Evolution of the Downgoing Lithosphere and the Mechanisms of Deep Focus Earthquakes

M. Nafi Toksoz, Norman H. Sleep and Albert T. Smith

(Received 1973 April 25)†

Summary

The thermal evolution of the lithospheric slab at subduction zones and its geophysical effects are numerically calculated. An alternating-direction, implicit, finite-difference scheme is used to compute the thermal models taking into account all heating sources and phase boundaries. These models, with the appropriate spreading rates and dip angles, are compared with different island arc systems. Temperatures inside the slab are strongly controlled by the conductivity and by the time elapsed since the initiation of descent. The depth to which temperature anomalies persist is generally about 700 km or less.

The thermal results are used to construct the seismic velocities and ray paths, density anomalies, and the resulting stress distribution. Comparing the theoretical stress distribution and the focal mechanism studies of intermediate- and deep-focus earthquakes indicates the importance of both the mantle’s rheology and the temperature dependence of the slab’s elastic properties. The intermediate and deep focus earthquakes are located along the coolest region of the slab. The theoretical results explain the source mechanisms and the orientation of principal stresses under major island arcs.

Introduction

The interest in the distribution and nature of deep-focus earthquakes has been intensive and the circum-Pacific belt has been extensively studied (Honda 1934; Wadati 1935; Gutenberg & Richter 1954; Oliver & Isacks 1967; Sykes 1966; Kondorskaya & Postolenko 1959). A renewed interest in and much better understanding of island arcs and deep seismicity have emerged in recent years with the sea-floor spreading and the global tectonic hypotheses (Isacks & Molnar 1971; Griggs 1972; Smith & Toksoz 1972).

In this paper we discuss the thermal regime and the stress field inside a descending lithospheric plate. First we present theoretical models of the thermal evolution of a downgoing slabs. Later, we compute the resulting stress fields and compare these with the distribution and mechanisms of intermediate and deep focus earthquakes.

Thermal evolution of a downgoing slab

The motion of the descending slab is a continuation of the lithospheric spreading. Once the kinematics of the motion are specified, the calculation of the temperature field becomes practical without solving the full convection problem for the Earth.

† Received in original form 1972 December 15.
Studies of the thermal regimes of downgoing slabs have been carried out by several investigators using different models and techniques. McKenzie (1969) used a two-dimensional analytic formulation which assumed the top and the bottom of the slab were isotherms. Griggs (1972) included the phase changes in a numerical solution with similar boundary conditions. Turcotte \& Oxburgh (1968) obtained a solution using boundary layer theory. Minear \& Toksöz (1970a, b), Hasebe, Fujii \& Uyeda (1970), and Toksöz, Minera \& Julian (1971), used two-dimensional numerical models in which the mantle surrounding the slab was explicitly included so energy was conserved.

In this section we briefly review the numerical calculations following Toksöz et al. (1971). We especially focus our attention to effects of various physical properties and to the questions related to geometry and the rate of descent.

![Diagram of computational scheme and boundary conditions](https://academic.oup.com/gji/article-abstract/35/1-3/285/615604)

**Fig. 1.** Schematic representation of the computational scheme, input parameters, and boundary conditions used to calculate thermal models of slabs (modified after Minear \& Toksöz 1970a).

**Computation of temperature field**

The computational scheme used in this study is similar to the method described in detail in an earlier paper (Minear \& Toksöz 1970a). The basic model consists of a slab of material moving downward into the mantle at a specified angle. The surrounding mantle material is assumed to be fixed (Fig. 1). The computational scheme consists of translating temperatures downward in the slab and then allowing the slab to warm up over the time interval $\Delta t$, which corresponds to the vertical movement $d$. In essence we assume the dynamics and compute the temperature field given this motion field.

Temperatures are computed from the conservation of energy equation

$$C_p \rho \frac{\partial T}{\partial t} = \nabla \cdot (K \nabla T) + H$$

(1)
where
\[ C_p = \text{specific heat at constant pressure}, \]
\[ \rho = \text{density}, \]
\[ T = \text{temperature}, \]
\[ K = \text{conductivity}, \]
\[ H = \text{heat generation rate per unit volume}. \]

The alternating-direction, implicit, finite-difference scheme (Peaceman & Rachford 1955) is used to solve (1) numerically. The method was described earlier by Minear & Toksöz (1970a).

**Physical parameters and energy sources**

Certain physical parameters (conductivity, specific heat, density, mantle geotherm) and the energy sources that are incorporated in \( H \) (radioactivity, adiabatic compression, phase changes and shear-strain heating) must be specified for solving (1). Some parameters can be specified and remain nearly constant as a function of position. Others strongly depend on temperature and need to be updated at each step and every grid point.

**Physical parameters**

In these calculations we used the average density profile of the oceanic model (Press 1970) and a constant specific heat of \( C_p = 1.3 \times 10^7 \text{erg g}^{-1} \text{°C}^{-1} \).

In the thermal conductivity the contributions of both the lattice conduction and radiative heat transfer were taken into account. One set of models was based on the formulation of MacDonald (1959).

\[
k = k_1 + \frac{(16n^3sT^3)}{(n\epsilon_0 + 120\pi\sigma_0 \exp (-E/kT))}
\]

where
\[
k_1 = \text{lattice conductivity}, \]
\[ 0.25 \times 10^6 \text{erg cm}^{-1} \text{s}^{-1} \text{°C}^{-1}, \]
\[ T = \text{absolute temperature}, \]
\[ K = \text{Boltzmann's constant}, \]
\[ E = \text{width of energy gap for electronic conduction}, 3 \text{ eV}, \]
\[ s = \text{Stefan-Boltzmann constant}, \]
\[ n = \text{index of refraction} (n = 1.7), \]
\[ \epsilon_0 = \text{low temperature opacity} (\epsilon_0 = 10 \text{ cm}^{-1}), \]
\[ \sigma_0 = \text{electrical conductivity} (\sigma_0 = 10 \text{ ohm}^{-1} \text{cm}^{-1}). \]

The other set of models used the experimental results of Schatz & Simmons (1972) on olivine. In this case the lattice conductivity is the larger of

\[
k_1 = (30.6 + 0.21T)^{-1}
\]

\[
k_1 = 0.003 + (3 \times 10^{-6}) z
\]

where \( z \) is in kilometres, \( T \) in °K and the unit of conductivity is \( \text{cal cm}^{-1} \text{s}^{-1} \text{°C}^{-1} \). The radiative conductivity is given by

\[
k_r = 0 \quad T \leq 500 \text{°K}
\]

\[
k_r = 5.5 \times 10^{-6}(T - 500) \quad T > 500 \text{°K}.
\]

The conductivity change associated with the phase change at 400 km is probably not very large since ferrous iron, which increases opacity and reduces radiative conductivity, enters the common phases on both sides of the transition zone (Ringwood 1970). If iron-free oxides such as MgO, Al₂O₃ or SiO₂ exist as minerals below the 600-km discontinuity, the radiative conductivity would be extremely high.

In earlier studies the unperturbed geotherm of the mantle was either assumed to be adiabatic (McKenzie 1969), the conductivity geotherm (Minear & Toksöz 1970a, b;
Toksoz et al. (1971) or to be controlled by the solidus temperature of peridotite (Griggs 1972). In this paper an approximation to the average geotherm in a convecting Earth was also used. The unperturbed geotherm is indicated in each temperature figure in the following section.

**Heat sources**

In addition to conduction of heat from the surrounding mantle, the slab is heated by internal heat sources. These consist of radioactivity, adiabatic compression, phase changes, and the shear strain heating along the slab–mantle boundaries.

**Radioactive heating.** The average radioactivity of the oceanic lithosphere is probably about equal to the radioactivity of the upper mantle, because differentiation at mid-oceanic ridges occurs mainly above 30 km depth (Kay, Hubbard & Gast 1970).

Radioactive heating in the upper mantle of the Earth is too low to seriously affect the evolution of the slab. A more important effect of radioactive heating is on the unperturbed geotherm of the upper mantle. The estimates of heat generation due to radioactivity cover a wide range: $2.3 \times 10^{-7}$ erg g$^{-1}$ s$^{-1}$ (MacDonald 1959, 1963), $5.4 \times 10^{-8}$ erg g$^{-1}$ s$^{-1}$ (Armstrong 1968), and $1.5 \times 10^{-8}$ erg g$^{-1}$ s$^{-1}$ (Hurley 1968a, b). The larger radioactivity of MacDonald would increase the temperature only 6°C in the 10 My descent of the slab. In these calculations we have used heat production values of $2.3 \times 10^{-7}$ erg g$^{-1}$ s$^{-1}$ for the upper mantle and $1.67 \times 10^{-8}$ erg g$^{-1}$ s$^{-1}$ for the mantle below 465 km.

**Adiabatic heating.** As the lithosphere descends into the mantle, it is compressed and heated. The adiabatic temperature gradient is given by

$$\frac{\partial T}{\partial z} = g \alpha T/C_P$$

where $g$ is the gravitational acceleration and $\alpha$ is the volume coefficient of thermal expansion. The rate of thermal energy release at depth $h$ due to adiabatic compression is

$$\frac{dQ}{dt} \bigg|_h = C_P \rho (\partial T/\partial t) = C_P \rho (\partial T/\partial z) V_z = \rho g \alpha T V_z$$

where $V_z$ is the vertical component of the velocity of the downgoing slab. The value $dQ/dt$ can be evaluated at each point by using appropriate values for $\rho$, $g$, $\alpha$, and $T$ (Hanks & Whitcomb 1971). $\rho$ and $g$ are known and $T$ is computed from the previous time step. Laboratory measurements of $\alpha$ tabulated by Skinner (1966) indicated a slight increase in $\alpha$ with increasing temperature and a decrease in $\alpha$ with increasing pressure. For these calculations we expressed the depth dependence of $\alpha$ by

$$\alpha = \exp(3.58 - 0.0072z)$$

where $z$ is depth in kilometres and $\alpha$ is measured in $10^{-6}$ °C$^{-1}$. Near the surface ($z = 0$) the value of $\alpha = 36 \times 10^{-6}$ °C$^{-1}$ corresponds to that of olivine at about $T = 500$ °C (Skinner 1966). The exponential coefficient 0.0072, for depth dependence, is somewhat lower than estimates of Birch (1968) and slightly higher than those of Verhoogen (1951). The above expression gives a value of $\alpha = 18 \times 10^{-6}$ °C$^{-1}$ at $z = 1000$ km, in agreement with the value used by Griggs (1972).

Since $\alpha$ decreases with depth while the temperature is increasing inside the slab, the adiabatic heat generation $(dQ/dt)$ varies slowly, increasing slightly with depth.
Phase changes. Several phase changes may occur in the upper mantle. Seismic velocity profiles indicate the presence of second-order discontinuities at depths of about 350 and 650 km (Toksoz, Chinnery & Anderson 1967; Johnson 1967; Julian & Anderson 1968; Archambeau, Flinn & Lambert 1969) that are most likely caused by olivine to spinel and spinel to post-spinel changes (Anderson 1967). In addition to these, a phase change corresponding to a basalt-eclogite or a plagioclase peridotite-garnet peridotite reaction is considered at shallower depth.

If the vertical velocity through a phase change is fast enough to disturb the isotherms, the phase boundary will be moved from its normal position (Schubert & Turcotte 1971). To calculate the effects of phase changes, all points in the model were examined at the end of each translation step and the change in the amount of each phase from the previous step recorded. Latent heat was calculated from this change and included as a heat source in the model. The routine for translating temperature was also used to determine the amount of each phase with the slab. This calculation scheme was used for phase changes going both ways whether or not the slab is assumed to move.

Three phase changes were considered in the models. Phase change energies were computed from entropy \( S \) and volume \( V \) changes. Because the temperatures vary inside the slab, the phase boundaries were determined from the \( dP/dT \) slopes for each phase change, starting from the shallowest. The \( \Delta S \) values for the three phase changes are taken as \(-11.7 \times 10^5\), \(-7.13 \times 10^5\), and \(-5.94 \times 10^5\) erg g\(^{-1}\) C\(^{-1}\). Corresponding \( \Delta V \) values are \(-6.5 \times 10^{-2}\), \(-2.65 \times 10^{-2}\), and \(-2.62 \times 10^{-2}\) cm\(^3\) g\(^{-1}\). These values are averages of those given by Verhoogen (1965), Akimoto & Fujisawa (1968), Ringwood (1970, 1972), Sclar, Carrison & Schwartz (1964). The parameters for the spinel–post-spinel phase changes are the least reliable. These reactions were studied indirectly with chemical analogues (Akimoto 1970; Ringwood & Major 1970). Ringwood (1972) takes this reaction to be exothermic (and \( dP/dT > 0 \)) while others (D. L. Anderson 1972, private communication) suggest that it may be endothermic. Because of these uncertainties, we constrained the phase boundary to remain at 650 km depth, but adopted a positive heat contribution. We also calculated some models without any heat contribution from this phase boundary. No allowance was made for any changes in heat balance due to possible melting of the material.

Shear-strain heating. An accurate estimate of the shear-strain heating is extremely difficult, because we know neither the mechanisms involved in the descent of the slab nor the parameters, such as the thickness of the shear zone and its viscosity. The complexity of the problem even for an idealized fluid model has been illustrated by Turcotte & Oxburgh (1968).

In the calculations we used several values for shear-strain heating. In all cases viscous heat generation was confined to 14 km thick layers along the top and bottom of the slab. At the top edge, the maximum shear-strain heat generation rates are \(1.6 \times 10^{-4}\) erg cm\(^{-3}\) s\(^{-1}\), unless specified otherwise. At the lower boundary, the shear-strain heating is \(1.6 \times 10^{-5}\) erg cm\(^{-3}\) s\(^{-1}\) (a factor of 10 less than the top) and it has a negligible effect on the thermal regime.

At the 8 cm/yr subduction rate, the frictional heating used at the top of the slab is equivalent to a stress of a few kilobars (in an extreme case it could be as high as 4 kb). Although this is high compared to apparent stress drop associated with earthquakes, there is evidence that high stresses prevail in the lithosphere at the convergence zones. For example, to produce the observed uplift at the outer rise of the Japan trench, stresses of a few kilobars are needed (Hanks 1971).

Temperature models

A series of temperature models were computed to demonstrate the evolution of the thermal regime. Only a few models will be shown in this paper in addition to
those in the previous papers to demonstrate the effects of important parameters and varying geometry (Minear & Toksoz 1970a, b; Toksoz et al. 1971).

The development of the thermal regime in a downgoing slab is shown in Figs 2 and 3 as a function of time. Temperature fields shown after 3.6 and 7.1 My (Fig. 2) and 10.7 My (Fig. 3) from the start of downward motion demonstrate the relative effects of time and crossing of phase boundaries. The phase boundaries are elevated by as much as 150 km as a result of the lower temperatures. At the rate of 8 cm/yr, the interior of the slab remains cooler relative to the surroundings to a depth of at least 600 km. In the coolest zone, the temperature difference is more than 500°C.

Fig. 2. Temperatures inside the base slab model after 3.6 My and 7.1 My were calculated using the parameters given in the text. The velocity of the slab is 8 cm/yr.
The rapid heating of the slab interior below 600 km is primarily due to the higher thermal conductivity (because of increased contribution of radiative heat transfer at higher temperatures) and to the contribution of spinel–post-spinel phase transformation. The thermodynamic parameters of this transformation are not well known and that is why we constrained the phase boundary to a constant depth. However, if the reaction is exothermic, the heat generated tends to raise the temperature to the ambient. When the internal slab temperature approaches the mantle temperature, the heat generated internally cannot be transferred to cooler portions and the source temperature increases rapidly. As a result the slab loses its integrity and is assimilated into the mantle. Even without any heating from the lower phase boundary, the slab reaches thermal equilibrium at about 750–800 km depth (Toksoz et al. 1971).

Parameters that strongly influence the temperatures inside and around the down-going lithosphere include the conductivity, and the descent rate. The effect of using the Schatz & Simmons (1972) conductivity is shown in Fig. 4. Comparing this with Fig. 3 (based on conductivity given by (2)) illustrates that in the case of lower effective conductivity (Fig. 4) temperatures are lower inside the slab.
Fig. 4. Temperature field of the slab calculated using Schatz & Simmons (1972) conductivity. All other parameters same as those in Fig. 3. Due to lower conductivity, this slab penetrates the 600 km phase change.

The temperatures inside the slab are strongly controlled by the time elapsed after the initiation of the descent. A slower moving slab will tend to heat up more because of the increased conduction of heat from the surrounding mantle. For a descent velocity corresponding to 1 cm/yr spreading rate and a total time of 100 My, Toksoz et al. (1971) showed that with parameters similar to those of Fig. 3 the slab reaches thermal equilibrium at a depth of about 400 km. Here we expanded these calculations to include stop-and-start type slab motions and differing angles at which the slab penetrates into the mantle.

The temperature field of a slab dipping 29 degrees was computed using the physical parameters of the base model (i.e. Fig. 3) and the results are shown in Fig. 5. Although 16 My, as opposed to 11 My for the base model, was required for this slab to penetrate to 650 km, the maximum depth of penetration of both slabs is controlled by the lower phase change. The maximum penetration of mantle isotherms into the slab is about 30–100 km less in the shallower dipping model.

It is conceivable that for a period of time a slab may become hung up at a trench such that no subduction occurs. Once subduction ceases it is likely that the slab will detach and sink as it may have done in the New Hebrides. Detachment could also occur if new material were being subducted more slowly than the separation rate. The thermal effects of stopping the motion after the slab penetrated to 650 km are
FIG. 5. Temperature field of the slab after 16 My. All parameters same as those in Fig. 3, except dip.
Fig. 6. The thermal fields of the slab in Fig. 5, after additional rest periods. The temperature anomaly becomes progressively broader with time.
shown in Fig. 6. In these calculations the same slab shown in Fig. 5 was allowed to remain stationary. After 16 My (total age 32 My) the slab was still evident but the temperature anomaly was spread over a larger region. After 48 My only a very broad temperature anomaly remained. It is not likely that a seismic zone could remain active for a long time after subduction ceases. The most likely event is that, once the slab is partially heated, detachment would occur and the cooler slab would sink.

The use of a cooler ('convective') geotherm and the effect of the 600 km phase boundary is shown in Fig. 7. The use of the 'convective' geotherm does not greatly alter the temperature field (Fig. 7). The lower phase change was not included in constructing the geotherm of the model. Unlike the models using the strongly superadiabatic MacDonald (1959) geotherm, the temperature field is not greatly influenced by the kinematics assumed for the slab, since motions in the nearly adiabatic regions of the mantle effect temperatures only slightly. It should also be noted that the effect of a 400 km phase change is evident, although this reaction was explicitly included.

![Fig. 7. Temperature field in a slab penetrating into a mantle having a 'convective' geotherm. The parameters used for this slab are self-consistent with the geotherm. Penetration of the slab was stopped at 10.7 My, and the lower right slab was allowed to equilibrate with the mantle for 3.6 My. The phase boundary at 650 km was not taken into account in this model, but the higher phase (350 km) boundary was included.](https://academic.oup.com/gji/article-abstract/35/1-3/285/615604)
when constructing the geotherm. The 600 km phase change was not included in these calculations.

The omission of the phase change at 600 km depth increases the depth at which the slab reaches thermal equilibrium. Even in this case, however, the equilibration occurs at about 800 km depth (Toksöz et al. 1971). At these great depths, the contribution of radiative heat transfer becomes substantial and the slab interior warms up rapidly. Continuing the calculations to greater depths results in temperatures inside or around the slab higher than the ambient mantle temperatures. It is reasonable to assume that the slab assimilates into the mantle and becomes part of the general mantle convection pattern after it reaches thermal equilibrium. As long as the subduction continues at some uniform rate, the temperatures inside the slab above the depth at which the slab’s interior reaches equilibrium change very little with time. For an 8 cm/yr subduction rate for example, the thermal regime reaches a nearly stable state after about 10 or 12 My above a depth of 700 km.

Convergence of continental lithospheres. In a separate yet related area to the subduction of the oceanic lithosphere, we investigated the thermal regime for the convergence of two continental lithospheres. Because of the generally shallower angle of penetration (for example, the dip of the thrust belt is between 10° and 20° for the Zagros convergence zone in Southern Iran (Nowroozi 1972)), slower rate of subduction, continental geotherm, and higher radioactivity in the continental crust, the thermal evolution is expected to be different from that in the oceanic case.

The large scale structure in the region of a continent–continent collision may be in part inherited from the pre-existing active continental margin and in part due to the subducted continental crusta. For continental crust to have much effect on temperature during its subduction, either the shear strain heating would have to be high or the subduction rate would have to be low for prolonged radioactive heating. The average radioactive heat production of continental crust is about 4.1 × 10^-6 erg g^-1 s^-1 (Hurley 1968a, b; Armstrong 1968) and this can cause a maximum temperature increase of 10°C/My.

A numerical model of continental convergence is shown in Fig. 8. The subduction rate is 1 cm/yr. A high value of frictional heating (1.6 × 10^-4 erg cm^-2 s^-1) is used along the upper edge. Radioactive heat generation is 4.1 × 10^-6 erg g^-1 s^-1 in the crust (hatched region) and 1.5 × 10^-8 erg g^-1 s^-1 in the mantle. Although this model shows high geothermal gradients above a subducted continent this does not seem to be sufficient to cause extensive melting or magmatization associated with orogenies.

Fig. 8. Temperature field in subducted continent after 10 My with extreme model of frictional heating along the fault plane. The highly radioactive continental crust is indicated by hatching.
Another continental convergence model with lower frictional heating ($1 \times 10^{-5}$ erg cm$^{-3}$ s$^{-1}$ above 35 km depth) is shown in Fig. 9 for two time intervals. Other parameters are the same as those of Fig. 8. After 7.7 My, the temperatures are slightly lower inside the slab, except at the boundary below 100 km depth. After about 50 My (lower figure) the temperature returns to steady state. We may conclude that orogeny is most likely to occur during the collision or after about 50 My when the subducted continental material has become heated. In the former case the heat of the orogeny would be related to volcanic activity along the active margin prior to the collision. In the latter it would be radioactive decay in the subducted material.

**Geophysical effects**

The strongly varying temperature field of the upper mantle inside and in the vicinity of the slab affects the surface heat flux, volcanism, density, seismic wave velocity and attenuation, as well as the stress field and the occurrence and mechanisms of earthquakes. Most of these effects have been discussed in previous papers (Toksoz et al. 1971; Sleep 1973). Not all surface data are equally revealing about the properties of the deep lithosphere. For example, heat flow is primarily controlled by the temperature field at relatively shallow depth and by convective heat transfer underneath and behind island arcs (Hasebe, Fuji & Uyeda 1970).

The gravity anomalies associated with downgoing slabs are broad regional anomalies (Minear and Toksoz 1970a, b; Griggs 1972). These are generally obscured by the much sharper anomalies over trenches and island arcs that result from topo-
graphy and crystal structure. Observed gravity anomalies have been fitted with a model including a slab in the Aleutians (Grow 1972).

The most pronounced effects of the temperature fields of descending slabs are on seismic velocities, $Q$ structure, and intermediate and deep focus earthquakes. Velocity and amplitude anomalies of seismic waves have been used to demonstrate the existence of the slab (Davies & McKenzie 1969; Sorrells, Crowley & Veith 1971; Mitronovas & Isacks 1971; Toksoz et al. 1971; Jacob 1970, 1972; Abe 1972a, b; Davies & Julian 1972). The observed travel-time anomalies are in close agreement with the theoretical models based on the calculated temperature profiles (Sleep 1973).

The distribution and stress fields associated with these deep focus earthquakes will be discussed in detail in the next section.

**Distribution and mechanisms of deep focus earthquakes**

Since their discovery the origin and mechanisms of intermediate and deep focus earthquakes have been of great interest. The sea-floor spreading and global tectonic hypothesis with descending slabs has provided an explanation for these deep earthquakes. It is generally accepted that the slab behaves like a stress guide. However, the mechanisms of the intermediate and deep earthquakes show remarkable differences between regions not explained by any simplified hypothesis (Isacks & Molnar 1971). In this section we treat this problem in two steps. First we determine the hypocentres relative to the slab boundaries. Then, computing the stresses arising from subduction, we investigate the mechanisms of these earthquakes.

**Distribution of deep focus earthquakes**

The shallow earthquakes near island arcs are concentrated in the convergence zones between plates, mostly along the upper boundary of the underthrusting plane of the oceanic lithosphere. For the Aleutians and Japan both the distribution of hypocentres and the source mechanisms of earthquakes support this conclusion (Barazangi & Dorman 1969; Stauder 1968; Utsu 1971; Isacks & Molnar 1971; Plafker 1972; Engdahl 1971). For intermediate and deep focus earthquakes the precise location of hypocentres becomes more difficult because of the effect of velocity anomalies within the slab on the seismic ray paths.

In order to further constrain the locations of intermediate earthquakes with respect to the slab, we investigated the central Aleutians. Nuclear explosions LONGSHOT, MILROW and CANNIKIN provided sources of precise location and time for

![Fig. 10. Observed and calculated $P$-wave travel-time anomalies for LONGSHOT explosion. The observed delays are station corrected and averaged over intervals of azimuth and distance as given by Abe (1972a). The fit is obtained by assuming the strike of the arc is N82W and adjusting the slab relative to the LONGSHOT location.](https://academic.oup.com/gji/article-abstract/35/1-3/285/615604)
Evolution of the downgoing lithosphere

Fig. 11. Temperature regime and the distribution of earthquake hypocentres in the Central Aleutians near Amchitka. The thermal model is obtained by combining the isotherms from two slabs dipping at different angles. The earthquake hypocentres are from Engdahl (1971). The focal mechanisms of two representative earthquakes show normal faulting in the oceanside of the trench and thrusting in the convergence zone beneath the arc. Note how intermediate earthquakes are in the coolest region of the slab and not in the shear zone.

calibrating the velocity models and slab location relative to the trench. Furthermore the network of seismic stations in the area provided very good locations of the earthquakes. The theoretical ray paths from surface focus events were calculated using seismic velocities deduced from a temperature model of the Aleutian slab. A most likely location of LONGSHOT relative to the slab was determined by comparing the computed travel-time delays (Fig. 10) and shadow zone with the observations (Abe 1972a; Jacob 1972; Sleep 1973). The preferred slab location based on the above data is then compared with the hypocentres of earthquakes (Engdahl 1971) in Fig. 11. It is very clear that the intermediate focus earthquakes are located along the coolest region of the slab. Similar distributions are implied for the earthquakes under Japan (Katsumata 1967; Utsu 1971) and the Tonga Kermadec region from travel times and attenuation characteristics of P and S waves (Sykes, Isacks & Oliver 1969; Mitronovas, Isacks & Seeber 1969; Toksoz et al. 1971; Sleep 1973).

Mechanisms of deep focus earthquakes

Observational data on source mechanisms of intermediate and deep focus earthquakes have accumulated in recent years from fault-plane studies. A comprehensive summary is given by Isacks & Molnar (1971). In this section we describe the calculation of stress field in a downgoing slab and compare this with the earthquake distribution and mechanisms.

Computation of stress fields. The theoretical studies of stress regime (McKenzie 1969; Griggs 1972; Smith & Toksoz 1972) are hampered by both the computational difficulties and the limited knowledge of mantle rheology. Here, following Smith &
FIG. 12. Density anomaly in g cm\(^{-3}\) within the descending slab using a typical thermal regime (in this case Fig. 2 of Smith & Toksöz 1972). Only thermal volume contraction is used for the density calculations.
Toksöz (1972) we utilized a kinematic approach for computing stresses in a subducting slab. We included the direct effects of subduction (i.e. body forces due to density anomalies resulting from cooler temperatures) in the calculations.

In our calculations a slice of mantle, which includes the slab, was modelled as a heterogeneous, temperature-dependent, viscoelastic medium. Using the correspondence principle for viscoelasticity (Christensen 1971), the operators of the viscoelastic problem are replaced by an equivalent elastic modulus which incorporates the temperature and stress dependence for that increment in time. The much simpler elastic problem was then solved numerically for the stresses (Zienkiewicz 1971). For sufficiently small time increments, this also applies to a non-linear problem. The model for the slab assumed quasistatic behaviour after application of the loads. The time since loading was taken as approximately 100 years; this approximates the frequency of transient phenomena (i.e. deep earthquakes) within the slab. Thus the problem assumes a certain temperature dependent rheology and application time for the loads; the resulting operator $E$, an effective Young’s modulus, is solved as an elastostatic problem for the strains resulting from creep since application of the load.

Using the thermal models in the preceding section and the thermal expansion coefficient (7), the density anomalies were computed relative to the mantle isotherm at each depth. These specify the vertical body forces ($F = g \rho \Delta T$) at all points inside the slab. For a temperature model similar to Fig. 4, the corresponding body forces are shown in Fig. 12. The density changes and the resulting forces due to phase changes were separately taken into account.

A finite difference scheme incorporating the body forces was then used to compute the stresses for the elastostatic analogy. The scheme uses an integral formulation of the equations together with over-relaxation to obtain a solution (Tillman 1971; Smith & Toksöz 1972); thus, it is essentially a finite-element computation. A 20-km grid spacing was used to solve the problem for an 840 by 680 km region enclosing the slab and the adjacent mantle. The imposed boundary conditions were that the top face (the Earth’s surface) was free, the vertical boundaries at the sides of the region and the bottom boundary were specified to be rigid. Other boundary conditions were also used for the side boundaries, but the effects were found to be insignificant for the calculated stresses inside the slab.

The rheology of the slab and the surrounding mantle were specified in terms of temperature and pressure dependent constants. For the mantle, the operator $E$ was obtained using effective mantle viscosities (McConnell 1968; Cathles 1971) and the elastic-viscoelastic analogy (Christensen 1971). The slab rheology was assumed to be temperature dependent. For long-term stresses, the temperature dependence of the slab properties are most likely determined by creep processes (Weertman 1970). Two separate relationships were used to express the variation in the effective Young’s modulus inside the slab:

$$E_{\text{slab}} = E_{\text{mantle}} + 5.0 \times 10^{10} \exp (-0.36 \times 10^{-2} \Delta T)$$  \hspace{1cm} (8)

and

$$E_{\text{slab}} = E_{\text{mantle}} \exp \left[-(1.2 \times 10^{-2} + 5.5 \times 10^{-5} p) \Delta T \right]$$  \hspace{1cm} (9)

where $\Delta T$ is the lateral temperature difference between the mantle and slab, and $p$ is the pressure in kilobars. The latter represents a much stronger temperature dependence. Numerical models calculated with both of these relationships did not alter the stresses significantly (Smith & Toksöz 1972).

The marked effect that the mantle’s support has upon the stress distribution inside the slab is illustrated in Fig. 13. Both slab models are identical and include only the body forces due to density anomalies resulting from the thermal contraction. The stress regimes are quite different. With little support or resistance to the descending lithosphere, the tension axis follows the trend of the slab along its full length. On the other hand, when the mantle’s support increases with depth (top diagram,
The density anomalies due to phase changes introduce additional stresses in the region of the transformation. Above the phase change, the stress is down-dip tension, and below, it is down-dip compression. For the olivine-spinel phase change, this can add about 200 bars of shear stress (Smith & Toksöz 1972). The effect of these phase changes may be important for variations in deep seismicity. If the post-spinel phase change at 650 km rises within the slab, additional stress would occur below the transformation and help account for the sudden rise in seismicity at about 600 km observed in the Tonga-Kermadec region.

The stresses from convection on the slab in the mantle can be incorporated as surface traction without including body forces. These are then superimposed on the
previous solutions. A uniform 50 bar resistance along the upper face of the slab introduces primarily compressive stresses parallel to the axis of the slab. The resulting shear stress does not exceed 100 bars throughout the length of the slab (Smith & Toksoz 1972). These models, however, do not properly simulate the shallow stresses at the island arcs, and that was not their intent. For shallow stresses the boundary conditions at the sides of the region and the fault zone between the two lithospheric plates are important and require further refinements. Yet as formulated they are adequate for the problem at hand: the intermediate and deep stresses within the slab. Thus there is no inconsistency between the relatively high stresses (1 kb or more) used for shear heating calculations and the models of the stress distribution within the slab’s interior. To resolve those problems associated with the bending and thrusting of the descending lithosphere, additional models are necessary.

**Fig. 14.** Comparison of theoretical model (top model, Fig. 13) to the seismicity of Japan given by Utsu (1971).
Comparison with earthquake mechanisms. The comparison of the calculated stress fields with the distribution and mechanisms of intermediate and deep focus earthquakes can now be made in regions where the seismicity is best understood. For this we chose Japan (Honshu) and the Kermadec region.

In Fig. 14 the theoretical stresses and the earthquake hypocentres given by Utsu (1971) for the Japanese island arc are compared. The shallow earthquakes are concentrated primarily along the shear zone between the underthrusting oceanic lithosphere and the continental lithosphere. As far as it can be determined from travel times and attenuation characteristics, intermediate and deep earthquakes are located in the interior of the slab along a band corresponding to the maximum shear stresses.

The correspondence between the theoretical and observed directions of maximum stresses is shown in Fig. 15. In the Honshu case the model dips 30° and has mantle properties as shown in Fig. 13(a). The rheology of the slab is defined by equation...
(9). On the right the directions of maximum principal stress determined by the earthquake focal mechanisms (Isacks & Molnar 1971) are shown. Both the observed and the theoretical maximum stresses are down-dip compression, which are in good agreement with each other. In the case of Kermadec, where the slab has a higher dip and rheology specified by equation (8), the observed principal stresses are down-dip compression for deep earthquakes and 'tension' for intermediate earthquakes. Again, the observations are in good agreement with the theoretical model.

The calculated maximum shear stresses inside the descending slab are approximately 500 bars. Estimates of stress drop are available for a limited number of deep focus earthquakes and are near 100 bars (Wyss & Molnar 1972). This figure is reasonable for an initial (in situ) stress of about 500 bars assuming that a relationship similar to those of shallow earthquakes exists between the initial stress and stress drop.

*Faulting* mechanisms of deep earthquakes. The fault plane solutions obtained for deep focus earthquakes indicate that double-couple type source functions fit the observed radiation pattern (Isacks & Molnar 1971; Stauder 1968). The 650 km deep Colombian earthquake studied in detail by Mendiguren (1972; epicentre 1°5’S, 72°6’W, origin time 1970 July 31, 17h08m05.4s, magnitude 7.1) using both first motion polarities and free oscillations verifies a double-couple source. Any volumetric component of this earthquake source contributes no more than 10 per cent to the radiated seismic energy.

With the descent of the isotherms inside the slab, could the material fail in a mode seismically equivalent to brittle fracture at a depth of 600 km or more? In the absence of direct laboratory data on rock failure at 200 kb confining pressure, this problem can only be explored indirectly. If we assume diffusion creep to be the mechanism of deformation (Gordon 1965; Orowan 1967; Carter & Ave’ Lallemant 1970; Raleigh & Kirby 1970; Weertman 1970; Goetze 1971), we can calculate the strain rate $\dot{\varepsilon}$:

$$\dot{\varepsilon} = 5.37 \times 10^{10} \times \exp\left(-\frac{H + V_{ac}P}{RT}\right) \sigma^{0.3},$$

(10)

where $H$ = activation enthalpy, $V_{ac}$ = activation volume, $P$ = hydrostatic pressure, $R$ = gas constant, $T$ = temperature (°K), and $\sigma$ = maximum shear stress. We can also compute the effective viscosity $\eta$:

$$\eta = \frac{\sigma}{\dot{\varepsilon}}.$$

(11)

Taking our calculated value for stress (i.e. $\sigma = 500$ bars), and $H = 100$ kcal/mole, $V_{ac} = 40$ cm$^3$/mole (Goetze 1971, and private communication) and the temperature in the centre of the slab from Fig. 3, we find within 20 km of the centre the effective viscosity is greater than $\eta = 10^{27}$ poise for virtually the full length of the slab. This value is indeed very high, and it implies a very low creep rate inside the slab. The slab behaves nearly elastically, and uniform creep cannot relieve the stresses. It is necessary to have either accelerated creep or brittle fracture to occur in response to stress concentrations. In the case of accelerated creep, however, the maximum deformation must take place over a very narrow (low aspect ratio) volume to produce the equivalent of a dislocation model for the seismic source. This mechanism is in agreement with the creep model described by Griggs (1972) based on experiments with dunite.

The maximum depth of earthquakes is also of great interest. No earthquakes have been observed below a depth of 730 km, and many seismic zones terminate at shallower depths. The absence of well-defined travel-time anomalies for deep focus earthquakes indicates that a well-defined slab does not exist for a significant distance below the earthquakes (Toksoz *et al.* 1971; Mitronovas & Isacks 1971; Sen Gupta...
A broad region of somewhat reduced temperature or a weak continuation of the slab below the depth of maximum earthquakes cannot be excluded by the data.

The limited depth of earthquakes may be explained by either the absence of stresses necessary to produce earthquakes or by the absence of material which would undergo brittle failure. These two are related since the temperature plays a strong role in both cases. Below a depth of 650 km, the interior of the descending slab rapidly reaches thermal equilibrium because of efficient radiative heat transfer. Yet equilibration through thermal conduction is not likely to be the sole cause of a maximum depth for earthquakes, as earthquakes in some seismic zones such as Tonga do not become progressively rarer with depth (Isacks et al. 1968). It is likely that the 650-km phase change rises to shallower depths in the slab and produces the stress for earthquakes below 650 km (Smith & Toksoz 1972). The seismicity of the small detached segments of the slabs under New Hebrides and under Tonga can also be explained by the stresses due to this phase boundary and the resistance of the lower mantle.

Mechanisms of shallow earthquakes in convergence zones

There are a large number of shallow earthquakes that also occur in the convergence zones, under or near the trenches. There is a general agreement regarding the mechanisms of these shallow earthquakes. On the basis of locations and source mechanisms these can be divided into two categories:

1. The shallow earthquakes associated with normal faults in the oceanic side of the trenches are probably due to tension in the outer layer of the lithosphere arising from the bending of the slab and the upward buckling (Stauder 1968b; Hanks 1971; Kanamori 1971). Generally these earthquakes are not very large and do not cut across the lithosphere (the Sanriku earthquake of 1933 may be an exception). The net stress in the lithosphere may be compressive as evidenced from the earthquakes discussed below.

2. The second group of shallow earthquakes associated with subduction occurs along the shear zone between the downgoing slab and the continental lithosphere. Many major earthquakes are associated with thrust zones and fall in this category. The dip of the fault plane is generally less than about 30°. Some examples of these are the Chile earthquake of 1960, the Alaska earthquake of 1964 (Plafker 1972), and the Rat Island earthquake of 1965 (Stauder 1968a). This type of earthquake is also prevalent in regions where continental lithospheric plates converge. The earthquakes in the southern Iran thrust belt (Nowroozi 1972) are good examples of this category.

We should clarify that in all the above areas some events not matching the two generalized categories do occur (e.g. strike slip faults). Considering the immensely complicated crustal features, tears in the plates, and localized stress concentrations, faults along other orientations and earthquakes of differing source mechanisms would and do occur. Thus the previous generalizations about earthquake characteristics represent the predominant types rather than all earthquakes in the convergence areas.

Conclusions

Our understanding of deep earthquakes and the evolution of the descending lithosphere relies upon what inferences can be extracted from observations and from predictions using theoretical and laboratory work. Here theoretical calculations of the thermal regime, the seismic velocity structure, and the stress in the descending lithosphere provide the predictions; and these are in good agreement with relevant observations. A major contribution of the theoretical calculations—in addition to verifying the hypothetical models—has been their ability to identify the significant parameters that control the evolution and properties of the descending lithosphere. Some major conclusions of this study are:
1. The temperatures inside the slab are strongly controlled by the time elapsed after initiation of the descent and by the conductivity, especially the radiative term. The gross structure of the thermal regime is insensitive to reasonable variations in the parameters.

2. The maximum depth of penetration is strongly influenced by the 650-km phase change, and by the increase in the radiative contribution to effective thermal conductivity. Temperature anomalies do not extend beyond 700 km if the above phase change is exothermic. In all cases, however, the slab is thermally assimilated into the mantle at some depth shallower than about 800 km.

3. For convergence of continental plates, the underthrusting lithosphere generally does not penetrate very deep. Orogeny associated with continental collisions most likely occurs either during the collision or later when the subducted continent becomes heated.

4. From the comparison of travel-time anomalies associated with the LONGSHOT explosion and the calculated values based on theoretical temperatures and velocities, the intermediate depth earthquakes are located within the coolest region of the slab under the Aleutians. This also seems to be the case in the Tonga region and under Japan as inferred from the velocity anomalies and attenuation of seismic waves.

5. Stress calculations show that the maximum shear stress occurs within the coolest region of the slab for a reasonable temperature-dependent rheology.

6. The regional stress patterns depend upon the temperature field, the slab rheology, dip of the slab, and mantle support. The last item, the mantle's support of the slab, may be influenced by the descent velocity, convection, phase changes, and lateral differences within the mantle such as under oceans or continents.

7. The magnitude of the calculated shear stress within the slab is about 500-1000 bars. This is a reasonable value as inferred from estimates of the stress drop for deep-focus earthquakes. The orientation of the theoretical principal stresses are in very good agreement with those determined from focal mechanism studies of intermediate and deep-focus earthquakes. The centre of the slab is sufficiently cold to allow either accelerated creep or brittle fracture to occur in response to stress concentrations.

Acknowledgments

This research was supported by the Advanced Research Projects Agency of the Department of Defense and was monitored by the Air Force Office of Scientific Research under Contract No. F44620-71-C-0049 and by NSF Grant GA-29358. One of the authors (NHS) was supported by a NSF fellowship during part of the research, and another (ATS) was supported by a Hertz Foundation Fellowship.

Department of Earth and Planetary Sciences, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

References


Evolution of the downgoing lithosphere


