1. INTRODUCTION

The Himalaya is the most impressive example on earth of an active collisional orogen. It combines rapid crustal shortening and thickening, intense denudation driven by the monsoon climate, and frequent very large earthquakes along an incomparably long and high mountain arc. It has therefore been the focus of a variety of investigations that have addressed various aspects of mountain building on various timescales. Geological and geophysical studies give some idea of the structure of the range and physical properties at depth. The long-term geological history of the range, over say several millions to a few tens of millions of years, has been documented by structural, thermobarometric, and thermochronological studies. Morpho-tectonic investigations have revealed its evolution over several thousands or tens of thousands of years; and geodetic measurements and seismological monitoring have revealed the pattern of strain and stress built-up over several years. This chapter is an attempt to show that the results of these investigations can be assembled into a simple and coherent picture of the structure and evolution of the range. The author also intends to illustrate the interplay between these various processes operating at different timescales. One important example of processes that interact via feedback mechanisms is particularly clear in the Himalaya: the thermal structure of the range, which is a result of the long-term crustal deformation and pattern of exhumation, governs, through its influence on rheology, the pattern of deformation as well as the seismic behavior of the range-bounding thrust fault.

In this chapter the key role played by surface processes is emphasized. These processes have carved morphologic features that can be used to deduce vertical displacements. They also have generated the molasse deposits that have filled the subsiding foreland basin, providing a record of mountain building. In addition, they have participated actively in the evolution of the
range by influencing the thermal structure and the stress field through redistribution of mass at the earth surface. Surface processes must therefore be taken into account in any analysis of the mechanics of mountain building. They are also probably the major factors that differentiate intracontinental megathrust from subduction zones.

This chapter is not a comprehensive review of the Himalaya. For pedagogic reasons the author mainly describes studies carried out across the Himalaya of central Nepal because he is most familiar with this area that has attracted particular attention over the last decade.

The author first briefly introduces in Section 1 the geodynamic setting and presents in Section 2 a model of the development of the Himalayan orogen and foreland basin. In Section 3 the structure and kinematic evolution of the Himalaya as constrained from surface geology, geochronology, and geophysical investigations is reviewed in more detail. Section 4 describes how the kinematics of thin-skinned deformation along the Himalayan foothills can be determined from the deformation of abandoned river terraces. In Section 5 the pattern of river incision across the whole range is described and it is shown that it basically reflects the kinematics of overthrusting. Section 6 discusses geodetic measurements of crustal deformation, historical seismicity, and the seismic cycle along the Himalaya. Section 7 discusses some general questions about continental deformation and seismicity:

- How is deformation distributed throughout the range, and where are the faults capable of producing very large recurrent earthquakes?
- What can we learn about future large earthquakes from seismicity and deformation monitored over a limited period of time?
- What proportion of crustal deformation is expressed in the seismicity?
- Starting with recent deformation, measured over a decade with geodetic techniques, can we extrapolate backwards to explain the long-term history of the range as expressed in its structural geology?
- How does the erosion rate compare with tectonic uplift?

2. An Active Collisional Orogen

2.1. Geodynamical Setting and Key Structural Features

The Himalayan arc is one of the major zones of deformation that have absorbed the indentation of India into Eurasia (e.g., Powell and
Conaghan, 1973). The collision started about 50 Myr ago and produced a combination of lateral escape and crustal thickening that has given rise to the highest topographic features on earth (e.g., Molnar and Tapponnier, 1975; Harrison et al., 1992; Tapponnier et al., 2001) (Fig. 1). Subsequently, India and stable Eurasia continued converging at a rate of about 5 cm/year (Patriat and Achache, 1984). At present, the 4–5 cm/year of northward displacement of India relative to stable Eurasia is still being absorbed by a combination of horizontal shear and crustal shortening. This is demonstrated both by the pattern and kinematics of active faults in Asia (Molnar and Tapponnier, 1975; Avouac and Tapponnier, 1993) and by GPS measurements (Larson et al., 1999) (Fig. 2a), although the respective contribution of these two mechanisms to the overall deformation remains a matter of debate (Tapponnier et al., 2001; Wang et al., 2002). Across the Himalaya of central Nepal a fraction of this convergence (estimated at about 2 cm/year) is absorbed by crustal shortening, as shown from GPS geodetic campaigns carried out over the last decade (Fig. 2b). Ongoing crustal shortening across the Himalaya is also manifested by recurring large earthquakes with magnitude Mw
above 8, such as the Bihar–Nepal earthquake of 1934 or the Kangra earthquake of 1905 (Fig. 3, Table 1).

Relics of the Tethys ocean that used to separate the northern margin of India from the active southern margin of Eurasia can now be traced along the Indus-Tsangpo suture zone (ITSZ) (e.g., Burg, 1983; Searle et al., 1987) well north of the Himalayan summits (Figs. 1 and 4). To the south, Cambrian to Eocene Tethyan sediments deposited on the northern passive margin of the Indian continent were sutured to the volcanic and plutonic rocks of the once-active margin of Eurasia (Burg et al., 1987; Searle et al., 1987). Now lying at elevations around 5000 m, they were intensely deformed (Ratschbacher, 1994), probably in the early “Alpine” period of the collision which lead to the development of the North

---

Fig. 2. (a) Velocities relative to stable Eurasia measured from GPS over about the last 10 years, compiled in 2001 (data from Wang et al., 2001). All velocities were expressed in ITRF 97. (b) Velocity relative to India determined from GPS measurements by IDYL-HIM (Jouanne et al., 1999), LDG (Avouac et al., 2001), and CIRES (Larson et al., 1999). All velocities were expressed in ITRF 97. Insert shows all measurements projected on section AA’ with accounting for the arc curvature as defined from the arcuate shape of the front of the high range (Bilham et al., 1997).
Himalayan Nappe zone (Burg et al., 1984; Yin et al., 1999). There is evidence that thrust faulting and crustal thickening in southern Tibet persisted until mid-Miocene times (Yin et al., 1999). Extension perpendicular to the range has also been documented along a normal fault that separates the Tethyan sedimentary cover from the High Himalayan crystalline units. This fault is called the North Himalayan Normal fault (Burg et al., 1984) or the South Tibetan Detachment (Burchfiel et al., 1992). In this chapter, these extensional processes that are thought to have resumed between about 15 and 20 Myr ago are not discussed (Searle et al., 1997); rather, the focus is on the subsequent deformation that affected the Himalayan orogen. To the south, crustal shortening chiefly resulted from deformation on a limited number of major thrust faults: from north to south, these are the Main Central Thrust fault (MCT), the Main Boundary Thrust fault (MBT), and the Main Frontal Thrust fault (MFT) (Fig. 4) (e.g., Gansser, 1964; Le Fort, 1975a; Nakata, 1989; Yeats et al., 1992). To the first order, these major thrust faults were activated in a forward propagation sequence. This geometry has prompted the view that the Himalaya should be seen as a crustal-scale accretionary prism.
Davis et al., 1983), with a structure like that produced when a rigid backstop overthrusts a rigid basement with overlying sand layers (Fig. 5). According to this model, a mountain range is a critical wedge with steady-state geometry reflecting the balance between frictional stresses at base of

![Fig. 3. Map of Himalayan arc and Gangetic plain with location of section AA' and INDEPTH seismic profiles (e.g., Hauck et al., 1998). The Main Frontal Thrust fault (MFT) emerges along the front of the Siwalik hills that form the sub-Himalaya. The front of the high range, defined here as the 3500-m elevation contour line, lies about 100 km north of MFT. Shaded areas show proposed ruptured areas of the M > 8 earthquakes of 1905 and 1934 (see parameters in Table 1).](image)

TABLE 1. Estimated Parameters of Major Himalayan Earthquakes (Mw > 8) since 1897 (Pandey and Molnar, 1988; Chander, 1989; Molnar, 1990; Bilham, 1995; Ambraseys and Bilham, 2000, Bilham and England, 2001)

<table>
<thead>
<tr>
<th>Date</th>
<th>Epicenter</th>
<th>L (km)</th>
<th>I (km)</th>
<th>$M_0 \times 10^{21}$ (Nm)</th>
<th>Mw</th>
</tr>
</thead>
<tbody>
<tr>
<td>1897</td>
<td>$\sim 26^\circ N$</td>
<td>$\sim 91^\circ E$</td>
<td>200–300</td>
<td>$\sim 100$</td>
<td>03</td>
</tr>
<tr>
<td>1905</td>
<td>33°N</td>
<td>76°E</td>
<td>280 (&gt; 120)</td>
<td>80</td>
<td>2</td>
</tr>
<tr>
<td>1934</td>
<td>27.6°N</td>
<td>87.1°E</td>
<td>250–300</td>
<td>100</td>
<td>4</td>
</tr>
<tr>
<td>1950</td>
<td>28.4°N</td>
<td>96.8°E</td>
<td>200</td>
<td>120</td>
<td>8</td>
</tr>
</tbody>
</table>

(Davis et al., 1983), with a structure like that produced when a rigid backstop overthrusts a rigid basement with overlying sand layers (Fig. 5). According to this model, a mountain range is a critical wedge with steady-state geometry reflecting the balance between frictional stresses at base of
the wedge and shear stress induced by the topographic slope (Chapple, 1978; Davis et al., 1983; Dahlen and Suppe, 1988; Dahlen, 1990). Crustal thickening results from frontal accretion by southward propagation of the deformation front and from internal thickening to
maintain a constant critical slope. If erosion is taken into account, then the whole wedge might reach some steady-state geometry with particle trajectories like those pictured in Fig. 6. Horizontal shortening, and hence active thrust faults, would be distributed throughout the prism. If erosion is assumed to be distributed uniformly, the pattern of horizontal displacement can be expressed analytically based on a simple mass balance yielding,

\[ V(x) = \frac{V_0 h x}{(h + (x + W)\tan\varphi)}, \]

where the origin of abscissa is taken at the backstop, assumed to be fixed. \( V_0 \) is the convergence rate, hence the velocity of the leading edge of the wedge. Fig. 6 shows the predicted horizontal surface velocity for \( h = 15 \) km, \( W = 200 \) km, \( \tan(\alpha + \beta) = 4\% \), and \( V_0 = 20 \) mm/year. This set of parameters corresponds approximately to the geometry and size of the Himalayan
wedge of central Nepal, assuming that about half of the underthrusting Indian crust is incorporated into the Himalayan wedge. For the wedge to be in steady state, a 2-D erosion flux of 300 m$^2$/year is required (per unit length of the Himalayan arc). This is in the range of possible value inferred from sediment budget over the last 2 Myr at the scale of the whole Himalayan range (Métivier et al., 1999) or from denudation rates derived from river sediment discharge in central Nepal (Lavé and Avouac, 2001). This value corresponds to an erosion rate of about 1.5 mm/year on average over the section.

Over the last few years several models have been proposed, all based on assumption that the crust is detached from a subducting upper mantle with prescribed kinematics (e.g., Willet et al., 1993). Various rheologies, possibly with some depth dependence, as well as various distributions of erosion across the surface, have been considered (Williams et al., 1994; Willet, 1999). Although they differ as to the details of the predicted kinematics or steady-state wedge geometry, these various models would predict a more- or less-distributed shortening throughout the wedge, as expressed by Eq. (1). The geodetic measurements in Fig. 2b show distributed contraction similar to what is predicted from Eq. (1). However, it should be realized that this comparison does not make any physical sense. The geodetic measurements represent the accumulation of interseismic strain, part of which is elastic deformation that will ultimately be released by coseismic deformation. Equation (1), in contrast, represents unrecoverable strain due to brittle deformation of the sand layers. So, Eq. (1) should rather be compared to the averaged displacement field that results, over the long term, from accumulated interseismic and coseismic deformation. This analysis therefore requires identification of the major active faults across the Himalaya and the determination of the slip rates on these faults.

2.2. Foreland Deposition During Overthrusting

Sedimentation in the foreland basin provides indirect information on the growth of an orogen. As the Himalayan wedge grows, it overthrusts and flexes down the Indian basement forming a “foreland flexural basin” in which a fraction of the material eroded from the range accumulates with a stratigraphic organization that depends on the kinematics of overthrusting (Lyon-Caen and Molnar, 1985) (Fig. 7). Several kilometers
of Cenozoic molasse (Murrees and Siwaliks formations) have thus accumulated on the Precambrian Indian basement. These molasse crop out along the foothills where they were scraped off the basement and folded as schematized in Fig. 9. They consist of an upward coarsening sequence (claystone, siltstone, sandstone, and conglomerates) of upper Miocene to Pleistocene age. As shown in Fig. 7, the development of the flexural basin during underthrusting implies some progradation away from the mountain front of sediment facies and of the contact between the basement and the most distal sediments. This model implies, as observed, an upward gradation from distal to more proximal facies.

If the age of the oldest sediments lying over the pre-Tertiary basement is plotted as a function of the distance perpendicular to the range, the rate
of sediment progradation can be deduced (Fig. 7). The southward progradation over about the last 15 Myr was estimated in this way to be between 10 and 15 mm/year (Lyon-Caen and Molnar, 1985), based on a compilation of well data in the Gangetic foreland.

In Fig. 8, only the well data are close to the study area shown in Fig. 3. The front of the high range (defined from the 3500-m elevation contour line shown in Fig. 3) is also indicated because the topographic load over the Indian plate has probably been the main factor controlling the shape of the flexed basement. This plot is not very well constrained since the age of the youngest sediments on top of the basement was estimated with very few chronological constraints and with the assumption of constant sedimentation rates (Lyon-Caen and Molnar, 1985). Two additional points were added based on the stratigraphic sections along the Surai Khola (Appel et al., 1991a) and along the Bakeya (Harrison et al., 1993), as explained in Fig. 9 (modified from Lavé, 1997).
(Appel et al., 1991b) and along the Bakeya (Harrison et al., 1993) (see location in Fig. 3); these were both dated from magnetostratigraphy, and can be used to infer sedimentation rates and hence progradation rates (Lavé, 1997). If the geometry of the foreland basin is assumed to be steady state, the stratigraphic thickness, \( h^* \), of the foreland sediments deposited before \( t_{1-2} \) might be used to estimate the distance at time \( t_{1-2} \) of that section from the front of the high range. As shown in Fig. 9, this estimate requires some restoration of the fault slip on the MFT. The position of the transition from the Lower to the Middle Siwalik and from the Middle to the Upper Siwalik was both used. Altogether, the plot in Fig. 8 indicates a rate of progradation, \( V_{pr} \), of about 15 mm/year over the last 15 Myr.

Assuming that the geometries of the Himalayan wedge and of the flexed Indian plate have not changed with time, as initially proposed by Lyon-Caen and Molnar (1985), \( V_{pr} \) might be taken as equal to the overthrusting rate \( V_0 \) (the convergence rate between India and southern Tibet) (Fig. 7). Actually this reasoning ignores how the mountain front might have retreated or advanced as a result of erosion or of crustal thickening. Later, Molnar (1987) proposed that internal deformation of the crustal wedge would account for about 5 mm/year of additional shortening, and revised the thrusting rate to about 15–20 mm/year.

In fact, a more general formulation should account for internal deformation of the range as well as for possible erosional retreat of the
front of the high range, taken to represent the edge of the overriding load that flexes down the plate (Fig. 7). We should then write,

\[ V_{pr} = V_{HR} + \frac{dW}{dt}, \]

where \( V_{HR} \) represents the velocity of the front of the high range with respect to the foreland, and \( W \) is the width of the accretionary prism (Fig. 7). To relate \( V_{pr} \) to the convergence rate \( V_0 \), it is necessary to take into account erosion of the mountain front. The retreat (if positive) or advance (if negative) of the front of the high range with respect to the hanging wall (southern Tibet being considered as the back stop) is then

\[ V_{er} = V_0 - V_{HR}. \]

This term might actually be large and could have varied during the Himalayan orogeny. In the absence of any crustal thickening or isostatic response, if eroded at a vertical rate \( e \) the mountain front would apparently retreat by

\[ V_{er} = e/tg\alpha. \]

The denudation rate in the High Himalaya is estimated to be of the order of 4–8 mm/year (Lavé and Avouac, 2001) and the slope of the front of the high range is about 10%. This would yield a very rapid retreat by as much as 40–80 mm/year. In reality, erosion is compensated by some uplift, \( u \), due to isostatic response and crustal thickening, so that the apparent retreat is less and should be written,

\[ V_{er} = \frac{(e - u)}{tg\alpha}. \]

Some estimate of \( V_{er} \) is therefore needed to fully describe the kinematics of mountain building and its relation to rates of sediment progradation and deposition in the foreland. Thus, the southward progradation rates of sediments and subsidence rates in the foreland depend on overthrusting, but a direct estimate of the thrusting rate is not straightforward and some idea of the distribution of erosion and of internal shortening of the wedge is needed.

3.1. Geological and Geomorphologic Setting

The major Himalayan thrust faults separate domains with contrasted geology and also control the present morphology of the range (Figs. 4 and 10). The Indo-Gangetic plain is the 200–300 km wide foredeep at front of the range, where a fraction of the material eroded from the rising topography is trapped, the rest of the material being transferred by the Ganges drainage system and ultimately delivered to the Bengal fan (e.g., France-Lanord et al., 1993). Several kilometers of Cenozoic molasse, essentially the Miocene to Quaternary Siwalik formation, have accumulated over the Precambrian Indian basement. The Indian basement is mainly composed of Archean to Early Proterozoic metamorphic rocks. North of the Indo-Gangetic plain, the sub-Himalaya is a zone of thin-skinned tectonics bounded to the south by the MFT and to the north by the MBT (e.g., Delcaillau, 1986; Mugnier et al., 1999; Lavé and Avouac, 2000). It consists of Tertiary siltstones, sandstones, and conglomerates that have been scraped off the basement, folded, and faulted at the front of the advancing range. Although highly erodible, the molasse form steep reliefs which attest to the on-going tectonic activity of this fold and thrust belt. In the study area of central Nepal, the sections along Surai Khola (Corvinus, 1988; Appel and Rossler, 1994), Tinau Khola (Gautam and Appel, 1994), Arung Khola (Tokuoka et al., 1986), and Bakeya Khola (Harrison et al., 1993) show nearly the same sedimentary sequence, 3500–5500 m thick, that was deposited between 14 and 1 Myr ago according to paleontological and magnetostratigraphic studies. It forms subdued reliefs with elevations of the order of a few hundred meters. In places, the fold belt involves slices of pre-Tertiary sediments, reddish-maroon quartzites and gray shales with some doleritic intrusions (Gautam et al., 1995; Lavé and Avouac, 2000), very similar to Vindhyan units drilled at Raxaul (Sastri et al., 1971) (see location in Fig. 4). Pre-Tertiary units involved in the sub-Himalaya fold belt are also reported 50 km east of the section along the Bagmati, north of the Kamla Khola (Mascle and Hérail, 1982). The décollement underlying the fold belt must therefore lie on top of the Indian basement, at some 5–6 km depth along the section across central Nepal. This hypothesis is fairly consistent with the comparison of the balanced cross-sections in the sub-Himalaya south of the Katmandu basin and the geological well log of Raxaul, which cuts
through the whole undeformed Tertiary cover over the basement (Lavé and Avouac, 2000).

Some authors have proposed that the deformation front might extend farther south of the MFT as a blind detachment below the Indo-Gangetic plain. This view was initially proposed by Seeber and Armbruster (1983),
based on the intensity distribution of the large historical earthquakes along the Himalaya, in particular during the 1924 Bihar–Nepal event. The author has never seen any convincing evidence for this possibility. There is no geomorphic or subsurface geological evidence of any zone of deformation south of the MFT along the Himalaya of central Nepal (Lavé and Avouac, 2001). The zone of high macroseismic effects south of the MFT most probably resulted from site effects, i.e., amplification of ground motion and liquefaction within the loose, water-saturated sediment cover of the Gangetic plain.

North of the MBT, the Lesser Himalaya units consist of low-grade metasediments: phyllite, quartzite, and limestone of Devonian or older ages (Upreti, 1999). At some places overlying Tertiary units have been preserved, in particular the sandstones and siltstones of the Dumri formation (Sakai, 1985). This formation consists of Himalayan foreland sediments that were deposited between about 16 and 21 Myr ago, as indicated from ~20-Myr-old detritic muscovites and from preliminary results from a magnetostratigraphic study (DeCelles et al., 1998). The section across the Himalaya of central Nepal near Katmandu is characterized by a crystalline sheet thrust on top of the Lesser Himalayan units (Stöcklin, 1980). It forms the Katmandu klippe which is the chief reason for the impressive relief of the Mahabarat range which reaches elevations up to 2500–3000 m just north of the sub-Himalayan fold belt (Fig. 10). The crystalline units, consisting mainly of schist and gneiss intruded by Late Cambrian to Ordovician granites (Schärer and Allegre, 1983), are overlain by Cambrian to Eocene “Tethyan” sediments (Stöcklin, 1980). Thermobarometric studies indicate that this crystalline sheet overrode the Dumri formation, in particular in the Tansen area (Bollinger, 2002). The basal “Mahabarat” thrust can be interpreted as the southern extension of the MCT, although the possibility that it is a distinct thrust fault rooting below the MCT cannot be excluded (e.g., Upreti, 1999). North of the Klippe, the foliation in the Lesser Himalaya schist depicts a large antiformal structure, the Pokhara–Gorkha anticlinorium (Pêcher, 1989). This structure has been interpreted as a hinterland-dipping duplex structure as shown in Fig. 10 (Brunel, 1986), that has been recognized on most sections across the Nepal Himalaya (Schelling and Arita, 1991; Schelling, 1992; DeCelles et al., 2001) and in Kumaon, India (Srivastava and Mitra, 1994). The structural control of the morphology in the Lesser Himalaya is generally inverted, with synclinal crests and anticlinal valleys, suggesting a typically “mature” relief (Valdiya, 1964).

North of a line trending about N108°E, that roughly coincides with the trace of the MCT, the topography rises abruptly from elevations around
1000 m to more than 6000 m. The abrupt break-in-slope marks the front of the High Himalaya that can be traced all along the Himalayas of Nepal (Fig. 3). The High Himalaya units consist of amphibolite-grade schist intruded by large leucogranitic plutons. Neodymium isotopic provenance suggests that Lesser Himalaya metasediments were derived from the Indian craton, while the High Himalayan rocks most probably correspond to an exotic terrane accreted onto India in the early Paleozoic (Robinson et al., 2001).

A singular feature of the range morphology is the position of the front of the high-range well to the north of the main bounding thrust faults along the foothills. This feature has inspired a variety of interpretations. Because this area coincided with a zone of localized ongoing uplift revealed from leveling data, it has been attributed to active thrusting at the front of the high range (Bilham et al., 1997). Some authors interpret it as the preserved topographic signature of a Late Miocene reactivation of the MCT zone (Harrison et al., 1997). Others have proposed that headward regressive erosion along the rivers cutting the edge of the Tibetan Plateau would have induced uplift of the Himalayan peaks through isostatic rebound, enhancing orographic precipitation, and hence denudation (Molnar, 1990; Burbank, 1992; Masek et al., 1994; Montgomery, 1994). Such a process would have driven accelerated uplift, independently of the kinematics of active thrust faulting, in response to climate change during the Cenozoic.

3.2. Crustal-Scale Structural Models of the Himalaya

Based on surface geology, a variety of crustal-scale sections across the Himalaya have been proposed. One early view is that all the major thrust faults are crustal-scale faults that developed following a forward propagation sequence after the collision along the ITSZ (e.g., Le Fort, 1975b; Mattauer, 1975, 1986; Molnar and Lyon-Caen, 1988; Molnar, 1990). According to that view, all faults would be parallel to one another and reach to the Moho as schematized in Fig. 4.

Alternatively, some authors have proposed that all the major faults root in a common midcrustal décollement (Brunel, 1983; Schelling and Arita, 1991). Such a geometry was initially inferred from the Lesser Himalayan antiformal structure, and it appears to be consistent with barometric studies. Indeed, peak metamorphic pressures documented in the Nepal Himalaya never exceed about 8 kbar (see Guillot, 1999 for a review), suggesting that
rocks were exhumed from depths of about 30–35 km at most. This is the hypothesis used to construct the cross-section in Fig. 10. This section bears some similarity with the more elaborated “balanced” sections that have been constructed along several transects across the Himalaya of Kumaon (Srivastava and Mitra, 1994), far-western Nepal (DeCelles et al., 2001), and eastern Nepal (Schelling and Arita, 1991). However, it should be realized that the geometric rigor used to balance these sections might give a misleading impression of accuracy. The basic assumptions—that deformation occurs by bedding slip with maintenance of constant width and length of the various units, and that the footwall remains rigid—certainly do not hold in the case of the Lesser Himalayan units, because these have developed an intense foliation and have experienced a significant amount of pure shear. Although the details of the balanced sections should therefore be regarded with caution, they have the advantage of integrating a logical model of the structural evolution and do provide a reasonable basis for kinematic interpretation. The main point is that these balanced sections all suggest some duplex structure in Lesser Himalaya, with all the major thrust faults rooting in a midcrustal décollement beneath the high range. The duplex must have developed after the emplacement of the crystalline thrust sheets.

3.3. Moho Geometry

The seismic experiments conducted in southern Tibet all suggest a crustal thickness of the order of 70–80 km (Hirn et al., 1984; Zhao et al., 1993; Kind et al., 1996) about twice the crustal thickness of the Indian shield which has been estimated at about 40 km on the basis of teleseismic receiver functions (Saul et al., 2000) (Fig. 12). The Bouguer anomaly in India and in Tibet primarily indicates local Airy compensation in keeping with those estimated crustal thicknesses. By contrast, important deviations from Airy isostasy are observed below the Himalayan range and its foreland, as seen in particular from the gravity data across the Himalaya of central Nepal (Fig. 11) (Lyon-Caen and Molnar, 1983; Lyon-Caen and Molnar, 1985; Jin et al., 1996; Cattin et al., 2001). The observed values are more negative than expected from local isostasy over the Gangetic plain, indicating some mass deficit there. By contrast they indicate some mass excess below the range (Fig. 14). These deviations are the typical signature of a flexural support of the range, meaning that the weight of the Himalaya is supported by the strength of the underthrusting Indian plate. Moreover, the steep gravity gradient of the
order of 1.3 mGal/km beneath the High Himalaya suggests a locally steeper Moho, coincident with the position of the midcrustal ramp along the Main Himalayan Thrust fault (MHT) (but not necessarily parallel). Flexural modeling in terms of a thin elastic plate overlying an inviscid fluid (Lyon-Caen and Molnar, 1983; Lyon-Caen and Molnar, 1985; Jin et al., 1996) can successfully reproduce the observed gravity anomalies, but this requires that some forces or momentum be exerted on the flexed plate in addition to the load due to the high topography. This kind of modeling also requires an abrupt decrease in the strength of the Indian plate beneath the high range to account for the locally steeper gradient of gravity anomalies. Actually this northward decrease in the apparent flexural rigidity might reflect the thermal structure of the range and its influence on crustal rheology. Due to the kinematics of underthrusting, the thermal structure implies relatively high temperatures at midcrustal depths favoring ductile flow within the crust (Fig. 12). This induces some decoupling within the upper crust and upper mantle resulting in an abrupt decrease of the flexural rigidity (Burov and Diament, 1995). It turns out that, if the thermal structure and its influence on crustal rock rheology are taken into account, there is no need for additional forces other than the weight of the topography. The gravity profile is indeed relatively well reproduced by a 2-D mechanical model in which the Indian lithosphere is flexed down by the advancing topography of the range and sedimentation in the foreland (Cattin et al., 2001) (Fig. 11). The computed Moho fits seismological constraints and is consistent with the main trends in the observed Bouguer anomaly. Although the picture in Fig. 11 shows a relatively smooth Moho, there is some indication that the structure of the lower crust might be quite complex with imbrications of crustal and upper mantle rocks (Hirn et al., 1984) (Fig. 12).

3.4. Geophysical Constraints on the Geometry of the Main Himalayan Thrust Fault (MHT)

The structure of the crust across the central Himalaya has been investigated through a variety of means including gravimetric, magnetotelluric, and seismic techniques. A major feature revealed by the seismic experiments in southern Tibet is a strong midcrustal reflector at a depth of 35–40 km. The existence of this reflector was first suggested by wide-angle seismic reflexion studies (Hirn and Sapin, 1984); later, it was better imaged from the common midpoint (CMP) deep seismic profiles run during the INDEPTH experiment (Zhao et al., 1993; Brown et al., 1996; Nelson et al., 1996).
As shown in Fig. 13, this conspicuous reflector was found to coincide with the midcrustal décollement inferred from the structural sections across the range and was therefore termed the Main Himalayan Thrust fault, MHT (Brown et al., 1996).

Fig. 11. Bouguer anomalies along a N18°E profile across the Himalaya of central Nepal (modified from Cattin et al., 2001). All the data within a 30-km wide swath were projected onto the section. Several data sets were merged with accuracies ranging from about 0.5 to 7 mGal [Bureau Gravimétrique International database (Van de Meulebrouck, 1983; Abtoux, 1987; Sun, 1989)]. The Moho is set to 40 km beneath the Indian shield according to teleseismic receiver functions (Saul et al., 2000); white line shows Moho according to INDEPTH seismic profiles (Zhao et al., 1993; Brown et al., 1996), and thick vertical bars show Moho picks on receiver functions (Kind et al., 1996). The thick line shows the expected Bouguer anomaly in the case of isostatic compensation of a crust with uniform 2.67 density (computed from the topography smoothed with a 50-km wide Gaussian). The thin line is the result of a mechanical model that accounts for the low-density sediments in the foreland, the rheological layering of the crust, and its dependence on the thermal structure (Cattin et al., 2001).
FIG. 12. Steady-state thermal structure of the range computed from a 2-D finite element model (Henry et al., 1997) (from Cattin and Avouac, 2000). Erosion is assumed to exactly balance tectonic uplift, as computed from thrusting of the hanging wall over the flat–ramp–flat geometry of the MFT–MHT, assuming a rigid footwall. Boundary conditions are a constant surface temperature of 0°C, a bottom heat flow of 15 mW/m², and a heat production in the upper crust of 1.5 μW/m³. Sensitivity tests show that, insofar as the model-fit thermobarometric constraints [650°C above the MCT (e.g., Hubbard, 1989; Hubbard and Harrison, 1989), the computed thermal structures show only small differences, so the model shown here is probably a robust estimate (Cattin and Avouac, 2000; Bollinger, 2002).

FIG. 13. Geophysical constraints on the crustal structure across central Nepal. The conductivity section was obtained from a magnetotelluric experiment carried out along the section AA' across central Nepal (Lemonnier et al., 1999). Also reported are the INDEPTH seismic sections run about 300 km east of section AA' (see location in Fig. 3) (Zhao et al., 1993; Brown et al., 1996; Nelson et al., 1996). All the thrust faults are inferred to root at depth in a subhorizontal ductile shear zone that would correspond to the prominent midcrustal reflector.
The deep electrical structure of the central Nepal Himalaya was imaged from magnetotelluric sounding (Lemonnier et al., 1999) (Fig. 13). Variations of electrical conductivity in the crust can result from changes in fluid content, pore geometry, or lithology. Conductive zones in the crust are generally thought to reflect well-connected conductive phases (brines, melts) or conductive minerals such as graphite (Marquis et al., 1995). The section shows a high conductivity in the foreland (\(\sim 30 \, \Omega \, m\)) consistent with the geometry of the molassic foreland basin, which contrasts with the resistive Indian basement (\(>300 \, \Omega \, m\)) and Lesser Himalayan units (\(>1000 \, \Omega \, m\)). A continuous shallow-dipping conductor can actually be traced northward and coincides relatively well with the position of décollement beneath the Lesser Himalaya inferred from structural studies. It may well reflect some fluid-rich sediments dragged along the thrust fault, a process expected as the rugged topography of the Indian basement underthrusts the Lesser Himalaya.

Farther north, the magnetotelluric section show a major conductive zone (\(\sim 30 \, \Omega \, m\)) at about 15 km depth under the front of the High Himalaya. This zone coincides with the position of the midcrustal ramp beneath the front of the High Himalaya, as well as with a zone of intense microseismic activity (Fig. 13) revealed from local seismic monitoring (Pandey et al., 1995, 1999; Cattin and Avouac, 2000).

These geophysical data thus point to the MHT being a major thrust fault that can be traced nearly continuously from the MFT, along the foothills, to beneath southern Tibet. Contrary to the early views depicted in Fig. 4, the MCT, MBT, and MFT should not be seen as equivalent major thrust faults cutting through the whole crust. They more probably represent splay faults rooting in a single midcrustal décollement. The geometry of the MHT is characterized by a ramp and flat geometry, probably with two major ramps. One is very shallow and corresponds to where the fault emerges with a dip angle around 30° at the surface along the MFT. The other one, more conjectural, lies at midcrustal depth beneath the front of the high range and is estimated to dip north by about 15°.

3.5. Chronological Constraints on the Development of the Thrust Package

The chronology of the deformation across the Himalaya has been constrained from cross-cutting relationships, provenance data from the Tertiary molasse (Siwalik and Dumri formations), and thermochronology.
A complete discussion of this issue is well beyond the purpose of this review, although a few key results are pointed out here:

- compressional deformation north of the Himalaya started in the early “Alpine” times of the collision, and there is evidence that crustal thickening in southern Tibet persisted until mid-Miocene times (e.g., Ratschbacher, 1994; Yin et al., 1999). Farther south, deformation and anatexis in the MCT zone were occurring at about 22 Myr (Copeland et al., 1991; Hodges et al., 1996), an age consistent with slightly younger cooling ages in the rocks of the hanging wall, attributed to the emplacement of the MCT (Hubbard, 1989; Copeland et al., 1991; Hubbard et al., 1991; Hodges et al., 1996). Such a timing is also consistent with conventional and isotopic provenance data from the Dumri formation that indicate an early Miocene time for the onset of erosion of the High Himalayan crystalline rocks (Robinson et al., 2001). Out-of-sequence reactivation of MCT may have occurred by late Miocene to Early Pliocene times as indicated from monazite ages of rocks exhumed from midcrustal depths (Harrison et al., 1997; Catlos et al., 2001). This reactivation has been proposed to be responsible for the inverse thermal gradient and for the steep morphologic front of the High Himalaya range (Harrison et al., 1997, 1998).

- there is no direct, reliable estimate of the timing of thrusting on the MBT. Age and provenance of Siwalik sandstones in northern India, west of our study area, show that exhumation of LH have started some 10 Myr ago (Meigs et al., 1995). This observation was interpreted to indicate motion on the MBT. However, exhumation of LH rocks could alternatively have resulted from the development of the LH duplex, so a much younger activation of the MBT, possibly around 5 Myr, could also be possible (DeCelles et al., 1998; Robinson et al., 2001).

- deformation in the sub-Himalaya has proceeded from the mid-Miocene to present.

The chronological constraints discussed above, together with restoration of the various balanced cross-sections across the central Himalaya (Schelling and Arita, 1991; Srivastava and Mitra, 1994; DeCelles et al., 2001), provide some idea of the kinematics over approximately the last 20 Myr. Figure 14 presents the kinematic model proposed for far-western Nepal (Robinson et al., 2001) which also holds with some minor variations for central Nepal (Bollinger et al., submitted-a).

The development of the crustal wedge, over a geological period of time thus appears as a combination of frontal accretion at the toe,
FIG. 14. Schematic model of the structural evolution of the Himalayan orogen (modified from Robinson et al., 2001). This scheme is based mainly on observations along a section across the Himalaya of western Nepal. MCT, Main Central Thrust; DT, Dadeldhura Thrust; RT, Ramgarh Thrust; MBT, Main Boundary Thrust; MFT, Main Frontal Thrust; STD, Southern Tibetan Detachment. (A) 45–16 Ma: Emplacement of the MCT and then of the DT crystalline thrust sheets, coeval with motion on the STD. Deposition of the Dumri and Bhainskati foreland formations. (B) 15–6 Ma: Emplacement of the Lesser Himalaya RT thrust sheet and development of the duplex; deposition of the Lower and Middle Siwaliks. (C) 5–0 Ma: Motion on the MBT and MFT; deposition of the Upper Siwaliks; major phase of exhumation of the LH duplex.
internal deformation of the wedge, and underplating. Underplating associated with duplexation of the Lesser Himalayan units at midcrustal depth has probably been the dominant process in the growth of the Himalayan wedge since about the Middle Miocene. This kinematic scheme is similar to the pattern of deformation expected for a critical brittle taper subjected to erosion except that maintenance of the critical slope was probably achieved without any significant thickening of the upper crustal units.

3.6. Kinematics Model of the Structural Evolution and Exhumation of the Orogen

If topography and the thermal structure are assumed to be in a steady-state regime, the pattern of exhumation should provide direct information on the kinematics of the deforming orogenic wedge.

The most complete picture of the chronology of exhumation in the area considered here comes from the determination of $^{39}\text{Ar}^{40}\text{Ar}$ ages of muscovites (Fig. 15). This technique provides an estimate of the age of cooling of the rock sample through the muscovite-blocking temperature which is estimated at about 350°C. Several authors, in particular, Copeland et al. (2003) have analyzed samples collected along the central Nepal section from the high range to the southern edge of the Katmandu klippe (Fig. 15). These data indicate that the trailing edge of the Katmandu thrust sheet cooled below 350°C about 20–22 Myr ago, an age consistent with the onset of deformation and anatexis in the MCT zone as discussed above. Ages decrease gradually northwards to about 5 Myr at the front of the high range, following a nearly linear trend corresponding to a slope of about 0.23 Myr/km (Fig. 15). Given that the 350°C steady-state isotherm crosses the MHT at the midcrustal ramp (Fig. 16), this linear trend indicates that the hanging wall overthrusts the midcrustal ramp at a rate that corresponds to the inverse of the age gradient,

$$V_{cr} = V_0 - V_{HR} \sim 4.3 \text{ mm/year}.$$  \hfill (6)

As a first attempt at interpreting this trend, we may ignore accretion (Fig. 16). In that case, according to equation (2) and the data in Fig. 8, the foreland would underthrust the Himalayan topographic wedge by

$$V_{fr} = V_{HR} \sim 15 \text{ mm/year}.$$ \hfill (7)
This is a value consistent with the data in Fig. 9. Moreover, these two values add to the total estimated shortening rate across the range of $V_0 \approx 20$ mm/year. This simple kinematics reconciles the rate of sediment progradation and subsidence of the foreland with the retreat of the mountain front driven by erosion.

The kinematics may be compared with two end-member cases of a steady-state regime without accretion:

- If the hanging wall is assumed to overthrust an undeformable footwall by 20 mm/year, there would be no sediment progradation and no subsidence of the foreland, and we should observe a diachronic exhumation corresponding to,

$$V_{cr} \sim 20 \text{ mm/year},$$

$$V_{pr} = V_{HR} \sim 0 \text{ mm/year}.$$
– In the opposite case, if the convergence is assumed to be entirely accommodated by underthrusting of the Indian basement below a rigid hanging wall without any erosion at the surface, we get,

\[ V_{cr} = V_0 \sim 0 \text{ mm/year}, \quad \text{(10)} \]
\[ V_{pr} = V_{HR} \sim 20 \text{ mm/year}. \quad \text{(11)} \]

The kinematic deduced from the geochronological data is clearly closer to the second end-member. So both the structural and geochronological data require the footwall to be deforming because it must underthrust the MHT (Fig. 16), before it gets accreted to the footwall. The kinematics in Fig. 16 might be modified to fit the structural evolution model of Fig. 14 by assuming that some accretion occurs by underplating due to the development of a duplex at midcrustal depths (Fig. 17). The resulting model is shown in Fig. 18. To draw this picture, a continuous process of accretion is assumed. The top of the underthrusting plate is accreted while the lower part underthrusts the orogenic wedge. In reality the process of accretion could be discontinuous.

**Fig. 16.** Kinematic model of overthrusting at rate \( V_0 - V_{HR} \), and underthrusting at rate \( V_{HR} \), if the topography and thermal structure are assumed to be in a steady state. These kinematics predict diachronic exhumation (through the Muscovite-blocking temperature which is estimated at 350°C) with an age gradient of \( 1/V_0 - V_{HR} \).
The assumption of steady-state topography would imply $\frac{dW}{dt} = 0$ mm/year in Eq. (2), so that Eqs. (7) and (8) are not modified. Thus, the model still reconciles the rate of sediment progradation and subsidence of the foreland with the retreat of the mountain front driven by erosion. It also
fits the structural evolution model, which requires accretion. For a more
detailed discussion of this model and its comparison to structural,
petrological, thermometric, and thermochronologic data, the reader may
refer to Bollinger (2002), Bollinger et al. (submitted-a) and Bollinger et al.
(submitted-c). In the next sections, we compare this kinematic model with
data on deformation and exhumation of the Himalayan orogen over the
Holocene.

4. Kinematics of Active Folding from Holocene River
Terraces across the Sub-Himalaya

The kinematics of thrusting along the front of the Nepal Himalaya can
be documented from the study of deformed river terraces. Here the
author reviews the methodology and presents results obtained from the
terraces along the Bagmati river, which lies approximately south of
Katmandu Basin along section AA′ (see location in Fig. 3). Similar results
have been obtained along the Bakeya river, which cuts across the same
rising anticline about 10 km west of the Bagmati, but these results are
not reviewed here. For more details the reader should refer to Lave´ and
Avouac (2000).

4.1. Structural Section across the sub-Himalaya

Several structural cross-sections were realized along the Bagmati river and
nearby rivers (Lavé and Avouac, 2000). In this area, the Siwalik hills are
divided in two fold belts (Fig. 19). The southern one is the topographic
expression of the MFT that is inferred to reach the surface along the
southern limb of the fold belt (Nakata, 1989). The other fold belt, about
15 km north of the MFT fold belt, is associated with the Main Dun Thrust.
At some places in Nepal the Siwalik hills are more complex with several
fault zones between the MFT and MBT that are assumed to merge with a
single décollement at depth (Mugnier et al., 1992). The MDT fold belt
makes a steep monocline with dips around 65° to the NNW. This belt
involves about 500 m of pre-Tertiary sediments, mainly reddish-maroon
quartzites and gray shales in the lower part. The incompetent shales show
meter-scale folds. The MFT fold belt makes a gently inclined monocline
with dips varying between 25 and 50° to the NNW (Fig. 19). At the front,
Lower Siwalik clay- and siltstones are affected by meter-scale faults and folds. The sections can be balanced by interpreting both folds associated with the MFT and MDT as fault-bend folds (Suppe, 1985) associated with curved ramps that root in the same décollement (Figs. 19 and 10). Palinspastic restorations of the Bagmati and Bakeya sections imply minimum shortening of 11 and 10 km, respectively at the MFT, and 12 km at the MDT. The relief of the fold shows that about 90% of the material uplifted due to slip on the MFT and MDT has been eroded away. So, on average over a geological period of a few million years, denudation balances tectonic uplift.

Fig. 19. Structural section with elevation of abandoned terraces along the Bagmati river across the Siwalik hills south of Katmandu basin, approximately along the section AA’ (see location in Fig. 2) (modified from Lave and Avouac, 2000). Also indicated are ages of terrace abandonment from C14 dating after calibration to calendar ages. The age of T3 (in italics) was indirectly inferred from the age of the same terrace level along the nearby Bakeya river.
4.2. River Incision across the sub-Himalaya

Abandoned fluvial terraces are ubiquitous in the sub-Himalaya of central Nepal, and have been most extensively surveyed along the Bagmati, Bakeya, Narayani, and Ratu rivers (Delcaillau, 1986; Lave’ and Avouac, 2000). They were first recognized and mapped from air photos and Landsat and SPOT images. In the field, they were classified according to their geomorphic nature, facies of the fill material, degree of weathering, and thickness of the soil profile. Nearby disconnected treads were also correlated on the basis of their elevations. The determination of fluvial incision requires elevation measurements of datable levels attributable to former riverbeds. Several different levels may be of interest. If the bedrock shows evidence for fluvial abrasion, in particular if it is still capped by fluvial gravel, the contact between the bedrock and the overlying gravel is called strath surface (Bull, 1991). Sometimes, strath surfaces are preserved only in the form of bedrock benches. Such features are difficult to exploit due to the lack of datable material or criteria for lateral correlation. If the strath surface is capped by a thin gravel cover (typically <10 m), the terrace is called a strath terrace. If the gravel thickness is much larger, it may be called a fill terrace (Bull, 1991). The difference in elevation between various strath surfaces and the present riverbed provide some estimate of the incision rate. This estimate is the most useful for tectonic analysis because it is not susceptible of being influenced significantly by climatic effects. The difference in elevation between the top of the terrace and the present riverbed can also be informative, but it might not relate simply to long-term fluvial erosion because the deposition of the fill and its subsequent incision probably a transient fluvial response to a climatic change (e.g., Weldon, 1986; Bull, 1991). Such fill terraces can therefore be diachronous, so that the precise chronology of their deposition and abandonment is difficult to constrain tightly.

In the Siwalik hills, the strath surfaces are generally covered by a few meters of gravel, subrounded boulders, and cobbles (<30 cm) grading upward into coarse sands. This gravel is very similar to the bedload of the modern rivers. It is overlain by overbank sands and silts, with, at places, mixed colluvium or fanglomerates fed from adjacent valley slopes. These Holocene terraces are characterized by a low degree of weathering with beige to orange color of the silt and sand. Charcoals were found at several sites within the terrace material and could be dated from $^{14}$C. A few samples were found within fluvial bedload, and yielded a lower bound on the time when the river was flowing at this level. Most of the samples were found in
the overlying, less coarse, overbank deposits and providing a lower bound on the age of the underlying strath level.

The chronologic data obtained in this way indicate four major episodes of strath-terrace abandonment, dated to about $9.2 \pm 0.15$, $6.1 \pm 0.15$, $3.7 \pm 0.1$, and $2.2 \pm 0.2$ ka ($^{14}$C ages converted to calendar ages). These dates probably correspond to wet-to-dry climatic transitions during the Holocene period (Lavé, 1997; Lavé and Avouac, 2000). The uppermost level, labeled $T_0$ in Fig. 19, is generally the most prominent in the landscape because it corresponds to particularly wide terrace treads. The second most prominent level is $T_3$ (dated to about $2.2$ ka). Along the Bagmati river the $T_2$ terrace is not well preserved and could not be mapped and measured easily in the field. Figure 19 shows the elevation above the present riverbed of the three other strath surfaces along the Bagmati river, with the locations of the sampling sites where good dating was possible. These data provide an estimate of the average incision, $I(x, y, t)$, at any point $(x, y)$ along the river since the abandonment of terrace $T$ at time $t$ incision,

$$I(x, y, t) = T(x, y, t) - R(x, y), \quad (12)$$

where $R(x, y)$ is the present river profile, and $T(x, y, t)$ is the present elevation profile of the abandoned river terrace.

4.3. Converting Incision Rates to Uplift Rates in the sub-Himalaya

The geometry of the abandoned terraces in Fig. 19 clearly demonstrates that they were abandoned and warped as a result of fold growth as depicted in Fig. 20. These data provide evidence for active thrusting along the MFT. The terraces are not affected by any clear deformation across the MDT. They also can be traced farther upstream across the MBT, without any evidence for tectonic activity. The profiles of the various abandoned terraces along the Bagmati and Bakeya rivers indicate that the MFT has been the major active thrust fault across the sub-Himalaya of central Nepal over the Holocene period. To quantify uplift rates across the MFT fold, a model of the geometry of the terraces before they were deformed is needed. It may simply be assumed that the rivers have maintained a constant profile during deformation, with river incision having counterbalanced tectonic uplift, and that the fold has been growing steadily during the Holocene period. This latter approximation makes sense, given that the mean recurrence interval of the earthquakes that could activate the fold is at most a few hundred years.
Accordingly, the tectonic uplift since terrace abandonment would simply be equal to the incision, i.e., the difference of elevation between the abandoned strath and the present river,

\[
U(x, y, t) = I(x, y, t).
\]  

(13)

The term that would arise from the apparent uplift of the riverbed due to the horizontal component of the displacement field can be neglected due to the generally low stream gradient of the rivers cutting across the sub-Himalaya.

Given that, on average over many seismic cycles, the fold might be assumed to grow more or less steadily, \(U(x, y, t)\) may be assumed to be proportional to the time elapsed since the terrace abandonment. It follows
that in that case the incision deduced from two different terraces abandoned at times \( t \) and \( t' \) should match,

\[
I(x, y, t)/t = I(x, y, t')/t'.
\]  

The incision profiles deduced from \( T_0 \) and \( T_3 \) along the Bagmati section are similar but some mismatch on the northern limb of the fold, due to an incision deficit, was observed (Lavé and Avouac, 2000). This suggests that the simple assumption of Eq. (13) does not hold.

Actually the cumulative uplift since terrace deposition, at time \( t \), with respect to a point attached to the Indian lithosphere just south of the MFT, ought to be written at point \( M(x, y) \),

\[
U(x, y, t) = I(x, y, t) + D(x, y, t) + P(x, y, t),
\]  

where \( D(x, y, t) \) is the local base-level change since time \( t \), and \( P(x, y, t) \) is the change in elevation at point \( M(x, y) \) due to possible changes of the river gradient and sinuosity. The term \( D(x, y, t) \) is the sedimentation rate (or possibly incision rate) at the front of the MFT (Lavé and Avouac, 2000). There are no uplifted Holocene terraces south of the MFT, and rivers clearly aggrade where they come out of the Siwalik hills. The sedimentation rate therefore must have remained positive during the Holocene. Based on the magnetostratigraphic sections (Appel et al., 1991b; Harrison et al., 1993), its value can be estimated to \( 0.45 \pm 0.5 \text{ mm/year} \).

In the absence of any way to assess possible changes of the stream gradient during the Holocene, this parameter is assumed to be constant and equal to the present \( 0.27\% \) value. It should be realized that the Holocene trend toward a drier climate in this part of Asia, may have forced hydrological modifications, including gradient changes, but these are not easily analyzed.

The paleo-channel corresponding to each terrace level can be estimated from the terrace remnants within the steep flanks of the canyon (Lavé and Avouac, 2000). The study of these channels reveals that the river sinuosity has increased during the Holocene. The river was least sinuous (\( s = 1.8 \)) at the time of \( T_0 \) strath beveling. At the time of \( T_1 \) abandonment the sinuosity had increased to about 2.0. At time of \( T_3 \) abandonment, it was about 2.1. The present riverbed shows the narrowest and most sinuous channel (\( s = 2.4 \)). This sinuosity increase must have resulted in a gradual increase of the apparent slope of the river, and hence an incision deficit, on a projection
perpendicular to the fold axis. This might explain why no straths but only fill terraces are found north of the MFT fold zone, where tectonic uplift has probably been too slow to compensate the gradual elevation of the riverbed.

To account for sinuosity changes $P(x, y, t)$ might be written as,

$$P(x, y, t) = S(dL(x, y, t)/dt)$$

(16)

where $L(x, y, t)$ is the longitudinal distance along the river at time $t$ and $S$ is the channel gradient, which is assumed to be constant along that particular river reach and unchanged with time.

The uplift rate profiles in Fig. 21 were computed from Eq. (15), taking sinuosity and base-level changes into account. The two curves are identical within the error bars. Together with the uplift rate profiles deduced from other terrace levels along the Bagmati and Bakeya rivers (Lavé and Avouac, 2000), these data suggest that the anticline has been growing more or less steadily during the Holocene. The zone of active uplift is 1–14 km wide and the uplift rate reaches a maximum of about 1 cm/year (Fig. 21).

Although the terraces we have surveyed were primarily abandoned because the rivers entrenched to keep pace with tectonic uplift, it should be emphasized that river incision may not simply relate to tectonic uplift. River entrenchment must be analyzed not only in terms of base-level changes and tectonic uplift, but also of sinuosity changes. In addition, under certain circumstances hydrological changes can drive very rapid and episodic river incision. If so, this factor will contribute to an apparent ‘base level change’ possibly overweighting the tectonic term in equation (15). For example, Holocene climate change has indexed as much as 1 cm/year of incision along major rivers flowing across the northern piedmont of the Tian Shan range in Central Asia (Poisson and Avouac, in press). So tectonic uplift can be retrieved from river entrenchment, but some caution is needed to correctly account for nontectonic factors. The effects due to fluvial adjustment, hydrological changes, or variations of sediment supply may be dominant in areas with moderate rate of tectonic deformation.

4.4. Converting Uplift Rates to Horizontal Shortening from Area Balance

A simple way to deduce horizontal shortening from uplift profiles across the MFT fold is to assume conservation of mass (Molnar et al., 1994).
Fig. 21. Uplift rate (top) along the Bagmati section (bottom) deduced from terrace T₀ and T₃ after correction for 0.45 mm/year base-level change due to sedimentation south of the MFT and sinuosity changes. Dashed line shows the predicted uplift rate pattern assuming fault-bend folding. For a cylindrical fold, it would correspond to \( U(x, y) = d(t) \sin \theta(x, y) \). It yields a good fit to the observed uplift rates for a horizontal shortening rate across the MFT of 21 mm/year.
Let $A$ be the area between the present profile of an abandoned terrace tread and the profile of the same terrace level at the time $t$ it was abandoned (Fig. 22). Let $h$ be the thickness of the units detached from the basement. Assuming plane strain and conservation of mass, we derive the mean horizontal displacement, $d$, necessary to account for the deformation of the terrace, as

$$d(t) = A(t)/h = \left( \int_0^x U(x, t) \, dx \right)/h.$$  \hspace{1cm} (17)
This approach requires a continuous terrace tread across the growing fold. The only such terrace is the $T_0$ level along the Bagmati river. Deformation of that level corresponds to an area of $A(T_0) = 1.05 \pm 0.25 \text{ km}^2$. According to the structural description above, the depth of the basement is $5.0 \pm 0.3 \text{ km}$ beneath the MFT and dips northward with a slope of 2.5%. Given that the MFT ramp merges with the décollement about 18 km north of its trace at the surface, the depth to the décollement beneath the fold is $5.5 \pm 0.3 \text{ km}$. We get a shortening $d = 192 \text{ m} \ (\pm 60/\sim 50 \text{ m})$ since $9.2 \pm 0.2 \text{ cal. kyr}$, hence a shortening rate of $21 \pm 7/\sim 6 \text{ mm/year}$ (Fig. 23).

Fig. 23. Plot of horizontal shortening deduced from the various terrace treads along the Bagmati and Bakeya section as a function of age of terrace abandonment (Lavé and Avouac, 2000). These data are consistent with a uniform $21 \pm 1 \text{ mm/year}$ shortening rate over the Holocene period. If the fold is assumed to have been growing by incremental deformation during large earthquakes, the mean shortening rate is slightly modified since some amount of elastic straining at the large scale may yet remain to be released, or may not have been released by the time of emplacement of the different terraces. Assuming that these increments can be as large as $5 \text{ m}$ of horizontal shortening and that all values between 0 and $5 \text{ m}$ are equiprobable, we obtain a long-term averaged shortening rate, offering a best fit to the terrace record, of $21.5 \pm 1.5 \text{ mm/year}$. 
4.5. Converting Uplift Rates to Horizontal Shortening from the Fault-Bend Fold Model

Inspection of Fig. 21 suggests that the uplift profiles recorded by the Holocene terraces are closely related to the geological structure of the fold. Indeed, the small variation of the uplift rate within about 1 km in the middle of the fold correlates with variation of bedding dip angles, hence of the dip of the ramp at depth. The correlation between structural geology and recent uplift is in fact expected for a fault-bend fold. According to this model, the hanging wall accommodates deformation imposed by the fault geometry at depth by bedding plane slip, with no change in the length and width of the beds (Fig. 22). It follows that, on the back limb, uplift rate depends on the dip of the fault at depth, which equals the local bedding dip angle. Assuming cylindrical geometry, the local uplift since time $t$, $U(x, t)$, and the related horizontal shortening $d(t)$, have a simple geometric relationship. In the case of a fold with cylindrical geometry (with beds striking perpendicular to the section taken to be parallel to the $y$ axis), and assuming that the slip increment since terrace abandonment is small, one may write

$$U(x, y) = U(x) d(t) \sin \theta(x),$$

where $\theta(x)$ is the dip angle at point $(x, y)$. The equation holds insofar as the dip angle does not vary significantly over a distance of about $d$, $d \ll dx/d(\sin \theta(x))$.

In the case of a noncylindrical geometry, Eq. (18) should be replaced by (Lavé and Avouac, 2000),

$$U(x, t) = d(t) \sin \theta'(x),$$

where,

$$\sin \theta'(x) = \sin \theta'(x) \sin(\phi(x) - \phi_0).$$

Here $\phi(x, y)$ is the bedding strike at point $M(x, y)$ and $\phi_0$ is the mean direction of transport.

The fault-bend model thus leads to a testable correlation between structural geology and the terrace warping. The uplift profiles deduced from
terraces $T_0$ and $T_3$ along the Bagmati river are in close agreement with the bedding dip angles as expressed in Eq. (19) (Fig. 22). The model was also successfully tested from the terraces along the Bakeya river (Lavé and Avouac, 2000).

This model makes it possible to estimate the cumulative shortening since the abandonment of each terrace by least-squares adjustment of the model to the uplift profiles, $d(t)$ being the only adjustable variable. Each abandoned terrace yields a point on the plot in Fig. 23, which shows $d$ as a function of elapsed time since the abandonment of each terrace. The points are seen to follow a linear trend. A least-squares fit yields $20.4 \pm 1$ mm/year, which may be taken to represent the rate of fault slip on the MFT, or alternatively to the horizontal shortening rate across the fold over the Holocene period. This rate is much better constrained than, but consistent with, the estimate deduced from conservation of mass in the previous paragraph. However, this estimate ignores the fact that the slip is not simply a linear function of time. Slip on the MHT has most probably resulted from recurring coseismic slip events. When this stick-slip behavior is taken into account, assuming slip events of up to $5$ m, the long-term slip rate is revised to $21 \pm 1.5$ mm/year (Lavé and Avouac, 2001).

5. HOLOCENE KINEMATICS OF OVERTHRUSTING AND RIVER INCISION ACROSS THE WHOLE RANGE

5.1. Fluvial Incision across the Whole Range

The kinematics of uplift and overthrusting along a whole section across the Himalaya of central Nepal can be inferred from the pattern of river incision (Lavé and Avouac, 2001). As shown above, the strath terraces provide good constraints on incision rates in the sub-Himalaya where rivers are forced to cut down into the rising anticlines and have abandoned numerous strath terraces. In the Lesser Himalaya strath terraces are sparse. There are, however, prominent fill terraces, more than about $100$ m thick, probably of Pleistocene ages (Iwata, 1976; Fort et al., 1983). Close to the front of the high range, the valleys become steep and narrow and the terraces are no longer preserved. Such a terrace pattern suggests a slow rate of fluvial incision in the Lesser Himalaya and an abrupt increase at the front of the high range. Due to the difficulty of using the terrace record to determine incision rates, the geometry of modern channels along major
rivers draining across the range may be used instead (Lavé and Avouac, 2001). Fluvial incision can be estimated from the shear stress exerted by the flowing water at the bottom of the channel, which can be shown to be a good proxy for river-incision rate. The model was calibrated from a few sites in the Lesser Himalaya and sub-Himalaya where Holocene strath terraces could be dated. This approach has been shown to yield results consistent with the terrace record of river incision, based on cases where the two approaches could be applied to a same reach. Figure 24 shows average profiles obtained from river incision estimated along six major rivers across the Himalaya of central Nepal. It turns out that river incision in the Lesser Himalaya is minor, less than a few mm/year, and rises abruptly at the front of the high range to reach values of about 7–8 mm/year within a 50-km wide

![Graph showing fluvial incision across the Himalaya of central Nepal.](image-url)
zone that coincides with the High Himalayan peaks and with the position of the midcrustal ramp.

5.2. Denudation across the Whole Range

In the sub-Himalaya, the analysis of the river terraces makes it clear that river incision is keeping pace with tectonic uplift. The present topography only accounts for at most 10% of the total volume of rock uplifted by active folding and thrusting, implying that denudation also must be in equilibrium with tectonic uplift (Hurtrez et al., 1999). Denudation should equal fluvial incision if fluvial downcutting is the rate-limiting process that drives hillslope processes. This should be expected if, as suggested for the High Himalaya of northern Pakistan, hillslopes are near their critical slopes for mass movement (Burbank et al., 1996). Mass processes are probably the dominant factor in the Himalaya of Nepal, especially in the high range, where the steep hillslopes often show bare evidence for recent landsliding. The highly dissected topography seems to be totally controlled by the fluvial network. In that case an estimate of the sediment yield can be obtained from the fluvial incision rates along the major rivers. Such estimation was done for 11 catchments in the Himalaya of central Nepal, within the Naryani and Sapt Kosi watersheds, for which measurements of suspended load are available (Lavé and Avouac, 2001). The values obtained this way were found to be too large by about 20%. This misfit is probably insignificant in view of the large uncertainties in both the measurements and estimates, and might partly be offset by the contribution of the bedload.

Denudation rates in the Himalaya of Nepal might also be compared to sediment yield delivered to the foreland basins and to the Bengal fan. The pattern of fluvial incision of Fig. 24 predicts a sediment flux of 375 km²/Myr. This may be compared with the $1 \pm 0.5 \times 10^6$ m³/year eroded from the Himalaya over the last 2 Myr and deposited in the Gangetic plain and Bengal fan (Métilvier et al., 1999). Given that the Ganges and Brahmaputra drain the Himalaya over an approximately 1800 km long stretch of the Himalaya, this value implies about $280 \pm 110$ km²/Myr of denudation on average across a section of the range. Although it is probable that river incision and denudation of hillslopes have varied due to climate changes (Burbank et al., 1996; Goodbred and Kuehl, 2000), over the long term the pattern of denudation must resemble the curve proposed in Fig. 24. We conclude that, to the first order, denudation of the whole landscape matches the pace of fluvial downcutting, and that fluxes have not varied dramatically over the Quaternary period.
5.3. Holocene Kinematics of Overthrusting along the MFT–MHT

The average slip rate on the MFT over the Holocene thus appears to be very close to the total shortening rate across the range as measured from GPS. It also compares well with geological estimates of the shortening rate across the Lesser Himalaya of Nepal, which fall in the range of 16–25 mm/year (Hauck et al., 1998; DeCelles et al., 2001) (point 2 in Fig. 25, top). This estimate also compares well with the Quaternary rate of shortening across the Himalaya deduced from E–W extension in southern Tibet. If this Quaternary extension is assumed to accommodate the lateral variation of the direction of thrusting along the arcuate Himalayan front (Baranowski et al., 1984; Armijo et al., 1986; McCaffrey and Nabelek, 1998), the estimated extension rate of about 10 mm/year implies a Late Quaternary shortening rate of $20 \pm 10$ mm/year (Armijo et al., 1986) (point 3 in Fig. 25, top).

The observed pattern of erosion is found to closely mimic uplift as predicted by slip along the flat–ramp–flat geometry of the MHT. The theoretical pattern of uplift might be simply estimated by assuming that the hanging wall is thrust along the MHT and only accommodates the MHT geometry by vertical shear (Molnar, 1987). If the topography is approximately steady state, as argued in the previous section, the flexural loading of the Indian plate does not vary, so that this model does not require any computation of the flexural response of the lithosphere. This kinematic approach predicts an uplift pattern that fits reasonably with the fluvial-incision pattern of Fig. 24 (Lavé and Avouac, 2001). In Fig. 25 the observed incision pattern is compared with the patterns of tectonic uplift and erosion computed from the 2-D mechanical model described in the next section. It thus seems that over the Holocene period, all the shortening across the Himalaya has been absorbed by slip along one single major thrust fault. It should be noticed at this point that these kinematics, which represent an average over the Holocene period, differ from those expected from the critical brittle wedge model.

5.4. Mechanical Model of Overthrusting: Evidence for Low Friction on the Décollement and Coupling between Denudation and Uplift

Figure 25 shows the results of a mechanical model of long-term overthrusting, designed to fit the average kinematics of the deformation over the Holocene period (Cattin and Avouac, 2000). More details about the
Fig. 25. Velocity, horizontal, and vertical displacement rates (Cattin and Avouac, 2000).
(a) Velocity field relative to India computed from a finite element model which accounts for
the dependency of rheology on local temperature and for erosion–sedimentation at the surface.
The model assumes frictional sliding along the MHT across the brittle upper crust.
(b) Comparison between computed uplift and erosion profiles, and measured river incision
profile of Fig. 24. (c) Horizontal velocity relative to India of (1) sub-Himalaya as derived from
the slip rate on the MFT, (2) the high Himalaya as derived from the rate of sediment
progradation in the foreland and balanced cross-section, and (3) southern Tibet derived from
the E–W extension rate of southern Tibet. These data indicate that deformation of the hanging
wall of the MHT is negligible. The model (continuous line) fits the pattern providing that
friction on the flat portion of the MHT does not exceed 0.2, or 0.3 if a high pore fluid pressure is
assumed.
model are given in Appendix A. This model assumes a prescribed geometry for the MFT, with its characteristic ramps and flats geometry, along which quasistatic frictional sliding is allowed. Elsewhere, the medium is assumed to deform according to a combination of brittle failure and thermally activated ductile flow. Deformation thus depends on the assumed rheology and local temperature. Various temperature fields and rheological laws have been tested and yielded similar results, showing that the long-term pattern of deformation does not provide critical information about these model parameters. The model considers a section of the lithosphere subjected to 20 mm/year of horizontal shortening. According to this model the MFT–MHT roots in a subhorizontal shear zone that fits the prominent mid-crustal reflector imaged from the INDEPTH profiles (Fig. 26). This is not a surprise, however. It is a result of the position of the downdip end of the prescribed fault’s geometry and of the thermal structure that has been obtained assuming \textit{a priori} that the fault continues northwards as a subhorizontal shear zone. The modeling should therefore be seen primarily as a test of the internal consistency of the various data sets and hypotheses used to constrained the model's geometry and thermal structure.

The mechanical model does provide important additional insights. First, it turns out that the friction along the MFT–MHT must be low. Indeed, the observation that the overhanging wall does not shorten during overthrusting requires that the friction at the base of the thrust sheet induces deviatoric stresses that are balanced by the slope of the topography. In other words, the slope of the wedge must be overcritical. If the friction was too high, the wedge would be undercritical and the hanging wall would deform, so that there would be no slip on the MFT at the surface until the wedge topography reached its critical value. If the possible effect of pore pressure is neglected, we found that the friction on the flat portion of the MFT must be smaller than about 0.13. A classical interpretation of low basal friction in such an overthrusting context invokes fluid pressurization in the fault zone (Hubbert and Rubey, 1959). Even if a high-pore pressure is assumed, up to 0.9 the lithostatic pressure, we calculate that the friction on the flat portion of the MFT must be lower than about 0.3. This value is significantly smaller than the standard value for intact rocks of about 0.6 (Byerlee, 1978), but is in the range of empirical values inferred for some intracontinental weak faults or along subduction zone. By contrast, the friction on the ramp could be up to about 0.6, because friction at the base of the thrust sheet is balanced there by the steep slope of the front of the high range.

Second, the model leads to some natural balance between denudation and uplift rate. This balance is expressed in Fig. 25, in which it can be
seen that denudation rate, modeled from a simple linear diffusion equation, approximately equals the uplift rate. It should be recalled that dynamic equilibrium of the topography requires, in fact, that erosion rate, $e$, balances local uplift rate as well as the apparent uplift due to the advancing topography, as expressed from Eq. (5). In the case where incision rate, $I$, is the observed quantity, the slope of the topography, $tga$, should be replaced by the stream gradient, $S$, so that Eq. (5) becomes

$$ I = U + V_0 S. $$

(21)

This equation should be used to assess dynamic equilibrium of the topography and might be used conversely to relate tectonic uplift to denudation when a steady-state topography is assumed a priori. From this approach, the incision rates along the major rivers across the Himalaya of

Fig. 26. Shear–strain rates over the long term computed by assuming a quartz rheology (a) or a diabase rheology (b) (modified from Cattin and Avouac, 2000). In both cases a subhorizontal shear zone develops that closely follows the midcrustal reflector imaged by the INDEPTH experiment (Zhao et al., 1993).
central Nepal were shown to yield approximately the same pattern of
tectonic uplift, and this pattern appears to match the prediction of the
mechanical model of Fig. 25 (Lavé and Avouac, 2001). Sensitivity test shows
that the midcrustal ramp is required for the model to match the pattern of
uplift obtained from Eq. (21) (Cattin and Avouac, 2000; Lavé and Avouac,
2001). The dynamic equilibrium between denudation and tectonic uplift
arises naturally from the coupling between surface processes and ductile
flow at depth and contributes to localizing the deformation (Avouac and
Burov, 1996).

5.5. Do the Kinematics of Deformation Determined over the Holocene
Period Explain the Morphology and Structure of the Range?

The morphology of the Himalaya of central Nepal reflects a dynamic
equilibrium between present tectonics and surface processes. Active
tectonics in the Himalaya are primarily controlled by localized slip along
the MHT and the topography is close to steady state because denudation,
driven by fluvial downcutting, balances tectonic uplift. This scenario implies
that the present morphology of the front of the high range is also the result
of present tectonics and erosion. The sharp relief together with the high
uplift rates in the High Himalaya is interpreted here to reflect thrusting over
the midcrustal ramp rather than an isostatic response to reincision of the
Tibetan Plateau driven by late Cenozoic climate change (e.g. Burbank,
1992), or a Late Miocene reactivation of the MCT (Harrison et al., 1997).
Similarly, it is not necessary to invoke deformation near the front of the
high range (Bilham et al., 1997). Such a scenario is improbable because it
would imply shortening beyond the 21 mm/year absorbed at the MFT, and
the total shortening rate across the range would therefore exceed the value
obtained from GPS measurements.

Now, we may test whether the kinematics of deformation established for
the Holocene period may be extrapolated back in time. This kinematics does
not involve any accretion and cannot account for the fact that the
Himalayan wedge was built through accretion of several thrust sheets
detached from the Indian plate. In addition, it predicts a diachronous
exhumation corresponding to $V_{HR} = V_0 \sim 21$ mm/year instead of the
5 mm/year determined from Fig. 15. Finally, the position of the bulge
would be fixed with respect to the Indian plate and $V_{pr} \sim 0$ mm/year, so
that there should be no sediment progradation on the Indian plate nor
subsidence (for a steady-state geometry of the flexed Indian plate). Clearly this kinematics cannot reflect the long-term process. The simplest explanation is that accretion occurs due to migration of the midcrustal ramp, as described in Section 3.6 and Fig. 17, but that over the Holocene period no episode of accretion has occurred in central Nepal. So although the pattern of exhumation ages in Fig. 15 shows no evidence for this, the accretion process is most probably discontinuous. An alternative explanation would be that the midcrustal ramp does not exist. The zone of high uplift rates north of the Lesser Himalaya could then be interpreted as due to underplating. It would then indicate the position of the zone of accretion of Fig. 17. This explanation seems difficult to reconcile with the structural section and, although the existence of the midcrustal ramp might appear conjectural, the first interpretation is favored.

6. The Seismic Cycle in the Nepal Himalaya

6.1. Large Historical Earthquakes in the Nepal Himalaya

The Himalaya has produced four earthquakes with magnitude larger than eight since the end of the 19th century. Table 1 lists plausible values for the parameters of these events. Considerable uncertainty remains as to the exact location and size of the 1934 Bihar–Nepal earthquake, the most recent large historical event in the study area (e.g., Chen and Molnar, 1977; Pandey and Molnar, 1988). The epicenter was probably located in the Lesser Himalaya east of Katmandu (Chen and Molnar, 1977). Assuming a thrust-fault dipping by about $5^\circ$ to the north, long-period seismograms indicate a released seismic moment of about $4.1 \times 10^{21}$ N m, corresponding Mw~$8.4$ (Molnar and Deng, 1984). Macroseismic intensities and subsidence of the foreland revealed from leveling data suggest that the earthquake ruptured a 250–300 km along-strike segment of the arc (Bilham et al., 1998) (Fig. 3). The rupture area may have extended up to the MFT but probably not farther to the south (Chander, 1989). The northward extent of the rupture is not constrained at all.

There are historical indications that Katmandu has been struck by repeated severe earthquakes in the past (Rana, 1935; Pant, 2002). Major destructions were reported in 1833, 1681, 1408, 1344, 1255, and 1223. The 1833 event might have ruptured about the same arc segment as the 1934
quake (Bilham, 1995), but the extent of damage in Katmandu suggests a slightly smaller or more distant event. The 1255 was a major disaster that caused the death of King Abhaga Molla and killed about one-third of the population of Katmandu valley (Pant, 2002). Very little is known about the 1344 event, which may have caused the death of King Ari Molla (Pant, 2002). When assessing these historical earthquakes it should be pointed that some palaces and temples in Katmandu valley built after the mid-15th century remained intact until the 1934 event. These constructions are characterized by deep foundations (up to 15 m deep) and were reinforced by wooden frames so that they could resist significant shaking. The temple of Shiva in Bakthapur was such a building, dated to 1458 AD, which was totally destroyed in 1934 (Pandey and Molnar, 1988). Some other temples built by the end of the 14th century or in the early 15th century were preserved until they were destroyed in 1934. The 1681 and 1408 events are therefore probably smaller than the 1934 earthquake. On this basis, the return period of earthquakes similar to or larger than the 1934 event may fall between 100 and 475 years.

Among the large Himalayan earthquakes of the last century, the 1905 Kangra event was probably the most similar to the 1934 event (Fig. 3). It ruptured a fault plane whose structural position was probably similar (Molnar, 1987; Chander, 1989). The analysis of the seismogram has lead to a wide range of estimates corresponding to a surface wave magnitude between 7.5 and 8.2, with the lower estimates being more probable (Ambraseys and Bilham, 2000). The area between the locations of the 1934 and 1905 events represents a ∼800-km long seismic gap. The most recent event that ruptured its eastern portion is probably the 1833 event. A large magnitude event along its western portion was reported in 1803 (Oldham, 1883). Although interpretation of the historical data is difficult, due to the political situation in the area at that time (Bilham et al., 1995), it may have been a very large earthquake with magnitude around 8 that ruptured an arc segment between Dehra Dun and far-western Nepal (Bollinger, 2002).

6.2. Microseismic Activity and Geodetic Deformation in the Interseismic Period

Geodetic measurements over the last 10 years (Bilham et al., 1997; Jouanne et al., 1999; Larson et al., 1999) have revealed that, all along the Himalayan front in Nepal, relative displacements between the Gangetic
plain and the Lesser Himalaya MFT (MHT) have been small. This is particularly clear from the 10-point GPS network emplaced along section AA’ (Fig. 27). The results indicate horizontal contraction perpendicular to the strike of the range and essentially confined to a 50-km wide zone at the front of the High Himalaya, some 100 km north of the MFT. Velocities, relative to northern India, rise from about 3 mm/year around Katmandu basin to about 15-mm/year 50 km to the north. Data collected in southern Tibet some 150 km east of the study area (Bilham et al., 1997; Larson et al., 1999) suggest that velocities taper to about 20 mm/year north of the High Himalaya (Fig. 28).

Local seismic monitoring in the Katmandu area (Pandey et al., 1995, 1999) has also revealed intense microseismic activity in this area, with two M1~6 events over the last 5 years (Fig. 27). Most events clustered in a narrow zone beneath the front of the high range at depths between
30 and 5 km (Fig. 28). This zone also coincides with high uplift rates evidenced from the leveling measurements (Jackson and Bilham, 1994) (Fig. 28). A casual observation is that the seismicity belt abruptly ends to the north as soon as the elevation exceeds about 3500 m (Fig. 27). At places the seismicity cut-off can be seen to follow quite closely the sinuous shape of this elevation contour line, suggesting some cut-off effect due to the topography (Bollinger et al., submitted-b).

These data suggest that over the last few years the MFT–MHT has remained locked from where it emerges along the foothills to the midcrustal ramp, while aseismic deformation has proceeded beneath the High Himalayas and southern Tibet. This process would have resulted in stress buildup around the downdip end of the locked fault zone, triggering seismic activity.

The mechanical consistency of the simple model described above was tested by the calculations described in Appendix A (Cattin and Avouac, 2000). If the friction on the fault is increased to a value high enough (about 0.17), frictional sliding along the fault is inhibited while ductile shear extends along the subhorizontal shear zone beneath the high range and southern Tibet. The model then predicts horizontal displacement and uplift rates consistent with the geodetic measurements (Fig. 28), as well as stress accumulation near the northern edge of the locked fault. It turns out that most of the micro-earthquakes fall within this area of enhanced Coulomb stress (Cattin and Avouac, 2000). A simple explanation for the seismicity cut-off corresponding to a particular elevation (3500 m) is that the vertical stress increase induces a change of the stress tensor such that the deviatoric stresses are no longer affected, or even decrease during interseismic strain buildup. The seismicity releases an accumulated moment that amounts to less than 1% of the upper crustal strain revealed from geodetic monitoring (Avouac et al., 2001). These earthquakes thus provide a strain release that is insignificant compared to crustal deformation over the long term. The measured geodetic deformation is therefore either permanent and aseismic, or elastic. The second hypothesis should be preferred because, as discussed in the previous section, long-term shortening across the range is entirely accommodated by localized slip on the MFT–MHT. The deformation measured over the last few years across the Katmandu section is necessarily due to elastic straining of the upper crust that has to be ultimately released by slip on the MHT. This stress transfer is probably chiefly the result of recurring major earthquakes such as the Bihar–Nepal earthquake.
FIG. 28. (a) Structural section with seismicity located from three three-component seismic stations installed in 1996 near section AA'. (b) Uplift rates derived from the measurements along the leveling line in Fig. 2b (Jackson and Bilham, 1994) and computed from the mechanical model of Cattin and Avouac (2000) (continuous line). (c) Horizontal velocities
6.3. Lateral Variations of Interseismic Straining

Interseismic stress buildup by elastic straining of the upper crust is thus probably the main process responsible for the observed belt of microseismicity that can be traced along the front of the high range all along the Himalayas of Nepal, including the “seismic gap” west of Katmandu. Geodetic measurements also provide clear evidence that the MFT is essentially locked all along the front of the Himalaya of Nepal (Jouanne et al., 1999; Larson et al., 1999; Bilham et al., 2001) and of northwest India (Banerjee and Burgmann, 2002). It seems highly probable that this portion of the Himalayan arc also produces large recurrent earthquakes similar to the 1934 and 1905 events.

There may however be some along-strike variations of elastic straining and stress buildup. These variations could result from lateral variations in the geometry of the fault zone, or from seismic coupling with possible nonstationary loading during the interseismic period. Such variations have been observed along subduction zones possibly reflecting stress buildup around major asperities (Dmowska et al., 1996). In the case of an intracontinental megathrust, such as along the Himalaya, heterogeneous interseismic straining might be more easily detected than along a subduction zone because of the intracontinental setting and shallow dip of the fault, and also because interseismic straining is thought to trigger the seismicity in the belt along the front of the high range. The seismic activity over the last few decades along the Himalayan arc does show some lateral variations. At places along the Himalayan clusters of arc shallow-dipping thrust events with relatively large magnitudes, up to Ms 6.5–7, have been detected (Baranowski et al., 1984). These events falling along the seismicity belt at front of the high range can also be interpreted to result from interseismic stress buildup. The most recent such events are the 1991 Uttar Kashi (Ms~7.2) earthquake and the 1999 Chamoli (Ms~7.3) earthquake. Quite a few slightly smaller events have occurred in far-western Nepal between about 81.5 and 82°E (Fig. 27). These clusters may correspond to some lateral geometric complexities of the Himalayan arc or alternatively reflect relative to India determined from GPS measurements and computed from the mechanical model of Cattin and Avouac (2000) (continuous line). Black dots indicate LDG measurements along section AA’. Other measurements (Jouanne et al., 1999; Larson et al., 1999) were projected on section AA’ with account for the arc curvature. To account for possible lateral irregularities the location of the sites not located close to section AA’ are assigned a standard uncertainty of 20 km.
different stage in the seismic cycle. In fact, 3-D modeling shows that the pattern of seismicity and the geodetic data can be fitted by a very simple model in which the fault is locked everywhere from the MFT at the surface to beneath the front of the high range, over a distance of about 100 km, without any significant lateral variation in the rate or direction of shortening (Bollinger, 2002). Variations in the pattern of interseismic straining are therefore probably rather subtle. In particular, there is no evidence for heterogeneous straining such as might be expected if asperities were building up in the interseismic period. Lateral variations in the pattern of seismicity thus most probably reflect variations of the stress field induced by the effect of topography or by the effect of past large earthquakes.

6.4. A Model of the Seismic Cycle in the Central Nepal Himalaya

Motion along the brittle portion of the MHT is thus probably stick-slip as a result of recurring large earthquakes (Fig. 29). The elastic stress accumulated in the upper crust during the interseismic period must be transferred to slip on the MFT. Some of the largest events along the front of the high range probably do participate in this stress transfer and activate a portion of the MFT–MHT. This may be argued, for example, for the Uttar Kashi event based on the analysis of the accelerometric data (Cotton et al., 1996). The author’s interpretation of such M ≤ 7.5 events is that they did not propagate all the way to the Himalayan piedmont probably because elastic stresses in the surrounding elastic medium were too low to sustain the propagation of the seismic rupture. It might then be conjectured that the magnitude 7 events contribute to loading of the shallow-dipping décollement below the Lesser Himalaya, raising elastic stresses there until one event ruptures all the way to the earth surface. Such M~8–8.5 events would then account for most of the stress transfer to the front of the thrust-fault system (Fig. 28). If so, the average return period of large earthquakes on the MHT in the study area could also be estimated from a comparison of slip rate on the MHT and coseismic slip. For an average coseismic slip between 4 and 6 m, a value that has also commonly been estimated for other large Himalayan earthquakes (Molnar, 1990), and considering the slip rate on the MFT as well as the shortening rate across the range, it should take at least 180–330 years to accumulate the elastic strain released during such an event. Given that preseismic, postseismic, or possibly aseismic creep events may also contribute to some fraction of the
FIG. 29. Model of the seismic cycle in the Nepal Himalaya.
fault slip, the average return period of such an event might actually be larger.

7. DISCUSSION

7.1. Some Implications of the Himalayan Case for the Mechanics of Mountain Building

The pattern of river incision and the deformation of abandoned river terraces has provided some key constraints on the Holocene kinematics of deformation across the Nepal Himalaya. It turns out that crustal shortening has essentially resulted from overthrusting of the Himalaya over the MFT–MHT thrust system with little internal deformation. During that process erosion seems to have maintained a balance between denudation and tectonic uplift, probably due to coupling between surface processes and crustal flow at depth. The kinematics of deformation over a longer geological time period require in addition an accretion of material, probably as the result of underplating near the brittle–ductile transition, a process documented elsewhere from the study of exhumed thrust system (Dunlap et al., 1997). The structure of the range suggests that underplating has resulted from southward migration of a midcrustal ramp along the thrust-fault system. So, contrary to previous conceptions, the Himalayan wedge has not grown by frontal accretion due to a forward propagation of the thrust sequence. It does not owe its critical wedge geometry to distributed brittle thickening of the whole crust but to a combination of brittle–ductile underplating and erosion at the surface. For a more general discussion of the impact of underplating on the dynamics of orogenic wedges, the reader might refer to Platt (1986). The thermal structure is a key factor in determining these kinematics, because of its influence on the rheology of the crust. The distribution of erosion is another fundamental factor due to its influence on the thermal structure and its role in mass redistribution at the surface. Any analysis of the mechanics of mountain building must take these factors into account.

7.2. Some Limitations of the Model

The mechanical modeling described here is based on a number of very simple assumptions: erosion is modeled from linear diffusion; the fault geometry is prescribed; fault motion is modeled from static stable sliding;
the thermal structure is assumed \textit{a priori}; boundary conditions are expressed in terms of prescribed displacements; the mantle below the lithosphere is assumed to simply maintain some hydrostatic support. These assumptions place significant limitations on our ability to answer a number of questions:

- Because the fault geometry is imposed, the model shown here does not provide any particular insight as to the mechanics of accretion over the long term. In that respect, future investigations should be based on more elaborated thermomechanical modeling allowing for large strains, as well as strain localization.
- As a mountain grows, the increasing crustal thickness modifies the stress field. If far-field tectonic boundary forces are assumed to be constants, the change in thickness should modify the force balance and thus influence strain rates. This effect is not included here. So the model might be appropriate to simulate a steady-state regime, but cannot be used to assess the long-term evolution of a mountain range.
- The model does a good job in predicting sediment yields that are comparable to observations. However, it leads to a steady-state parabolic topography that bears little resemblance to the real topography of the Himalaya. This is due to the linear diffusion equation used to simulate erosion. More insight on the problem of the evolution, and eventual dynamic equilibrium, of the topography would require more sophisticated models of hillslope erosion, river transport, and downcutting, with proper account of the coupling between these processes. Such models would be required to get some idea of the transient response of the morphology of a mountain belt, and of the sediment discharge to the lowlands, to possible changes in climate conditions or tectonic boundary forces.
- The present model does not adequately describe a possible time dependence of interseismic deformation, since the rupture dynamics and the stress transfers associated with the seismic cycle are not taken into account.

7.3. Evidence for Low Friction on the MFT–MHT

The observation that the footwall does not experience any significant shortening during thrusting along the shallow dipping MFT–MHT implies a low friction of the order of 0.1, or 0.3 at most if a pore pressure is assumed. Another constraint on the friction along that fault zone comes from the observation that microseismic activity along the front of the
high range shows a cut-off at an elevation of 3500 m. This threshold effect probably corresponds to the point where Coulomb stresses no longer increase, or actually decrease, during the interseismic period. This is achieved if the NS deviatoric stress is of the order of magnitude of the vertical stress variation due to the 3500-m difference in elevation with the lowland. Unfortunately, estimating the effect of the topography on the stress field at depth is not simple due to isostatic balance and elastic support. To the first order, the 3500-m elevation difference should correspond to a $\sigma_{zz}$ increase of about 90 MPa. Given that the elevation in the Lesser Himlaya is of the order of 1000 m, we infer deviatoric stresses of at most 70 MPa, hence shear stresses on the underlying décollement of about 35 MPa at most. This is in keeping with a friction of the order of 0.1, given the depth of the décollement around 10 km. Such a low friction could result from highpore fluid pressure. This seems a reasonable scenario if some foreland sediments are dragged along the interface, as the MT data might suggest. Fluid pressurization would then result from sediment compaction during underthrusting. Dynamic friction effects might also play a role. If the thrust sheet only moves during large seismic events, the shear stress on the décollement may, in fact, always remain much lower than the value required to balance static friction. In that case the apparent low friction would actually represent dynamic friction during seismic rupture.

7.4. High Midcrustal Conductivity: Evidence for Metamorphism Dehydration?

Here the author briefly comes back to the interpretation of the electrical conductivity of Fig. 13. The comparison with the thermal structure (Fig. 12) is instructive. Because the midcrustal zone of high conductivity does not follow any particular isotherm, in particular since it does not extend to the north, a thermometamorphic control model (e.g., Marquis and Hyndman, 1992) can be excluded. This conductive body, which becomes prominent where the underthrusting footwall reach temperatures of the order of 350°C, may instead result from the presence of fluids released by metamorphic reactions, possibly chlorite breakdown. The released fluids would percolate upward through the zone where intense microseismic activity is triggered by interseismic straining (Cattin and Avouac, 2000), resulting in pathways for percolation through the brittle portion of the crust.
7.5. What Controls the Downdip End of the Locked Portion of the Fault?

The geodetic and seismic data provide some information on the position of the downdip end of the locked portion of the fault. The mechanical model of interseismic straining shown in Fig. 28 produces a smooth transition from a fully locked fault zone to ductile flow along its northward continuation, where temperatures exceed about 450°C. Close inspection of the velocity field reveals that a portion of the fault, near the base of the prescribed fault geometry, undergoes stable sliding. In fact the exact location of the transition cannot be constrained tightly from the GPS campaign data, and there is a variety of possible slip-rate distributions—including the one produced from the mechanical modeling—that would fit the data equally well.

To assess more accurately the position of the downdip end of the locked fault, and to avoid possible trade-offs among model parameters, we have adopted a simplified kinematic approach (Durette, 2002). The model assumes that stable sliding and ductile flow along the downdip continuation of the locked fault can be modeled using a slipping dislocation embedded in an elastic medium, an approach that has been adopted in some previous studies (Jackson and Bilham, 1994; Bilham et al., 1997). For that analysis, we considered only the GPS data, corresponding to the arc segment that probably broke in 1934, i.e., east of the Katmandu basin. We also included velocities determined at four continuous GPS stations along that section and the IGS station in Lhasa (Fig. 30). The model parameters are: the dip angle of the dislocation, the position of the updip end of the dislocation (corresponding to the downdip end of the locked fault zone). The best-fitting solution indicates a slip rate of $18.5 \pm 1.5$ mm/year on a shallow-dipping fault. The position of the updip end of the dislocation is relatively well constrained and, when compared to the thermal structure of Fig. 12, falls in a range of temperature between 300 and 400°C (although large, this range of values ignores the uncertainty of the thermal structure). Such a temperature range could correspond to the transition from slip-weakening friction to aseismic stable sliding. Indeed, according to laboratory experiments and/or field observations, this thermally activated transition occurs at a temperature around 325–350°C for quartzo-feldspathic rocks (Blanpied et al., 1991, 1995). This is also in keeping with the observation that the downdip extent of the seismogenic zone along subduction zones generally coincides with the 350°C isotherm if this temperature is reached above the Moho (Oleskevich et al., 1999).
FIG. 30. Geodetic measurements of horizontal uplift rates (a) and displacement rates (b) across the Himalaya of Nepal (Jackson and Bilham, 1994; Jouanne et al., 1999). Also shown are the velocities determined at four continuous GPS stations along section AA' that have been operated since 1977 in collaboration with the Department of Mines and Geology of Nepal, and at the IGS station in Lhasa (Courtesy of M. Flouzat and collaborators at LDG). These data were processed using Bernese and Addneq softwares and referenced to ITRF 97 using seven IGS stations (Durette, 2002). Continuous lines were obtained from least-squares adjustment.
So, the rationale proposed for subduction zone (Hyndman et al., 1997), for cases where the locked fault zone does not extend deeper than the forearc Moho, seems to hold also for intracontinental megathrust faults. The coincidence between the transition to stable sliding and the zone of higher conductivity images from the magnetotelluric sounding experiment may also suggest that the fluids and possible mineral changes caused by metamorphic reactions may also influence the position of the downdip end of the locked fault portion.

7.6. Is Interseismic Straining a Stationary or Nonstationary Process?

The results described in Section 6 suggest that interseismic straining measured over the last decade is representative of the average pattern and rate of straining during the interseismic period. The consistency, to better than about 10%, between long-term slip rate on the MHT and interseismic shortening rate across the range implies that interseismic straining is essentially constant, except possibly during relatively short periods of time compared to the recurrence interval of large earthquakes. This is also suggested by the fact that there is no resolvable difference in the pattern of interseismic straining measured east of Katmandu, which presumably broke during the 1934 event, and that measured in western Nepal, which has not broken for at least two centuries. Given the uncertainties in geodetic measurements and in the long-term slip rate, we cannot exclude the possibility that there are lateral variations that could reflect variations in the interseismic straining pattern and rate during the seismic cycle. If so, this information might provide important constraint on the rheology of the crust.

Indeed, simple consideration of the mechanics of the lithosphere imply that there should be some time variations in crustal straining in the period between two large recurring earthquakes. To illustrate this point some results obtained from a simple 1-D model of the seismic cycle is shown from a 2-D dislocation model (Durette, 2002). The model is based on a 2-D analytical approximation of surface deformation (Singh and Rani, 1993). Both GPS and leveling data were weighted according to their standard deviation, assuming a Gaussian distribution of errors. (c) Geometry of the creeping dislocation (continuous line) with 2σ uncertainty in the position of its end. The best-fitting solution corresponds to a slip rate of 18.5 ± 1.5 mm/year and a dip angle of 1 ± 0.5°. The depth of the updip end of the dislocation is estimated at 15 km, with a large 2σ confidence interval of 8–25 km. The horizontal distance from the MFT is estimated between 90 and 120 km. The reduced \( \chi^2 \) min is 4.
Perfettini and Avouac, manuscript in preparation) (Figs. 31 and 32). In this simple model, the behavior of the thrust-fault system is simplified and modeled as a simple springs and sliders system (Fig. 31). Three main domains with different rheologies are considered, and the springs allow stress transfer between them during the seismic cycle. The upper fault portion is assumed to undergo unstable frictional sliding and is modeled by slider 1 with a state-and-rate friction law (Dieterich, 1979; Rice and Tse, 1986). Its behavior is obtained by solving the following equations:

\[ \tau_1 = \sigma_1 \mu_1 \left\{ a_1 \log(V/V_\ast) + b_1 \log(\theta V_\ast/D_c) \right\} \]  \hspace{1cm} (22)

\[ \frac{d\theta}{dt} = 1 - (\theta V/D_c) \]  \hspace{1cm} (23)

FIG. 31. Simplified spring-and-slider model. This 1-D model is meant to illustrate the effect of stress transfer between the seismogenic fault portion, which is assumed to show stick-slip behavior (slider 1), the ductile shear zone at depth, which is modeled from a Newtonian viscosity (slider 3), and a transition zone where increasing ductility is assumed to promote stable sliding (slider 2). A constant force is applied to slider 3 so that the long-term velocity is 21 mm/year. The force corresponds to \( F = 96 \text{ MPa} \) for \( \eta = 10^{-20} \text{ Pa s} \), and to \( F = 84 \text{ MPa} \) for \( \eta = 10^{-19} \text{ Pa s} \). The ductile shear zone is assumed to have a thickness of 5 km. Other model parameters are: \( a_1 = 0.04, b_1 = 0.005, a_2 = 0.02, \mu_\ast = 0.1, \sigma_2 = 2\sigma_1, \sigma_1 = 270 \text{ MPa}, D_c = 10^{-2} \text{ m}, V_\ast = 10^{-10} \text{ m/s}, h_1 = 10 \text{ km}, h_2 = 2h_1, G = 30 \text{ GPa} \). These values were chosen so as to imply about 5 m of coseismic slip. Brittle creep parameters were assumed to correspond to the values obtained from the adjustment of postseismic relaxation following the Chichi earthquake in Taiwan (Perfettini and Avouac, in press).
Fig. 32. Slip motion of sliders 1, 2, and 3 for a Maxwell relaxation time of 211 (a) and 210 years (b).
where $\tau_1$ is the shear stress; $\sigma_1$ is the effective normal stress set to some arbitrary value of 250 MPa (corresponding to a depth of about 10 km); $\mu^*$ is the steady-state friction coefficient corresponding to some reference-sliding velocity $V^*$; $a_1, b_1,$ and $D_c$ are frictional parameters that were chosen in order to obtain a coseismic slip of the order of 5 m.

The deeper fault portion is assumed to undergo rate-strengthening brittle creep. This process is modeled from a rate-strengthening friction law associated with slider 2,

$$\tau_1 = \sigma_2\{\mu^* + a_2 \log(V/V^*)\}$$  \hspace{1cm} (24)

where $a_2 > 0$ to ensure stable sliding.

The fault roots at depth in a ductile shear zone with viscosity $\eta$, assumed to be about $w = 5$ km thick, which is modeled from a viscous slider,

$$\tau_3 = \eta^* V_3,$$ \hspace{1cm} (25)

where $\eta^* = \eta/w$.

Motion of the most frontal slider, slider 1, would represent the slip along the seismogenic fault portion. Motion of slider 3 represents the convergence across the fault system. To solve for the motion of the system with proper accounting for stress transfer during the seismic cycle, we apply a constant applied force, $F$, on slider 3. The force was adjusted in order to produce an average slip rate of 20 mm/year.

For simplicity we assume that the fault segment with brittle creep is two times longer than the seismogenic zone, and lies at a depth about twice as great. The elastic modulus is assumed to be constant within the crust. These simple assumptions imply that the stiffness of the second spring should be about four times smaller than that of the first one so that we consider,

$$k_1 = 4k_2 = k$$ \hspace{1cm} (26)

The effective normal stress $\sigma_1$ is set to an arbitrary value of 250 MPa, corresponding to 10 km depth, and $\sigma_1$ is taken to be twice as large.

The behavior of the system is then essentially determined from:

- the coseismic slip: $\Delta U$;
- the duration of the interseismic period: $\Delta t = \Delta U/V_0$;
- the Maxwell relaxation time defined here as: $t_M = k/\eta^*$;
- the brittle relaxation time defined here as: $t_R = a_2\sigma_2/k V_0$. 
The model parameters in the simulations of Fig. 32 were adjusted so that the model always predicts a coseismic slip, $\Delta U$, of the order of 5 m, and the frictional parameters were not varied. The brittle creep law associated with the second slider was chosen arbitrarily so as to correspond to a 8-year relaxation time, as deduced from postseismic deformation in Taiwan (Perfettini and Avouac, in press). Fig. 32 shows the slip history of all three sliders for two different viscosities that correspond to a Maxwell relaxation time between 200 and 20 years.

Slider 1 has stick-slip behavior with 5 m of slip every 250 years. Slider 2 produces some afterslip that decays rapidly over the first few years of postseismic relaxation. Later on its motion is controlled by the viscous slip of slider 3. If the viscosity is high, i.e., if the Maxwell time is of the order of the duration of the interseismic period, slider 3 has a nearly uniform motion (Fig. 32a). If the Maxwell time is much shorter than the duration of the interseismic period, then some significant variations are observed.

This simple 1-D periodic model is meant to illustrate the effect of stress transfers during the seismic cycle. In reality the fault slip and interseismic interval vary from one event to another, so that the real behavior is not perfectly cyclic. The stress-transfer mechanism assumes a fault with infinite horizontal extent. In reality the stress transfer occurs in 3-D and coseismic stress drop due to the rupture of a particular arc segment would be compensated by increased stresses on adjacent segment and elastic response of the whole surrounding medium (not only of the thrust sheet as modeled here).

The virtue of this model is essentially to illustrate that time variations of straining might be expected and, if detected, would provide unique constraints on the rheological properties of the crust.

7.7. Seismic Coupling and Implication for Seismic Hazard

Geodetic measurements and background seismicity can be used jointly to delineate the locked portion of the fault, providing important information on the potential location and size of future earthquakes. The major thrust fault along the Himalaya of Nepal appears to be locked from the earth surface, at the MFT, to beneath the front of the high range, a distance of about 100 km. Does this means that this locked portion breaks during major $M \sim 8-8.5$ earthquakes, the recurrence of which might be simply estimated from the ratio of the coseismic slip and long-term slip rate (as in the 1-D
model above)? This would be true if coseismic slip was the only mode of transient slip on the MFT.

This question might be addressed by considering the seismic moment released by the historical earthquakes along the Himalayan arc as a whole. Since 1897, the large Himalayan earthquakes in Table 1 have released a cumulative moment of about $17 \times 10^{21}$ N m. For comparison, the moment released from all the seismic events recorded along the Himalaya over the last 30 years amounts only to about $4 \times 10^{19}$ N m. So the large seismic events account for nearly all, to within a few %, of the seismic moment released over the long term. If we assume that the width of the locked fault zone is everywhere about 100 km and that the 21 mm/year slip rate in central Nepal is a reasonable average for the whole arc, we find that slip along the MFT–MHT should release about $1.5 \times 10^{18}$ N m/year for a 100% seismic coupling. So, the total seismic moment released since 1897 is close to the total amount to be released over a century for a 100% seismic coupling. However, if we now extend the same reasoning to a longer time period, say since 1803, we come up with a very different estimate. If the 1803 and 1833 earthquakes are assumed to have released seismic moments of the order of $2–4 \times 10^{21}$ N m, the large earthquakes could in fact amount to only about 70% of the value obtained assuming a 100% seismic coupling. Such an estimate would imply that the return period of large events like the Bihar–Nepal earthquake should be revised to about 340 years, a value more consistent with the historical record of earthquake damage in the Katmandu basin.

So, by considering the historical earthquake records, seismic slip on the MHT accounts for 70–100% of the geologically determined slip rate on the MHT. It is therefore possible that part of the slip on the MHT could be accommodated by some aseismic slip that might take place during the postseismic phase or by any kind of transient events.

Some insight might be gained from the comparison of the Himalayan context of central Nepal with the western flank of the central range in Taiwan, where the tectonic setting is very similar and where the Mw 7.6 Chichi earthquake occurred in 1999 (e.g., Kao and Chen, 2000). This earthquake broke a shallow-dipping fault along the piedmont of the range. GPS measurements show that this fault was previously fully locked over a distance of about 40 km from the surface to a depth where it also roots in a subhorizontal aseismic shear zone (Dominguez et al., 2003). The earthquake was followed by some afterslip along the downdip extension of the ruptured zone (Hsu et al., 2002), a process that can be modeled from the response of the stable-sliding portion of the fault to the coseismic stress change as
described using the springs-and-sliders system of the previous section (Perfettini and Avouac, in press). In fact, the earthquake ruptured only a fraction of the previously locked fault (Dominguez et al., 2003) with a rather heterogeneous slip distribution (Kao and Chen, 2000; Huang, 2001; Johnson et al., 2001). Although significant, afterslip was not sufficient to smooth out coseismic slip heterogeneities. Also it did produce only a small amount of slip along the portion of the fault that was locked before the earthquake but did not break during the main shock. This fault portion probably slips at a rate around 40 mm/year to transfer deformation from ductile fault portion to the shallow faults that splay upward including the Chelungpu and Changhua faults. It is unclear whether slip along that fault portion is ultimately taken up by seismic events by other modes of transient slip. A dominantly seismic mode of slip is possible, although no historical events would suggest that this fault portion might have produced any large earthquake in the past. The possibility for aseismic slow events in such a context, as has been observed on some occasion along subduction zones (Sacks et al., 1981; Dragert et al., 2001; Lowry et al., 2001) or on the San Andreas fault (Linde et al., 1996), should not be discarded. The point is that since asperities do not build up in the interseismic period, recurring transient slip events (slow events and earthquakes) must sum up so that heterogeneities are smoothed out in the long term.

We infer that on an intracontinental megathrust such as along the Himalaya of Nepal or along the central range of Taiwan, (1) a future earthquake could break only a fraction of the previously locked fault zone, and (2) seismic coupling might be small, possibly of the order of 70%. For seismic hazard assessment, a conservative hypothesis is to assume a 100% seismic coupling. Conversely, determination of crustal deformation based on the seismic moment release is highly uncertain, even in case where the geometry of the locked fault zone is known, due to the uncertainty in seismic coupling.

8. CONCLUSION

The Himalaya is probably a unique place to address a variety of geological problems associated with mountain building processes. Some progress has been made over the last decades due to various multi-disciplinary efforts that have provided important information on the subsurface structure of the range and on the kinematics of deformation over
different timescales. At this point, the available data can be assembled in a relatively consistent way in the light of our current understanding of the mechanics of the lithosphere and of the seismic cycle. However, most current models in seismotectonics or morphotectonics are based on rather “static” concepts: interseismic straining is considered uniform; the topography is assumed to be steady state due to a balance between erosion and tectonic uplift; denudation of hillslopes is assumed to equal river incision; fault zones geometry do not vary with time; and so on. We should be able to refine our understanding of the dynamics underlying these processes. It is believed that major advances will come from better observational constraints as well as improved physical models of the evolution over time of these various processes.

ACKNOWLEDGMENTS

My understanding of the Himalayan orogen has greatly benefited from interactions and discussions with many former students, postdoc and colleagues in my group (Jérome Lavé, Laurent Bollinger, Rodolphe Cattin, Hugo Perfettini, S. Dominguez, Mireille Fouzat, Thierry Hérilit, Frédéric Perrier) as well as with my Nepali colleagues, M. R. Pandey and his collaborators from the Department of Mines and Geology.

APPENDIX A. MECHANICAL MODELING

The mechanical simulation of deformation during the interseismic period or due to long-term overthrusting along the MFT–MHT, shown in Figs. 25, 26, and 28, was computed from a 2-D finite element code (Hassani et al., 1997) initially designed to account for the mechanical layering of the crust. The code was modified to incorporate surface processes and the dependency of rheology on local temperature (Cattin and Avouac, 2000).

A.1. Boundary Conditions

The model considers a 700-km long section (Fig. A1) that approximates the N18°E section through the Himalaya of central Nepal of Fig. 10. The model is initially loaded with the present average topography along the swath considered here. Only the brittle portion of the MHT is prescribed. It includes the flat beneath the Lesser Himalaya and the midcrustal ramp at the front of the High Himalaya. Boundary conditions mimic an Indian
mantle lid that underthrusts southern Tibet (Fig. A1). Vertical and horizontal displacements are excluded at the southern end of the model. At the northern end, vertical displacements are free, and for depths above 40 km a southward horizontal velocity of 20 mm/year is imposed. The upper crust is thus subjected to 20 mm/year of horizontal shortening. At the northern end and for depths below 40 km we exclude horizontal displacements so that the Indian mantle lid does not shorten. This discontinuity was arbitrarily placed at a depth of 40 km in the lower crust. The model is long enough that the results in the Himalayan portion of the model are insensitive to the exact position of this discontinuity. The model is loaded with gravitational body forces \( g = 9.81 \text{ m/s}^2 \) and is supported at its base by hydrostatic pressure to allow for isostatic restoring forces.

A.2. Surface Processes

Following Avouac and Burov (1996) denudation is modeled using a 1-D linear diffusion equation in the range, i.e., north of the MFT,

\[
\frac{\partial h}{\partial t} = k \frac{\partial^2 h}{\partial x^2},
\]

where \( k \) is the mass diffusivity coefficient, expressed in units of area per time (m\(^2\)/year), \( h \) is the elevation, and \( x \) is the distance from the MFT. This linear diffusion law assumes that the southward flux of sediments at the surface is proportional to the local slope. The value of \( k \) was chosen to insure a denudation rate consistent with the 250–675 m\(^2\)/year flow of sediments.
eroded from the range (Lavé and Avouac, 2001). In order to test the sensitivity of the model to erosion $k$ was varied between $10^3$ and $10^5$ m$^2$/year.

In the foreland ($x \leq 0$), as argued by Hurtrez et al. (1999), we assume that erosion balances tectonic uplift or subsidence. This implies that denudation exactly compensates uplift at the MFT and sedimentation maintains a flat foreland at a constant elevation south of the MFT. This model is only an approximation that was designed to account as simply as possible for the dependency of denudation on topography and sedimentation in the foreland (see Avouac and Burov, 1996 for discussion).

A.3. Rheology of the Continental Lithosphere

A depth-varying rheology has been incorporated with elasto-brittle deformation in the upper part of the crust and ductile deformation in the lower part. The empirical rheological equations and laboratory-derived material properties are used under the assumption that they can be extrapolated to geological conditions.

Under low stresses, rocks deform elastically. For isotropic materials the relationship between each component of strain ($\varepsilon_{ij}$) and stress ($\sigma_{ij}$) is written as

$$\varepsilon_{ij} = \frac{1 + \nu}{E} \sigma_{ij} - \frac{\nu}{E} \sigma_{kk} \delta_{ij}$$  \hspace{1cm} (A2)

where $E$ and $\nu$ are the Young’s modulus and the Poisson’s ratio, respectively.

Beyond the elastic domain, rocks deform brittlely or ductilely. We use an elasto-plastic pressure-dependent law with the failure criterion of Drucker–Prager (Hassani et al., 1997). Failure occurs if

$$\frac{1}{2} (\sigma_1 - \sigma_3) = c (\cot \phi) + \frac{1}{2} (\sigma_1 + \sigma_3) \sin \phi$$  \hspace{1cm} (A3)

where $c$ is the cohesion, $\phi$ is the internal friction angle.

Ductile flow in the lithosphere is empirically described by a law which relates the critical principal stress difference necessary to maintain a steady-state strain rate to a power of the strain rate (Carter and Tsenn, 1987; Kirby and Kronenberg, 1987; Tsenn and Carter, 1987). This power-law creep is written as

$$\dot{\varepsilon} = A_P (\sigma_1 - \sigma_3)^n \exp(-E_P/RT),$$  \hspace{1cm} (A4)
where $R$ is the universal gas constant, $T$ is the temperature, $E_P$ is the activation energy, and $A_P$ and $n$ are empirically determined “constants,” assumed not to vary with stress and $(p, T)$ conditions. The ductile-flow law is thus strongly dependent on rock-type and temperature.

We only distinguish the crust and upper mantle. For the upper mantle, an olivine-controlled rheology is assumed (see parameters in Table A1). For the crust we have considered two end members, by considering either a quartz-controlled or a diabase-controlled rheology (Table A1). The quartz-like rheology implies a relatively thin elastic core in the middle crust and a thick lower crust that favors decoupling between mantle and crust, while a diabase rheology results in a stronger crust with a higher-viscosity lower crust (Burov and Diament, 1995).

### A.4. Frictional Sliding along the Fault

The MHT is assumed to follow a simple static friction law,

$$|\sigma_T| - \mu'(\sigma_N - p) \leq 0$$

$$\Leftrightarrow |\sigma_T| - \mu \sigma_N \leq 0,$$

(A5)

where $\sigma_T$ and $\sigma_N$ are the normal and the shear stress on the fault, and $\mu$ is the friction coefficient, $\mu'$ is the effective friction coefficient, and $p$ is the pore pressure.
REFERENCES


Perfettini, H., and Avouac, J.-P. Postseismic relaxation driven by brittle creep: A possible mechanism to reconcile geodetic measurement and the decay rate of aftershocks, application to the Chichi earthquake, Taiwan. *J. Geophys. Res.*, in press.


Weldon, R. J. (1986). Late Cenozoic geology of Cajon Pass; implications for tectonics and sedimentation along the San Andreas fault. California Institute of Technology.


