

Nonlocal Theories in Continuum Mechanics

M. Jirásek

The purpose of this paper is to explain why the standard continuum theory fails to properly describe certain mechanical phenomena and how the description can be improved by enrichments that incorporate the influence of gradients or weighted spatial averages of strain or of an internal variable. Three typical mechanical problems that require such enrichments are presented: (i) dispersion of short elastic waves in heterogeneous or discrete media, (ii) size effects in microscale elastoplasticity, in particular with the size dependence of the apparent hardening modulus, and (iii) localization of strain and damage in quasibrittle structures and with the resulting transitional size effect. Problems covered in the examples encompass static and dynamic phenomena, linear and nonlinear behavior, and three constitutive frameworks, namely elasticity, plasticity and continuum damage mechanics. This shows that enrichments of the standard continuum theory can be useful in a wide range of mechanical problems.

Keywords: Damage mechanics, dispersion, elasticity, enriched continuum, gradient models, nonlocal models, plasticity, quasibrittle materials, size effect, strain localization, wave propagation.

1 Introduction

In standard continuum mechanics, a solid body is decomposed into a set of idealized, infinitesimal material volumes, each of which can be described independently as far as the constitutive behavior is concerned. Of course, this does not mean that the individual material points are completely isolated, but their interaction can take place only on the level of balance equations, through the exchange of mass, momentum, energy and entropy.

In reality, however, no material is an ideal continuum. Both natural and man-made materials have a complicated internal structure, characterized by microstructural details whose size typically ranges over many orders of magnitude. The expression “microstructure” will be used as a generic denomination for any type of internal material structure, not necessarily on the level of micrometers. Some of these details can be described explicitly by spatial variation of the material properties. But this can never be done simultaneously over the entire range of scales. One reason is that such a model would be prohibitively expensive for practical applications. Another, more fundamental, reason is that, on a small enough scale, the continuum description *per se* is no longer adequate and needs to be replaced by a discrete model (or, ultimately, by interatomic potentials based on quantum mechanics).

Constructing a material model, one must select a certain resolution level below which the microstructural details are not explicitly “visible” to the model and need to be taken into account approximately and indirectly, by an appropriate definition of “effective” material properties. Also, one should specify the characteristic wavelength of the imposed deformation fields that can be expected for the given type of geometry and loading. Here, the notion of characteristic wavelength has to be understood in a broad sense, not only as the spatial period of a dynamic phenomenon but also as the length on which the value of strain changes substantially in static problems. If the characteristic wavelength of the deformation field remains above the resolution level of the material model, a conventional continuum description can be adequate. On the other hand, if the deformation field is expected to have important components with wavelengths

below the resolution level, the model needs to be enriched so as to capture the real processes more adequately. Instead of refining the explicit resolution level, it is often more effective to use various forms of the so-called **enriched** or **generalized continuum** formulations.

A systematic overview and detailed discussion of generalized continuum theories has been given e.g. in the recent review paper by Bažant and Jirásek (2002). The aim of the present study is to demonstrate by specific examples how the need for enriched continuum formulations arises from discrepancies between experimental observations and theoretical predictions based on the standard theories, and also how the model performance can be improved by adding carefully selected enrichment terms.

The enrichments to be discussed here are in general referred to as nonlocal, but this adjective must be understood in the broad sense, covering both **strongly nonlocal** and **weakly nonlocal** formulations. Precise mathematical definitions of strong and weak nonlocality were given by Rogula (1982) and are also explained in Bažant and Jirásek (2002). Here we only note that strongly nonlocal theories are exemplified by integral-type formulations with weighted spatial averaging or by implicit gradient models, while weakly nonlocal theories include for instance explicit gradient models. The meaning of these expressions will become clear from the examples to follow. We will start from enriched formulations of the theory of elasticity, and then proceed to elastoplasticity and damage mechanics.

The paper is organized as follows. Section 2 treats the **dispersion** of short elastic waves in heterogeneous or discrete media. It is shown that the standard homogenization procedure erases the information on dispersive properties. Dispersion laws are then derived for a host of generalized continuum models, including strain-gradient elasticity, models with mixed spatial-temporal derivatives, and integral-type nonlocal elasticity. Advantages and drawbacks of individual formulations are discussed, and a general framework of nonlocal strain-gradient elasticity is outlined.

Section 3 deals with **size effects** in microscale elastoplasticity, in particular with the size dependence of the apparent hardening modulus. Using the academic example of a

semi-infinite shear layer, it is shown that stiffer behavior of smaller structures can be reproduced with explicit or implicit gradient plasticity if appropriate boundary conditions are enforced. The general trends are discussed and compared to experimental measurements of size effect in plastic torsion of thin wires.

Section 4 is concerned with **localization** of strain and damage in quasibrittle structures and with the resulting transitional size effect. Mathematical and numerical difficulties related to the objective description of strain localization due to softening are explained using the one-dimensional example of a tensile bar. It is shown that if a stress-strain law with softening is incorporated in the standard continuum theory, the numerical results suffer from pathological sensitivity to the discretization parameter such as the size of finite elements. This can be remedied by special enrichments acting as localization limiters, e.g. by a nonlocal damage formulation. The onset of localization is studied analytically and relations to dispersion analysis are pointed out. It is also shown that the nonlocal model can correctly reproduce the transitional size effect on the nominal strength of a quasibrittle structure.

All generalized theories presented here introduce a model parameter with the dimension of length that reflects the intrinsic material length scale. The response of a material point depends not only on the strain and temperature history at that point but also on the history of a certain neighborhood of that point or even of the entire body. For this reason, such theories are classified as nonlocal in the broad sense.

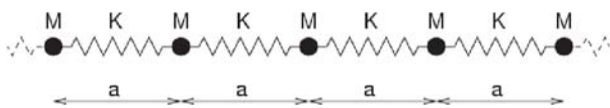
2 Dispersion of elastic waves

2.1 Continuum versus discrete models

In the **standard continuum theory**, propagation of waves in a homogeneous one-dimensional linear elastic medium is described by the hyperbolic partial differential equation

$$\rho \ddot{u} - Eu'' = 0, \quad (1)$$

where ρ is the mass density, E is the elastic modulus, $u(x, t)$ is the displacement and, as usual, overdots stand for derivatives with respect to time t and primes for derivatives with respect to the spatial coordinate x . Since ρ and E are constant coefficients, equation (1) admits solutions of the form



(a)

$$u(x, t) = e^{i(kx - \omega t)} = e^{ik(x - ct)}, \quad (2)$$

where i is the imaginary unit, ω is the circular frequency, k is the wave number, and $c = \omega/k$ is the wave velocity. Substituting (2) into (1), we find the condition

$$-\rho\omega^2 + Ek^2 = 0 \quad (3)$$

which implies that the magnitude of the circular frequency is proportional to the magnitude of the wave number. The signs of ω and k are irrelevant – if both change, the real part of (2) does not change, and if one of them changes, the wave propagates in the opposite direction but otherwise remains the same. Therefore, we will restrict ourselves to nonnegative values and write the solution of (3) as

$$\omega = k \sqrt{\frac{E}{\rho}}. \quad (4)$$

Due to the direct proportionality between ω and k , the velocity of a harmonic elastic wave, $c = \omega/k = \sqrt{E/\rho}$, is constant, independent of the wave number k . Since a wave of a general shape can be represented by a linear combination of harmonic waves, even such a general wave propagates at constant velocity c and its shape remains invariant.

The situation is different in a **discrete mechanical system**, which can be best exemplified by a regular infinite one-dimensional array of mass points connected by linear springs (Fig. 1a). The governing equations of motion form a system of ordinary differential equations

$$M\ddot{u}_j + K(u_j - u_{j-1}) - K(u_{j+1} - u_j) = 0, \quad j \in Z, \quad (5)$$

where M is the mass of each mass point, K is the spring stiffness, u_j is the displacement of mass point number j initially located at x_j , and Z is the set of all integer numbers. For an assumed harmonic wave of the form

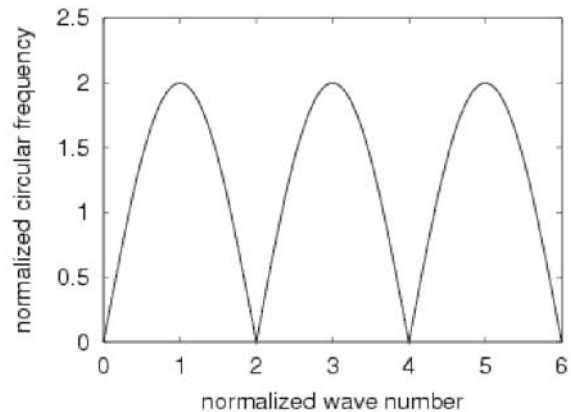
$$u_j(t) = e^{i(kx_j - \omega t)} \quad (6)$$

we obtain the condition

$$-M\omega^2 + K(1 - e^{ik(x_{j-1} - x_j)}) - K(e^{ik(x_{j+1} - x_j)} - 1) = 0, \quad j \in Z. \quad (7)$$

The spacing of mass points is assumed to be regular, i.e., $x_j - x_{j-1} = x_{j+1} - x_j = a$. The circular frequency corresponding to wave number k is thus

$$\omega = \sqrt{\frac{2K}{M}(1 - \cosh ika)} = \sqrt{\frac{2K}{M}(1 - \cos ka)} = 2\sqrt{\frac{K}{M}} \left| \sin \frac{ka}{2} \right|. \quad (8)$$



(b)

Fig. 1: (a) Discrete mass-spring model with nearest-neighbor interaction and (b) the corresponding dispersion diagram with the normalized frequency, $\omega\sqrt{M/K}$, plotted as a function of the normalized wave number, ka/π

For this mass-spring system, the relationship between ω and k is nonlinear and the wave velocity depends on the wave number. In such a case, one must distinguish between the phase velocity, $c_p = \omega/k$, and the group velocity, $c_g = d\omega/dk$. In the long wave limit (i.e., for k approaching zero), both the phase velocity and the group velocity tend to $c_0 = a\sqrt{K/M}$.

As is clear from the $\omega-k$ diagram in Fig. 1b, c_0 is the upper bound on the velocity of waves at any wave number. As k increases from 0 to $2\pi/a$, the corresponding phase velocity decreases from c_0 to 0. This means that shorter harmonic waves propagate at slower velocities. For a general wave, propagation of different harmonic components at different velocities leads to changes of the wave shape. This phenomenon is known as **dispersion**, the equation relating ω and k is called the dispersion equation, the resulting function $\omega(k)$ is the dispersion law, and its graph is the dispersion curve.

It is natural to expect that a discrete mechanical model consisting of mass points M regularly spaced at distance a and connected by springs of stiffness K is in some sense equivalent to a continuum characterized by mass density $\rho = M/a$ and elastic modulus $E = Ka$. However, this “equivalence” has its limits. A standard homogeneous linear elastic continuum has a linear dispersion curve with slope $c = \sqrt{E/\rho} = \sqrt{Ka^2/M}$, which coincides with the initial slope $c_0 = a\sqrt{K/M}$ of the nonlinear dispersion curve of the discrete model. Both dispersion curves are almost identical for long waves, but for wave numbers comparable to π/a (i.e., for wavelengths comparable to $2a$) they differ substantially. The standard continuum can be considered as a long-wave approximation of the discrete model (or vice-versa). If the actual physical system is close to the mass-spring model, the “equivalent” continuum model does not capture the phenomenon of dispersion of short elastic waves. On the other hand, if a homogeneous elastic continuum is discretized in space by finite differences (or by finite elements with a lumped mass matrix), the resulting set

of equations has the form (5) with $M = \rho a$ and $k = E/a$, where a is the grid parameter of the numerical method (e.g., size of finite elements). If these ordinary differential equations are integrated exactly in time, the solution captures correctly long-wave phenomena but introduces an artificial (numerical) dispersion of short waves with wavelengths comparable to the element size. Note that the numerical dispersion disappears if equations (5) are integrated in time using the central difference scheme with time step $\Delta t = a/c$, which is just at the limit of numerical stability. Consequently, propagation of shocks and other wide-spectrum phenomena are not represented accurately.

Discrete mass-spring systems are quite realistic models for crystalline materials on a scale of observation close to the atomic spacing. Of course, real crystal lattices are three-dimensional and interaction forces arise not only between immediate neighbors but also at longer distance. Still, a one-dimensional lattice is an acceptable model for plane waves propagating perpendicular to a certain set of crystallographic planes. In this case, each mass point actually represents one plane of densely packed atoms instead of one single atom. Interactions at longer distance can easily be incorporated by adding springs of stiffness K_2, K_3, \dots, K_N , connecting pairs of mass points spaced by $2a, 3a, \dots, Na$. A straightforward extension of the foregoing dispersion analysis yields the dispersion relation

$$-M\omega^2 + \sum_{n=1}^N K_n(1 - e^{-ikna}) - \sum_{n=1}^N K_n(e^{-ikna} - 1) = 0 \quad (9)$$

from which

$$\omega = \sqrt{\frac{2}{M} \sum_{n=1}^N K_n(1 - \cos kna)} = \frac{2}{\sqrt{M}} \sqrt{\sum_{n=1}^N K_n \sin^2 \frac{kna}{2}} \quad (10)$$

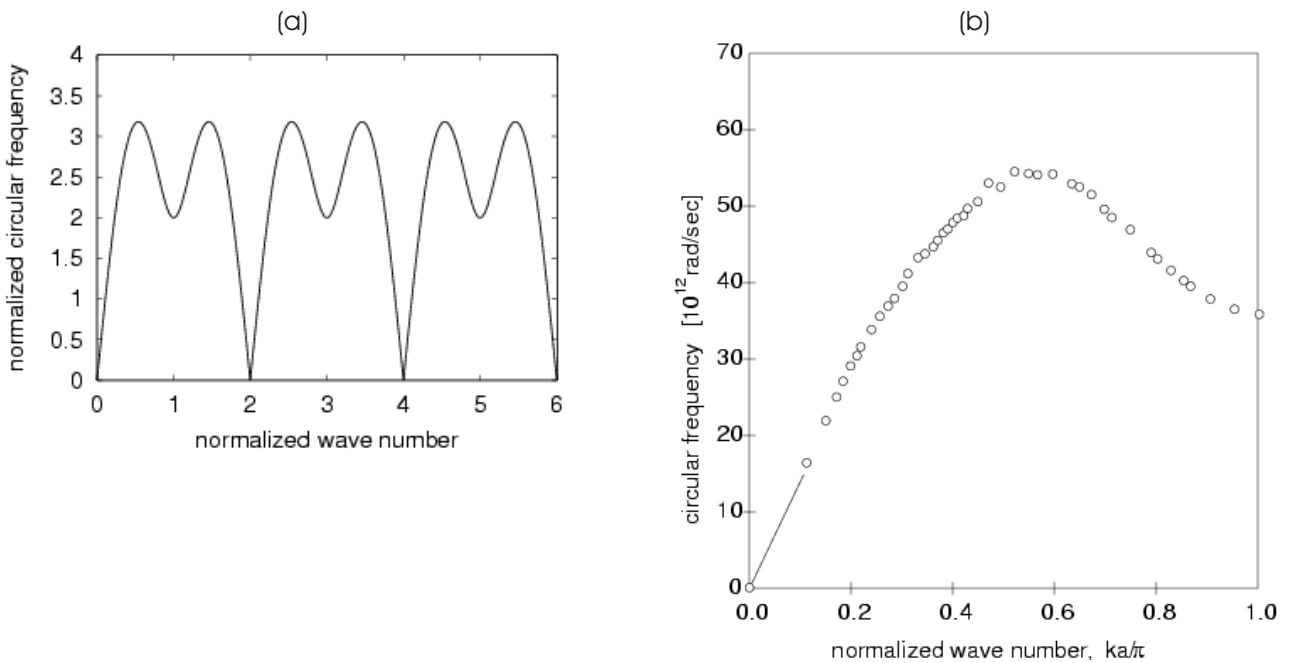


Fig. 2: (a) Dispersion curve of a mass-spring model with interaction up to distance , (b) dispersion curve of aluminum for longitudinal waves in the direction (after Yarnel et al., 1965)

In the long-wave limit ($k \rightarrow \infty$), the phase velocity $c_p = \omega/k$ tends to

$$c_0 = \frac{a}{\sqrt{M}} \sqrt{\sum_{n=1}^N K_n n^2}. \quad (11)$$

An example of a dispersion curve constructed with $N=3$ and $K_1=K_2=K_3=K$ is shown in Fig. 2a. This curve has a more general shape than for the nearest-neighbor interaction only (cf. Fig. 1b), but the wave number $2\pi/a$ still corresponds to zero frequency. This is natural, because the displacements u_j generated by a harmonic wave with wavelength a are the same at all the lattice sites. The associated mode is a uniform translation of the lattice points, which would not lead to any vibrations.

A striking property of the dispersion law (10) is its periodicity. This is closely related to the discrete and periodic nature of the underlying mechanical model. In fact, wave numbers that differ by an integer multiple of π/a correspond to the same physical state of the mass-point chain. Consequently, the dispersion curve of this chain is uniquely characterized by its initial part for wave numbers between 0 and π/a . This range of wave numbers is in the theory of crystal vibrations called the first Brillouin zone. Another interesting property is that the frequency range of harmonic waves is limited. The dispersion law maps the first Brillouin zone onto a certain interval $[0, \omega_{\max}]$ where ω_{\max} is a limiting circular frequency. If certain mass point is externally excited with a circular frequency larger than ω_{\max} , the vibration does not propagate through the entire chain and remains localized to the neighborhood of the excited mass point. The discrete chain therefore acts as a low-band filter.

Dispersion of elastic waves in crystals is a real physical phenomenon that can be observed and studied experimentally. Fig. 2b shows an example of a dispersion curve of aluminum, measured by Yarnel, Warren and Koenig (1965). This specific curve corresponds to waves propagating in the crystallographic direction (110), and it can be well approximated by a function of the form (10).

2.2. Strain-gradient elasticity

If dispersive phenomena were limited to the atomistic scale, elastic wave propagation could be described by discrete atomistic models on that scale and by standard continuum models on any coarser scale. However, dispersion arises not only due to the discrete character of the crystal lattice, but in general due to any type of material heterogeneity. Leaving aside the ideal case of a perfect monocrystal, the internal structure of all materials exhibits heterogeneities on various scales. Some defects in crystals can still be treated by atomistic models, but in most other cases the material needs to be considered as a continuum (because the relevant scale is already above the atomistic one) with a certain heterogeneous microstructure.

For elastic materials, there exist sophisticated and mathematically well-founded homogenization techniques providing the effective elastic moduli of a homogeneous material that is in a certain sense equivalent to the heterogeneous one and can replace it in large-scale simulations. Again, this equivalence is limited and holds with reasonable accuracy for long-wave phenomena only. In the present context, waves are

considered as long if the wavelength is much larger than the characteristic size of major heterogeneous features in the internal structure of the material. If there is a need to describe shorter waves in a realistic manner, it is in principle possible to explicitly resolve the details of the heterogeneous internal structure, but this approach often leads to extreme demands on the computational resources. Also, since the particular microstructure is usually not known exactly but only in terms of a stochastic description, the method of explicit resolution would need to exploit a Monte-Carlo type of technique or use stochastic differential equations, which again complicates the procedure and makes it computationally expensive. As an elegant and efficient alternative, it is possible to construct enrichments of the standard continuum theory that reflect the main features of the microstructure without using fast oscillating material properties.

In standard continuum elasticity it is assumed that the density of elastic energy stored per unit volume, w , depends only on the strain tensor, which is directly related to the deformation gradient, i.e., to the first gradient of the displacement field. The elastic energy stored by the entire body, W , is then evaluated as the spatial integral of the elastic energy density. In the one-dimensional setting, one can write

$$W = \int_L w(u'(x)) dx, \quad (12)$$

where $u' = du/dx$ is the strain, further denoted as ε , and L is the interval representing geometrically the one-dimensional body. In linear elasticity, the elastic energy density

$$w(\varepsilon) = \frac{1}{2} E \varepsilon^2 \quad (13)$$

is a quadratic function of strain.

One class of enrichments is based on the incorporation of higher gradients of the displacement field (Toupin, 1962; Mindlin 1964, 1965). In general, the elastic energy density can be assumed to depend on u'' , u''' , u^{IV} , etc. The simplest **strain-gradient theory of elasticity** uses enrichment by the second displacement gradient, u'' , which is equal to the strain gradient, ε' , further denoted as η . If we consider one single material point only, the strain gradient is locally independent of the strain value. In the linear case, the enriched elastic energy density potential is written as

$$w(\varepsilon, \eta) = \frac{1}{2} E \varepsilon^2 + \frac{1}{2} C \eta^2, \quad (14)$$

where C is a higher-order elastic modulus. The variation of elastic energy density is given by

$$\delta w = \frac{\partial w}{\partial \varepsilon} \delta \varepsilon + \frac{\partial w}{\partial \eta} \delta \eta = \sigma \delta \varepsilon + \chi \delta \eta, \quad (15)$$

where $\sigma = \partial w / \partial \varepsilon = E \varepsilon$ is the (Cauchy) stress and $\chi = \partial w / \partial \eta = C \eta$ is the so-called double stress.

Based on the extended form of the principle of virtual work, it is possible to derive the static equilibrium equation

$$(\sigma - \chi')' + b = 0, \quad (16)$$

where b is the body force density. In dynamics, b is replaced by the inertial force density, $-\rho \ddot{u}$. Combining this with the constitutive equations $\sigma = E \varepsilon$ and $\chi = C \eta$ and with the kinematic equations $\varepsilon = u'$ and $\eta = u''$, we obtain the wave equation of strain-gradient elasticity,

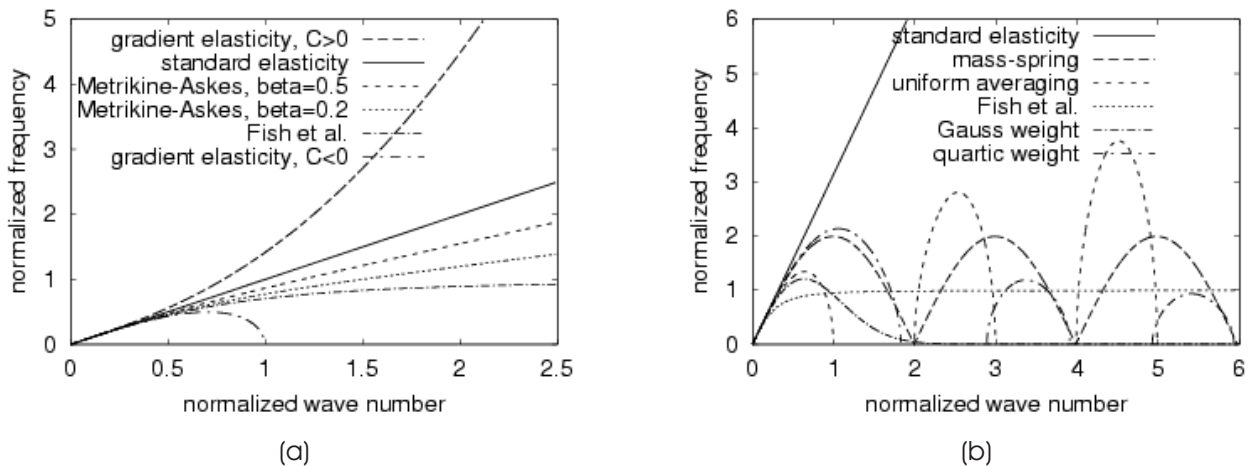


Fig. 3: Dispersion curves of (a) strain-gradient elasticity and gradient models with mixed derivatives, (b) nonlocal elastic models with different weight functions; the normalized circular frequency is $\omega l/c_0$ in case (a) and $\omega a/c_0$ in case (b), and the normalized wave number is kl in case (a) and ka/π in case (b)

$$\rho \ddot{u} - Eu'' + Cu^{IV} = 0, \quad (17)$$

which differs from the standard wave equation (1) by the presence of a term with the fourth spatial derivative of displacement.

Substituting the assumed harmonic wave solution (2) into (17), we obtain the dispersion equation

$$-\rho\omega^2 + Ek^2 + Ck^4 = 0 \quad (18)$$

from which

$$\omega = k \sqrt{\frac{E + Ck^2}{\rho}}. \quad (19)$$

The dispersion curve is plotted in Fig. 3a. The phase velocity

$$c_p = \frac{\omega}{k} = \sqrt{\frac{E + Ck^2}{\rho}} \quad (20)$$

is not constant, except for the case when $C = 0$ and the model reduces to standard elasticity. In strain-gradient elasticity it is usually assumed that the higher-order modulus C is positive. This assumption leads to a convex energy potential and permits to generalize certain uniqueness theorems known from standard elasticity. However, for $C > 0$ the phase velocity c_p increases with increasing wave number k . We have seen that the discrete mass-spring model exhibits the opposite trend, and this is also confirmed by measurements of dispersion curves in real crystals. Even for heterogeneous continua, the dispersion curves (determined experimentally or by analytical solution of some simple cases) typically have negative curvature. So the strain-gradient theory gives a reasonable approximation of the dispersion effect only if the higher-order modulus C is negative. Convexity of the elastic potential is then lost and uniqueness cannot be guaranteed. Indeed, if $C = -El^2$ where l is a model parameter with the dimension of length, the phase velocity of a harmonic wave with wave number $k_{crit} = 1/l$ vanishes. This means that the equation of motion (17) is satisfied by a stationary wave of wavelength $2\pi/k_{crit} = 2\pi l$. A similar result was found for the discrete mass-spring model, but in that case the stationary wave in reality represented a uniform translation, because the values of the displacements had physical meaning only at discrete

points with spacing equal to the critical wavelength. In contrast to that, a stationary wave in a continuous elastic medium is physically inadmissible. The problem is aggravated by the fact that, for wave numbers exceeding k_{crit} , the circular frequency ω solved from the dispersion equation (18) becomes imaginary. This means that harmonic modes with wavelengths shorter than $2\pi l$ would spontaneously blow up. The source of this instability becomes apparent if one realizes that, for short waves, the negative higher-order part the elastic energy, $-El^2(u'')^2$, exceeds in magnitude the positive standard part, $E(u')^2$, and so the total energy density becomes negative. If the amplitude of the wave grows, energy is released instead of being consumed.

2.3 Models with mixed spatial-temporal derivatives

Due to the unstable behavior of short waves, equation (17) is sometimes called the “**bad Boussinesq problem**”. This equation can describe dispersion of waves with “moderate” wave numbers but leads to instabilities if waves shorter than the critical wavelength $2\pi l$ are involved. If the body of interest is discretized by finite elements, the minimum wavelength that can be captured by the numerical approximation is proportional to the element size. Therefore, for meshes that are sufficiently coarse with respect to the material length parameter l , the numerical solution leads to reasonable results. However, upon mesh refinement, the solution becomes polluted by unstable modes rapidly oscillating in space.

Several modifications of the bad Boussinesq problem were proposed in the literature. Fish, Chen and Nagai (2002a) replaced the term with the fourth spatial derivative, u^{IV} , by a term with a mixed derivative, \ddot{u}'' . Their arguments can be rephrased and expanded as follows: For small wave numbers, the fourth-order term in (17) is negligible with respect to the second-order terms, so we can write $Eu'' \approx \rho\ddot{u}$. Differentiating this twice with respect to x , we obtain $Eu^{IV} \approx \rho\ddot{u}''$. Finally, replacing in (17) u^{IV} by $(\rho/E)\ddot{u}''$ and C by $-El^2$ yields a modified wave equation

$$\rho \ddot{u} - Eu'' - \rho l^2 \ddot{u}'' = 0 \quad (21)$$

which was called by Fish et al. (2002a) the “good Boussinesq problem”. This problem can be expected to have similar solutions to the original bad Boussinesq problem (17) at low wave numbers but a different asymptotic behavior for high wave numbers. Indeed, the usual procedure leads to the dispersion equation

$$-\rho\omega^2 + Ek^2 - \rho l^2 \omega^2 k^2 = 0 \quad (22)$$

from which

$$c_p = \frac{\omega}{k} = \sqrt{\frac{E}{\rho(1+k^2l^2)}} = \frac{c_0}{\sqrt{1+k^2l^2}}. \quad (23)$$

For this model with enrichment by a mixed derivative, the phase velocity remains real and positive for all wave numbers. The dispersion curve, plotted in Fig. 3a, is monotonically increasing, has a negative curvature, and for $k \rightarrow \infty$ approaches a horizontal asymptote at circular frequency $\omega = c_0/l$. The model can reasonably reproduce dispersion of waves at moderate wavelengths and does not give rise to instabilities for very short waves. Its extension to multiple dimensions is relatively complicated (Fish, Chen and Nagai, 2002b; Nagai, Fish and Watanabe, 2004).

Metrikine and Askes (2002) used a different line of reasoning and arrived at an equation of motion with two enrichments terms, proportional to \ddot{u}'' and u^{IV} . In its most general form, this equation can be written as

$$\rho\ddot{u} - Eu'' - l^2(\rho\ddot{u} - \beta Eu'') = 0, \quad (24)$$

where l is an internal length and β is an additional model parameter. For $l=0$ or for $\beta=1$, the model reduces to the standard elastic continuum, while for $\beta=0$ it reduces to the good Boussinesq problem of Fish et al. (2002a). Metrikine and Askes (2002) used certain heuristic arguments to link l and β to the microstructure and then proposed a parameter identification procedure based on a reflection-transmission test (Askes and Metrikine, 2002). Of course, one can also consider l and β as free model parameters and determine them by optimal fitting of the dispersion curve for a given material. The dispersion equation corresponding to (24),

$$-\rho\omega^2 + Ek^2 + l^2(-\rho\omega^2 + \beta Ek^2)k^2 = 0 \quad (25)$$

yields the phase velocity

$$c_p = \frac{\omega}{k} = c_0 \sqrt{\frac{1 + \beta k^2 l^2}{1 + k^2 l^2}}, \quad (26)$$

where $c_0 = \sqrt{E/\rho}$, as usual. If parameter β is nonnegative, the phase velocity remains real and positive for all wave numbers. For $0 < \beta < 1$, the dispersion curve has a negative curvature; see Fig. 3a. So, with a proper choice of parameters, the model can reasonably approximate dispersion and does not suffer by unstable behavior of short waves. Its disadvantage is that the presence of the fourth derivative u^{IV} requires either a C_1 -continuous finite element approximation (which is hard to construct on general meshes in multiple dimensions) or a mixed approach with independent approximations of several fields (e.g., of the displacement field and the strain field). Also, nonstandard higher-order boundary conditions are needed on the physical boundary of the investigated body.

2.4 Integral-type nonlocal elasticity

Another class of enrichments is based on weighted spatial averaging. The simplest model of this kind can be derived from the elastic potential

$$W = \frac{1}{2} \int_L \int_L E(x, \xi) \varepsilon(x) \varepsilon(\xi) dx d\xi, \quad (27)$$

where $E(x, \xi)$ is a function describing the generalized elastic modulus. The variation of elastic energy is evaluated as

$$\begin{aligned} \delta W &= \frac{1}{2} \int_L \int_L E(x, \xi) \delta \varepsilon(x) \varepsilon(\xi) dx d\xi + \\ &+ \frac{1}{2} \int_L \int_L E(x, \xi) \varepsilon(x) \delta \varepsilon(\xi) dx d\xi = \\ &= \frac{1}{2} \int_L \int_L [E(x, \xi) + E(\xi, x)] \varepsilon(\xi) d\xi \delta \varepsilon(x) dx. \end{aligned} \quad (28)$$

This can be written in the usual form $\delta W = \int_L \sigma(x) \delta \varepsilon(x) dx$ if the stress is defined as

$$\sigma(x) = \int_L E_s(x, \xi) \varepsilon(\xi) d\xi, \quad (29)$$

where

$$E_s(x, \xi) = \frac{1}{2} [E(x, \xi) + E(\xi, x)] \quad (30)$$

is the elastic modulus function symmetrized with respect to its arguments. The corresponding equilibrium equations derived from the principle of virtual work then keep their standard form, $\sigma' + b = 0$. Consequently, the wave equation for this model reads

$$\rho \frac{\partial^2 u(x, t)}{\partial t^2} - \frac{\partial}{\partial x} \int_L E_s(x, \xi) \frac{\partial u(\xi, t)}{\partial \xi} d\xi = 0. \quad (31)$$

Since function $E_s(x, \xi)$ reflects the strength of long-distance interaction between points x and ξ , its value can be expected to be negligible if the distance between x and ξ is large compared to the internal length of the material (which corresponds to the characteristic size and spacing of major heterogeneities). For functions E_s with a sufficiently fast decay, the integrals in (29) and (31) make sense even if the integration domain L is considered as the entire real axis. If the body infinite and macroscopically homogeneous, function $E_s(x, \xi)$ should depend only on the distance between x and ξ . Bearing in mind these restrictive assumptions, we present the modulus function in the form

$$E_s(x, \xi) = E_0 \alpha_0(x - \xi), \quad (32)$$

where E_0 is a reference value of the elastic modulus and α_0 is a dimensionless even function, further called the **nonlocal weight function**. The second term in the wave equation (31) can then be transformed as follows:

$$\begin{aligned} \frac{\partial}{\partial x} \int_{-\infty}^{\infty} E_s(x, \xi) \frac{\partial u(\xi, t)}{\partial \xi} d\xi &= E_0 \int_{-\infty}^{\infty} \frac{\partial \alpha_0(x - \xi)}{\partial x} \frac{\partial u(\xi, t)}{\partial \xi} d\xi = \\ &= -E_0 \int_{-\infty}^{\infty} \frac{\partial \alpha_0(x - \xi)}{\partial x} \frac{\partial u(\xi, t)}{\partial \xi} d\xi = E_0 \int_{-\infty}^{\infty} \alpha_0(x - \xi) \frac{\partial^2 u(\xi, t)}{\partial \xi^2} d\xi. \end{aligned} \quad (33)$$

Substituting the assumed harmonic form of an elastic wave (2) into the transformed wave equation

$$\rho \frac{\partial^2 u(x,t)}{\partial t^2} - E_0 \int_{-\infty}^{\infty} \alpha_0(x-\xi) \frac{\partial^2 u(\xi,t)}{\partial \xi^2} d\xi = 0 \quad (34)$$

we obtain the dispersion equation

$$-\rho\omega^2 + E_0 k^2 \alpha_0^*(k) = 0 \quad (35)$$

in which

$$\alpha_0^*(k) = \int_{-\infty}^{\infty} \alpha_0(r) e^{-ikr} dr \quad (36)$$

is the Fourier image of the nonlocal weight function $\alpha_0(r)$. Finally, the phase velocity is evaluated as

$$c_p = \frac{\omega}{k} = \sqrt{\frac{E_0 \alpha_0^*(k)}{\rho}} = c_0 \sqrt{\alpha_0^*(k)}. \quad (37)$$

Relation (35) shows that there is a unique correspondence between the dispersion law and the Fourier image of the nonlocal weight function. If the dispersion law of a certain material is given in the form $\omega(k) = kc_p(k)$ where $c_p(k)$ is a known function, it is possible to construct a nonlocal elasticity model that exactly reproduces the dispersion properties. For this, it suffices to set $E_0 = \rho c_0^2$ where $c_0 = c_p(0)$, and to evaluate the weight function by the inverse Fourier transform, taking into account that the phase velocity does not depend on the sign of the wave number:

$$\alpha_0(r) = \frac{1}{2\pi} \lim_{k \rightarrow \infty} \int_{-k}^k \alpha_0^*(k) e^{ikr} dk = \frac{1}{\pi c_0^2} \int_0^{\infty} c_p^2(k) \cos kr dk. \quad (38)$$

Of course, this is possible only if the dispersion law to be reproduced has reasonable properties, such that the inverse Fourier transform exists. For instance, for the dispersion law corresponding to the bad Boussinesq problem, $c_p(k)$ is given by (20) and the integral in (38) does not converge (independently of the sign of C). On the other side, for the good Boussinesq problem (21) we have $c_p(k) = c_0/\sqrt{1+k^2 l^2}$ and the inverse Fourier transform (38) yields

$$\alpha_0(r) = \frac{1}{\pi} \int_0^{\infty} \frac{\cos kr}{1+k^2 l^2} dk = \frac{1}{2l} e^{-\frac{|r|}{l}}. \quad (39)$$

A nonlocal elasticity model with this particular weight function gives exactly the same dispersion law as the model with enrichment by a mixed derivative proposed by Fish et al. (2002a). On an infinite domain, both models are equivalent.

The transformation of the model of Fish et al. (2002a) into a integral-type nonlocal model can also be performed directly. At a fixed time instant, equation (21) can be written as

$$\ddot{u} - l^2 \ddot{u}'' = \frac{E}{\rho} u'' \quad (40)$$

and interpreted as an ordinary differential equation for the unknown acceleration \ddot{u} , with the current displacement u considered as known. Equation (40) has the form of the so-called **Helmholtz equation**, and its solution satisfying conditions of boundedness (which play the role of boundary conditions at plus and minus infinity) can be expressed as

$$\ddot{u}(x,t) = \frac{E}{\rho} \int_{-\infty}^{\infty} G(x,\xi) u''(\xi,t) d\xi, \quad (41)$$

where $G(x,\xi)$ is the **Green function** of the Helmholtz equation, formally obtained as the solution of this equation with the Dirac distribution $\delta(\xi)$ on the right-hand side. It turns out that the Green function is in this case given by

$$G(x,\xi) = \frac{1}{2l} e^{-\frac{|x-\xi|}{l}} \quad (42)$$

and so equation (41) is in fact equivalent with (34) if the nonlocal weight function α_0 is selected according to formula (39).

2.5 Combination of nonlocal averaging and strain gradients

For the dispersion law corresponding to the gradient model proposed by Metrikine and Askes (2002), the integral in the inverse Fourier transform (38) does not converge. So this model is not equivalent to any integral-type nonlocal elastic model derived from the potential (27). Still, using the alternative procedure based on the Green function, it is possible to construct a more general nonlocal model equivalent to the original enriched gradient formulation. Indeed, rewriting (24) as

$$\ddot{u} - l^2 \ddot{u}'' = \frac{E}{\rho} (u'' - \beta l^2 u^{IV}) \quad (43)$$

and “solving” for the acceleration, we obtain

$$\ddot{u}(x,t) = \frac{E}{\rho} \left[\int_{-\infty}^{\infty} G(x,\xi) u''(\xi,t) d\xi - \beta l^2 \int_{-\infty}^{\infty} G(x,\xi) u^{IV}(\xi,t) d\xi \right]. \quad (44)$$

The result resembles the wave equation of strain-gradient elasticity (17), but with the derivatives u'' and u^{IV} replaced by their weighted spatial averages. This observation motivates the development of a nonlocal strain-gradient model with the elastic energy potential given by

$$W = \frac{1}{2} \int_L \int_L E(x,\xi) \varepsilon(x) \varepsilon(\xi) dx d\xi + \frac{1}{2} \int_L \int_L C(x,\xi) \eta(x) \eta(\xi) dx d\xi. \quad (45)$$

Taking the variation and using the same line of reasoning as for the basic version of nonlocal elasticity, we identify the constitutive equations

$$\sigma(x) = \int_L E_s(x,\xi) \varepsilon(\xi) d\xi \quad (46)$$

$$\chi(x) = \int_L C_s(x,\xi) \eta(\xi) d\xi \quad (47)$$

in which $C_s(x,\xi) = [C(x,\xi) + C(\xi,x)]/2$ is the symmetrized higher-order modulus. Substituting this into the equation of motion of the strain-gradient theory, $\rho \ddot{u} - (\sigma - \chi)' = 0$, and using the kinematic relations $\varepsilon = u'$ and $\eta = u''$, we obtain the wave equation

$$\rho \frac{\partial^2 u(x,t)}{\partial t^2} - \frac{\partial}{\partial x} \int_L E_s(x,\xi) \frac{\partial u(\xi,t)}{\partial \xi} d\xi + \frac{\partial^2}{\partial x^2} \int_L C_s(x,\xi) \frac{\partial^2 u(\xi,t)}{\partial \xi^2} d\xi = 0 \quad (48)$$

Table 1: Special cases of nonlocal strain-gradient elasticity

model	$\alpha_0(r)$	$\alpha_1(r)$
standard elasticity	$\delta(r)$	0
strain-gradient elasticity	$\delta(r)$	$\delta(r)$
Fish et al.	$\frac{1}{2l} e^{- r /l}$	0
Metrikine and Askes	$\frac{1}{2l} e^{- r /l}$	$\frac{\beta}{2l} e^{- r /l}$
mass-spring chain	$\frac{\langle a-r \rangle}{a^2}$	0

The brackets $\langle \dots \rangle$ denote the positive part operator, defined by $\langle x \rangle = (x + |x|)/2$.

that generalizes equation (31). For the moduli functions in the special form

$$E_s(x, \xi) = E_0 \alpha_0(x - \xi) \quad (49)$$

$$C_s(x, \xi) = E_0 l^2 \alpha_1(x - \xi) \quad (50)$$

dispersion analysis provides the following expression for the phase velocity:

$$c_p(k) = c_0 \sqrt{\alpha_0^*(k) + k^2 l^2 \alpha_1^*(k)}. \quad (51)$$

Here, α_1^* denotes the Fourier image of function α_1 .

Table 2: Nonlocal weight functions and their Fourier images

model	$\alpha_0(r)$	$\alpha_0^*(k)$
standard elasticity	$\delta(r)$	1
mass-spring (neighbors)	$\frac{\langle a-r \rangle}{a^2}$	$\frac{2(1 - \cos ka)}{k^2 a^2}$
mass-spring (long interaction)	$\frac{\sum K_n \langle na-r \rangle}{a^2 \sum n^2 K_n}$	$\frac{2 \sum K_n (1 - \cos kna)}{k^2 a^2 \sum n^2 K_n}$
Eringen	see eq. (54)	$\frac{2(1 - \cos ka)}{k^2 a^2}$ if $k < \frac{\pi}{a}$
uniform averaging	$\frac{1}{2a}$ for $ r \leq a$	$\frac{\sin ka}{ka}$
Fish et al.	$\frac{1}{2a} e^{- r /a}$	$\frac{1}{1 - k^2 a^2}$
Gauss weight function	$\frac{1}{a\sqrt{\pi}} e^{-r^2/a^2}$	$e^{-k^2 a^2/4}$
quartic weight function	$\frac{15}{16a} \left\langle 1 - \frac{r^2}{a^2} \right\rangle^2$	$\frac{15}{k^5 a^5} [(3 - k^2 a^2) \sin ka - 3ka \cos ka]$

The nonlocal strain-gradient model just presented is quite general and covers as special cases all the other models discussed so far. The special choices of weight functions α_0 and α_1 are summarized in Table 1. The model permits to reproduce a very wide class of dispersion laws. Which terms need to be activated depends on the asymptotic behavior of the dispersion curve at wave numbers approaching infinity. If the dispersion law $\omega(k)$ is bounded, it is sufficient to use a regular weight function α_0 obtained by the inverse Fourier transform of $(\omega/c_0 k)^2$. If $\omega(k)$ tends to infinity but remains of order $O(k)$, it is possible to use either a weight function α_0 with a singular part of the Dirac type, or regular functions α_0 and α_1 . Finally, if $\omega(k)$ grows superlinearly but remains of order $O(k^2)$, function α_1 must have a singular Dirac-type part. Still faster growth could be reproduced by models with second (Mindlin, 1965) or still higher (Green and Rivlin, 1964) strain gradients, but it seems that dispersion laws of real materials do not exhibit such behavior, so this question is purely academic. In fact, all dispersion laws with superlinear growth at $k \rightarrow \infty$ are suspicious because the phase velocity becomes unbounded and disturbances can propagate at an arbitrary velocity if the wavelength is selected as sufficiently short.

2.6 Nonlocal model reproducing dispersion of discrete lattice

It is interesting that the nonlocal elastic model with a weight function α_0 that linearly decreases from its maximum value at $r=0$ to zero at $r=a$ and vanishes for $r > a$ leads to exactly the same dispersion law as the simplest mass-spring model with nearest-neighbor interactions and with spacing a between neighboring mass points. Table 2 shows several other nonlocal weight functions and their Fourier images that can be used to construct the corresponding dispersion curves,

which are plotted in Fig. 3b. The dispersion law of a mass-spring model with long-distance interactions can be reproduced by a nonlocal model with a piecewise linear weight function whose characteristics uniquely depend on the spring stiffnesses. If all stiffnesses used by the discrete model are positive, the weight function is concave (for nonzero r). The dispersion law gives real frequencies for all wave numbers, but for wave numbers that are integer multiples of $2\pi/a$ the frequency vanishes, i.e., the model admits stationary waves of wavelength a , $a/2$, $a/3$, etc. This is natural for the discrete model, as already explained, but the same property is shared by the nonlocal continuum model.

For the simplest weight function, constant for r between 0 and a and vanishing for $r > a$, the dispersion law gives real frequencies only for wave numbers in intervals $[0, \pi/a]$, $[2\pi/a, 3\pi/a]$, $[4\pi/a, 5\pi/a]$, etc. Between these bands, the frequency becomes imaginary, which indicates an instability. The potential appearance of periodic modes that carry no energy for the nonlocal model with uniform strain averaging over a finite neighborhood was mentioned by Bažant and Chang (1984). It is clear that every periodic function with period $2a$ is mapped by the nonlocal operator onto a zero function, and therefore has no influence on the nonlocal average. Each such function can be decomposed into a sum of harmonic functions of wavelengths $2a$, $2a/2$, $2a/3$, etc., which correspond to zero frequencies. Thus the static solution can be modified by a periodic function with period $2a$ without disturbing equilibrium. The dispersion analysis indicates that in dynamics the situation is even worse, because harmonic modes with wave numbers in the intervals $(\pi/a, 2\pi/a)$, $(3\pi/a, 4\pi/a)$, etc., are associated with imaginary frequencies and would blow up.

To avoid the potential appearance of unstable modes, it is sufficient to use weight functions with positive values of their Fourier images for any positive wave number k . This is the case for instance for the Green function of the Helmholtz equation, and also for the Gauss-type weight function

$$\alpha_0(r) = \frac{1}{a\sqrt{\pi}} e^{-r^2/a^2} \quad (52)$$

that is often used by nonlocal models. On the other hand, the Fourier image of the truncated quartic polynomial function

$$\alpha_0(r) = \frac{15}{16a} \left\langle 1 - \frac{r^2}{a^2} \right\rangle^2 \quad (53)$$

is positive only for wave numbers smaller than $5.763/a$. Consequently, instabilities could develop for fine meshes with element size in the order of a .

The relationship between atomic lattices and the nonlocal integral models was studied by Eringen (1972), who proposed a weight function that should correspond to the mass-spring model. However, Eringen did not use the complete inverse Fourier transform but integrated (38) only in the limits $-\pi/a \leq k \leq \pi/a$. Consequently, his nonlocal model would reproduce the dispersion law of the mass-spring model only for wavelengths larger than $2a$. All smaller wavelengths would be associated with a zero frequency, i.e., stationary waves could easily appear. Eringen's weight function

$$\alpha_0(x) = \frac{4 + \pi^2 - 2\pi \operatorname{si} \pi}{2\pi^3 a \left(1 + \frac{1}{2} \pi^2 - \frac{1}{2} \pi \operatorname{si} \pi\right)} \times \left[\cos \frac{x}{2a} + \frac{a\pi^2}{2x} \sin \frac{x}{2a} + \frac{x}{4a} \operatorname{si} \frac{x}{2a} - \frac{1}{4} \left(\pi + \frac{x}{2a} \right) \operatorname{si} \left(\pi + \frac{x}{2a} \right) - \frac{1}{4} \left(\pi - \frac{x}{2a} \right) \operatorname{si} \left(\pi - \frac{x}{2a} \right) \right] \quad (54)$$

(where $\operatorname{si} x \equiv \int_0^x \eta^{-1} \sin \eta d\eta$) is also much more complicated

than the piecewise linear function reproducing the dispersion of the discrete model exactly for arbitrary wave numbers, and does not seem to be very practical.

3 Size effects in microscale plasticity

3.1 Experimental observations

Among phenomena that are hard to model and predict by standard continuum theories, one finds also various forms of size effects on the apparent "material" properties. Such effects can be observed already in the elastic range. For instance, according to standard elasticity, the torsional stiffness of a prismatic beam with a circular cross section should be proportional to the shear modulus of elasticity, G , and to the third power of the sectional diameter, D . However, if the torsional stiffness is evaluated experimentally as the ratio between the torque and the relative twist angle, it turns out that the expected proportionality to D^3 holds with sufficient accuracy only for diameters larger than a certain threshold value (Morrison, 1939; Lakes, 1986; Fleck, Muller, Ashby and Hutchinson, 1994). If the results obtained for thick wires are extrapolated to thin wires, the actual stiffness is underestimated. In the context of the standard theory, this can be interpreted as a dependence of the elastic modulus on the size of the sample. However, such an explanation is not satisfactory from the theoretical point of view.

In a more fundamental approach, it is admitted that the standard continuum elasticity theory provides only a large-size approximation to the static torsion problem, just as it provides only a long-wave approximation to the dynamic wave dispersion problem. If the size of the structure is comparable to a certain internal length scale of the material, higher-order effects appear and the classical concept of a homogeneous local continuum needs an adjustment. In elasticity, such an adjustment that accounts for the principal features of the microstructure is provided by various theories enriched by higher-order gradients or integral-type nonlocal terms, already exposed in Section 2.

In general, by size effect we mean a situation when a certain parameter normally considered as a material property appears to be dependent on the size of the sample or specimen for which it is evaluated. The reason for this unexpected behavior is usually that the specimens are either too small or too big and the underlying theory is not adequate on the extreme scale. Important size effects in microscale plasticity have been detected in indentation tests (Nix, 1989; Ma and Clarke, 1995; Poole, Ashby and Fleck, 1996), bending of thin sheets (Stolken and Evans, 1998), plastic torsion (Fleck et al.,

1996), and void growth. From the practical point of view, proper understanding and modeling of size effects on the small scale is essential for applications of simulation methods in the analysis and design of micro- and nano-devices used in modern technology.

3.2 Illustrative example: shear layer

To illustrate the power of plasticity theories extended by gradient or integral-type nonlocal terms, we will analyze a rather academic but instructive problem of a material layer of thickness L , placed between stiff loading plates and loaded statically by shear. This problem can be modeled in one spatial dimension, which facilitates its analytical solution. The choice of the coordinate system and the type of loading are shown in Fig. 4. It is assumed that the elastic properties of the material are isotropic and plastic flow is isochoric (preserves volume), so that volumetric and deviatoric effects can be decoupled. If this is the case, the relevant components of stress and strain are the shear stress $\sigma_{xy} \equiv \tau$ and the engineering shear strain $2\varepsilon_{xy} \equiv \gamma$. Both of them are considered as independent of spatial coordinates y and z . From the equilibrium equation with vanishing body forces, it follows that τ is also independent of x , i.e., it is uniformly distributed in space and varies only as a function of the pseudo-time parameterizing the loading process.

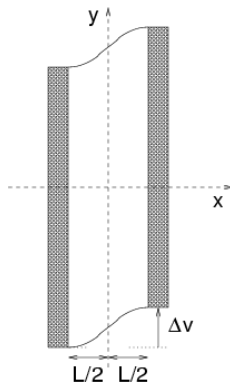


Fig. 4: Material layer between stiff plates loaded by shear

Classical plasticity with linear isotropic hardening is described by the basic equations

$$\tau = G(\gamma - \gamma_p) \quad (55)$$

$$\dot{\gamma}_p = \dot{\kappa} \operatorname{sgn}(\tau) \quad (56)$$

$$\dot{\kappa} \geq 0, \quad f(\tau, \kappa) \leq 0, \quad \dot{\kappa} f(\tau, \kappa) = 0 \quad (57)$$

where γ_p is the plastic strain and κ is the cumulative plastic strain. The yield function f is given by

$$f(\tau, \kappa) = |\tau| - \tau_Y(\kappa) \quad (58)$$

where τ_Y is the current yield stress in shear, evaluated from the hardening law

$$\tau_Y(\kappa) = \tau_0 + H\kappa \quad (59)$$

with τ_0 = initial yield stress in shear and H = hardening modulus (in shear). During monotonic loading with positive value of shear stress τ , there is no difference between the plastic shear strain γ_p and the cumulative plastic strain κ . We will therefore replace γ_p by κ and call it simply the plastic strain.

If the hardening modulus is positive and the loading is monotonic, the stress τ uniquely determines the corresponding strain γ . Since the stress is uniform, the strain must be uniform as well. Therefore, the strain at any point is equal to the average strain determined as the ratio between the relative displacement of the loading plates, $\Delta v = v(L/2) - v(-L/2)$, and the layer thickness, L . The stress-strain curve can be directly determined from the measured dependence between v and the tangential traction on the boundary, t , and it should be independent of the layer thickness. Therefore, the standard plasticity model does not indicate any size effect.

3.3 Explicit gradient plasticity

If the layer thickness is comparable to the characteristic length of the material microstructure, the assumption of the local dependence of stress on the history of strain at the same material point becomes questionable. The reason is that the hardening process does not take place at each infinitely small material point separately and independently of the surrounding points. This can be taken into account by sophisticated models that consider the details of the hardening mechanisms, e.g., by discrete dislocation models. As a simpler alternative, one can use enriched continuum models that take into account the micromechanical processes only "on the average", but by terms of a higher order than in the standard continuum theory. Motivated by certain micromechanical considerations, Aifantis (1984) proposed a family of models with the yield stress dependent not only on the value of the cumulative plastic strain (internal variable driving the hardening process) but also on its first and second gradients. In the one-dimensional setting and for linear hardening, the simplest version of the Aifantis gradient plasticity model replaces the hardening law (59) by

$$\tau_Y(\kappa, \kappa'') = \tau_0 + H\kappa - Hl^2\kappa'' \quad (60)$$

where l is a parameter with the dimension of length.

The elastic part of the model remains unchanged, and so the strain is uniform in the elastic range. After the onset of yielding, the yield condition must be satisfied in the plastic zone. It is easy to show that in the present case the plastic zone must extend over the entire layer. The yield condition $f=0$ combined with the hardening law (60) then provides the ordinary differential equation

$$\kappa - l^2\kappa'' = \frac{\tau - \tau_0}{H} \quad (61)$$

for the unknown function κ . This equation should be supplemented by boundary conditions at the layer boundaries. The choice of the specific form of boundary conditions has a major influence on the solution and on the resulting size effect. For homogeneous Neumann boundary conditions, enforcing a vanishing normal derivative of κ at the boundary, the uniform solution obtained with the classical model would remain valid even for the gradient model. However, if the shear layer is fixed to rigid or very stiff loading plates and the bond between the two materials is perfect, the boundary (or rather the material interface) acts as an obstacle for dislocation motion, which is the main mechanism of plastic flow in crystalline materials. This has been confirmed by simulations based on discrete dislocation models (Shu, Fleck, van der Giessen and Needleman, 2001). In the extreme case, plastic flow is completely prevented and the cumulative plastic strain

κ must vanish at the boundary. For homogeneous Dirichlet boundary conditions, $\kappa(\pm L/2) = 0$, the solution of (61) is

$$\kappa(x) = \frac{\tau - \tau_0}{H} \left(1 - \frac{\cosh x/l}{\cosh L/2l} \right). \quad (62)$$

The distribution of plastic strain (normalized by the factor $(\tau - \tau_0)/H$ that would correspond to the value of uniform plastic strain in the standard theory) across the layer thickness is plotted in Fig. 5b for different relative sizes, L/l . If the layer thickness is substantially larger than the material length, the plastic strain is almost uniform, except for narrow boundary layers. With decreasing structural size L , the relative importance of these boundary layers with reduced plastic strain levels increases, which makes the overall response stiffer.

To characterize the overall response of the layer, the relative tangential displacement of one loading plate with respect to the other can be evaluated as

$$\begin{aligned} \Delta v &= \int_{-L/2}^{-L/2} \gamma(x) dx = \int_{-L/2}^{-L/2} \left[\frac{\tau}{G} + \kappa(x) \right] dx = \\ &= \frac{\tau L}{G} + \frac{\tau - \tau_0}{H} \left(L - 2l \tanh \frac{L}{2l} \right). \end{aligned} \quad (63)$$

Defining the average shear strain $\tilde{\gamma} = \Delta v/L$, we can transform (63) into the average stress-strain law

$$\tau = \tau_0 + \tilde{G}_{ep}(\tilde{\gamma} - \gamma_0), \quad (64)$$

where $\gamma_0 = \tau_0/G$ is the limit elastic shear strain and

$$\tilde{G}_{ep} = \left[\frac{1}{G} + \frac{1}{H} \left(1 - \frac{2l}{L} \tanh \frac{L}{2l} \right) \right]^{-1} \quad (65)$$

is the elastoplastic (tangent) shear modulus.

According to (64)–(65), the overall response of the specimen after the onset of yielding is linear and is equivalent to the response of the same specimen made of the standard elastoplastic material with hardening modulus

$$\tilde{H} = \frac{H}{1 - \frac{2l}{L} \tanh \frac{L}{2l}}. \quad (66)$$

This “apparent” hardening modulus depends not only on the material parameters H and l but also on the layer

thickness L , i.e., on the size of the specimen for which it is evaluated; see Fig. 5a. Therefore, it cannot be considered as an intrinsic material parameter.

If the actual behavior of the material is close to the gradient model but the experimental results are interpreted using the standard elastoplastic model, the value of the hardening modulus will appear to be **size-dependent**. The reason is that the model is oversimplified and the comparison between theory and experiments uses only global characteristics such as the measured relation between the loading force and the relative displacement of one loading plate with respect to the other. If detailed measurements of the strain field inside the specimen were available, they would reveal a discrepancy between the actual strain distribution and the theoretical solution based on the oversimplified assumptions. This would give a hint regarding the necessary refinement of the theoretical model by inclusion of higher-order terms. But even if such detailed measurements are not possible or not available, the development of an appropriate enriched continuum theory can be guided by the experimentally detected size effect. It turns out that the size effect for one specific type of test performed on a series of geometrically similar specimens can often be well reproduced by several different types of enriched models that are not necessarily equivalent in a general case.

Ideally, the enriched theory should be verified by several tests leading to different types of stress and strain distributions, and also supported by micromechanical arguments and confirmed by observations of the actual processes in the microstructure. Only then can the model be assumed to reasonably reproduce the actual material behavior and to have some predictive power. If only one series of experiments is fitted, the model can usually serve for reliable interpolation in the limits that have been covered by the experiments, but extrapolation to smaller or larger sizes can be dangerous.

The diagram linking the shear stress to the average shear strain for different thicknesses L of the shear layer is plotted in Fig. 6a, and the dependence of the apparent hardening modulus \tilde{H} on the specimen size (layer thickness) is shown in Fig. 5a. If the layer thickness is much larger than the material length l , the stress-strain curve is practically size-independent and the apparent hardening modulus \tilde{H} evaluated from the

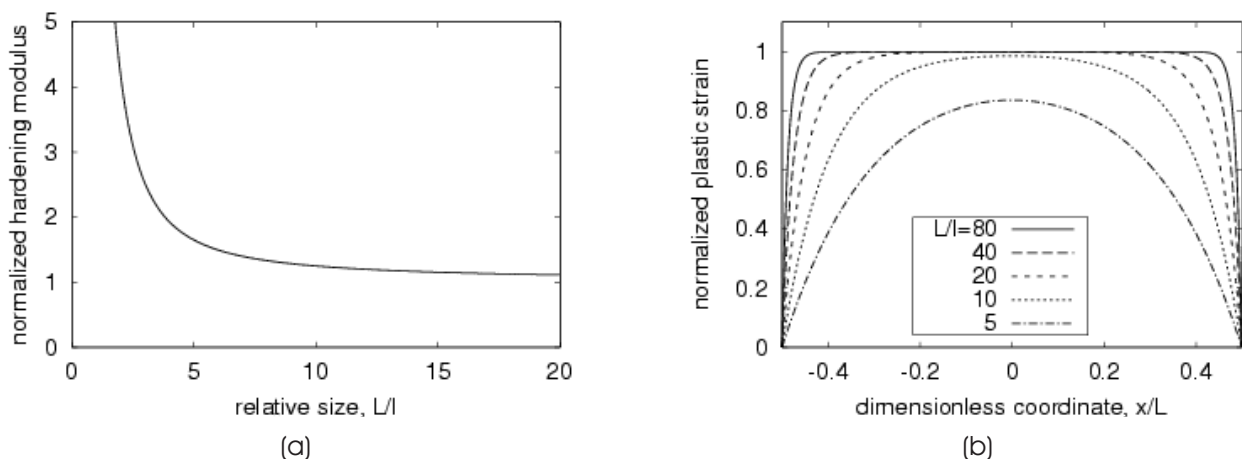


Fig. 5: Explicit gradient plasticity: (a) dependence of the apparent hardening modulus on the size of the specimen (layer thickness), (b) distribution of plastic strain across the layer

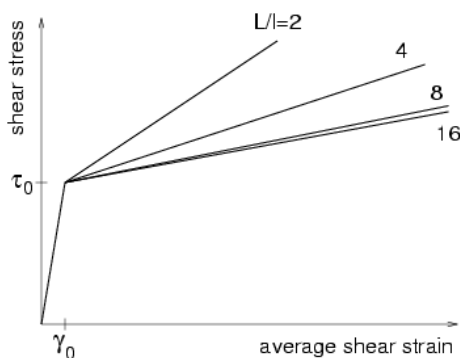
test is very close to the model parameter H , considered here as an intrinsic material property. So the standard theory can be used as a good approximation if the specimen is large. For layer thicknesses smaller than about $20l$, the stress-strain curve becomes size-dependent and it rises above the basic stress-strain curve valid in the large-size limit. The apparent hardening modulus increases with decreasing size. Even though the shear test of a layer is difficult to perform, this trend can be considered as realistic because stronger hardening for smaller specimens is indeed observed in experiments with plastic torsion of tiny wires or microbending of thin films. This is documented in Fig. 6b, adapted from Fleck et al. (1994). The figure shows the dependence of the normalized torque, M/D^3 , on the normalized angle of twist, ϕD , which is equal to the shear strain on the wire surface. According to any theory with stress at a point dependent on the history of strain at that point only, the resulting curve should be independent of the wire diameter D . In reality, this is true only for sufficiently thick wires. The figure clearly shows that for diameters in the order of dozens of micrometers, the actual response after the onset of yielding is stronger than expected. Aifantis (2003) and Fleck and Hutchinson (2001) have demonstrated that this size effect can be captured by gradient plasticity theories similar to the one presented here.

3.4. Implicit gradient plasticity

The model with yield stress dependent on the second gradient of cumulative plastic strain falls into the category of **explicit** gradient models. Recently it turned out that certain advantages can be gained by using **implicit** gradient formulations, which are closely related to integral-type nonlocal models. A prominent example is the so-called ductile damage model of Geers, Engelen and Ubachs (2001). This model introduces the dependence of the yield stress on the nonlocal cumulative plastic strain, $\bar{\kappa}$, defined implicitly as the solution of a Helmholtz-type differential equation with the local cumulative plastic strain on the right-hand side. For a one-dimensional problem, this equation reads

$$\bar{\kappa} - l^2 \bar{\kappa}'' = \kappa, \quad (67)$$

where l is, as usual, a model parameter with the dimension of length. In the simplest case, one could replace κ in (58)–(59)



(a)

by $\bar{\kappa}$, but such a formulation would have certain deficiencies. Micromechanical arguments based on the idea of a plastically hardening matrix weakened by growing voids lead to the following expression for the yield stress:

$$\tau_Y(\kappa, \bar{\kappa}) = (\tau_0 + H\kappa)[1 - \omega_p(\bar{\kappa})]. \quad (68)$$

Here, the term $\tau_0 + H\kappa$ represents the yield stress of the matrix, which exhibits hardening driven by the local value of the cumulative plastic strain, while $1 - \omega_p(\bar{\kappa})$ is an integrity factor taking into account the reduction of the effective area by voids that carry no stress. Void propagation is assumed to be driven by the nonlocal cumulative plastic strain, $\bar{\kappa}$, and is reflected by the ductile damage function ω_p that vanishes for $\bar{\kappa} = 0$ and increases later on.

Due to the nonlinear format of the hardening law (68), analysis of the plastic strain distribution in a general state would lead to a nonlinear differential equation. To allow for an analytical treatment, we restrict attention to the initial distribution of the plastic strain rate. Differentiating (68) with respect to time and using the consistency condition $\dot{f} = \dot{\tau} - \dot{\tau}_Y = 0$, we obtain the differential equation

$$H(1 - \omega_p)\dot{\kappa} - (\tau_0 + H\kappa) \frac{d\omega_p}{d\bar{\kappa}} \dot{\bar{\kappa}} = \dot{\tau}. \quad (69)$$

At the onset of yielding, κ , $\bar{\kappa}$ and ω_p vanish in the entire layer and the derivative $d\omega_p/d\bar{\kappa}$ evaluated at $\bar{\kappa} = 0$ has the same value ω_p^* at all points x . Taking all this into account and substituting the rate form of (67) into (69), we get the linear differential equation with constant coefficients,

$$(H - \tau_0 \omega_p^*) \dot{\bar{\kappa}} - Hl^2 \bar{\kappa}'' = \dot{\tau}. \quad (70)$$

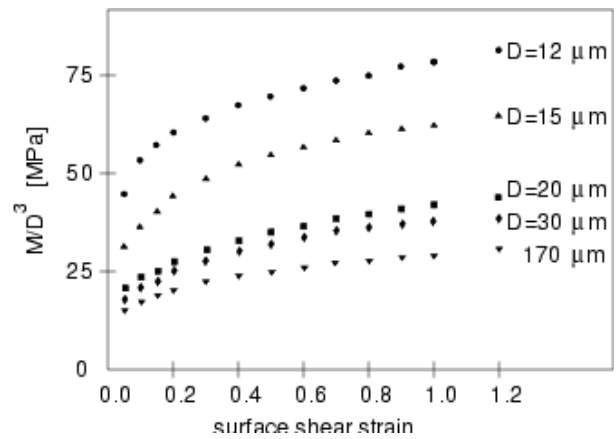
In the absence of body forces, static equilibrium implies that the shear stress τ must be uniform, and so the right-hand side in (70) is constant. If $H - \tau_0 \omega_p^* > 0$, the general solution reads

$$\dot{\bar{\kappa}}(x) = \frac{\dot{\tau}}{H\alpha^2} + C_1 \cosh \frac{\alpha x}{l} + C_2 \sinh \frac{\alpha x}{l}, \quad (71)$$

where C_1 and C_2 are integration constants and

$$a = \sqrt{1 - \tau_0 \omega_p^* / H}$$

is a dimensionless parameter introduced for convenience.



(b)

Fig. 6: (a) Stress versus average strain for different thicknesses of the shear layer, (b) normalized torque, M/D^3 , versus surface shear strain for twisted copper wires of various diameters D (after Fleck et al., 1994)

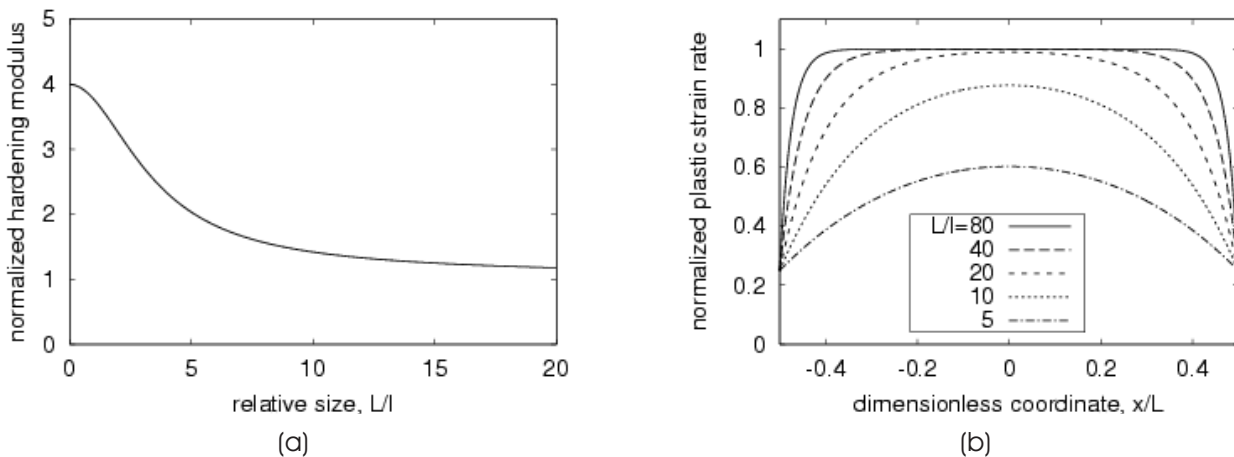


Fig. 7: Implicit gradient plasticity: (a) dependence of the apparent hardening modulus on the size of the specimen (layer thickness), (b) initial distribution of plastic strain rate across the layer

The particular solution depends on the boundary conditions, which should be formulated in terms of the nonlocal cumulative plastic strain and its normal derivative. For homogeneous Neumann boundary conditions, $\bar{\kappa}'(\pm L/2) = 0$, the integration constants vanish and the solution is uniform. Consequently, no size effect is predicted by the model. This is similar to what happens for the Aifantis model with the homogeneous Neumann boundary conditions $\kappa'(\pm L/2) = 0$. A stiff boundary with a perfect bond inhibits propagation of voids, which can be expressed by the condition that $\bar{\kappa}$ vanishes at the boundary. With this homogeneous Dirichlet condition applied at both parts of the boundary, the particular solution of (70) becomes

$$\dot{\bar{\kappa}}(x) = \frac{\dot{\tau}}{H\alpha^2} \left[1 - \frac{\cosh(\alpha x/l)}{\cosh(\alpha L/2l)} \right] \quad (72)$$

and the corresponding local cumulative plastic strain rate is

$$\dot{\kappa}(x) = \frac{\dot{\tau}}{H\alpha^2} \left[1 - (1 - \alpha^2) \frac{\cosh(\alpha x/l)}{\cosh(\alpha L/2l)} \right]. \quad (73)$$

Following the same procedure as in Section 3, we find that the shear stress rate is linked to the average shear strain rate by the linear relation

$$\dot{\tau} = \tilde{G}_{ep} \dot{\bar{\gamma}}, \quad (74)$$

where

$$\tilde{G}_{ep} = \left(\frac{1}{G} + \frac{1}{\tilde{H}} \right)^{-1} \quad (75)$$

is the elastoplastic shear modulus and

$$\tilde{H} = \frac{H\alpha^2}{1 - (1 - \alpha^2) \frac{2l}{\alpha L} \tanh \frac{\alpha L}{2l}} \quad (76)$$

is the size-dependent hardening modulus. For very large sizes L , the hardening modulus tends to $H\alpha^2$, which is the value corresponding to the local theory with ω_p computed from κ instead of $\bar{\kappa}$. For values of L comparable to l , the hardening modulus increases with decreasing size, which means that the average response becomes stiffer. So the present model predicts a qualitatively similar trend as the explicit version of gradient plasticity.

The ratio between the size-dependent hardening modulus and its large-size limit is plotted in Fig. 7a as a function of the relative size, L/l , for parameter $\alpha = 0.5$. Fig. 7b shows the distribution of plastic strain rate across the layer for different ratios L/l . Even though the general trends are the same as for the explicit gradient model, certain differences can be revealed by comparing Fig. 7 with Fig. 5. For the explicit model, the hardening modulus tends to infinity as the layer thickness approaches zero, while for the implicit model it tends to a finite value. The local plastic strain on the boundary is zero for the explicit model (as dictated by the boundary condition), while for the implicit model it is positive (because the boundary condition is formulated in terms of **nonlocal** plastic strain). For the explicit model, the profile of plastic strain distribution across the layer thickness keeps the same shape and grows proportionally during hardening, while for the implicit model the analytical solution covers only the initial distribution of the plastic strain rate and later the shape of the profile would change.

4 Strain localization due to softening

4.1 Problems with objective description of softening

In many structures subjected to high levels of solicitation, the initially smooth distribution of strain at a certain critical stage abruptly changes into a localized pattern with large strains concentrated in relatively narrow regions. This phenomenon, called **strain localization**, can be caused for instance by the softening character of the material response. The general definition of **softening** is more involved, but in the one-dimensional case softening means decreasing stress at increasing strain. The physical source of softening usually resides in the growth and coalescence of defects such as voids or cracks. From the micromechanical point of view, this means that the internal structure of the material is evolving and the approximate description of the material as a macroscopically homogeneous one may become questionable. Indeed, softening incorporated into standard inelastic continuum models leads to serious mathematical and numerical problems, and enriched theories are needed to provide an objective description of the softening process.

The essence of the localization problem will be explained using a one-dimensional example, which could be interpreted as localization of shear strain in a layer of elastoplastic material between two rigid plates (already studied in the previous section in the context of size effects). However, to better show various facets of nonlocal theories and their broad application field, we will discuss the closely related problem of a prismatic bar of length L and cross-sectional area A under uniaxial tension. The bar is made of a softening material described by a continuum damage model rather than by an elastoplastic model. Damage mechanics is frequently used for quasibrittle materials such as concrete under predominantly tensile loading. Many results and conclusions of the present section could be directly reinterpreted in terms of shear bands in ductile materials such as metals or confined soils.

In the one-dimensional setting, the stress-strain law used by the damage model reads

$$\sigma = (1 - \Omega)E\varepsilon, \quad (77)$$

where σ and ε are the (normal) stress and strain, E is Young's modulus of elasticity, and Ω is a scalar damage variable characterizing the current size and density of defects that reduce the effective area capable of transmitting stress. In the "virgin", undamaged state of material with no defects (or with very small initial defects that are incorporated in the elastic modulus), the value of the damage parameter is zero, and it remains zero throughout the elastic stage of loading. When the elastic limit is exceeded, damage starts growing and the elastically computed stress $E\varepsilon$ is reduced by the integrity factor $1 - \Omega$. The limit value $\Omega = 1$ corresponds to a fully damaged material that can no longer carry any stress. The growth of damage must be described by an appropriate damage evolution law for the internal variable Ω . This law could be postulated in the rate form, but a particularly simple and practical formulation is obtained with a damage law in the total form

$$\dot{\Omega} = g(\kappa), \quad (78)$$

where κ corresponds to the maximum level of strain reached in the previous history of the material. Mathematically, the internal variable κ is defined by the loading-unloading conditions

$$\dot{\kappa} \geq 0, \quad \varepsilon - \kappa \leq 0, \quad (\varepsilon - \kappa)\dot{\kappa} = 0. \quad (79)$$

During monotonic loading, κ is equal to the strain ε , and so the damage evolution function g that appears in (78) can easily be identified from the monotonic stress-strain curve.

Now suppose that the stress-strain diagram is linear up to a certain strain level ε_0 , after which stress decreases as a linear function of strain until the zero stress level is reached at strain ε_f (Fig. 8a). This linear softening model can be considered as the simplest description of concrete cracking under tension. Due to the heterogeneous and quasibrittle nature of the material, a contiguous stress-free crack across the entire section of a bar does not form instantaneously but is obtained as the final result of propagation and coalescence of many smaller cracks. Consequently, even after the onset of cracking the bar can still transmit a certain force but its residual strength decreases as the cracks evolve.

Under static loading and in the absence of body forces, the stress in the bar must be uniformly distributed. In the elastic range, strain is a unique function of stress, and so the strain distribution must be uniform as well. When the peak stress (tensile strength) f_t is attained, the uniqueness of the response is lost. Stress must still remain uniform and it can only decrease, but a given stress level can be reached either by further stretching the material into the softening regime, or by elastic unloading. Consequently, there are many different spatial distributions of the strain increments that lead to the same uniform stress decrement and thus represent a valid solution satisfying the equilibrium equation and the constitutive law. The compatibility equations do not represent any important constraint in the one-dimensional case, because the strain field can always be integrated to yield the displacement field. For example, the material can be softening in an interval of length L_s and unloading everywhere else. When the stress is completely relaxed to zero, the strain in the softening region is ε_f and the strain in the unloading region vanishes; the total elongation of the bar is therefore $u = L_s \varepsilon_f$. The length L_s remains undetermined, and it can have any value between 0 and L . This means that the problem has infinitely many solutions; the corresponding post-peak branches of the load-displacement diagram fill the fan shown in Fig. 8b.

The ambiguity is removed if imperfections are taken into account. Real material properties and sectional dimensions cannot be perfectly uniform. Suppose that the strength in a small region is slightly lower than in the remaining portion of the bar. When the applied stress reaches the reduced strength, softening starts and the stress decreases. Consequently, the material outside the weaker region must unload elastically, because its strength has not been exhausted. This leads to the conclusion that the size of the softening region cannot exceed the size of the region with minimum strength. Such a region can be arbitrarily small, and the corresponding softening

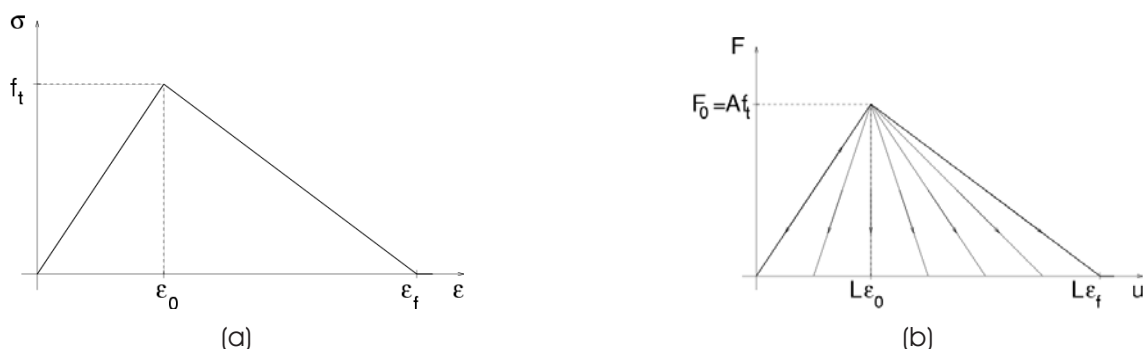


Fig. 8: (a) Stress-strain diagram with linear softening, (b) fan of possible post-peak branches of the load-displacement diagram

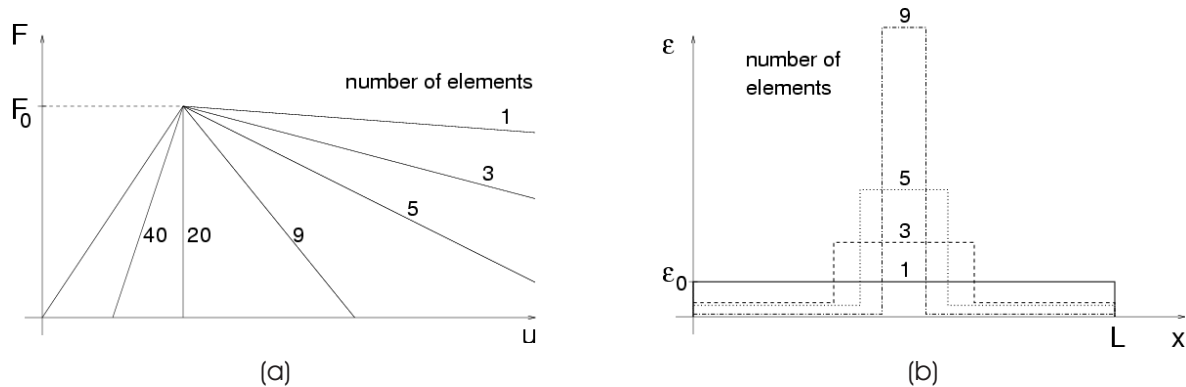


Fig. 9: Pathological effects of mesh refinement on the numerical results obtained with the local damage model: mesh-dependence of (a) load-displacement diagrams, (b) strain profiles

branch can be arbitrarily close to the elastic branch of the load-displacement diagram. Thus the standard strain-softening continuum formulation leads to a solution that has several pathological features: (i) the softening region is infinitely small; (ii) the load-displacement diagram always exhibits snapback, independently of the structural size and of the material ductility; (iii) the total amount of energy dissipated during the failure process is zero.

From the mathematical point of view, these annoying features are related to the so-called **loss of ellipticity** of the governing differential equation. In the present one-dimensional setting, loss of ellipticity occurs when the tangent modulus ceases to be positive, i.e., it coincides with the onset of softening (this is not always the case in multiple dimensions). The boundary value problem becomes ill-posed, i.e., it does not have a unique solution with continuous dependence on the given data. From the numerical point of view, ill-posedness is manifested by pathological sensitivity of the results to the size of finite elements. For example, suppose that the bar is discretized by N_e two-node elements with linear displacement interpolation. If the numerical algorithm properly captures the most localized solution, the softening region extends over one element, and we have $L_s = L/N_e$. The slope of the post-peak branch therefore strongly depends on the number of elements, and it approaches the initial elastic slope as the number of elements tends to infinity; see Fig. 9a, constructed for a linear softening law with $\varepsilon_f/\varepsilon_0 = 20$. The strain profiles at $u = 2L\varepsilon_0$ for various mesh refinements are plotted in Fig. 9b (under the assumption that the weakest spot is located at the center of the bar). In the limit, the profiles tend to $2L\varepsilon_0\delta(x - L/2)$ where δ denotes the Dirac distribution. The limit solution represents a displacement jump at the center, with zero strain everywhere else.

4.2 Nonlocal formulation serving as localization limiter

In real materials, inelastic processes typically localize in narrow bands that initially have a small but nonzero width. Propagation and coalescence of microdefects in the localization band can eventually lead to the formation of a displacement discontinuity, e.g., of a macroscopic stress-free crack or a sharp slip line. The initial thickness of the localization band depends on the material microstructure and is usually of the same order of magnitude as the characteristic material length, determined by the size or spacing of dominant het-

erogeneities. Therefore, it is natural to expect that enriched continuum theories can better reflect the actual deformation and failure processes and restore mathematical well-posedness of the boundary value problem. Indeed, when properly formulated, nonlocal or gradient enrichments regularize the problem in the sense that the resulting strain field is highly concentrated in certain narrow zones but remains continuous. The corresponding numerical solutions converge upon mesh refinement to a physically meaningful limit, and the numerical results do not suffer by pathological sensitivity to the discretization. Enrichments that prevent localization of strain into arbitrarily small regions are called **localization limiters**.

Nonlocal material models of the integral type were first exploited as localization limiters in the 1980s. After some preliminary formulations exploiting the concept of an imbricate continuum (Bažant, Belytschko and Chang, 1984), the nonlocal damage theory emerged (Pijaudier-Cabot and Bažant, 1987). Nonlocal formulations were then developed for a number of constitutive theories, including softening plasticity, smeared cracking, microplane models, etc. For a list of references, see e.g. Bažant and Jirásek (2002) or Chapter 26 in Jirásek and Bažant (2002).

Generally speaking, the nonlocal approach consists in replacing a certain variable by its nonlocal counterpart obtained by weighted averaging over a spatial neighborhood of each point under consideration. In nonlocal elasticity, the averaged quantity is usually the strain. Nonlocal elastic models can correctly reflect the experimentally observed dispersion of short elastic waves. However, in typical structural applications, the strain in the elastic regime remains relatively smoothly distributed (with the exception of stress concentrations and singularities around specific points, e.g., tips of pre-existing sharp cracks). Steep strain gradients appear only after the onset of localization and are accompanied by highly nonuniform distribution of damage. Therefore, most nonlocal models serving as localization limiters reduce to standard local elasticity at low strain levels, and the nonlocal effects are considered only in the inelastic regime. For instance, one widely used nonlocal damage formulation replaces the strain in the loading-unloading conditions (79) by its nonlocal average, $\bar{\varepsilon}$, while strain entering the stress-strain law (77) is still considered as local. According to the modified loading-unloading conditions,

$$\dot{\kappa} \geq 0, \quad \bar{\varepsilon} - \kappa \leq 0, \quad (\bar{\varepsilon} - \kappa) \dot{\kappa} = 0, \quad (80)$$

the internal variable κ has the meaning of the largest previously reached value of nonlocal strain.

If strain has a tendency to localize in a very narrow region, e.g., in one single cross section, the nonlocal strain becomes high not only in this region but also in its close neighborhood. This leads to damage growth in that neighborhood, and the local strain at the damaged points must be increased in order to keep the stress distribution uniform. By this mechanism, strain and damage are prevented from localizing into a single cross section. The localized region always has a certain finite size, controlled by the length parameter that appears in the definition of the nonlocal weight function.

4.3 Localization analysis

The initial bifurcation from a uniform state can be studied analytically under the simplified assumption that the strain keeps increasing at all points both for the fundamental uniform solution and the bifurcated nonuniform solution. If the bifurcated solution is considered as a small perturbation of the uniform solution, the stress perturbation $\delta\sigma$ can be linked to the strain perturbation $\delta\varepsilon$ by the linearized equation

$$\delta\sigma = (1 - \Omega)E \delta\varepsilon - \delta\Omega E\varepsilon \quad (81)$$

and the perturbation of the damage field $\delta\Omega$ is linked to the strain perturbation by the nonlocal relation

$$\delta\Omega(x) = g^* \int_L \alpha(x, \xi) \delta\varepsilon(\xi) d\xi, \quad (82)$$

where g^* is the derivative of the damage function g with respect to its argument κ evaluated for the fundamental (uniform) solution and therefore independent of the spatial coordinate.

Even though the fundamental solution is considered as static, we will analyze the evolution of the perturbation as a dynamic process. This approach provides more insight into the localization phenomena. In dynamics, the stress perturbation and the displacement perturbation must satisfy the equation of motion

$$\rho\delta\ddot{u} - \delta\sigma' = 0. \quad (83)$$

For an infinite body and nonlocal weight function in the form $\alpha(x, \xi) = \alpha_0(x - \xi)$, the assumed solution for δu in the form of a harmonic wave substituted into (81)–(83) leads to the dispersion equation

$$-\rho\omega^2 + (1 - \Omega)Ek^2 - g^* E\varepsilon k^2 \alpha_0^*(k) = 0, \quad (84)$$

where ω is the circular frequency of the wave (not to be confused with the damage variable Ω), k is the wave number, and α_0^* is the Fourier image of the weight function. The resulting dispersion law reads

$$\omega = c_0 k \sqrt{1 - \Omega - g^* \varepsilon \alpha_0^*(k)}. \quad (85)$$

The Fourier image α_0^* has a unit value at $k = 0$ and smaller values at positive wave numbers k . So if $1 - \Omega - g^* \varepsilon > 0$, all wave numbers have real positive frequencies and a small perturbation of an arbitrary shape propagates through the body but does not grow in magnitude. If $1 - \Omega - g^* \varepsilon < 0$, there exists a band of low wave numbers between zero and a positive limit k_{crit} for which the frequencies are imaginary. This indicates an instability if the perturbation contains a component with a sufficiently long wavelength. In statics, a stationary

wave of wavelength $2\pi/k_{crit}$ can be superimposed on the fundamental uniform solution without violating the equilibrium condition.

The critical wave number can be determined from the condition $\omega = 0$. In a monotonic loading process, the damage variable Ω is a function of strain, $\Omega = g(\varepsilon)$, and also the derivative $g^* = dg/d\varepsilon$ is a function of strain. The expression under the square root in (85) vanishes for the wave number satisfying the condition

$$\alpha_0^*(k_{crit}) = \frac{1 - g(\varepsilon)}{\varepsilon \frac{dg(\varepsilon)}{d\varepsilon}}. \quad (86)$$

For a given damage function g and nonlocal weight function α_0 , this equation can be solved (either analytically or numerically) and the critical wavelength $L_{crit} = 2\pi/k_{crit}$ can be determined as a function of strain.

For instance, for an exponential softening curve (more realistic for concrete than the linear one), the damage function is given by

$$g(\varepsilon) = 1 - \frac{\varepsilon_0}{\varepsilon} \exp\left(-\frac{\varepsilon - \varepsilon_0}{\varepsilon_f - \varepsilon_0}\right) \quad (87)$$

where ε_0 is the limit elastic strain and ε_f is another parameter controlling the slope of the softening curve. Substituting (87) and the Fourier image $\alpha_0^*(k) = \exp(-k^2 a^2/4)$ of the Gauss-like weight function (52) into (86), we obtain

$$L_{crit} = \frac{2\pi}{k_{crit}} = \frac{\pi a}{\sqrt{\ln\left(1 + \frac{\varepsilon}{\varepsilon_f - \varepsilon_0}\right)}}. \quad (88)$$

Of course, this expression is valid only in the inelastic range, i.e., for $\varepsilon \geq \varepsilon_0$. For the present damage function (87), the elastic limit coincides with the peak of the stress-strain curve. If, for instance, $\varepsilon_f = 11\varepsilon_0$, the critical wavelength for the uniform state at peak stress is $L_{crit} = \pi a / \sqrt{\ln 1.1} = 10.176 a$. This example shows that the critical wavelength is proportional to the model parameter that plays the role of the internal material length (and is often denoted as l).

In an infinite bar at peak stress, the appearance of a stationary wave of wavelength L_{crit} corresponds to a particular periodic localization pattern. In reality, there exists an energetically more favorable localization pattern that is not periodic but is concentrated into one single interval of length slightly below the critical wavelength. The exact size of the localized zone could be solved from a Fredholm integral equation of the second kind combined with the complementarity conditions, but since the solution is not available in closed form, we will not elaborate on that. If the bar is finite but longer than L_{crit} , it can be expected that at peak stress the subsequent increments of strain localize into a band of thickness approximately equal to L_{crit} . For shorter bars, localization can be delayed and the actual behavior depends on the specific form of the nonlocal averaging operator around the boundaries. However, in general it can be expected that localization occurs when the critical wavelength becomes approximately equal to the bar length. As shown in Fig. 10a, the critical wavelength monotonically decreases with increasing strain and asymptotically tends to zero. This property has important implications for the evolution of the localized strain

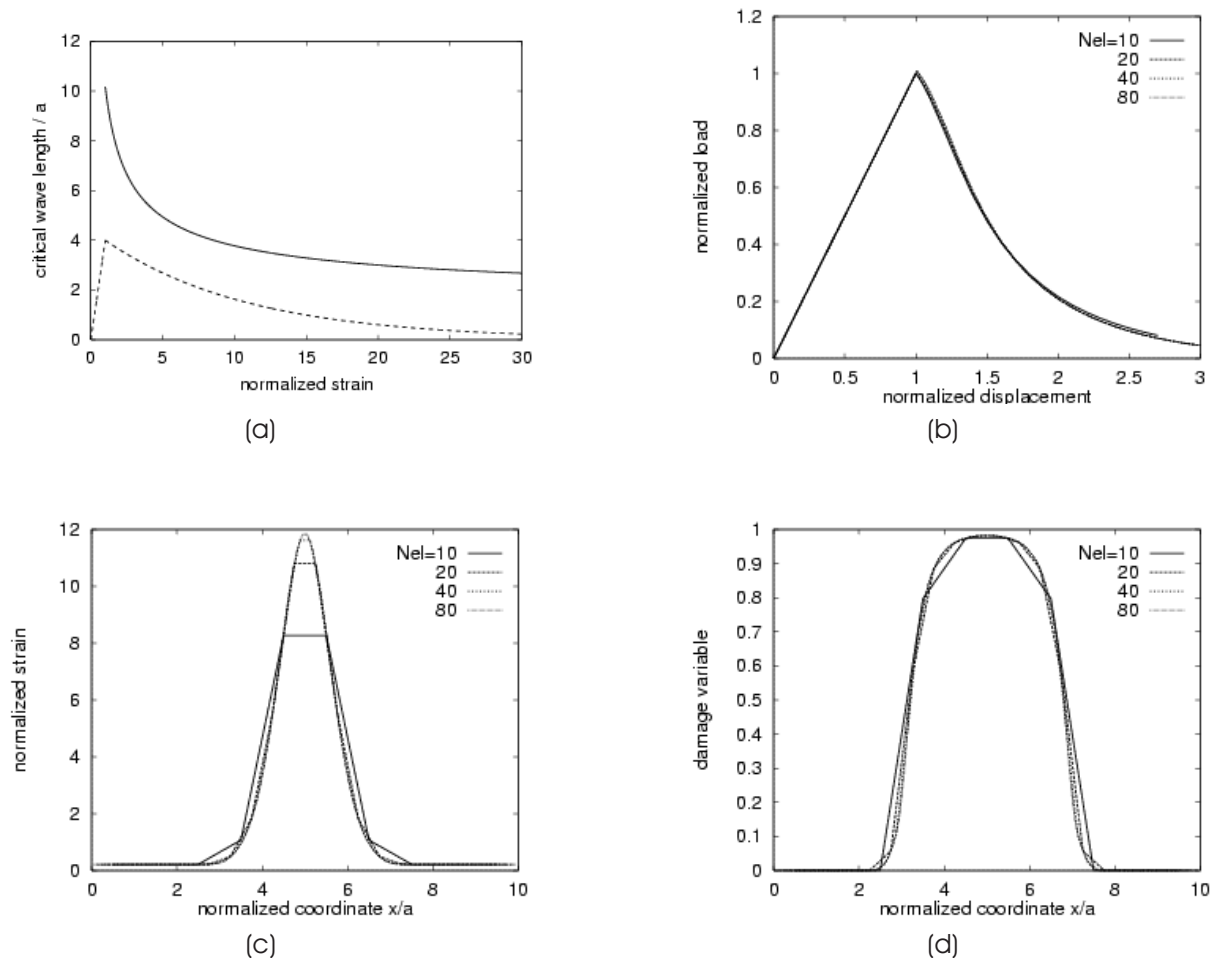


Fig. 10: Nonlocal damage model: (a) critical wavelength as a function of strain (for reference, the stress-strain curve is plotted by the dashed line), (b) convergence of load-displacement diagram, (c) convergence of strain profile, (d) convergence of damage profile; Nel = number of elements

profile – it indicates that the active zone in which strains are increasing tends to shrink during the loading process. Such a trend has indeed been confirmed by numerical simulations.

4.4 Mesh sensitivity and size effect

Fig. 10b shows the load-displacement diagram for strain localization in a bar under uniaxial tension, calculated by the finite element method using a nonlocal damage model with the weight function (53) and with the exponential damage function (87). As the number of elements N_{el} increases, the load-displacement curve rapidly converges to the exact solution. Convergence of strain and damage profiles generated by an applied displacement $u = 2L\varepsilon_0$ is documented in Fig. 10c, d. In contrast to the local model, the process zone does not shrink to a single point as the mesh is refined. Its size is controlled by parameter ℓ that sets the internal length scale. This example demonstrates that the nonlocal formulation serves as a localization limiter and provides an objective description of the localization process, with no pathological sensitivity of the numerical results to the discretization.

Another important advantage of nonlocal softening models is that they can realistically describe the size effect on nominal strength of quasibrittle materials. The **nominal strength** is understood as the peak load divided by a characteristic area of the structure. According to the standard (local) version of

perfect plasticity theory, the nominal strength for a set of geometrically similar structures of different sizes should be the same, independent of the size. For instance, for a beam of a rectangular cross section, subjected to three-point bending, the plastic collapse load is

$$F_0 = \frac{4M_p}{L} = \frac{bD^2\sigma_0}{L}, \quad (89)$$

where b is the width and D the depth of the cross section, L is the span, M_p is the plastic limit moment of the cross section, and σ_0 is the yield stress. The nominal strength defined as the peak load divided by the cross-sectional area,

$$\sigma_N = \frac{F_0}{bD} = \frac{D\sigma_0}{L} \quad (90)$$

is equal to the material property σ_0 multiplied by the geometrical factor D/L , which depends only on the shape of the structure but not on its size (assuming proportional scaling of all structural dimensions for three-dimensional similarity, or at least of the in-plane dimensions L and D for two-dimensional similarity). In contrast to elastic-perfectly plastic structures, structures made of quasibrittle materials often exhibit a strong size dependence of the nominal strength. For certain specimen geometries with initial notches scaled proportionally to other structural dimensions, the large-size limit is adequately described by linear elastic fracture mechanics,

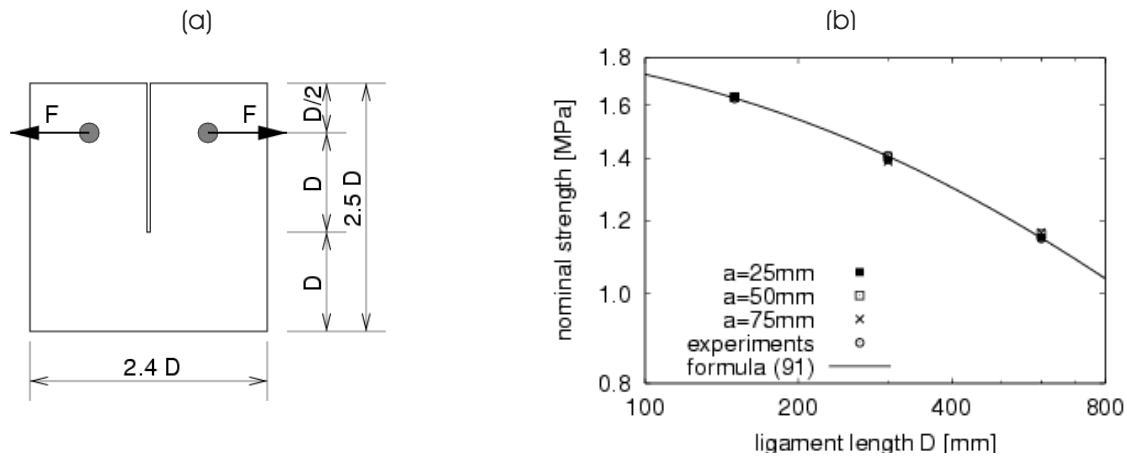


Fig. 11: (a) Compact tension specimen, (b) size effect diagram

which predicts proportionality of the nominal strength to $1/\sqrt{D}$. The transition between no size effect for small sizes and strong size effect for large sizes can be approximated by Bažant's (1984) formula

$$\sigma_N = \frac{\sigma_{N0}}{\sqrt{1 + \frac{D}{D_0}}}, \quad (91)$$

where D_0 and σ_{N0} are parameters to be determined by fitting of the experimental results.

The transitional size effect characteristic of quasibrittle structures is nicely reproduced by nonlocal softening models, if the model parameters are chosen correctly. As an example, consider the compact tension test of concrete depicted in Fig. 11a. The experimental results obtained by Wittmann, Mihashi and Nomura (1990) are shown in the logarithmic size effect diagram in Fig. 11b, along with the optimal fit by formula (91) and the results of numerical simulations. The role of the characteristic structural size D is played by the ligament length, and the nominal strength is defined as the peak load divided by the ligament area, bD , where b is the out-of-plane thickness of the specimen. The experimentally measured size effect can be accurately reproduced by a nonlocal isotropic damage model with different combinations of internal length parameter a and parameter ϵ_f of the damage function (86). If one parameter is fixed, the other can be determined by optimal fitting, but both parameters cannot be determined simultaneously from the size effect on nominal strength only. A unique parameter identification requires additional information such as the distribution of strain or damage in the process zone (Geers, de Borst, Brekelmans and Peerlings, 1999) or the size effect on fracture energy (Jirásek, Rolshoven and Grassl, 2004).

5 Concluding remarks

The common denominator of all examples presented in the preceding sections is that the characteristic wavelength of the deformation field becomes comparable to the characteristic size of the internal material structure. Here, the notion of characteristic wavelength has to be understood in a broad sense, not only as the spatial period of a dynamic phenomenon but also as the length on which the value of strain

changes substantially in static problems. Such a more general definition could be based e.g. on a suitably normalized ratio between the maximum strain and the maximum strain gradient (both in absolute values). Thus the characteristic wavelength is necessarily close to the internal material length if the size of the specimen is not much larger than the size and spacing of major heterogeneities, or if strain localizes due to softening.

The enrichment terms introduced by various generalized continuum theories have a differential or integral character, but all of them can be considered as nonlocal, at least in the weak sense. They always introduce a model parameter with the dimension of length, which reflects the internal length scale of the material.

Three typical cases covered here encompass static and dynamic phenomena, linear and nonlinear behavior, and three classes of material laws, namely elasticity, plasticity and continuum damage mechanics. This shows that the nonlocal enrichments can be useful in a wide range of mechanical problems. Unfortunately, so far there is no general and universally accepted theory covering this entire range within one unified framework. Even though the first nonlocal theories were pioneered in the 1960s, there many problems remain open and many issues unresolved. Some of the most challenging questions include the correct formulation of boundary conditions, micromechanical justification of models with nonlocal internal variables, or identification techniques for the internal length parameter and its possible evolution.

References

- [1] Aifantis E. C.: "On the microstructural origin of certain inelastic models." *Journal of Engineering Mechanics and Technology, ASME*, Vol. **106** (1984), p. 326–330.
- [2] Aifantis E. C.: "Update on a class of gradient theories." *Mechanics of Materials*, Vol. **35** (2003), p. 259–280.
- [3] Askes H., Metrikine A. V.: "One-dimensional dynamically consistent gradient elasticity models derived from a discrete microstructure. Part 2: Static and dynamic response." *European Journal of Mechanics A*, Vol. **21** (2002), p. 573–588.

- [4] Bažant Z. P.: “Size effect in blunt fracture: Concrete, rock, metal.” *Journal of Engineering Mechanics, ASCE*, Vol. **110** (1984), p. 518–535.
- [5] Bažant Z. P., Chang, T.-P.: “Instability of nonlocal continuum and strain averaging.” *Journal of Engineering Mechanics, ASCE*, Vol. **110** (1984), p. 2015–2035.
- [6] Bažant Z. P., Jirásek M.: “Nonlocal integral formulations of plasticity and damage: Survey of progress.” *Journal of Engineering Mechanics, ASCE*, Vol. **128** (2002), p. 1119–1149.
- [7] Bažant Z. P., Belytschko T. B., Chang, T.-P.: “Continuum model for strain softening.” *Journal of Engineering Mechanics, ASCE*, Vol. **110** (1984), p. 1666–1692.
- [8] Eringen A. C.: “Linear theory of nonlocal elasticity and dispersion of plane waves.” *International Journal of Engineering Science*, Vol. **10** (1972), p. 425–435.
- [9] Fish J., Chen W., Nagai, G.: “Nonlocal dispersive model for wave propagation in heterogeneous media. Part 1: One-dimensional case.” *International Journal for Numerical Methods in Engineering*, Vol. **54** (2002a), p. 331–346.
- [10] Fish J., Chen W., Nagai G.: “Nonlocal dispersive model for wave propagation in heterogeneous media. Part 2: Multi-dimensional case.” *International Journal for Numerical Methods in Engineering*, Vol. **54** (2002b), p. 347–363.
- [11] Fleck N. A., Hutchinson J. W.: “A reformulation of strain gradient plasticity.” *Journal of the Mechanics and Physics of Solids*, Vol. **49** (2001), p. 2245–2271.
- [12] Fleck N. A., Muller G. M., Ashby M. F., Hutchinson J. W.: “Strain gradient plasticity: theory and experiment.” *Acta Metallurgica et Materialia*, Vol. **42** (1994), p. 475–487.
- [13] Geers M. G. D., de Borst R., Brekelmans W. A. M., Peerlings R. H. J.: “Validation and internal length scale determination for a gradient damage model: Application to short glass-fibre-reinforced polypropylene.” *International Journal of Solids and Structures*, Vol. **36** (1999), p. 2557–2583.
- [14] Geers M. G. D., Engelen R. A. B., Ubachs R. J. M.: “On the numerical modelling of ductile damage with an implicit gradient-enhanced formulation.” *Revue européenne des éléments finis*, Vol. **10** (2001), p. 173–191.
- [15] Green A. E., Rivlin R. S.: “Simple force and stress multipoles.” *Archive for Rational Mechanics and Analysis*, Vol. **16** (1964), p. 325–353.
- [16] Jirásek M., Bažant Z. P.: “Inelastic Analysis of Structures.” John Wiley and Sons, Chichester, U.K., 2002.
- [17] Jirásek M., Rolshoven, Grassl, P.: “Size effect on fracture energy induced by nonlocality.” *International Journal for Numerical and Analytical Methods in Engineering*, Vol. **28** (2004), in press.
- [18] Lakes R. S.: “Experimental microelasticity of two porous solids.” *International Journal of Solids and Structures*, Vol. **22** (1986), p. 55–63.
- [19] Ma Q., Clarke D. R.: “Size dependent hardness in silver single crystals.” *Journal of Materials Research*, Vol. **10** (1995), p. 853–863.
- [20] Metrikine A. V., Askes H.: “One-dimensional dynamically consistent gradient elasticity models derived from a discrete microstructure. Part 1: Generic formulation.” *European Journal of Mechanics A*, Vol. **21** (2002), p. 555–572.
- [21] Mindlin R. D.: “Micro-structure in linear elasticity.” *Archive for Rational Mechanics and Analysis*, Vol. **16** (1964), p. 51–78.
- [22] Mindlin R. D.: “Second gradient of strain and surface tension in linear elasticity.” *International Journal of Solids and Structures*, Vol. **1** (1965), p. 417–438.
- [23] Morrison J. L. M.: “The yield of mild steel with particular reference to the effect of size of specimen.” *Proceedings of the Institute of Mechanical Engineers*, Vol. **142** (1939), p. 193–223.
- [24] Nagai G., Fish J., Watanabe K.: “Stabilized nonlocal model for wave propagation in heterogeneous media.” *Computational Mechanics*, Vol. **33** (2004), p. 144–153.
- [25] Nix W. D.: “Mechanical properties of thin films.” *Metallurgical Transactions*, Vol. **20A** (1989), p. 2217–2245.
- [26] Pijaudier-Cabot G., Bažant Z. P.: “Nonlocal damage theory.” *Journal of Engineering Mechanics, ASCE*, Vol. **113** (1987), p. 1512–1533.
- [27] Poole W. J., Ashby M. F., Fleck N. A.: “Microhardness of annealed and work-hardened copper polycrystals.” *Scripta Metallurgica et Materialia*, Vol. **34** (1996), p. 559–564.
- [28] Rogula D.: “Introduction to nonlocal theory of material media”, in D. Rogula (ed.), *Nonlocal Theory of Material Media*, No. 268 in CISM Courses and Lectures, Springer Verlag, Wien and New York, 1982, p. 125–222.
- [29] Shu J. Y., Fleck N. A., van der Giessen E., Needleman A.: “Boundary layers in constrained plastic flow: comparison of nonlocal and discrete dislocation plasticity.” *Journal of the Mechanics and Physics of Solids*, Vol. **49** (2001), p. 1361–1395.
- [30] Stolken J. S., Evans A. G.: “A microbend test method for measuring the plasticity length scale.” *Acta Metallurgica et Materialia*, Vol. **46** (1998), p. 5109–5115.
- [31] Toupin R. A.: “Elastic materials with couple-stresses.” *Archive for Rational Mechanics and Analysis*, Vol. **11** (1962), p. 385–414.
- [32] Wittmann F. H., Mihashi H., Nomura N.: “Size effect on fracture energy of concrete.” *Engineering Fracture Mechanics*, Vol. **35** (1990), p. 107–115.
- [33] Yarnel J. L., Warren J. L., Koenig S. H.: In *Lattice Dynamics*, R. F. Wallis (ed.), Pergamon Press, 1965, p. 57.

Milan Jirásek

Laboratory of Structural and Continuum Mechanics
Swiss Federal Institute of Technology at Lausanne (EPFL)
1015 Lausanne, Switzerland