions are commonly performed by complex multiplication in the frequency domain, e.g. $s(t) = F^{-1}(G(f)M(f))$, where F^{-1} denotes the inverse Fourier transform. To deconvolve $m(t)$, simply divide $S(f)$ by $G(f)$, $M(f) = S(f)/G(f)$, and $m(t)$ is then the inverse Fourier transform of $M(f)$, $m(t) = F^{-1}(S(f)/G(f))$. Unfortunately, problems arise at the step of dividing by $G(f)$. If $G(f) = 0$ at particular values of f, then $1/G(f)$ is unbounded and $M(f)$ can be unbounded or finite depending upon the details of $S(f)$ at these points. Thus, $M(f)$ is unstable with respect to infinitesimal variations in $S(f)$ at the zeroes of $G(f)$. Practically, this instability of $M(f)$ appears whenever $G(f)$ is smaller than the noise in $S(f)$.

The difficulties with $1/G(f)$ are resolved in two ways: (i) a filter is used to smooth $1/G(f)$, which results in a time limited $m(t)$, (ii) $M(f)$ is simply set to some predetermined value where $G(f) \approx 0$, which results in spurious frequency components in $m(t)$ that are not resolved.

5.2. *Time domain deconvolution*

A more general approach is that of the Backus–Gilbert geophysical inverse theory (Backus and Gilbert, 1970; see also Parker, 1977; Oldenberg, 1981). Given the equation

$$s(t) = \int_0^\infty g(t - t')\, m(t')\, dt' \tag{7}$$

relating the model, $m(t)$, to the data, $s(t)$, we ask: what properties of $m(t)$ can be determined from observations of $s(t)$? In the following discussion it is assumed that $g(t)$ is known. In our particular application, $g(t)$ will be the half-space Green's function convolved with the instrument. The uncertainties in $m(t)$ arise from the properties of $g(t)$ and inadequacies and errors in $s(t)$. If a solution exists, it will be unique if the only solution to the homogeneous equation, $\int g(t - t')\, m_N(t')\, dt' = 0$, is the trivial solution. In frequency domain, this equation is transformed to $G(f)M_N(f) = 0$. This relation is satisfied by a non-trivial $M_N(f)$ if $G(f)$ has zeroes. That is, $M_N(f)$ is arbitrary at values of f where $G(f) = 0$. This is the same difficulty as discussed above. Let $m_1(t')$ be a solution to eq. 7, then contributions from the null space of operator $\int g(t - t')$ can always be added, that is $m_2(t') = m_1(t') + m_N(t')$, and $m_2(t')$ also satisfies eq. 7. If a trial-and-error solution fits the data, there could be arbitrary unresolved contributions from the null space of $\int g(t - t')$. A stable deconvolution method must control the null space contributions to the solutions.

To construct solutions to eq. 7, it is necessary to discretize $s(t)$, $g(t)$, and $m(t)$, thereby changing the integral into the following summation (for the simplest case of equal sampling and $g_j = 0$ for $j < 1$)

$$s_i = \sum_{j=1}^{i} g_{i-j+1} m_j \tag{8}$$

However, the Backus–Gilbert method does not immediately discretize the entire problem. Instead, $m(t')$ and $g(t-t')$ are considered as continuous functions of t', while it is assumed that the data are available at only discrete values of t.

$$s(t_i) = \int_0^\infty g(t_i - t')\, m(t')\, dt' \tag{9}$$

Discretizing $s(t)$ causes $m(t')$ to be non-unique. The Backus–Gilbert method states that although $m(t)$ cannot be uniquely determined at a point, say t_0, a localized average of $m(t)$ around t_0 can be determined. That is, $\bar{m}(t_0)$ is the average m in the interval $(t_0 - y/2)$ to $(t_0 + y/2)$ where y is the averaging length. The trade-off between resolution and stability can be stated as follows: the average m, $\bar{m}(t_0, y)$, is less stable for a smaller y (higher resolution). There will be a minimum averaging length $y_{\min}$ associated with maximum resolution. Thus, the sampling of $m(t)$ should not be less than $y_{\min}$. The sampling interval that we use for $m(t)$ is larger than $y_{\min}$. Typical values of the sampling are: 1 s sampling of the seismogram and 4–6 s sampling of the time function.

The $g(t, t')$ function includes the instrument, which causes the spectrum to tend toward zero at very long periods as well as at short periods. While the uncertainties at short periods can be avoided by the sampling of $m(t)$, the lack of a zero frequency component is bothersome. It should take on positive values (unless the focal mechanism is drastically incorrect) because of the physical interpretation of the time function. Thus, after deconvolving the time function, a half-sine wave component, with a period twice the time function duration, is added so as to make the time function predominantly positive. The duration of $m(t)$ is typically set at 180 s, so the addition of the half-sine component flattens the spectrum from $T = 360$ s to zero frequency.

5.2.1. Generalized inverse

We have deconvolved $m(t)$ by solving eq. 8 using the damped generalized (Lanczos) inverse (Lanczos, 1961; Jackson, 1972; Wiggins, 1972). With the seismogram sampling interval smaller than that for the time function, eq. 8 is an overdetermined system, $A_{ij}m_j = s_i$ with $i = 1$ to n, and $j = 1$ to $k (n > k)$. Smoothing conditions can be explicitly included by adding equations that set the first differences of m_j to zero

$$\begin{Bmatrix} & A & \\ \text{———————} & & \\ 1-1 & & \\ \gamma \quad \begin{matrix} \diagdown & \diagup & \\ & 1-1 & \diagup \diagdown \\ \diagup & \diagdown & \\ & 1-1 & \end{matrix} & & \end{Bmatrix} \quad m = s \tag{10}$$

The parameter γ multiplies the lower submatrix and weights the smoothing conditions relative to fitting the data. In practice, it is not necessary to enforce smoothness as the sampling of $m(t)$ results in adequately smooth solutions.

There are basically two methods available for constructing a stable inverse to $\mathbf{A}$, the spectral (eigenvalue) method and the damped Lanczos inverse. We have used the damped Lanczos inverse as it is remarkably simple and computationally efficient. The undamped Lanczos inverse is

$$\mathbf{A}^* = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T, \tag{11}$$

where superscript T denotes the transpose of $\mathbf{A}$. This formula can be derived by minimizing the data mismatch in a least-squares sense, i.e. $\min(\mathbf{e}^T\mathbf{e})$ with $e_i = A_{ij}m_j - s_i$. With the above formula, only the symmetric $\mathbf{A}^T\mathbf{A}$ matrix needs to be inverted. As the Lanczos inverse results in the least squares estimate for $\bar{m}_j$, it provides an exact answer, if any exist and also gives the maximum resolution of $m(t)$. Our estimate of the model $(\bar{m}_j)$ is

$$\bar{m}_j = A^*_{ji}A_{ik}m_k = A^*_{ji}s_i \tag{12}$$

The resolution matrix is $\mathbf{R} = \mathbf{A}^*\mathbf{A}$, and for the Lanczos inverse, $\mathbf{R}$ is the identity matrix. The Lanczos inverse can produce an unstable estimate for $m(t)$ if the $\mathbf{A}^T\mathbf{A}$ matrix is "almost" singular. This situation is analogous to admitting null space contributions. Poorly resolved components of the solution can be eliminated by partially minimizing the length of the solution vector, $\bar{m}_j$. That is, instead of minimizing the data mismatch, minimize the quantity $(1-r)\mathbf{e}^T\mathbf{W}\mathbf{e} + r\mathbf{m}^T\mathbf{V}\mathbf{m}$ where $\mathbf{W}^{1/2}$ weights the data (usually, the diagonal elements of $\mathbf{W}$ are $1/\mathrm{var}(s_i)$), $\mathbf{V}$ represents an a priori

estimate of the model variance, and r is an adjustable parameter ($0 \leq r \leq 1$). This parameter indicates the relative weight between fitting the data and forcing a "shorter" solution. The damped Lanczos inverse is

$$\mathbf{A}^* = (\mathbf{A}^T\mathbf{W}\mathbf{A} + d\mathbf{V})^{-1}\mathbf{A}^T\mathbf{W} \qquad (13)$$

where d is the damping parameter, $d = r/(1-r)$. Experimenting with different assumptions for $\mathbf{V}$, it was found that various weighting schemes produce solutions generally similar to the solution when $\mathbf{V}$ is the identity matrix. Note that for non-zero d, the resolution matrix will no longer exactly equal the identity matrix.

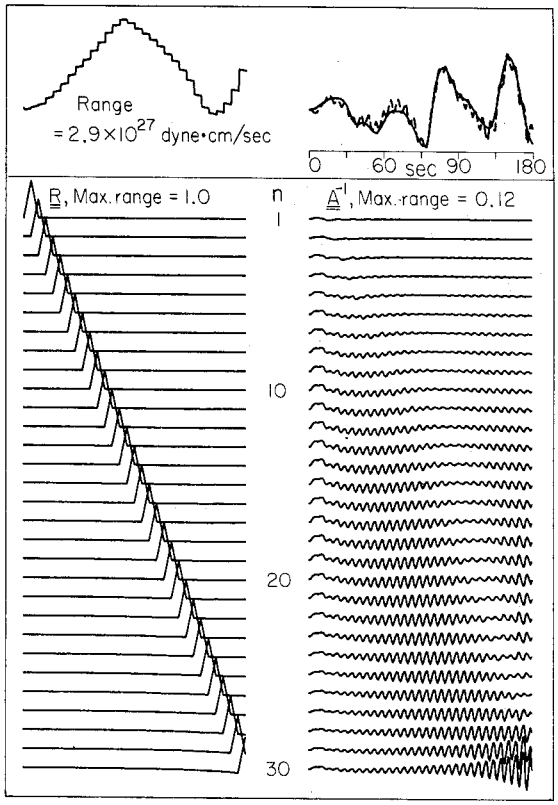

Fig. 13. The deconvolved Alaskan time function for the damping constant $d = 10^{-5}$. The solid trace at upper right is the seismogram recorded at BUL. The deconvolved time function is at upper left and the corresponding synthetic seismogram is the dashed trace. The time function is composed of 30 points with a spacing of 6 s. The rows of the resolution matrix are shown on the left, and the rows of the inverse matrix are shown on the right.

As d is an important parameter that strongly affects the solution, we examine the change in the resolution and inverse matrices associated with a change in d. The Alaskan earthquake record at BUL is used. Figure 13 shows the solution, corresponding synthetic seismogram, observed seismogram and the $\mathbf{R}$ and $\mathbf{A}^*$ matrices for a very small value of d: $d = 10^{-5}$. The slight damping allows long period components off the edge of the instrument pass-band to appear in the solution and the resolution matrix is essentially the identity matrix. The rows of $\mathbf{A}^*$ show how the seismogram is weighted to obtain the model values. Note that a particular value of $\overline{m}_j$ depends upon almost the entire seismogram. The variance of $\overline{m}_j$ is related to the data variance by the squared length of the row, i.e. $\mathrm{var}(\overline{m}_j) = \Sigma A^{*2}_{ij}\mathrm{var}(s_i)$ where A^*_{ij} are the elements of the inverse matrix. Hence, the variance of the time function in Fig. 13 increases with time for a uniform data variance. For example, if the standard deviation of the data is 10% of the maximum amplitude, the standard deviation of the time function in Fig. 13 would be 4% of the maximum amplitude for the first time point and 70% for the last time point. Figure 14 shows the effects of increasing d to 10^{-2} (the largest d value used in the analysis). The resolution is not significantly affected, indicating that the sampling interval is large, such that the elements of the time function will always be formally well resolved. Increasing the damping has a more significant effect on the inverse matrix, and consequently on $\overline{m}_j$. The peak to peak amplitudes of the rows of $\mathbf{A}^*$ are reduced by an order of magnitude. Also, the high frequency component seen in Fig. 13 has been seriously attenuated. Notice that the solution is now composed of a more local weighting of the seismogram. The relative variance of the model is smaller with the larger damping; for 10% standard deviation in the data, the model standard deviation is 2% at the first point and 1.5% at the last point.

These two figures also indicate the model features that are demanded by the data. Certainly the summed moment is highly sensitive to the value of d. By increasing the damping from $d = 10^{-5}$ to

$d = 10^{-2}$, the amplitude of $\mathbf{A}^*$ is reduced by an order of magnitude and the total moment of the time function is an order of magnitude lower. The addition of the half-sine component does not restore the moment to the value for $d = 10^{-5}$. The shape of the time function is of special interest. In general, a smaller damping allows oscillations to appear in the time functions. It is therefore striking that the Alaskan time function appears to be smoother when using a smaller damping. The time function for $d = 10^{-5}$ has a large amplitude for a long time interval. The increase in damping to $d = 10^{-2}$ partially minimizes the time function norm by decreasing the peak amplitude and introducing a "saddle" at ~ 100 s (Fig. 14). The decrease in the peak amplitude results in a poor match between the observed and synthetic seismograms in the first 60–80 s. Despite the adverse effects of the stronger damping, such as dropping the time function to negative values at the first point, notice that the first 70 s of the time function essentially consists of a rising ramp (similar to the time function for a smaller damping). With the above considerations, we can conclude that a smooth long-period time function (in particular, the feature of the rising-ramp in the first 70 s) is demanded by the data.

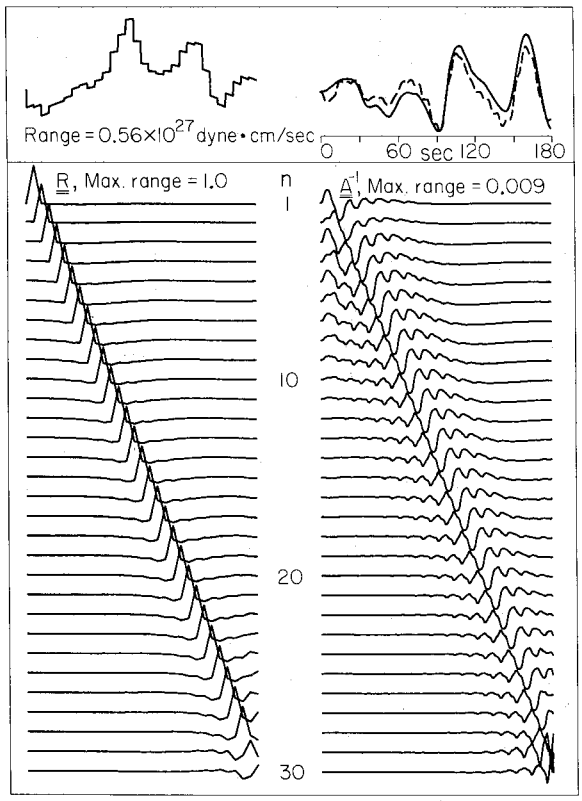

Fig. 14. The deconvolved Alaskan time function for the damping constant $d = 10^{-2}$. Otherwise the same as Fig. 13. The resolution is degraded only slightly, but the inverse matrix is severely attenuated.

The procedure used to model a seismogram is to first deconvole the time function with a large damping, $d = 10^{-2}$. If the synthetic seismogram for this time function does not match the data satisfactorily, then the damping is reduced. Hence, we are determining minimum seismic moments.

5.3. Inadequacies in g(t, t')

In the previous development it has been assumed that $g(t, t')$ is the true $g(t, t')$. Aside from approximations in the theory used to derive $g(t, t')$, there may be errors in the assumed average focal mechanism and source depth. If the errors in $g(t, t')$ are random, uncorrelated, normally distributed, and each row of the **A** matrix ($g_i(t')$) has the same variance, then we can accommodate the variance of g_i in the same manner as the variance of s_i. Each equation would be weighted by $1/\mathrm{var}(g_i)^{1/2}$ instead of $1/\mathrm{var}(s_i)^{1/2}$. This procedure is intuitively acceptable as the unexplained randomness in s_i can be considered as random inadequacies in g_i. However, the errors in $g(t, t')$ are probably not random. They could be highly correlated, as errors in the focal mechanism can cause systematic effects. It is difficult to correct for systematic errors as the character of the systematic error must be known, but then it would not be an error.

The approach followed here is to simply demonstrate the size of possible errors, with particular attention given to the longer periods. For a clearer presentation the effects will be shown in the frequency domain. Let the observed seismogram be $S(f) = G_0(f)M_0(f)$ where $G_0(f)$ and $M_0(f)$ are the "true" Green's function and time function. We choose a Green's function, $G(f)$, and decon-

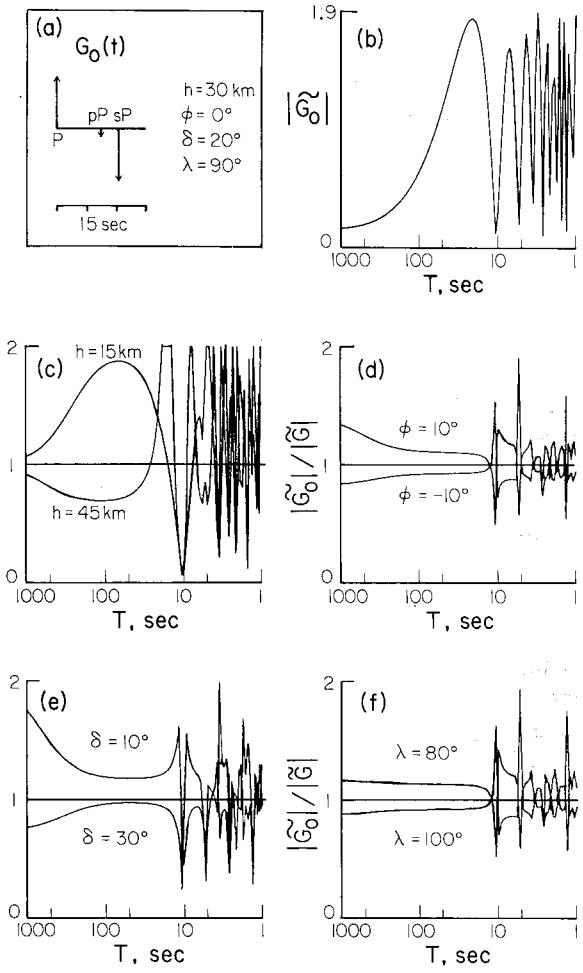

Fig. 15. Errors in the time function introduced by incorrect focal mechanism. In (a), the half-space Green's function (no instrument or Q) is shown for a station at an azimuth of 45° and a distance of $\Delta = 90°$. The reference focal mechanism is a point source at a depth of 30 km, with a fault strike of 0°, fault dip of 20°, and a slip vector angle of 90° (pure thrust). The spectral amplitude of $G_0(t)$ is plotted in (b) with a linear scale (as are (c) through (f)). If our assumed Green's function differs from $G_0(t)$, then we effectively multiply the spectrum of the true time function by the functions shown in (c) through (f) for various parameter perturbations.

volve $M(f)$ so that $M(f) = S(f)/G(f) = M_0(f)G_0(f)/G(f)$. If $G(f)$ is the true Green's function, then the true time function is recovered (here ignoring the deficiencies of $1/G$). However, if $g(t, t')$ does not equal $g_0(t, t')$, then M_0 is contaminated by the factor (G_0/G). In Fig. 15, this bias in spectral amplitudes is plotted for perturbations to the source description about a "standard" mechanism. The parameter perturbations change $G(f)$ by virtue of a slightly different interference between P, pP, and sP. These arrivals cause spectral zeroes in $G(f)$, and perturbing the mechanism causes an erratic behavior at periods less than the first spectral zero. At the longer periods the effects are less dramatic, less than a factor of 2 for the perturbations used. This bias in the longer period components would not significantly affect the conclusions regarding characteristic rupture length.

6. Source time functions

The time functions are calculated with focal mechanisms determined from long-period surface waves (Fig. 3). The longer periods indicate the average mechanism over the duration of the earthquake. In most cases the first motions are consistent with the long period mechanism. Figure 16 shows the time functions determined for a few Niigata seismograms with a point source depth of 10 km. Although changing the depth will change the total moment and alter the details of the time function, the basic shape and duration remain the same. The Niigata time function has a duration of ~ 20 s and represents the breaking of a single asperity. The duration time can be considered as the sum of the rupture time plus the particle rise time. Assuming a rise time of a few seconds and a rupture velocity of 2 km s^{-1}, the characteristic length of the Niigata asperity is 30 km if the rupture is unilateral, or 60 km for bilateral or circular rupture. Abe (1975) estimated the Niigata fault area to be 30×80 km. Hirasawa (1965) analyzed the P-waves of the Niigata earthquake and identified two or three events, and indeed some complications can be seen in Fig. 16. However, for the time scale of interest here, the Niigata earthquake can be considered as a single event.

The inversion results for the Kurile Islands earthquake are shown in Fig. 17 using the focal mechanism from the surface wave study of Kanamori (1970a), with a depth of 30 km. The

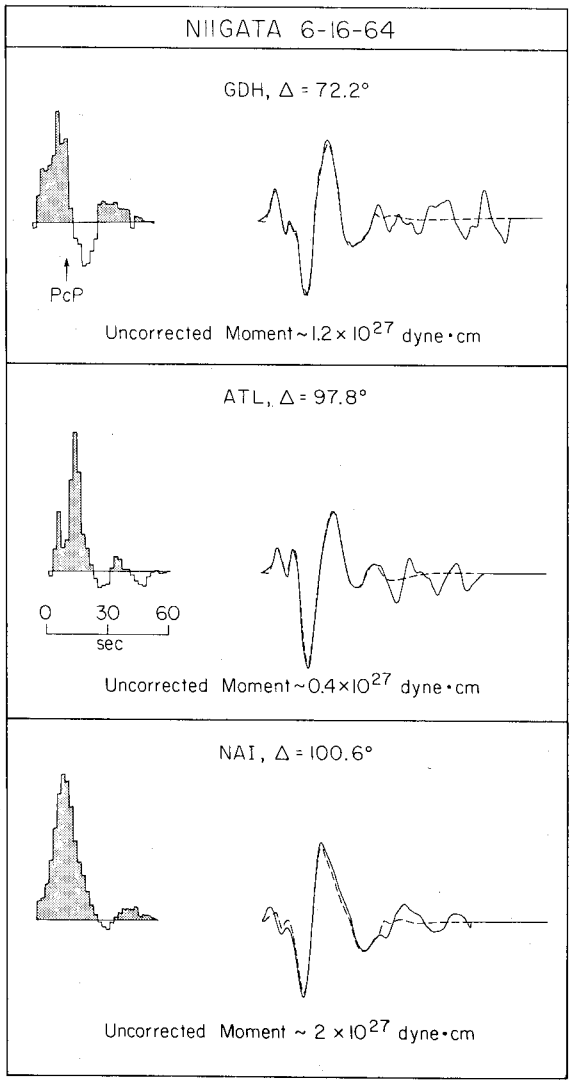

Fig. 16. Deconvolved time functions for the Niigata earthquake. The solid traces are the observed seismograms and the dashed traces are the synthetic seismograms for the time functions shown at left. The damping constant is 10^{-2}. The common feature of these time functions is a single event with a duration of 20 s. The time functions are not corrected for the core diffraction effect in Figs. 16–25.

most important feature is that the time function is composed of distinct events, each with a duration of 20–30 s. The moment of each of these pulses is approximately that of a $M_W = 7.5$ event. Although the total moment of these pulses does not equal the long period value of 70×10^{27} dyne · cm, this difference can be reduced by adding a long period component as previously discussed. This long period component would result from a slower deformation over the entire fault surface, while the individual pulses that dominate the P wave records are from the rupturing of asperities. The rupture of asperities is expected to cause a larger moment rate than rupture of the surrounding weaker area, due to a larger particle velocity and possibly a faster rupture rate (see eq. 5). Hence, the breaking of asperities causes a spatio-temporal variation in the moment rate. As the rupture front spreads across the fault plane breaking individual asperities, the time function (moment rate) will be composed of distinct pulses. The pulse duration depends upon the asperity size and the rupture velocity.

"Uncorrected moment" refers to the lack of the diffraction correction to the stations in the core shadow; therefore, the uncorrected moments are minimum estimates. Note that while the amplitude does decrease into the shadow, the shapes of the time functions are basically the same. It is possible to invert all of the seismograms simultaneously for a common time function. However, given the large fault areas of the earthquakes studied, we can expect real differences in the time functions due to directivity. The significance of any feature in a single station time function must be assessed by a comparison with the other single station solutions.

Figure 18 shows the time functions for the Rat Islands earthquake. The character of the time function, a sequence of pulses, is similar to that of the Kurile Islands earthquake. A closer examination indicates that the individual pulses have a longer duration and somewhat larger moment than the pulses of the Kurile Islands event. We will not pursue these differences, however, as the Alaskan time function is strikingly different from both the Kurile Islands and Rat Islands time functions.

6.1 Alaskan source time function

Figure 19 presents the most impressive results of the deconvolution study, though expected from our earlier discussion and the simple forward modeling. The first three minutes of the Alaskan re-

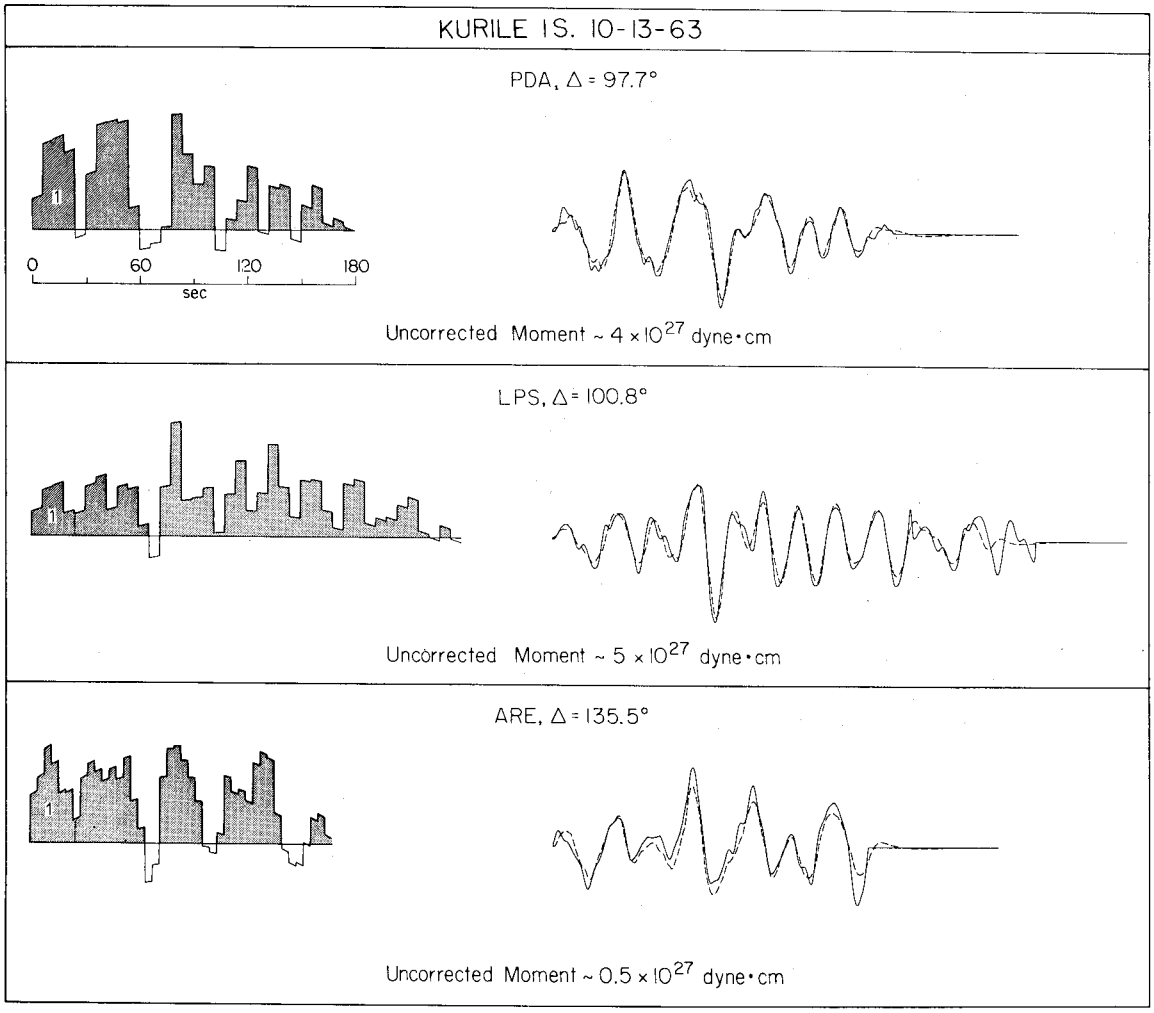

Fig. 17. Time functions for the Kurile Islands earthquake. The value of the damping constant is 10^{-2}. The common feature of these time functions is a sequence of events with a characteristic duration of 20–30 s, and the characteristic moment of each event is comparable to that of the Niigata event.

cords can be explained by one huge pulse, with a duration of at least two minutes. Although the forward modeling in Fig. 10 suggested that a rising-ramp of at least 40 s duration is required to match the first 30 s of the Alaskan records, the inversion results indicate that a rising-ramp of 60–70 s duration is more appropriate. A relatively smooth time function is clearly required by the data, as regardless of the damping used the time function is invariably smooth. In fact, it seems that our time function sampling is too coarse, thereby causing a poor fit in the first part of the records. Notice the moment of this huge pulse. Although the quoted moments can easily be in error by more than a factor of 2, recalling that these are minimum values, we can safely conclude that the Alaskan pulse has the largest moment, with the longest duration, ever reported from body wave investigations.

Before further discussion, or interpretation of the Alaskan time function, it is worthwhile to examine the stability of the solution with regard to

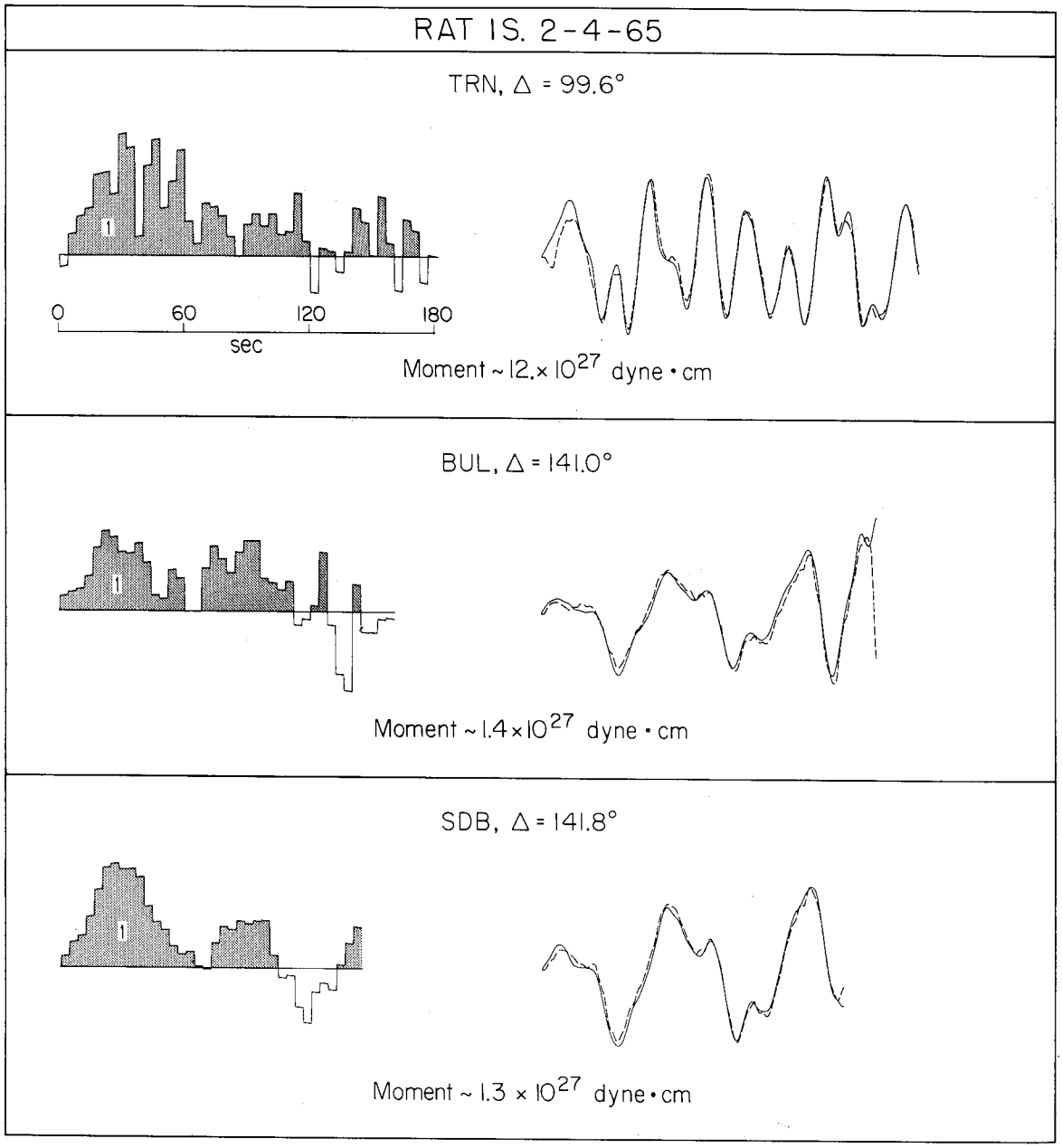

Fig. 18. Time functions for the Rat Islands earthquake. The damping constant is 10^{-3}. The time functions are similar to those for the Kurile Islands earthquake, except that the characteristic event duration is slightly longer and the characteristic moment is larger.

damping, focal depth and source finiteness. Figure 20 shows the time functions and synthetics for moderately damped solutions, $d = 0.01$; the same value used for the Kurile Islands solutions in Fig. 17. With this damping, the synthetics do not match the first part of the seismograms. Only by reducing

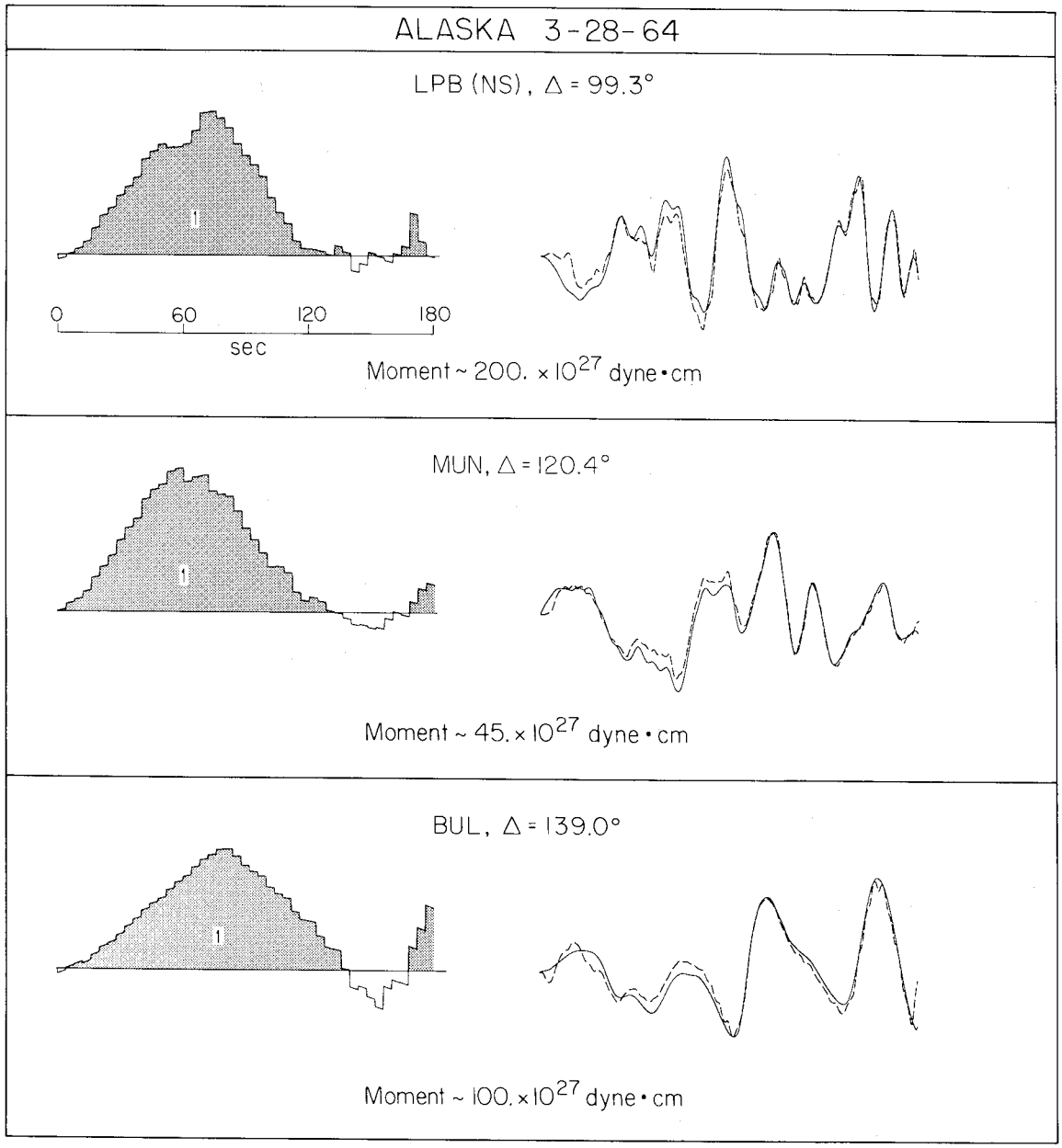

Fig. 19. Time functions for the Alaskan earthquake. The value of the damping constant is 10^{-4}. The Alaskan time function is characterized by one huge pulse with an initial ramp of 60 s duration. These time functions are deconvolved for the shallow source mechanism.

the damping to $d = 0.0001$ is an adequate fit obtained. Regarding the source depth, we have used two different focal mechanisms, corresponding to two acceptable mechanisms for the long period surface waves (Kanamori, 1970b). The deep mechanism at $h = 70$ km is inconsistent with first mo-

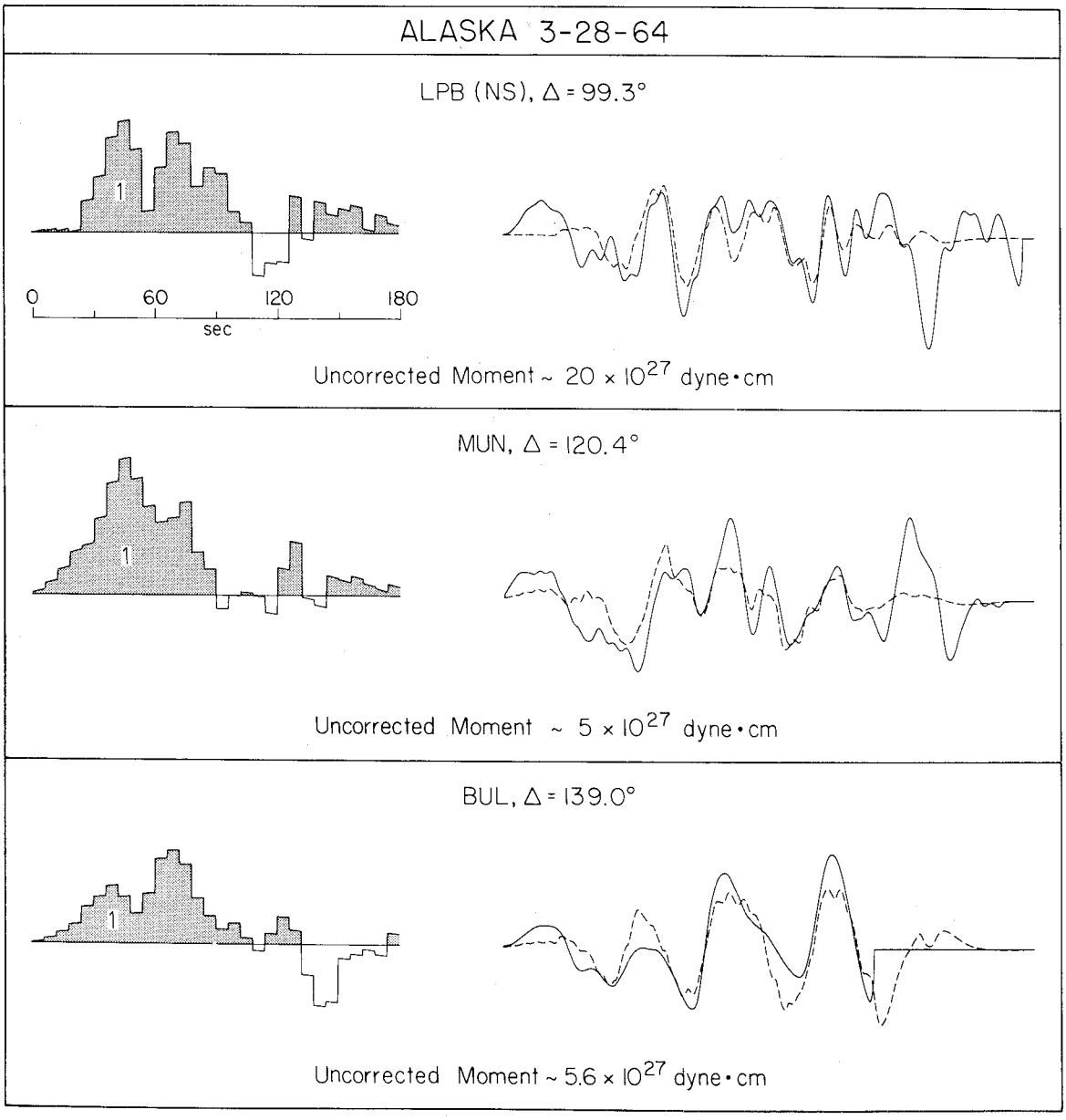

Fig. 20. Alaskan time functions for a damping of 10^{-2}. With this damping value, the first part of the seismograms cannot be matched. The data require a substantial long period component.

tions, yet inconsistent stations can be used by merely reversing the station polarity, and we then surprisingly obtain a compatible time function (Fig. 21). The shape of the time function is essentially the same as the shape in Fig. 19. The main difference is that the amplitude and moment are roughly half as large. Hence, the shape of the Alaskan time function is quite stable with regard to mechanism and depth, and this is further confirmed when using a distributed source. Figure 22

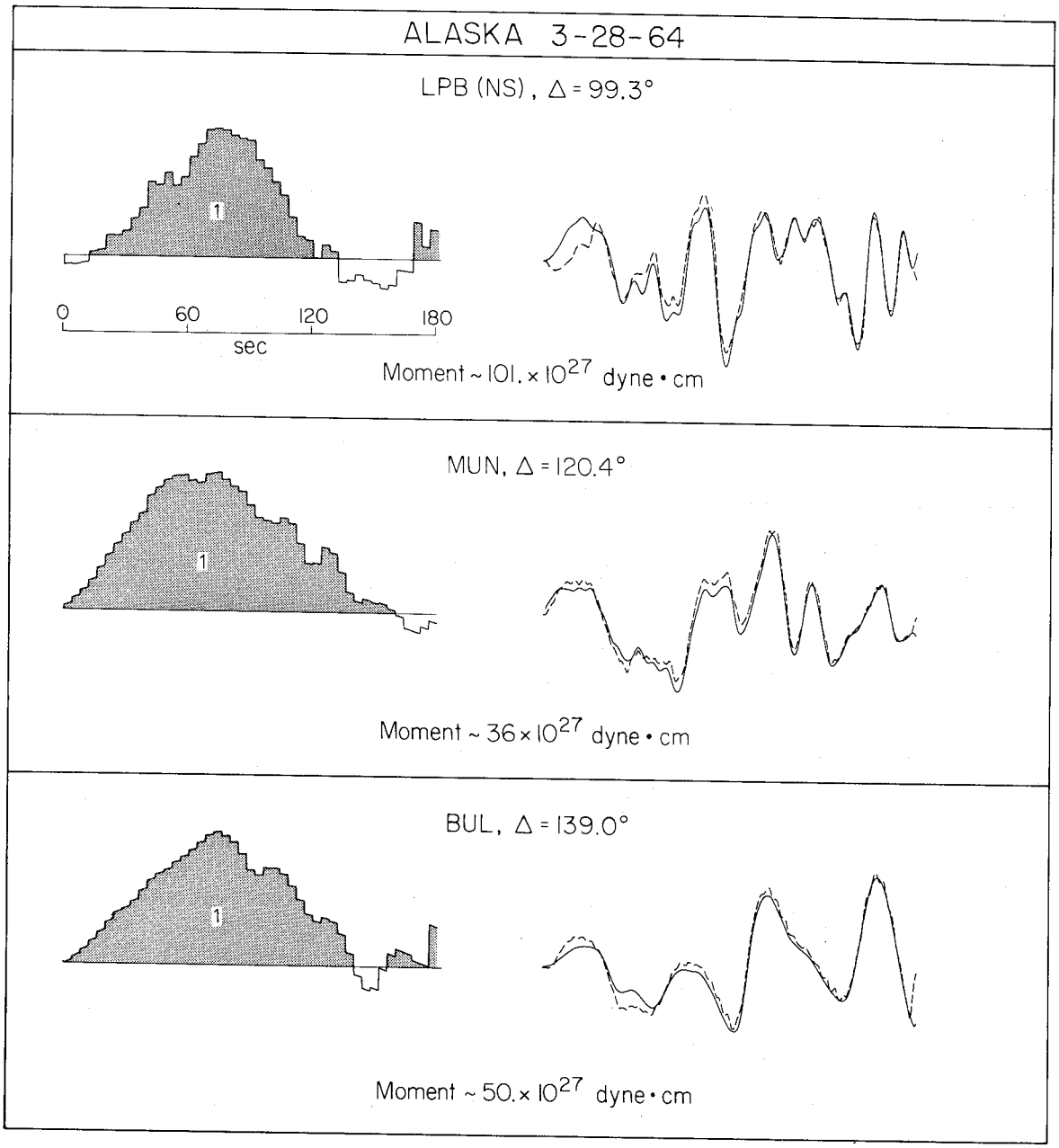

Fig. 21. Alaskan time functions for the deep ($h = 70$ km) source mechanism. The damping value is 10^{-4}. The time functions are essentially the same as those in Fig. 19 except that the moments are reduced by a factor of 2.

shows the time functions for BUL, calculated for the shallow fixed depth and a distributed source, and the two functions are quite similar in shape and amplitude. Therefore, we conclude that there are several important features of the Alaskan time function which are required to explain the observed seismograms: the "first event" of the Alaskan earthquake was a huge pulse with a ris-

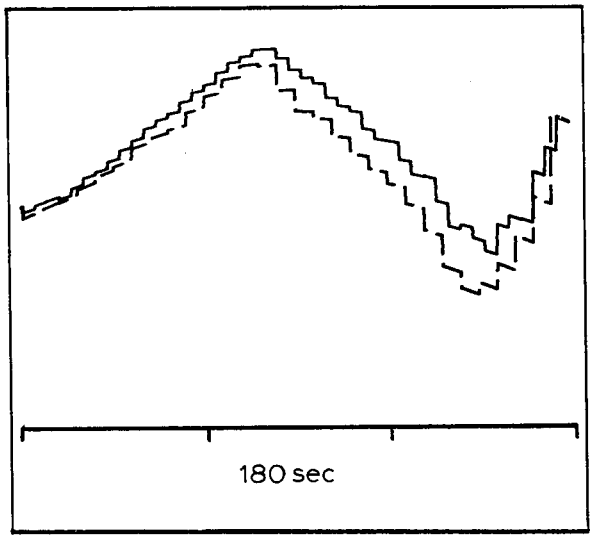

Fig. 22. Comparison of the Alaskan source time functions at BUL for point source and distributed source. The time function in Fig. 19 is replotted (solid). The dashed trace is the deconvolved time function for a source distributed between the surface and $h = 50$ km. The two functions are virtually identical in duration and moment.

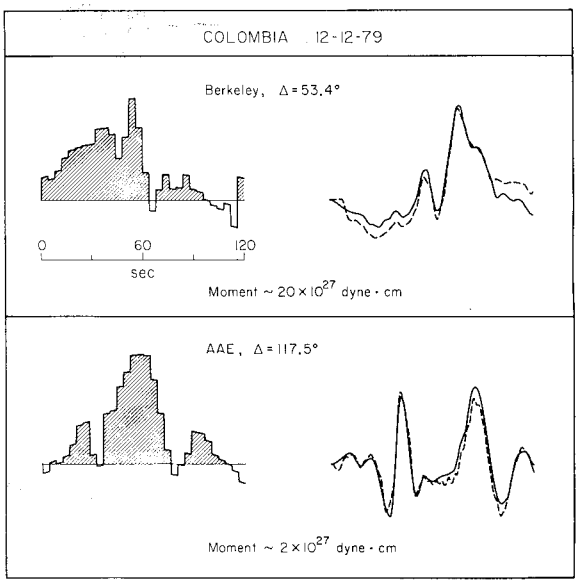

Fig. 23. Time functions for the 1979 Colombian earthquake. The damping level is 5×10^{-4}. The common features of the time function at these two stations are the long duration (60 s) and the abrupt termination at 1 minute.

ing-ramp of 1 minute duration and a total duration greater than 2 minutes, and the minimum moment release of this first pulse is 100×10^{27} dyne · cm.

6.2. Source time function of the 1979 Colombia earthquake

We have included the time function of the 1979 Colombian earthquake ($M_W = 8.2$) in Fig. 23. This is an interesting subduction zone event as it occurred in the region of the great earthquake of 1906 ($M_W \sim 8.8$, see Kanamori and McNally, 1982), and it also generated unusual P-waves (Fig. 23). Other subduction zone events of this size have waveshapes similar to those of the Niigata event, though perhaps of slightly longer duration (e.g., 1969 Kurile Islands and 1976 Philippines earthquakes). The WWSSN (15–100) recordings of the Colombian event show a peculiar double pulse character. This event was well recorded by the ultra-long period (100–300) seismograph at Berkeley. Though our results are tentative, the time functions shown in Fig. 23 are strongly suggestive of a circular type rupture, i.e. a ramp with a sharp cutoff. Notice that the duration is approximately 60 s and allowing for directivity, the rupture duration is at least 50 s. Hence, for a quarter-circular rupture (based on epicentral location) and 2 km s^{-1} rupture velocity, the asperity is ~ 100 km in size. Kanamori and Given (1981) estimated the total rupture length to be 230 km, thus the asperity is approximately half the fault area. Although the moment release of the Colombian earthquake is much smaller than the Alaskan earthquake, it is interesting that the Colombian time function indicates a large asperity size within the fault area of the great 1906 earthquake.

7. Interpretation

The time functions of the other events are plotted with the Alaskan time function at the same scale in Fig. 24. Though the Kurile Islands and Rat Islands earthquakes certainly have a long

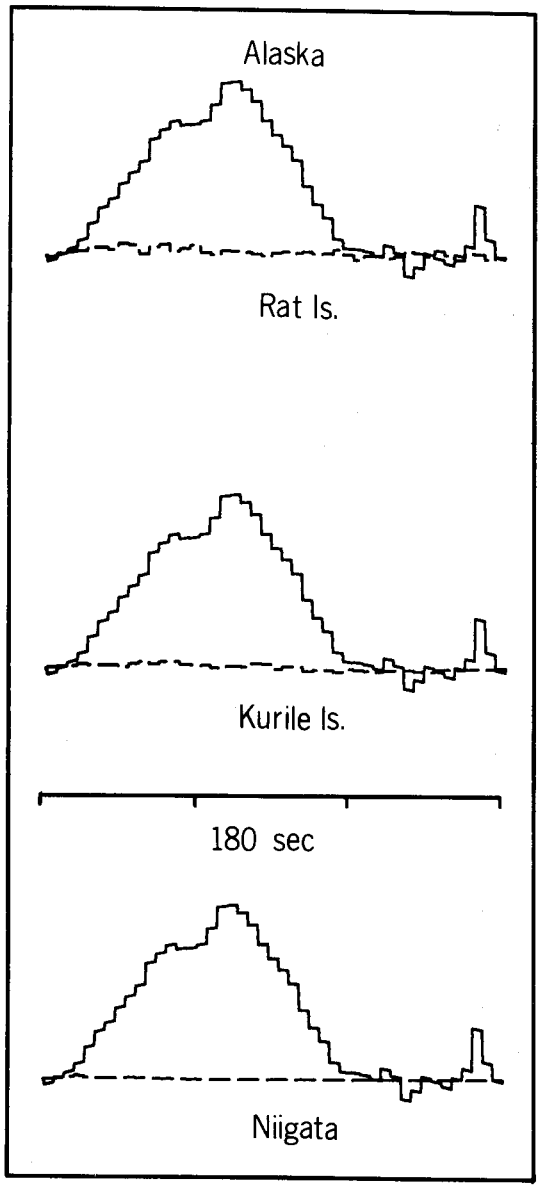

Fig. 24. Comparison of time functions, at approximately the same distance, at true relative amplitude scale. The Alaskan time function at LPB is plotted as the solid trace, while the Rat Islands (TRN), Kurile Islands (PDA), and the Niigata (ATL) time functions are plotted as dashed traces. Recalling that the amplitude scale is a moment rate, a substantially higher moment rate is resolved for the Alaskan earthquake.

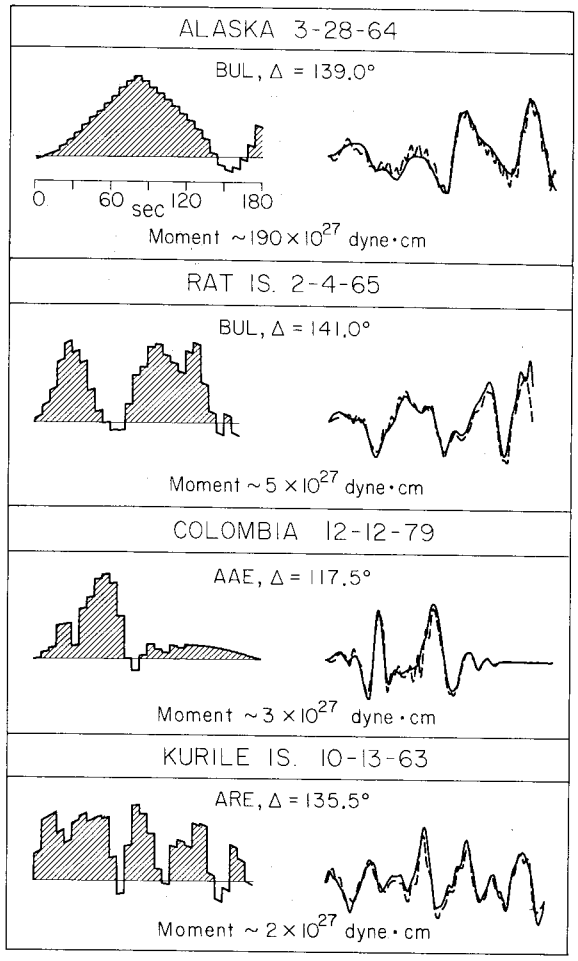

Fig. 25. Comparison of time functions, synthetic and observed seismograms at a large distance. The characteristic time scales can be directly compared, and varies from 2 minutes for the Alaskan earthquake to 30 s for the Kurile Islands earthquake. Thus, the large size of the Alaskan earthquake is associated with a longer characteristic time, hence a larger characteristic asperity length.

period component which would increase their amplitude, they would still appear small compared to the Alaskan time function. Of course, as the total moment of the Alaskan event is 750×10^{27} dyne·cm, it may not be surprising to recover a huge moment in the body waves. What may be more significant is the shape of the time function: smooth with an extremely long duration. The shape and duration of the time functions are contrasted in Fig. 25. Due to the unusual character and significance of the Alaskan time function, the translation of the time function into a length scale will be covered in some detail.

7.1. The Alaskan Asperity

The time functions for the Alaskan earthquake in Fig. 19 do not uniquely specify the rupture mode. The rupture is apparently a circular-type for the first minute, but, unlike the Colombian time function which ends abruptly, the Alaskan time function continues at some level after the ramp. The most reliable part of the Alaskan time function is the initial ramp, and we want to use this ramp to determine the length scale of the rupture. The simplest approach is to assume a rupture velocity, then the radius is $R \cong V_R t_R$, where V_R is the rupture velocity and t_R is the ramp duration. A typical value of the rupture velocity is 2 km s^{-1}, and with $t_R = 60$–70 s, the length scale is $R \cong 120$–140 km. If we consider the diameter to be the characteristic length scale, then $L = 2R \cong$ 240–280 km. We would like to estimate the length scale without simply assuming the rupture velocity. This can be done with additional information. For example, fixing the displacement allows an L_{min} to be determined.

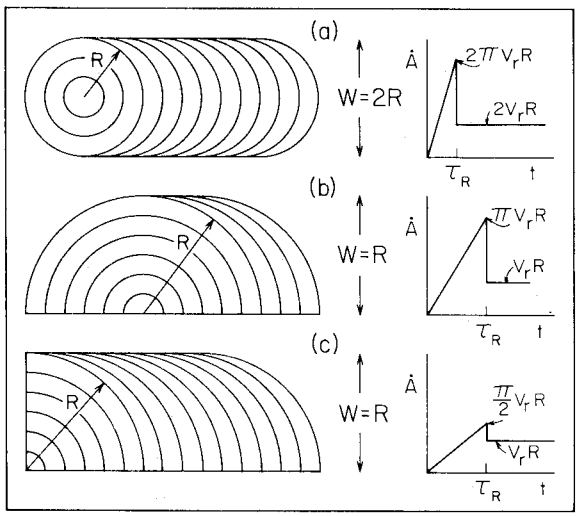

Fig. 26. Three examples of circular-type rupture modes. The fault planes are shown at left and the progress of the rupture front from the starting point is depicted by the circular arcs. The corresponding area rates are plotted as a function of time on the right. In (a), the rupture front is circular until it expands to the fault width, at which point it proceeds unilaterally down the fault. In (b), the rupture front is semi-circular initially, while in (c) it is quarter-circular. There are many other combinations that produce an initial circular-type rupture and thus a ramp time function.

There are islands in the vicinity of the Alaskan epicenter which were displaced horizontally and vertically by the earthquake (Hastie and Savage, 1970; Plafker, 1972). These static deformations have been modeled by Alewine (1974) and Miyashita and Matsúura (1978) to determine the final displacements on the main fault and a small imbricate fault. Though the solutions are not unique, an average static displacement of 15–18 m on the main fault in the epicentral region is indicated. Based on the long period surface waves, Kanamori (1970b) obtained a seismic moment of 750×10^{27} dyne · cm. Using a fault area of 15×10^4 km^2 (estimated from the aftershock distribution) and $\mu = 700$ kb, the average displacement is 7 m (or 10 m for $\mu = 500$ kb). Thus, the final displacement near the epicenter is twice the average. Some of the final displacement may have been post-seismic. However, using the upper bound of 15–18 m, displacement provides an estimate of the minimum scale length.

From eq. 6, the time function results from the convolution of the particle velocity and area rate, $m(t) = \mu \dot{D}(t) * \dot{A}(t)$. If the duration of $\dot{D}(t)$ is much less than the duration of $\dot{A}(t)$, then the main effect of the convolution is to multiply $\dot{A}(t)$ by the average displacement, $m(t) = \mu D \dot{A}(t)$. To obtain the moment released up to a particular time, say τ_R, the above expression is integrated from $t = 0$ to $t = \tau_R$ which gives $M_R = \mu D A(\tau_R)$. Thus the ruptured area at time τ_R is, $A(\tau_R) = M(\tau_R)/\mu D$, with minimum $M(\tau_R)$ and maximum D estimates providing a minimum estimate of the area. The Alaskan time function has a ramp time of $\tau_R = 60$–70 s and a moment release of 10^{29} dyne · cm at this time. With $\mu = 500$ kb, the fault area at τ_R is $A = 1.3$–1.1×10^4 km^2 for $D = 15$–18 m. Translating the area into a characteristic length as $L = A^{1/2}$, $L = 115$–105 km. A unilateral rupture across a square fault of dimension 115×115 km, with a rupture time of 60–70 s, yields a rupture velocity of 1.7–1.6 km s^{-1}. However, we know from the ramp shape of the time function that a unilateral rupture is not appropriate. Figure 26 shows three rupture modes which start as a circular-type rup-

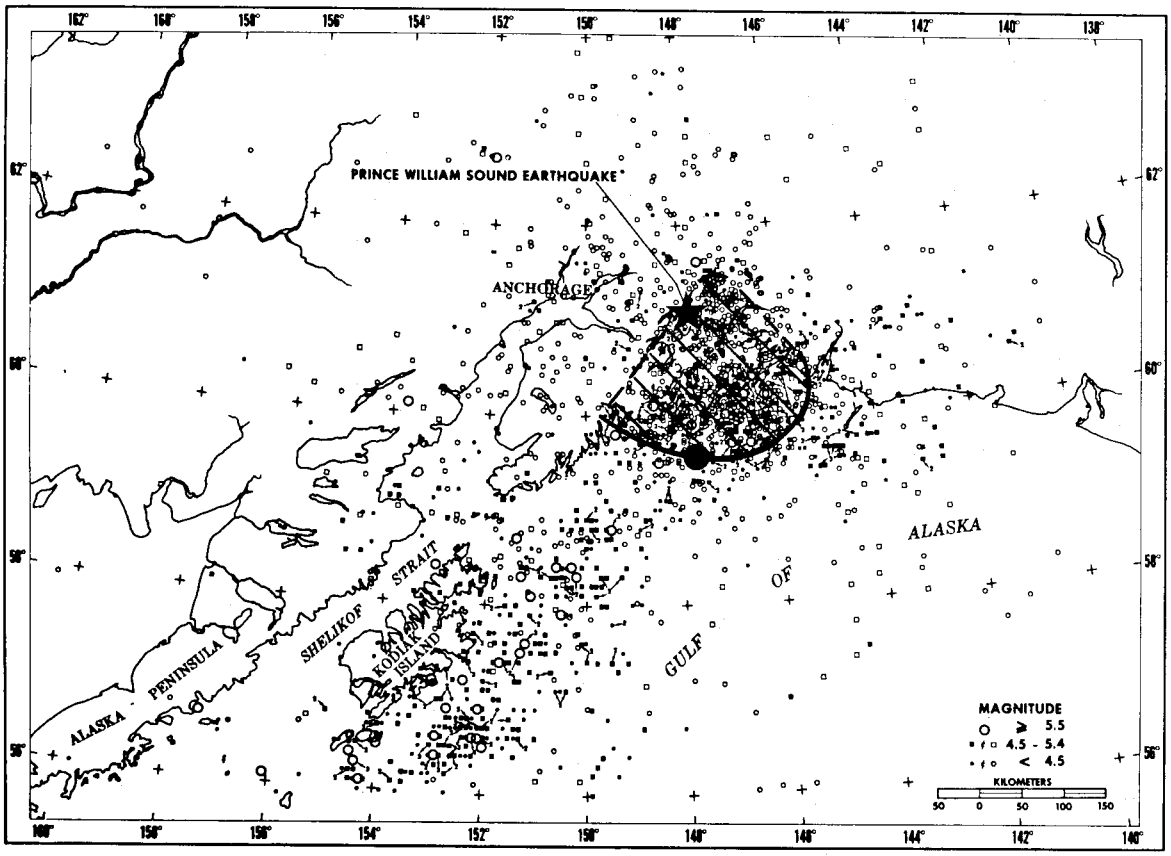

Fig. 27. The Alaskan asperity. The aftershocks of the 1964 Alaskan earthquake (epicenter indicated by star) are plotted for a 20 month period (from Algermissen et al., 1969). The small open squares and circles are aftershocks located with the temporary local network. Notice the concentration of aftershocks in the northeastern part of the fault zone (the rupture extended to the south of Kodiak Island). The heavy line, which delimits the high aftershock activity, is at 140–180 km from the epicenter. This distance is comparable to the length scale of the initial circular-type rupture estimated from the time function. Also, the location of event C (plotted as the large dot) from the study of Wyss and Brune (1967) provides additional evidence that the ruptured area extended to 180 km in the first minute. We conclude that the hachured region is a giant asperity.

ture and then continue along the fault in a unilateral fashion. The location of the Alaskan epicenter, relative to the aftershocks, suggests that the quarter-circular rupture in Fig. 26(c) is a reasonable representation of the Alaskan earthquake. In this case, $R = (4/\pi)^{1/2} A^{1/2} = 130$–$120$ km. With $\tau_R = 65$ s, we obtain a rupture velocity of 2.0–1.8 km s^{-1}. Therefore, assuming a seismic displacement of 18 m over the entire northeastern segment of the fault, we find that the minimum scale length ruptured in the first minute is 120 km. For a reasonable upper bound on the scale length: assume the seismic displacement is just 10 m and $M(\tau_R) = 2 \times 10^{29}$ dyne · cm, then $R = 240$ km. In support of this length scale, Kanamori (1970b) suggested that the fault width is at least 200 km on the basis of both the aftershock pattern and surface wave directivity. The distribution of aftershocks is quite interesting, as there seems to be a concentration of aftershocks within 200 km of the epicenter (Fig. 27). Although we do not know why the aftershocks are concentrated in the northeastern segment, this spatial variation indicates a characteristic length scale of ~ 200 km.

Regarding the initial rupture length scale, the results of Wyss and Brune (1967), using short period records to identify multiple events, are of interest. In view of the small short period amplitudes relative to the long period amplitudes, it appears that the short period multiple events are small glitches on the rising-ramp time function. However, a well correlated short period pulse could serve as a "marker". The largest and most dependable short period pulse (event C of Wyss and Brune, 1967) is located at the edge of the concentration of aftershocks (Fig. 27) and is delayed by 44 s. This supplies additional evidence that the initial rupture length scale extended to approximately 200 km in the southwest direction in the first minute of rupture.

8. Conclusions

The main conclusion of this paper is that the characteristic asperity scale length increases as M_W increases, with the extreme case of large scale length represented by the Alaskan earthquake. This conclusion stems directly from the large differences in the observed seismograms and is summarized in Fig. 28.

The seismograms are modeled to determine reliable quantitative features of the time functions. The Kurile Islands earthquake is characterized by a sequence of pulses, each pulse of 20–30 s duration with a seismic moment $< 10^{28}$ dyne · cm. Thus, we can view this rupture process as a sequence of magnitude 8 events plus an overall longer period component. Each pulse represents the breaking of a single asperity with a scale length of 40–60 km. The Rat Islands earthquake is similar in character to the Kurile Islands earthquake, except that the moment of the individual pulses is larger and the duration is somewhat longer, indicating an asperity scale length of 40–80 km. The 1979 Colombian earthquake began as a circular-type rupture with a single dominant pulse of 1 minute duration, and an associated length scale of 100–120 km.

The Alaskan earthquake is characterized by a smooth rupture for at least three minutes with the initial rupture consisting of a circular-type rupture

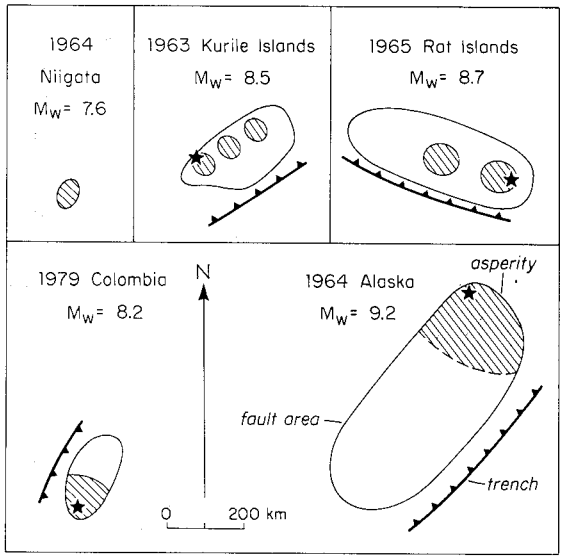

Fig. 28. Interpretation of the time functions as variations in the asperity distribution. For each event, the fault area and inferred asperities (hachured) are schematically graphed. The trench axes and epicenters (stars) are also shown. The Niigata event ruptured one asperity with a length scale of 50 km. The Kurile Islands earthquake ruptured at least three Niigata-sized asperities, while the Rat Islands earthquake ruptured larger asperities. The Colombian earthquake broke one large (~ 100 km) asperity, and the Alaskan earthquake started with the smooth rupture of a giant (150–200 km) asperity.

for at least the first minute. The relative lack of energy in the shorter periods allows us to directly observe the longer period components of the time function (e.g., Fig. 8). It seems that the length scale associated with the first minute of rupture is 140–200 km. Although the entire Alaskan earthquake may have been a relatively smooth rupture, the northeastern (epicentral) segment appears to be a giant asperity as the moment release was greater in this region, if the observed static deformation was co-seismic.

We conclude that the Alaskan earthquake occurred due to the failure of a smooth, strong asperity with a dominant length scale no less than 140 km. The rupture process of great earthquakes varies from a sequence of magnitude 8 events (e.g., the Kurile Islands earthquake) to the limiting case of a broad contact surface that is coupled by a giant strong asperity (e.g., the Alaskan earthquake).

Acknowledgments

This research was supported by the Division of Earth Sciences, National Science Foundation Grant No. 11973, the National Aeronautic and Space Administration Grant NSG-7610, and Exxon Production Research Agreement No. PR-7373. Contribution No. 3732, Division of Geological and Planetary Sciences, California Institute of Technology, Pasadena, California 91125.

References

Abe, K., 1975. Re-examination of the fault model for the Niigata earthquake of 1964. J. Phys. Earth, 23: 349–366.

Aki, K. and Richards, P.G., 1980. Quantitative Seismology. Freeman, San Francisco, 932 pp.

Alewine, R.W., 1974. Application of linear inversion theory toward the estimation of seismic source parameters. Ph.D. Thesis, California Institute of Technology, 303 pp.

Algermissen, S.T., Rinehart, W.A., Sherburne, R.W. and Dillinger, W.H., 1969. Preshocks and aftershocks of the Prince William Sound earthquake of March 28, 1964. In: L.E. Leipold (Editor), The Prince William Sound, Alaskan Earthquake of 1964 and Aftershocks, Vol. II. U.S. Department of Commerce, Coast and Geodetic Survey, Washington, DC, 350 pp.

Backus, G. and Gilbert, F., 1970. Uniqueness in the inversion of inaccurate gross Earth data. Phil. Trans. R. Soc. Ser. A, 266: 123–192.

Doornbos, D.J. and Mondt, J.C., 1979. P and S waves diffracted around the core and the velocity structure at the base of the mantle. Geophys. J.R. Astron. Soc., 57: 381–395.

Hastie, L.M. and Savage, J.C., 1970. A dislocation model for the Alaska earthquake. Bull. Seismol. Soc. Am., 60: 1389–1392.

Herrin, E. (Chairman), 1968. 1968 Seismological tables for P phases. Bull. Seismol. Soc. Am., 58: 1193–1241.

Hirasawa, T., 1965. Source mechanism of the Niigata earthquake of June 16, 1964, as derived from body waves. J. Phys. Earth, 13: 35–66.

Jackson, D.D., 1972. Interpretation of inaccurate, insufficient, and inconsistent data. Geophys. J.R. Astron. Soc., 28: 97–109.

Kanamori, H., 1970a. Synthesis of long-period surface waves and its application to earthquake source studies—Kurile Islands earthquake of October 13, 1963. J. Geophys. Res., 75: 5011–5027.

Kanamori, H., 1970b. The Alaska earthquake of 1964: radiation of long-period surface waves and source mechanism. J. Geophys. Res., 75: 5029–5040.

Kanamori, H., 1971. Great earthquakes at island arcs and the lithosphere. Tectonophysics, 12: 187–198.

Kanamori, H., 1977a. The energy release in great earthquakes. J. Geophys. Res., 82: 2981–2987.

Kanamori, H., 1977b. Seismic and aseismic slip along subduction zones and their tectonic implications. In: M. Talwani and W.C. Pitman III (Editors), Island Arcs, Deep Sea Trenches, and Back-Arc Basins. American Geophysical Union, Washington, DC, 480 pp.

Kanamori, H., 1981. The nature of seismicity patterns before large earthquakes. In: D.W. Simpson and P.G. Richards (Editors), Earthquake Prediction, an International Review. American Geophysical Union, Washington, DC, 680 pp.

Kanamori, H. and Anderson, D.L., 1975. Theoretical basis of some empirical relations in seismology. Bull. Seismol. Soc. Am., 65: 1073–1095.

Kanamori, H. and Stewart, G.S., 1976. Mode of the strain release along the Gibbs fracture zone, mid-Atlantic ridge. Phys. Earth Planet. Inter., 11: 312–332.

Kanamori, H. and Given, J.W., 1981. Use of long-period surface waves for rapid determination of earthquake-source parameters. Phys. Earth Planet. Inter., 27: 8–31.

Kanamori, H. and McNally, K.C., 1982. Variable rupture mode of the subduction zone along the Ecuador–Colombia coast. Bull. Seismol. Soc. Am., 72: 1241–1253.

Kelleher, J. and McCann, W., 1976. Buoyant zones, great earthquakes, and unstable boundaries of subduction. J. Geophys. Res., 81: 4885–4896.

Kelleher, J., Savino, J., Rowlett, H. and McCann, W., 1974. Why and where great thrust earthquakes occur along island arcs. J. Geophys. Res., 79: 4889–4899.

Lanczos, C., 1961. Linear Differential Operators, Van Nostrand, London.

Langston, C.A. and Helmberger, D.V., 1975. A procedure for modeling shallow dislocation sources. Geophys. J.R. Astron. Soc., 42: 117–130.

Lay, T. and Kanamori, H., 1981. An asperity model of great earthquake sequences. In: D.W. Simpson and P.G. Richards (Editors), Earthquake Prediction, an International Review. American Geophysical Union, Washington, DC, 680 pp.

Mikumo, T. and Miyatake, T., 1978. Dynamical rupture process on a three-dimensional fault with non-uniform frictions and near-field seismic waves. Geophys. J.R. Astron. Soc., 54: 417–438.

Miyashita, K. and Matsúura, Y., 1978. Inversion analysis of static displacement data associated with the Alaska earthquake of 1964. J. Phys. Earth., 26: 333–349.

Mula, A.H.G., 1981. Amplitudes of diffracted long-period P and S waves and the velocities and Q structure at the base of the mantle. J. Geophys. Res., 86: 4999–5011.

Oldenburg, D.W., 1981. A comprehensive solution to the linear deconvolution problem. Geophys. J.R. Astron. Soc., 65: 331–357.

Parker, R.L., 1977. Understanding inverse theory. Annu. Rev. Earth Planet. Sci., 5: 35–64.

Plafker, G., 1972. Alaskan earthquake of 1964 and Chilean earthquake of 1960: implications for arc tectonics. J. Geophys. Res., 77: 901–925.

Ruff, L.J. and Helmberger, D.V., 1982. The structure of the lowermost mantle determined by short period P-wave amplitudes. Geophys. J.R. Astron. Soc., 68: 95–119.

Ruff, L. and Kanamori, H., 1980. Seismicity and the subduction process. Phys. Earth Planet. Inter., 23: 240–252.

Ruff, L. and Kanamori, H., 1982a. Seismic coupling and uncoupling at subduction zones. Tectonophysics, (in press).

Sykes, L.R. and Quittmeyer, R.C., 1981. Repeat times of great earthquakes along simple plate boundaries. In: D.W. Simpson and P.G. Richards (Editors), Earthquake Prediction, an International Review. American Geophysical Union, Washington, DC, 680 pp.

Toksöz, M. and Hsui, A., 1978. Numerical studies of back-arc convection and the formation of marginal basins. Tectonophysics, 50: 177–196.

Uyeda, S. and Kanamori, H., 1979. Back-arc opening and the mode of subduction. J. Geophys. Res., 84: 1049–1061.

Wiggins, R.A., 1972. The general linear inverse problem: implications of surface waves and free oscillations for Earth structure. Rev. Geophys. Space Phys., 10: 251–285.

Wyss, M. and Brune, J.N., 1967. The Alaska earthquake of 28 March 1964: A complex multiple rupture. Bull. Seismol. Soc. Am., 57: 1017–1023.