Temperature beneath Tibet

Chi-yuen Wang a,*, Wang-Ping Chen b,1, Lee-Ping Wang c

a Department Earth & Planetary Science, University of California, Berkeley, CA 94720, USA
b Department of Ocean Science and Engineering, Zhejiang University, Hangzhou 310058, China
c Department of Chemistry, Stanford University, Stanford, CA 94305, USA

A R T I C L E   I N F O

Article history:
Received 26 February 2013
Received in revised form
28 May 2013
Accepted 29 May 2013

Editor: T.M. Harrison
Available online 10 July 2013

Keywords:
Tibet
temperature
underthrust Indian lithosphere
shear heating

A B S T R A C T

Temperature beneath Tibet is poorly understood, constituting a critical gap in understanding the dynamics of the most prominent case of active continental collision. Here we present results of numerical simulations to provide new insight into the thermal evolution of Tibet. We utilize collective constraints provided by several large-scale field experiments in the past 20 yr, and include recent rock physics data to fully account for important feedback processes among temperature, shear stress, shear heating, thermal conductivity and specific heat. We show that while the collision system as a whole is cooled by the northward-advancing Indian lithosphere beneath Tibet, the upper and middle crust is warmed by shear heating between the overlapping lithospheres. Such a thermal structure readily explains the longstanding enigma of a very warm upper crust over a cold upper mantle. Gradual northward warming of the system is also consistent with a bimodal distribution of seismicity in the upper crust and the upper mantle beneath southern Tibet and the absence of deep seismicity further north. We emphasize the localized nature of shear heating, which is self-sustaining yet self-limiting, therefore does not depend on precise values of various input parameters, such as the rate of convergence and the amount of radiogenic heating. This heat source may have further implications for late-stage magmatic activities and variations of crustal rheology under Tibet.

1. Introduction

Continental collision, a fundamental process in the Earth’s evolution, is on-going from the Alps to Papua New Guinea, covering a distance of about 15,000 km. A focal point along this massive chain of active collision is the Tibetan plateau, a unique tectonic feature on the Earth with a great elevation of about 5 km over a vast area of $3 \times 10^6$ km$^2$. While there is a general consensus that the plateau is built through a succession of collisions, the dynamics and the evolution of this immense plateau have been a focus of continual debate for almost a century (Argand, 1924; Tapponnier et al., 1982; Yin and Harrison, 2000; Royden et al., 2008).

Since temperature plays a critical role in determining the strength of the lithosphere (e.g., McKenzie et al., 2005; Burov and Watts, 2006), it is key to understanding the dynamics of continental collision. The processes controlling the temperature structure beneath Tibet, however, are poorly understood, with seemingly conflicting evidence in both its crust and mantle where important dynamic processes are at play.

In southern Tibet, for example, available data show high heat flow (Hu et al., 2000), accompanied by abundant hot springs (Zheng, 1997). Seismicity concentrates only in the topmost 5–10 km of the crust (Molnar and Chen, 1983; Scholz, 1998), and a horizon of strong impedance contrast detected on seismic arrivals at a depth of 32 km by associating P- and S-arrivals on seismic profiles with the $\alpha$- to $\beta$-quartz transition (Mechie et al., 2004). The evidence all indicates that the Tibetan crust is unusually hot.

On the other hand, major magmatic activity has clearly shifted to northern Tibet since ~15 Ma ago (Chung et al., 2005). Moreover, seismicity re-appears at depths of 80–100 km beneath southern Tibet (Chen and Molnar, 1983; Zhu and Helmberger, 1996; Chen and Yang, 2004; Priestly et al., 2008), indicating an uppermost mantle temperature below the brittle–ductile transition of ultramafic rocks, probably in the range of only 600–800 °C (Chen and Molnar, 1983; Wiens and Stein, 1983; McKenzie et al., 2005). Low temperature in the uppermost mantle is also consistent with the high speeds of seismic phases $Pn$ and $Sn$ beneath southern Tibet (Chen and Yang, 1990; Chung et al., 2005; McKenzie et al., 2005).
Tibet, akin to values of cold, cratonic region (Priestley and McKenzie, 2006; McNamara et al., 1995; Chen and Molnar, 1981; Ni and Barazangi, 1983).

Finally, recent high-resolution travel-time tomography across Tibet reveals conspicuous, positive anomalies of both P- and S-wave speeds ($V_p$ and $V_s$) between depths of about 100–250 km (Fig. 1b and c; Hung et al., 2011, 2010). This feature is laterally continuous and extends sub-horizontally over a distance of 500 km north of the Greater Himalaya (Fig. 1b and c). The most straightforward interpretation of these anomalies is that they represent the underthrust Indian lithosphere beneath Tibet ("Greater India," GI; Priestley and McKenzie, 2006; Hung et al., 2011; Chen et al., 2012), as previously suggested from studies of gravity anomalies (Jin et al., 1996) and S-wave birefringence (Chen and Özalaybey, 1998; Chen et al., 2010). A cold, cratonic GI beneath Tibet is a heat sink that would have cooled the Tibetan crust. Why then is the crust so hot?

We address this question through quantitative simulations of how temperature evolved beneath southern and central Tibet. Our simulations take into account several important, recent developments. First, we utilize the collective constraints provided by several large-scale, field experiments in the past 20 yr (e.g., Nelson et al., 1996; Mechie et al., 2004; Chen et al., 2012). In particular, important features deep in the mantle are constrained by the latest result of multi-scale, data-adaptive tomography of both P- and S-waves from Project Hi-CLIMB (Fig. 1; Hung et al., 2010, 2011; Chen et al., 2012). Second, we test different kinematic histories of Tibet (e.g., Molnar and Stock, 2009; Lippert et al., 2011; Clark, 2012; Meng et al., 2012). Third, we use recent high-temperature rock physics data that lead to important feedback processes among temperature, shear heating and thermal properties (Whittington et al., 2009). Finally, we test model sensitivity to modeling parameters in order to show that the key results of the present study are robust. Accounting for the important feedback processes, together with strong geometric and kinematic constraints on the model, are unique features of this study.

2. Model description and method of quantitative simulation

2.1. Geometric constraints

As a first approximation we simulate Tibet’s thermal evolution using a kinematic model that is broadly consistent with geological/geophysical constraints (Fig. 1a; Molnar and Tapponnier, 1975; Yin and Harrison, 2000) and whose geometry is constrained by the most recent, high-resolution tomographic images (Fig. 1b and c; Hung et al., 2011). These images reveal a conspicuous anomaly of high seismic wave speeds beneath Tibet, interpreted to represent the underthrust Indian lithosphere (Priestley and McKenzie, 2006; Hung et al., 2011; Chen et al., 2012). We simplify the investigation by focusing along a well-constrained, two-dimensional, vertical cross-section in the direction of convergence (green line in Fig. 1b), with the origin located at the surface near the Main Central Thrust where it dips steeply northward beneath the Greater Himalaya (Fig. 2a). We then simulate how temperature evolves in this system as the Indian plate progressively underthrusts along the MCT (or an adjacent fault system), starting from the initial collision at 50 Ma, and advances horizontally beneath Tibet at a prescribed velocity (Fig. 2b).

As a first approximation, we take the thick (~70 km) Tibetan crust as a duplex of the Asian crust, of a nominal initial thickness of about 35 km, over the underthrust Indian crust. With some variations, a duplexed crust has been proposed for some time (Argand, 1924; Zhao and Morgan, 1987; Powell, 1986; Owens and Zandt, 1997; Beaumont and Jamieson, 2004; Craig et al., 2012). Placing the boundary of slip, or the top of GI, near the middle of the thickened crust of the Tibetan plateau is also supported by additional seismic evidence. From dispersion of surface waves, using either earthquake sources or correlation of ambient noise to measure phase velocities, a mid-crustal "low velocity zone" (LVZ, i.e., an inversion of seismic wave speeds with increasing depth) centered near a depth of...
about 35 km has been reported to be widespread beneath Tibet (Li et al., 2009; Yao et al., 2010). Moreover, in the mid- to lower-crust, discrepancies between crustal models derived from Rayleigh and Love waves also indicate strong radial anisotropy or subhorizontal petro-fabric where the axis of symmetry is sub-vertical (Shapiro et al., 2004; Yang, 2011). These results are consistent with the notion that the LVZ marks the shear interface between underthrusting Indian plate and the Asian crust.

Unlike the Himalayas where large amounts of uplift and denudation have taken place following the collision (e.g., Henry et al., 1997; Herman et al., 2010), erosion and peneplanation in southern Tibet proceeded at low elevations prior to the collision; henceforth there has been little erosion as evidenced by the long-term preservation of peneplain in southern Tibet (Hetzel et al., 2011). We also assume as a first approximation that any temperature change due to block rotation (Allmendinger et al., 2007) or distributed deformation of the Tibetan lithosphere (Holt, 2000) is second order in affecting Tibet’s thermal evolution and may be neglected here.

Recent data from petrophysical studies also provide a dynamic basis for the presence of such a shear zone beneath Tibet. As will be explained in Section 2.4, these data predict enhanced thermal insulation and rheological weakening at high temperatures and high strain rates—effects that promote and localize the shear zone. At present, the nature and the exact position of this shear zone can only be taken as a testable hypothesis and we make specific suggestions of testing in Section 4.2.

2.2. Numerical implementation

We approximate the thermal evolution of overlapping lithospheres by solving the following equation:

$$\frac{\partial [\rho(T)C_p(T)\mathbf{T}]}{\partial t} = \nabla \cdot [K(T)\nabla T] - \mathbf{u} \cdot \nabla [\rho(T)C_p(T)\mathbf{T}] + Q_r + Q_s$$

where $t$ is the time, $T$ the temperature, $K$ the thermal conductivity, $C_p$ the heat capacity at constant pressure, $\rho$ the density, $Q_r$ the radiogenic heating, $Q_s$ the shear heating, and $\mathbf{u}$ the velocity of...
GI relative to the overlying Tibetan lithosphere. A well-verified finite element code (Comsol Multiphysics) is used to implement the simulation with a fine mesh-size (Fig. 2b); even finer meshes are implemented near regions where slips take place. We experiment with different mesh-sizes to ensure that results of the simulation are not affected by the choice of mesh-sizes.

For boundary conditions, we set the upper boundary, or the surface of Tibet, at an average altitude of 5 km and maintain the temperature there at 0 °C. The geotherm of the Indian shield (Fig. 3) is imposed on the left (southern) boundary beneath the origin. The right boundary has no net heat flux across it, and the lower boundary has an adiabatic gradient appropriate for a convective mantle. The boundary on the right is set far from the area of interest in order to avoid any numerical artifacts that may arise from the imposed boundary conditions.

### 2.3. Heat sources

A heat source of particular interest here is shear heating \( Q_s \) along the interface between the overlapping lithospheres. The relative motion between the underthrust GI and the Tibetan lithosphere is resisted by friction or viscous shear along their boundaries. Work done by shear resistance converts into heat. In shallow crust the controlling mechanism is friction, which is proportional to the frictional coefficient and overburden as follows:

\[
\tau = \mu g z (1 - \alpha)
\]

where \( \mu \) is the coefficient of friction, \( \rho \) is the rock density that is generally a function of depth \( z \), \( g \) is the gravitational acceleration, and \( \alpha \) is the ratio between pore pressure in the rock and the lithostatic pressure. Values of \( \mu \) lie between 0.1 for clay-rich fault gouges (Chu et al., 1981; Carpenter et al., 2011; Locker et al., 2011) and 0.6–0.8 for fresh rocks (Byerlee, 1978). Studies of active fault zones mostly suggest very low frictional coefficients (Brune et al., 1969; Lachenbruch and Sass, 1980; Zoback et al., 1987; Mount and Suppe, 1987), even though the cause for the low values is still a matter of debate (Wang, 2011). High pore pressure in fault zones may act to lower the frictional resistance (Rice, 1992; Byerlee, 1990), but the pore pressure along the MCT is unknown.

Although low effective friction may occur beneath the Himalayas (Davis et al., 1983; Hilley and Streecker, 2004; Herman et al., 2010), frictional sliding gives way to viscous flow at depths greater than that for the brittle–ductile transition. Crustal earthquakes terminate at a depth of ~25 km beneath the Himalayas (Sheehan et al., 2008) and ~10 km beneath southern Tibet (Chen and Molnar, 1983; Molnar and Lyon-Caen, 1989), suggesting that frictional sliding is replaced by viscous flow at greater depths. The mid-crustal shear zone beneath Tibet, as inferred from surface wave dispersion and seismic anisotropy (Section 2.1), is at a depth of ~35 km, where viscous shear dominates. The latter is a function of temperature and strain rate [Eq. (3)], but independent of the frictional coefficient in the upper crust. As demonstrated in Section 4.3, while different frictional coefficients clearly affect the simulated temperature beneath the Himalayas, they have little effect on the simulated temperature beneath Tibet—the focus of this study.

For viscous shear we use the well-known flow law for dislocation creep:

\[
\tau_v = \left( \frac{\dot{\epsilon}}{A} \right)^{1/n} \exp \left( \frac{E + p V}{n R T} \right)
\]

where \( \tau_v \) is the viscous shear stress, \( \dot{\epsilon} \) the shear strain rate, \( E \) the activation energy, \( p \) the pressure, \( V \) the activation volume, \( R \) the gas constant, and \( T \) the absolute temperature (Hirth and Kohlstedt, 2003; Karato, 2010). \( E, V, n \) and \( A \) are laboratory determined material constants. Because the range of depths involved in this study is only a few hundred kilometers, the term \( pV \) is small (within the experimental uncertainties of \( E \)) and thus negligible. Shear heating per unit volume is then given by

\[
Q_s = \tau_v \dot{\epsilon}
\]

Since

\[
\dot{\epsilon} \sim u/d
\]

where \( d \) is the thickness of the shear zone and \( u \) the overall slip rate across the shear interface, shear heating per unit horizontal area of the shear zone is

\[
Q_s d = \left( \frac{\dot{\epsilon}}{A} \right)^{1/n} \exp \left( \frac{E}{n R T} \right) d
\]

Therefore shear heating per unit horizontal area increases with \( \dot{\epsilon} \) raised to the power of \( (1/n) \) and decreases exponentially with \( T \).

At extremely fine grain sizes, a Newtonian rheology for diffusion creep may become important. However, unlike dislocation creep, diffusion creep cannot create lattice-preferred orientations that are necessary to form seismically detectable petrofabric (Karato, 2010). Furthermore, Homburg et al. (2010) showed field evidence that, although micro-mylonite zones occur within shear zones of mafic granulite, the micro-mylonites are not interconnected and overall the shear zones did not show wholesale reduction in grain size.

Granulite is commonly taken to be the major rock type in the middle to lower crust. It is also exposed in many parts of India (Prakash, 1999; Prakash et al., 2007; Dwivedi and Theunuo, 2011). Microprobe analysis of these rocks and applications of conventional geothermobarometry based on chemical exchanges yield an estimated P-T condition corresponding to that of the lower crust (Dwivedi and Theunuo, 2011). Here we use the experimentally determined rheological parameters of dry granulite for the upper shear zone along the top of the underthrusting GI (Table 1a). For the lower boundary of the GI, appropriate parameters are those of peridotite. High temperatures at such depths, however, render shear heating insignificant.

Radiogenic heat production, both in the Tibetan crust and in the Indian crust, is poorly constrained. Here we assume average radioactivity of 0.8 μW/m³ for the entire crust (Table 1b). Higher crustal radioactivity was assumed in earlier works (Henry et al., 1997; Beaumont and Jamieson, 2004; Bollinger et al., 2006; Craig et al., 2012). Sensitivity tests (Section 4.3) show that the simulated crustal temperature in this study is largely controlled by the combined effects of cooling from the underthrust GI and shear heating along the plate boundary, and relatively insensitive to the assumed radiogenic heat production.

### Table 1

Values of model parameters for the representative model.

<table>
<thead>
<tr>
<th>(a) Rheological parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lithology</td>
</tr>
<tr>
<td>Granulite (dry)</td>
</tr>
<tr>
<td>Peridotite (dry)</td>
</tr>
<tr>
<td>Peridotite (wet)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>(b) Radiogenic heat sources and depth-averaged densities</th>
</tr>
</thead>
<tbody>
<tr>
<td>Layer depth (km)</td>
</tr>
<tr>
<td>------------------</td>
</tr>
<tr>
<td>0–35</td>
</tr>
<tr>
<td>35–70</td>
</tr>
<tr>
<td>70–200</td>
</tr>
<tr>
<td>200–1000</td>
</tr>
</tbody>
</table>
2.4. Material properties at high temperatures and self-limiting, localized nature of shear heating

Eq. (1) explicitly specifies material properties as functions of temperature. This is an important point in that recent data from laboratory experiments show such properties can depend strongly on temperature. Most important are the changes of $K$ and $C_P$ with increasing $T$, halving $K$ and nearly doubling $C_P$, thus nearly nullifying their differential change (Whittington et al., 2009). Whittington et al. (2009) stressed the positive feedback between $\kappa$ and $T$ during strain heating at constant shear stress. In effect, a material at high temperatures impedes thermal diffusion.

In addition, shear stress in our simulations, instead of being a constant, is strongly dependent on both temperature and strain rate (Eq. (3)). Furthermore, the rate of shear heating is the product of shear stress and strain rate (Eq. (4)). The complex interdependence among temperature, thermal diffusivity, viscous shear stress and strain rate results in several coupled, feedback processes that are at the heart of this study.

Unlike the dispersed nature of radiogenic heating, shear heating tends to localize near the shear zone. Localization of this heat source is a result of positive feedback among three factors. First, rapid decrease in thermal diffusivity with increasing temperature effectively insulates the shear zone from rapidly losing heat to its surroundings (Whittington et al., 2009). Second, viscous deformation along the shear zone is self-sustaining: Since the ductile strength of rocks decreases exponentially with temperature (Eq. (3); Karato, 2010; Hirth and Kohlstedt, 2003), a localized zone of high temperature stays as a pronounced, weak zone of concentrated deformation. Third, since the effective viscosity, $\dot{\varepsilon}/\dot{\varepsilon}$, is proportional to $(\dot{\varepsilon})^{\beta/\beta}$ (Eq. (3)) and since experimental values of $\beta$ range from 3 to 4 for dislocation creep, the effective viscosity decreases with increasing strain rate. Thus, deformation tends to stay localized in a shear zone where both the initial strain rate and the temperature are the highest.

Even so, shear heating is also self-limiting, preventing the possibility of thermal runaway. This is due to the strong negative feedback between temperature and viscous shear stress, as discussed in Burg and Gerya (2005). High viscous shear stress leads to high rate of shear heating, which, in turn, raises the temperature. Meanwhile, high temperature rapidly relaxes any buildup of shear stress, thus reducing the rate of heating. We account for these inter-related feedback processes in a finite element scheme (Section 2.2), with temperature-dependent values of $K$, $C_P$, $\rho$ and $Q_s$ (Supplementary material) fully coupled at each time-step.

2.5. Initial distribution of temperature

The initial temperature of the Indian lithosphere is approximated by the geotherm of the Indian shield (Fig. 3; Roy and Mareshal, 2011). Although the geotherm was constructed strictly for the Darwar craton of southern India, existing seismic analyses (Shapiro and Ritzwoller, 2002; Bodin et al., in press) and heat-flow data (Roy and Rao, 2003; Roy et al., 2008) suggest that the cratonic structure may be representative of the Indian peninsula as a whole.

The initial geotherm beneath Tibet is much less well constrained; but widespread calc-alkaline volcanism in southern Tibet from Paleocene to early Eocene points to an initially hot mantle (Ding et al., 2003; Chung et al., 2005). Here we reconstruct this geotherm by approximating its crustal temperature with that beneath central Tibet north of the IMF (Mechie et al., 2004). At greater depths, we assume an adiabatic geotherm appropriate for a convective mantle, merging with the Indian geotherm at a depth of 200 km (Fig. 3).

The adiabatic geothermal gradient changes with depth ($z$) and may be estimated from the following relationship:

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

where $\alpha$ is the thermal expansivity, $g$ the gravitational acceleration, $T$ the temperature, and $C_P$ the heat capacity at constant pressure. Using reasonable values of these parameters, Fowler (1992) estimated adiabatic gradients of 0.5 °C/km and 0.3 °C/km in the upper and the lower mantle, respectively.

2.6. Kinematic considerations

The effect of the underthrust Indian lithosphere on the distribution of temperature beneath the orogen is simulated by advecting an initially cold GI across the collision front ($x = 0$) at prescribed relative velocity $u$ (Fig. 2b and Eq. (1)). The present direction of $u$ is N20°E (Zhang et al., 2004; Meade, 2007; Thatcher, 2007), with only moderate variations estimated to be from +15° to -10° in the past 50 Ma (Lee and Lawver, 1995).

GPS data also indicate that, to a first approximation, the surface of Tibet is currently under uniform shortening along N20°E (Zhang et al., 2004). Based mainly on the reconstruction of plate motions by Molnar and Stock (2009), Clark (2012) showed that such an approximate constant rate of strain ($\dot{\varepsilon}$), about 0.02 Myr$^{-1}$, may also have held since the onset of the Cenozoic collision. So the velocity, $u$, between the southern edge of the orogeny, the Himalayas, and the northern limit of GI, a point 500 km farther to the north-northeast, is $\sim$10 mm/yr (500 km times 0.02 Myr$^{-1}$). This relative velocity also equals the simplest average by taking the ratio between the horizontal extent of the cross-section of GI (≈500 km, Fig. 1B), as determined by seismic methods and modeling of gravity anomalies (Hung et al., 2010, 2011; Priestley and McKenzie, 2006; Jin et al., 1996; Chen et al. 2010), and an commonly accepted time of onset of collision, 50 Ma, recognizing that the final collision in the western Himalaya may have occurred −10 Ma later (Boivin et al., 2013). Finally, this relative velocity is also consistent with the average rate of convergence between the southern margin of Eurasia and the northern margin of India since the onset of collision, as reconstructed from new paleomagnetic data (Meng et al., 2012).

Uncertainties in $u$ are tied to the timing of the onset of the collision between India and Tibet. While most authors suggest the onset of collision occurred at $\sim$50 Ma, published ages range from 65 to 35 Ma (e.g., Jaeger et al., 1989; Aitchison et al., 2007; Najman et al., 2010). The lower estimate is motivated, in part, by noting that around 50 Ma ago, the paleo-latitude of what is currently northernmost India lay farther south of that of southern Tibet (Aitchison et al., 2007; Lippert et al., 2011; van Hinsbergen et al., 2011a, 2011b; Meng et al., 2012). Another way out is to extend the northern edge of the pre-collision Indian plate further north, filling the gap in paleo-latitudes. Using the extent of GI detected by geophysical means (Fig. 1) and the lower limit of the timing of collision leads to an estimate of $u = 17 \pm 3$ mm/yr. So 10–20 mm/yr seems to be a reasonable range for the averaged value of $u$.

While the precise timing of the collision and the amount of deformation within Tibet are important questions of their own right, thermal evolution of the orogen does not depend on a good estimate of $u$. This seemingly counter-intuitive notion can be qualitatively understood by the self-limiting and localized nature of shear heating. For a given, fixed thickness of a shear zone, $d$, the rate of shear heating is directly proportional to the strain rate $\dot{\varepsilon}$ (Eqs. (4) and (5)). A high value of $u$ leads to a high rate of shear heating, but the resulting rise in temperature is limited by a corresponding decrease in viscous shear stress through Eq. (3). Conversely, a low value of $u$ results in low rate of shear heating, yet
this effect is counteracted by a rise in viscous shear stress. This reasoning is borne out by numerical experiments: Within the range of reasonable estimates, precise values of $u$ and $d$ are unimportant in determining simulated temperature fields (see Section 4.3).

To choose a proper range for $d$, we turn to field observations of exhumed shear zones. These zones, originally formed at deep crustal levels, show typical thickness on the order of 1 km (Burgmann and Dresen, 2008). When the overall thickness of a shear zone is on the order of 10 km, shear strain is often accommodated by anastomosing networks of kinematically linked substructures of varying thicknesses, lengths, and displacements (Burgmann and Dresen, 2008). Here we use $d = 1000$ m and $u = 10$ mm/yr in the simulation of a representative model (Fig. 4). As discussed in Section 4.3 and shown in Fig. 8, uncertainties in these parameters do not significantly affect the main features in the simulated result.

In the present study, only the kinematics of the underthrust GI is prescribed. However, the mantle below and in front of GI must move in response. The analysis of such response, however, requires three-dimensional, dynamic simulation and is out of the scope of this study. It is likely that the mantle in front of the GI may simply be pushed out of the way by lateral flow (e.g., Owens and Zandt, 1997; Chen and Özalaybey, 1998; Jimenez-Munt et al., 2008) and the mantle below the GI also may undergo viscous shear. Such induced flow is subhorizontal and not expected to significantly impact the simulated results.

### 3. Results of numerical simulations

We conducted a suite of simulations that include a range of values for key parameters (see Section 4.3). Figs. 4a, b and 5 show the result of a representative simulation, with thermal, rheological and kinematic parameters summarized in Table 1a and b and Supplementary material. Two salient features stand out from the simulated temperature (Fig. 4a): first, the underthrust GI significantly cools a large region, extending to depths over 250 km and...
Comparison between Figs. 4b and 1c shows that, over a distance geotherm, shear heating warms the upper crust to sink, GI. In southern Tibet, superimposed over a generally cooled depth and distance. Overall, the system is cooled by the vast heat the positive anomaly in anomaly in the simulated temperature anomaly matches those of of $\sim$ average continental geotherm (Jaupart and Mareschal, 2011).

In order to compare the simulated thermal structure (Fig. 4a) with anomalies in $V_p$ and $V_S$ (Fig. 1c), we calculate the corresponding temperature anomaly (Fig. 4b) by subtracting from Fig. 4a an average continental geotherm (Jaupart and Mareschal, 2011). Comparison between Figs. 4b and 1c shows that, over a distance of $\sim$500 km, the overall shape and extent of the strong negative anomaly in the simulated temperature anomaly matches those of the positive anomaly in $V_S$. Beneath the Greater Himalaya, both anomalies extend from the surface to depth (upper left corner of Figs. 1c and 4b). Furthermore, the change of sign in anomalies across the IMF, and the gradual increase in the amplitude of anomalies to the south of the IMF, are at least qualitatively reproduced. Lastly, based on recent laboratory data (e.g., Li et al., 2004), the range of simulated temperature difference in the mantle, $-800$ °C, predicts a range of $-6\%$ in $\alpha V_S/V_p$, close to that deduced from seismic tomography (Fig. 1c).

Fig. 5a and b shows how the simulated geotherms vary over depth and distance. Overall, the system is cooled by the vast heat sink, GI. In southern Tibet, superimposed over a generally cooled geotherm, shear heating warms the upper crust to $>600$ °C at a depth of $\sim$35 km (Fig. 5a). At greater depths, a marked temperature inversion occurs in the lower crust, producing a minimum temperature $\sim$500 °C near the Moho. This result provides a simple explanation for the dichotomy of a warm crust overlying a cold upper mantle beneath southern Tibet. That is, the orogeny as a whole has been cooled markedly by the northward advancing GI, as expected; but at the same time, shear heating, a localized heat sources near the top of the advancing GI, warms the upper crust.

In central Tibet, the GI has been progressively warmed by its surroundings, and crustal temperature increases to $\sim$800 °C at a depth of 35 km (Fig. 5b), same as the estimates from associating $P$- and $S$-arrivals with the $\alpha$- to $\beta$-quartz transition (Mechie et al., 2004). In this case, there is no longer a pronounced temperature inversion in the crust (cf. Fig. 6b and a). These results suggest that substantial partial melting could occur beneath the northern Lhasa block, if sufficient amount of water is present (Thompson and Connolly, 1995; Holtz et al., 2001; Fig. 5b).

As a final note, heat flow data in Tibet are scarce (Hu et al., 2000) and none is available near the cross section presented here; thus we cannot directly compare the model with data. However, the simulated high crustal temperature, as discussed above, is broadly consistent with Tibet’s high heat flows.

4. Discussion

In this section we first test our result by the so-called earthquake thermometry and also suggest another test for the mid-crustal shear zone. We then address the sensitivity of our simulated thermal field to the assumed conditions and parameters.

4.1. Testing predicted temperatures with earthquake thermometry

A useful test of our results comes from earthquake thermometry. Generally seismicity occurs only below distinct limiting temperatures ($T_c$) of $350 \pm 50$ °C in the crust and $700 \pm 100$ °C in the mantle. This result came first from global studies of depths of intra-continental earthquakes (Chen and Molnar, 1983) and seismicity in the oceanic lithosphere away from plate boundaries (Chen and Molnar, 1983; Wiens and Stein, 1983; McKenzie et al., 2005). Subsequently, linking results of experimental rock physics (e.g., Scholz, 1998; Boettcher et al., 2007) with the analysis of frictional instability based on a rate and state-variable formulation (the “Dieterich–Ruina” relationship; e.g., Tse and Rice, 1986) placed
these values of $T_C$ on firm ground. So a bimodal distribution of focal depths beneath southern Tibet (Chen and Molnar, 1981, 1983; Chen and Yang, 2004), with one peak in the upper crust and the other in the uppermost mantle, are broadly consistent with the simulated temperature field (Fig. 6a): temperature rises quickly above the first critical temperature ($350 \degree C$) at a depth near 10 km in the crust, but does not reach over the other critical temperature ($700 \degree C$) until a depth of about 100 km in the upper mantle (Fig. 6a).

Detailed discussion of earthquake thermometry is beyond the scope of this paper and we refer readers to recent reviews on this subject by Chen et al. (2013, in press) and references therein. Suffice it to point out that the stick-slip phenomenon during frictional sliding in laboratory experiments is often taken to represent the mechanism of earthquakes (Brace and Byerlee, 1970; Tse and Rice, 1986; Scholz, 1990, 1998). Long intervals of "stick" correspond to seismic quiescence during the inter-seismic stage, while sudden slips that release stored elastic stress represent the seismic events.

Scholz (1998) noted that the transition from stick-slip to stable sliding in granite, an abundant rock type in the upper crust, is limited by a temperature of about 350 °C, about the same as $T_C$ for crustal seismicity. This temperature corresponds to the onset of plasticity in quartz, the most ductile major mineral in granite at crustal seismicity. This temperature corresponds to the onset of substantial crystal plasticity in the major crustal and mantle minerals, respectively.

Judging from empirical flow laws based on laboratory experiments, special assemblages of lower crustal rocks may have a $T_C$ that approaches the value for olivine. In particular, flow laws of synthetic anorthite aggregates imply an approximate $T_C$ of $600 \degree C$ near the temperature of brittle-ductile transition in olivine (Boettcher et al., 2007), the most abundant mineral in the upper mantle. In other words, the distinct values of $T_C$ mainly reflect the onset of substantial crystal plasticity in the major crustal and mantle minerals, respectively.

Meanwhile, intra-continental earthquakes do occur in the upper mantle. Some of them produced underside reflections off the Moho above the hypocenters (Yang and Chen, 2010; Chen et al., in press), and some others are as deep as 100 km beneath regions where the elevation is only about 2 km (Chen and Yang, 2004). In any event, for the purpose at hand, the precise relationship between the second peak of seismicity at depth beneath southern Tibet and the nature of the Moho there is beside the point. The key is that these unusually deep earthquakes indicate that at a depth of about 70–75 km, temperature near the crust–mantle transition is extremely complex (e.g., Nowack et al., 2010 and references therein); and it is impossible to determine if the lower crust is mainly made of anorthite or any other major minerals.

As noted earlier, crustal earthquakes terminate at a depth of about 70–75 km, temperature near the crust–mantle transition is low, only about 700 ± 100 °C.

At the same time, beneath central Tibet, temperatures at depths below ~30 km already exceed the limiting temperature for seismicity for both crustal and mantle rocks (Fig. 6b), consistent with the absence of seismicity below the upper crust (Molnar and Chen, 1983; Zhu and Helmberger, 1996; Chen and Yang, 2004).

4.2. Testable predictions of a localized mid-crustal shear zone

Two major results of our model are: a hot upper crust overlaying a cool upper mantle beneath southern Tibet, separated by a pronounced temperature inversion beneath the mid-crust, and localization of the source of crustal heating in a narrow shear zone. To sustain such a change in the geothermal gradient over time, the cause of high temperatures in the mid-crust must be part of the processes of on-going continental convergence. This is an important feature not found in previous studies (e.g., Henry et al., 1997; Beaumont and Jamieson, 2004; Bollinger et al., 2006; Craig et al., 2012) and leads to testable predictions.

As discussed in Section 2.4, both high temperature and rapid strain rate would concentrate deformation, hence localizing viscous shear heating along a ductile shear zone. In addition, notable decrease of thermal diffusivity with increasing temperature further insulates the hot shear zone. Indeed, additional numerical simulations show that main features of our models, i.e., the localized peak temperature in the mid-crust and the strong temperature inversion in the lower crust, are insensitive to the precise amount of radiogenic heating production in the crust (Section 4.3).

Existence of a mid-crustal shear zone and its thickness are testable predictions because high temperature associated with such a shear zone will cause a decrease in seismic wave speeds, causing a crustal "low-velocity" zone (LVZ). At the moment, evidence supporting a mid-crustal LVZ mainly come from the study of surface wave dispersion (e.g., Li et al., 2009; Yao et al., 2010; Shapiro et al., 2004; Yang, 2011). This type of approach has the advantage of wide spatial coverage but its resolution in depth is inadequate to determine the thickness of the LVZ—a critical test for our model. To this end, the so-called crustal receiver functions (RF) that use scattered wavefield to detect subsurface discontinuities below seismic stations are promising. Our models specifically call for a thin zone of high temperatures, whose thickness is on the order of 10 km. Such a sharp zone should be detectable by RF over both high- and low-frequencies. In contrast, a broad zone of temperature anomaly, caused either by a very broad shear zone or by radiogenic heat sources proposed in other studies, will be largely invisible to RF at high frequencies. This aspect of investigation is ongoing (Chunquan Yu and Robert van der Hilst, 2013, Mass. Inst. of Tech., personal communication).

4.3. Model sensitivity to modeling parameters

Through a series of numerical experiments, we explore model sensitivity to the assumed parameters including frictional coefficient, underthrust velocity, shear-zone thickness, radiogenic heat source and initial temperatures for the underthrust GI and the Tibetan mantle. Overall, because of the self-limiting and self-sustaining nature of viscous shear heating in the crust, the simulated crustal temperatures are largely unaffected by different values of parameters, as long as the values fall within reasonable ranges constrained by available geological and geophysical data. Specific, detailed discussions appear below.

4.3.1. Model sensitivity to frictional coefficient

As noted earlier, crustal earthquakes terminate at a depth of ~25 km beneath the Himalayas (Sheehan et al., 2008) and at ~10 km beneath southern Tibet (Chen and Molnar, 1983; Molnar and Lyon-Caen, 1989). At depths greater than the maximum depth of seismicity, viscous flow replaces frictional sliding as the dominant deformation mechanism. The mid-crustal shear zone beneath Tibet, as inferred from surface wave dispersion and seismic anisotropy (see Section 4.2), is at a depth of ~35 km, where viscous shear dominates. The latter is a function of temperature and strain rate [Eq. (3)] but independent of the frictional coefficient in the upper crust. Here we simulate temperatures using a range of frictional coefficients from 0.1 to 0.6 at depths down to 25 km and viscous shear at greater depths. The simulated result (Fig. 7) shows that, while different frictional coefficients do affect...
the simulated temperature beneath the Himalayas, they have little effect on that beneath Tibet.

4.3.2. Model sensitivity to underthrust velocity and shear-zone thickness

Testing model sensitivity to the underthrust velocity is important because shear heating is proportional to strain rate, \( k \sim u/d \), where \( u \) is the velocity of underthrust and \( d \) is the overall thickness of the shear zone. Because it is the ratio between \( u \) and \( d \), rather than each parameter alone, that affects strain heating, there is a tradeoff between these two parameters; i.e., the effect of a higher \( u \) at a fixed \( d \) may tradeoff with a smaller \( d \) at a fixed \( u \), etc. In Fig. 8 we plot simulated temperatures in southern Tibet for different values of \( u \) and \( d \). The initial and boundary conditions are kept the same as those in generating Fig. 5a, so these figures can be compared directly.

In the simulations, we increase \( u \) from 1 to 5 cm/yr, at the same \( d = 1 \) km. This range of change in \( u \) produces a difference of \( \sim 50 ^\circ C \) (8%) in the peak temperature in the mid-crust (Fig. 8). Thus the predicted depth of the upper crust seismicity is unchanged. Similarly, we changed the values of \( d \) – perhaps the least constrained parameter in the simulations – by three orders of magnitude. At the same velocity, smaller \( d \) corresponds to greater strain rate and therefore greater shear heating. But a three-order-of-magnitude decrease in \( d \) leads to an increase of \( \sim 80 ^\circ C \) (13%) in the predicted peak temperature of the crust (Fig. 8). These results further illustrate the self-limiting nature of shear heating: higher strain rate leads to higher shear heating and higher temperature; but high temperature exponentially reduces shear stress and thus diminishes shear heating.

Higher velocity of underthrust of the GI does lead to greater cooling with a drop in the minimal temperature in the lowermost crust and the uppermost mantle (Fig. 8), but not at depths greater than about 100 km. At the highest assumed velocity (\( u = 5 \) cm/yr), the predicted temperature near the base of the lower crust falls below \( T_C \) so that seismicity could also occur in the lowermost crust, as well as in the uppermost mantle (Fig. 8). In other words, different values of \( u \) and \( d \) mostly affect the exact value and distribution of temperatures in the lowermost crust where the bimodal distribution of seismicity does not offer tight constraints.

4.3.3. Model sensitivity to radiogenic heat sources

We test model sensitivity to radiogenic heat source by varying the latter from 0.4 to 1.6 \( \mu W/m^2 \) in the upper 35 km of the crust. The results (Fig. 9a) show that the peak temperature in the upper and mid-crust is mostly unaffected by the different radiogenic heat sources, again illustrating the self-limiting nature of shear heating: higher background temperature in the crust due to stronger radiogenic heating reduces the shear stress along the plate interface, thus reducing shear heating to compensate, in part, the effect of higher radiogenic heating, and vice versa.

4.3.4. Model sensitivity to initial geotherms

To test model sensitivity to the initial temperature, we use the geotherm of the Canadian shield (Fig. 3; McKenzie et al., 2005) for the initial temperature of the GI, instead of that of the Indian shield geotherm, and an upper mantle temperature \( -170 ^\circ C \) higher than that in generating the representative model (Fig. 5).
Comparing Figs. 9b and 5a shows that, while different initial temperatures affect the values of simulated geotherms, they do not affect the key features in the simulated thermal structure, which are produced by shear heating, a heat source insensitive to the initial conditions.

5. Concluding remarks

Our simulations predict a temporally evolving and spatially varying temperature field with a hot upper crust overlying a cold upper mantle in southern Tibet. Specifically, we show that while the collision system as a whole is cooled by the underthrust GI beneath Tibet, the crust is warmed by shear heating between the overlapping lithospheres. In regions under maximum cooling, shear heating along the plate interface creates a marked temperature inversion in the thickened crust. These features of our simulation resolve the enigmatic origin of bimodal distribution of earthquake depths beneath southern Tibet.

Furthermore, due to the self-sustaining nature of shear heating and the insulating effect of low thermal diffusivity at high temperatures, we predict a localized zone of heating in the mid-crust (Fig. 5a and 6a). This is a testable prediction, as a localized zone of high temperature (and thus low seismic velocity) should be detectable by receiver functions at both high and low frequencies. This prediction is robust because sensitivity tests based on numerical experiments show that the simulated thermal structure is insensitive to the uncertainties in the input parameters.

Cooling of the collision system by the underthrust GI corresponds to an increase in density up to $\sim 10$ kg/m$^3$ that could induce convective instability if not compensated (Wang and Chen, 2012). The longevity of the flat underthrusting GI beneath Tibet thus bears evidence of a compositional buoyancy (Hung et al., 2011). Based on experimental data of Schutt and Lesher (2006), we estimate a lower bound of melt depletion of 5–10% to counteract the negative buoyancy.

Finally, a comparison between simulated geotherms and the wet solidus for granite (Fig. 5a) shows that incipient partial melting seems plausible in the crust beneath the southern Lhasa block. Substantial partial melting may occur deep in the crust beneath the northern Lhasa block, provided that the wet solidus of granite is appropriate for the lower crust there. These inferences may have significant implications for the timing and locations of late magmatic activities and variations in the rheology of the Tibetan crust.

Acknowledgments

This study originated from a lecture given by the second author and attended by the first author during the 2011 CIDER Summer Program. We thank Michael Manga for reading the manuscript and providing helpful comments, Terry Plank for discussion, Shu-Huei Hung for tomograms, and two anonymous reviewers for providing constructive comments that helped to improve the paper. Supported by the US-NSF EAR Grants 9909362 (Hi-CLIMB), 0551995 and 0968979 (CIDER), and by the Chinese Ministry of Education and Zhejiang University (“985”).

Appendix A. Supplementary material

Supplementary data associated with this article can be found in the online version at http://dx.doi.org/10.1016/j.epsl.2013.05.052.

References


Fig. 9. (a) Simulated geotherms at 100 km north of the MCT, with radiogenic heat sources from 0.4 to 1.6 $\mu$W/m$^3$ in the upper crust. Results show that the peak temperature and other key features of the simulated geotherms are largely unaffected by uncertainties in the assumed radiogenic heat sources. (b) Simulated geotherm at 100 km north of the MCT, using the Canadian shield geotherm as the initial temperature of Greater India, and an initial mantle temperature for Tibet warmer by $\sim 170 \, ^\circ$C than that used in generating the representative model in Fig. 5a.