Journal of
Mechanics of
Materials and Structures

JOURNAL OF MECHANICS OF MATERIALS AND STRUCTURES

DYNAMIC RESPONSE OF MULTILAYER CYLINDERS: THREE-DIMENSIONAL ELASTICITY THEORY

ALEXANDER SHUPIKOV AND NATALIYA DOLGOPOLOVA

We suggest an analytic-numerical approach to solving the problem of vibrations of multilayer cylinders under impulse loading. The behavior of the cylinder is described by dynamic equations of three-dimensional elasticity theory. The number of layers in the pack and the thickness and mechanical characteristics of each layer are selected arbitrarily. The possibilities of the approach are proposed, and the validity of results obtained is demonstrated by numerical examples.

1. Introduction

Structural elements in the form of multilayer plates and shells are used extensively in different branches of machine building and civil engineering. The stressed-strained state (SSS) of real objects can be described most simply using the finite-element method, but the development of analytic and hybrid calculation methods is the focus of much current work.

In works where analytic and hybrid methods are used, the SSS is investigated most often by using different 2-dimensional discrete, continuous and discrete-continuous theories [Grigoliuk and Kogan 1972; Grigoliuk and Kulikov 1988; Reddy 1989; 1993; Noor and Rarig 1974; Noor and Burton 1989; Noor and Burton 1990a; 1996; Smetankina et al. 1995; Shupikov and Ugrimov 1997; Shupikov et al. 2004]. Using these theories, obtaining numerical results is relatively straightforward. In the process, one is faced with the complex problem of determining the limits within which these theories adequately describe the behavior of the object being investigated with a prescribed accuracy. At impulse and other nonstationary short-time actions, the solution of this problem is yet more challenging.

The behavior of a multilayer structure can be investigated most effectively in terms of three-dimensional elasticity theory. Pagano [1969] was one of the first to investigate this problem. He studied the cylindrical bending of a simply supported orthotropic infinite laminated strip under static loading. Later the exact solution of the bending problem for a finite-dimension plate under static loading was obtained [Pagano 1970a; 1970b; Little 1973; Noor and Burton 1990b].

Several other works concerned with the behavior of cylindrical shells under static loading can be mentioned. Thus, Ren [1987] studied the plane strain deformation of an infinitely long cylindrical shell subjected to a radial load changing harmonically along the circumference. In terms of three-dimensional elasticity theory, solutions were obtained for finite-length, cross-ply cylindrical shells, simply supported at both ends and subjected to transverse sinusoidal loading [Varadan and Bhaskar 1991], and for shell panels [Ren 1989] subjected to a transverse load that changes both in the axial and circumferential directions. Subsequent works [Bhaskar and Varadan 1993; 1994; Bhaskar and Ganapathysaran 2002; 2003] dealt with the three-dimensional analysis of cylindrical shells subjected to different kinds of loads acting both in the axial and circumferential directions. Liu [2000] presented static analyses of thick rectangular plane-view laminated plates, carried out in terms of the three-dimensional theory of elasticity using the differential quadrature element method.

Attention has also been given to the study of vibrations of multilayer cylindrical shells in terms of three-dimensional elasticity theory.

Noor and Rarig [1974] obtained equations of free vibrations of a simply supported laminated orthotropic circular cylinder based on linear three-dimensional elasticity theory. Kang and Leissa [2000] suggested a three-dimensional analysis method for determining the free vibrations and the form of the segment of a variable-thickness spherical shell. The displacement components in the meridional, normal, and circumferential directions were taken to be sinusoidal with respect to time, periodic in the circumferential direction, and were expanded into algebraic polynomials in the meridional and normal directions.

Weingarten and Reismann [1974] gave solutions for nonaxisymmetrical nonstationary vibrations of a uniform finite-length cylinder. They considered vibrations of uniform cylindrical shells within the framework of the three-dimensional theory of elasticity, and compared the results obtained with those given by other theories of shells. They showed that none of the two-dimensional theories could describe satisfactorily the wave-like character of the initial strain phase. Philippov et al. [1978] solved a similar problem, providing an analytic solution to the problem of axisymmetrical vibrations of a uniform infinite-length cylinder subjected to an impulse load.

Shupikov and Ugrimov [1999] have suggested an analytic-numerical method for solving the three-dimensional problem in the elasticity theory of nonstationary vibrations of multilayer plates subjected to impulse loads. The displacements in the tangential direction are expanded into a double Fourier series, and the partial derivatives in the transverse coordinate are replaced by their finite-difference version. As a result of these transformations, the problem of nonstationary vibration of a

multilayer plate under the action of an impulse load is reduced to integrating a system of ordinary differential equations with constant coefficients.

The objective of this work is to develop and generalize further the approach suggested by [Shupikov and Ugrimov 1999], and to investigate the vibrations of a thick multilayer closed cylindrical shell subjected to an impulse load.

2. Problem formulation

We consider a multilayer cylindrical shell of finite length A and radius R_0, which is composed of I uniform constant-thickness isotropic layers. The shell is referenced to the right-hand system of orthogonal curvilinear coordinates z, θ, r.

The coordinate surface is linked to the outer surface of the first layer, R_0 is the radius of inner surface of the shell (Figure 1). Contact between layers prevents their delamination and mutual slipping.

The behavior of each layer is described by Lamé's equations [Novatsky 1975]:

$$\mu^i\left(\nabla^2 u_r^i - \frac{u_r^i}{r^2} - \frac{2}{r^2}\frac{\partial u_\theta}{\partial \theta}\right) + (\lambda^i + \mu^i)\frac{\partial}{\partial r}\left(\frac{1}{r}\frac{\partial}{\partial r}(r u_r^i) + \frac{1}{r}\frac{\partial u_\theta^i}{\partial \theta} + \frac{\partial u_z^i}{\partial z}\right) - \rho^i\frac{\partial^2 u_r^i}{\partial t^2} = 0,$$

$$\mu^i\left(\nabla^2 u_\theta^i - \frac{u_\theta^i}{r^2} + \frac{2}{r^2}\frac{\partial u_r}{\partial \theta}\right)$$
$$+ (\lambda^i + \mu^i)\frac{1}{r}\frac{\partial}{\partial \theta}\left(\frac{1}{r}\frac{\partial}{\partial r}(r u_r^i) + \frac{1}{r}\frac{\partial u_\theta^i}{\partial \theta} + \frac{\partial u_z^i}{\partial z}\right) - \rho^i\frac{\partial^2 u_\theta^i}{\partial t^2} = 0,$$

$$\mu^i\nabla^2 u_z^i + (\lambda^i + \mu^i)\frac{\partial}{\partial z}\left(\frac{1}{r}\frac{\partial}{\partial r}(r u_r^i) + \frac{1}{r}\frac{\partial u_\theta^i}{\partial \theta} + \frac{\partial u_z^i}{\partial z}\right) - \rho^i\frac{\partial^2 u_z^i}{\partial t^2} = 0,$$
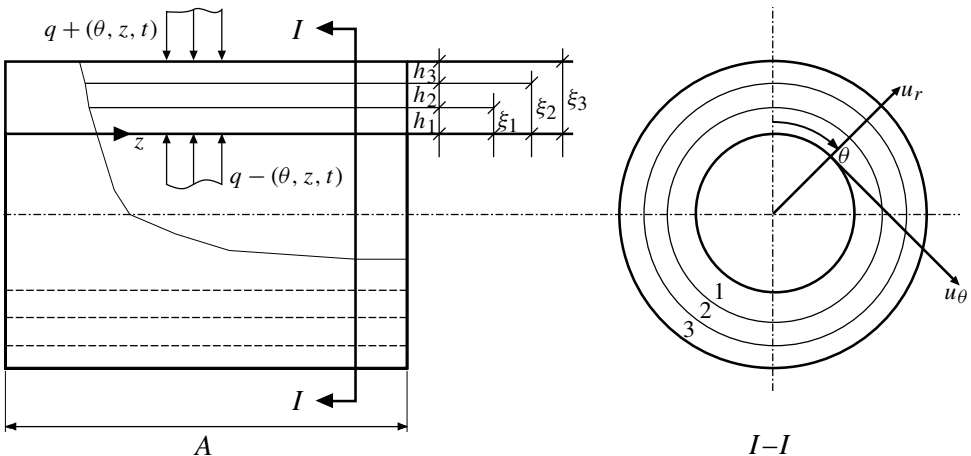
$$(1)$$

Figure 1. Multilayer cylindrical shell.

where $\nabla^2 = \partial^2/\partial r^2 + (1/r)\partial/\partial r + (1/r^2)\partial^2/\partial\theta^2 + \partial^2/\partial z^2$.

This system of equations is solved with the boundary conditions on the external surfaces of the first and I-th layers, namely

$$
\begin{aligned}
\sigma_{zr}^1 = \sigma_{r\theta}^1 = 0, \quad \sigma_{rr}^1 = -q^- \quad \text{for } r = R_0, \\
\sigma_{zr}^I = \sigma_{r\theta}^I = 0, \quad \sigma_{rr}^I = -q^+ \quad \text{for } r = R_0 + \xi^I, \ \xi^i = \sum_{j=1}^{i} h^j;
\end{aligned} \tag{2}
$$

the boundary conditions at the ends,

$$
\sigma_{zz}^i = u_r^i = u_\theta^i = 0 \quad \text{for } z = 0, L, \ i = \overline{1, I}; \tag{3}
$$

the contact conditions at adjacent layers,

$$
\begin{aligned}
u_r^i = u_r^{i+1}, \quad u_\theta^i = u_\theta^{i+1}, \quad u_z^i = u_z^{i+1}, \\
\sigma_{rr}^i = \sigma_{rr}^{i+1}, \quad \sigma_{r\theta}^i = \sigma_{r\theta}^{i+1}, \quad \sigma_{rz}^i = \sigma_{rz}^{i+1} \quad \text{for } r = R_0 + \xi^i, \ i = \overline{1, I-1};
\end{aligned} \tag{4}
$$

and the initial conditions

$$
\mathbf{u}^i(r, \theta, z, 0) = \frac{\partial \mathbf{u}^i(r, \theta, z, 0)}{\partial t} = 0 \quad \text{for } i = \overline{1, I}. \tag{5}
$$

Here i is the layer number, λ^i, μ^i are Lamé's coefficients, ρ^i is the specific density, and $\mathbf{u}^i = \{u_r^i, u_\theta^i, u_z^i\}$ is the displacement vector of a point in the i-th layer.

The stress tensor components are calculated from

$$
\sigma_{jk}^i = 2\mu^i \varepsilon_{jk}^i + \lambda^i \delta_{jk} \Delta^i, \tag{6}
$$

where δ is Kronecker's delta,

$$
\Delta^i = \varepsilon_{rr}^i + \varepsilon_{\theta\theta}^i + \varepsilon_{zz}^i \quad \text{for } i = \overline{1, I},
$$

and the components of the strain tensor have the form

$$
\begin{aligned}
\varepsilon_{rr}^i = \frac{\partial u_r^i}{\partial r}, \quad \varepsilon_{\theta\theta}^i = \frac{1}{r_i}\left(\frac{\partial u_i^\theta}{\partial \theta} + u_r^i\right), \quad \varepsilon_{zz}^i = \frac{\partial u_z^i}{\partial z}, \quad \varepsilon_{r\theta}^i = \frac{1}{2}\left(\frac{1}{r_i}\frac{\partial u_r^i}{\partial \theta} + \frac{\partial u_\theta^i}{\partial r} - \frac{u_\theta^i}{r_i}\right), \\
\varepsilon_{\theta z}^i = \frac{1}{2}\left(\frac{\partial u_\theta^i}{\partial z} + \frac{1}{r_i}\frac{\partial u_z^i}{\partial \theta}\right), \quad \varepsilon_{rz}^i = \frac{1}{2}\left(\frac{\partial u_z^i}{\partial r} + \frac{\partial u_r^i}{\partial z}\right), \quad \text{for } i = \overline{1, I}.
\end{aligned} \tag{7}
$$

Lamé's coefficients are linked to Young's modulus E^i and Poisson's coefficient ν^i by the relations

$$
\lambda^i = \frac{\nu^i E^i}{(1 + \nu^i)(1 - 2\nu^i)}, \quad \mu^i = \frac{E^i}{2(1 + \nu^i)}.
$$

3. Solution method

The displacements and the external load are expanded into double Fourier series with respect to the complete scheme of orthogonal functions satisfying the boundary conditions (3):

$$u_r^i = \sum_{m=1}^{M}\sum_{n=0}^{N} \Phi_{1mn}^i(r, t) B_{1mn}(z, \theta), \qquad u_\theta^i = \sum_{m=1}^{M}\sum_{n=1}^{N} \Phi_{2mn}^i(r, t) B_{2mn}(z, \theta),$$

$$u_z^i = \sum_{m=1}^{M}\sum_{n=0}^{N} \Phi_{3mn}^i(r, t) B_{3mn}(z, \theta), \quad q^{(\pm)} = \sum_{m=1}^{M}\sum_{n=0}^{N} q_{mn}^{(\pm)}(t) B_{3mn}(z, \theta), \tag{8}$$

for $i = \overline{1, I}$, where $B_{1mn}(z, \theta) = \sin\dfrac{m\pi z}{A}\cos n\theta$, $B_{2mn}(z, \theta) = \sin\dfrac{m\pi z}{A}\sin n\theta$, and

$$B_{3mn}(z, \theta) = \cos\frac{m\pi z}{A}\cos n\theta.$$

The partial derivatives of the functions $\Phi_{kmn}^i(r, t)$, $k = 1, 2, 3$, with respect to the coordinate r are replaced with their finite-difference presentations. For this we build a regular grid in each layer:

$$r^{i(l)} = R_0 + \xi^{i-1} + l\Delta^i, \qquad l = \overline{0, L^i}, \ \Delta^i = \frac{h^i}{L^i}, \ i = \overline{1, I}.$$

Here L^i is the number of nodes in the finite-difference grid in the i-th layer, $i = \overline{1, I}$.

The number of series terms M, N retained in expansion (8), and the number of nodes in the finite-difference grid in each of the layers L^i, is selected so as to ensure convergence of numeric results.

We set

$$\Phi_{kmn}^{i(l)} = \Phi_{kmn}^i(r^{i(l)}, t).$$

For approximation of partial derivatives, a three-point template is used [Forsythe and Wasov 1960]

$$\frac{\partial \Phi_{kmn}^{i(l)}}{\partial r} = \frac{\Phi_{kmn}^{i(l+1)} - \Phi_{kmn}^{i(l-1)}}{2\Delta^i}, \quad \frac{\partial^2 \Phi_{kmn}^{i(l)}}{\partial r^2} = \frac{\Phi_{kmn}^{i(l+1)} - 2\Phi_{kmn}^{i(l)} + \Phi_{kmn}^{i(l-1)}}{(\Delta^i)^2}. \tag{9}$$

As a result of these transforms, for $n = 0$ and $m = \overline{1, M}$ system (1) takes the form

$$(2\mu^i + \lambda^i)\left(\frac{1}{(\Delta^i)^2} - \frac{1}{r^{i(l)}}\frac{1}{2\Delta^i}\right)\Phi_{1m0}^{i(l-1)} + \frac{m\pi}{A}\frac{\lambda^i + \mu^i}{2\Delta^i}\Phi_{3m0}^{i(l-1)}$$

$$+ \left(-(2\mu^i + \lambda^i)\left(\frac{2}{(\Delta^i)^2} + \frac{1}{(r^{i(l)})^2}\right) - \mu^i\frac{m^2\pi^2}{A^2}\right)\Phi_{1m0}^{i(l)} \tag{10}$$

$$+ (2\mu^i + \lambda^i)\left(\frac{1}{(\Delta^i)^2} + \frac{1}{r^{i(l)}}\frac{1}{2\Delta^i}\right)\Phi_{1m0}^{i(l+1)} - \frac{m\pi}{A}\frac{\lambda^i + \mu^i}{2\Delta^i}\Phi_{3m0}^{i(l+1)} = \rho^i\frac{d^2\Phi_{1m0}^{i(l)}}{dt^2},$$

$$-\frac{m\pi}{A}\frac{\lambda^i+\mu^i}{2\Delta^i}\Phi_{1m0}^{i(l-1)}+\mu^i\left(\frac{1}{(\Delta^i)^2}-\frac{1}{r^{i(l)}}\frac{1}{2\Delta^i}\right)\Phi_{3m0}^{i(l-1)}$$

$$+\frac{m\pi}{A}\frac{\lambda^i+\mu^i}{r^{i(l)}}\Phi_{1m0}^{i(l)}-\left(\mu^i\frac{2}{(\Delta^i)^2}+\frac{m^2\pi^2}{A^2}(2\mu^i+\lambda^i)\right)\Phi_{3m0}^{i(l)} \tag{11}$$

$$+\frac{m\pi}{A}\frac{\lambda^i+\mu^i}{2\Delta^i}\Phi_{1m0}^{i(l+1)}+\mu^i\left(\frac{1}{(\Delta^i)^2}+\frac{1}{r^{i(l)}}\frac{1}{2\Delta^i}\right)\Phi_{3m0}^{i(l+1)}=\rho^i\frac{d^2\Phi_{3m0}^{i(l)}}{dt^2}.$$

Conditions (2), (4), and (5) become

$$(2\mu^1+\lambda^1)\frac{\Phi_{1m0}^{1(1)}-\Phi_{1m0}^{1(-1)}}{2\Delta^1}+\lambda^1\left(-\frac{m\pi}{A}\Phi_{3m0}^{1(0)}+\frac{1}{r^{1(0)}}\Phi_{1m0}^{1(0)}\right)=-q_{m0}^-, \tag{12}$$

$$(2\mu^I+\lambda^I)\frac{\Phi_{1m0}^{I(L^I+1)}-\Phi_{1m0}^{I(L^I-1)}}{2\Delta^I}+\lambda^I\left(-\frac{m\pi}{A}\Phi_{3m0}^{I(L^I)}+\frac{1}{r^{I(L^I)}}\Phi_{1m0}^{I(L^I)}\right)=-q_{m0}^+,$$

$$\frac{\Phi_{3m0}^{1(1)}-\Phi_{3m0}^{1(-1)}}{2\Delta^1}+\frac{m\pi}{A}\Phi_{1m0}^{1(0)}=0, \qquad\qquad \frac{\Phi_{3m0}^{I(L^I+1)}-\Phi_{3m0}^{I(L^I-1)}}{2\Delta^I}+\frac{m\pi}{A}\Phi_{1m0}^{I(L^I)}=0,$$

$$\Phi_{1m0}^{i(L^i)}=\Phi_{1m0}^{i+1(0)}, \qquad \Phi_{3m0}^{i(L^i)}=\Phi_{3m0}^{i+1(0)},$$

$$\frac{2\mu^i+\lambda^i}{2\Delta^i}\left(\Phi_{1m0}^{i(L^i+1)}-\Phi_{1m0}^{i(L^i-1)}\right)+\lambda^i\left(-\frac{m\pi}{A}\Phi_{3m0}^{i(L^i)}+\frac{1}{r^{i(L^i)}}\Phi_{1m0}^{i(L^i)}\right)$$

$$=\frac{2\mu^{i+1}+\lambda^{i+1}}{2\Delta^{i+1}}\left(\Phi_{1m0}^{i+1(1)}-\Phi_{1m0}^{i+1(-1)}\right)+\lambda^{i+1}\left(-\frac{m\pi}{A}\Phi_{3m0}^{i+1(0)}+\frac{1}{r^{i+1(0)}}\Phi_{1m0}^{i+1(0)}\right),$$

$$\mu^i\left(\frac{\Phi_{3m0}^{i(L^i+1)}-\Phi_{3m0}^{i(L^i-1)}}{2\Delta^i}+\frac{m\pi}{A}\Phi_{1m0}^{i(L^i)}\right)=\mu^{i+1}\left(\frac{\Phi_{3m0}^{i+1(1)}-\Phi_{3m0}^{i+1(-1)}}{2\Delta^{i+1}}+\frac{m\pi}{A}\Phi_{1m0}^{i+1(0)}\right),$$

$$\Phi_{1m0}^{i(s)}(0)=\frac{d\Phi_{1m0}^{i(s)}(0)}{dt}=\Phi_{2m0}^{i(s)}(0)=\frac{d\Phi_{2m0}^{i(s)}(0)}{dt}=\Phi_{3m0}^{i(s)}(0)=\frac{d\Phi_{3m0}^{i(s)}(0)}{dt}=0,$$

$$\text{all for } i=\overline{1,I}.$$

For pairs (m,n) with $n=\overline{1,N}$ and $m=\overline{1,M}$, we have instead the following form for system (1):

$$(13)$$

$$(2\mu^i + \lambda^i)\left(\frac{1}{(\Delta^i)^2} - \frac{1}{r^{i(l)}}\frac{1}{2\Delta^i}\right)\Phi_{1mn}^{i(l-1)} - \frac{n}{r^{i(l)}}\frac{\lambda^i + \mu^i}{2\Delta^i}\Phi_{2mn}^{i(l-1)} + \frac{m\pi}{A}\frac{\lambda^i + \mu^i}{2\Delta^i}\Phi_{3mn}^{i(l-1)}$$

$$-\left((2\mu^i + \lambda^i)\left(\frac{2}{(\Delta^i)^2} + \frac{1}{(r^{i(l)})^2}\right) + \mu^i\left(\frac{n^2}{(r^{i(l)})^2} + \frac{m^2\pi^2}{A^2}\right)\right)\Phi_{1mn}^{i(l)}$$

$$+\left(-\frac{n}{r^{i(l)}}(3\mu^i + \lambda^i)\right)\Phi_{2mn}^{i(l)} + (2\mu^i + \lambda^i)\left(\frac{1}{(\Delta^i)^2} + \frac{1}{r^{i(l)}}\frac{1}{2\Delta^i}\right)\Phi_{1mn}^{i(l+1)}$$

$$+\left(\frac{n}{r^{i(l)}}\frac{\lambda^i + \mu^i}{2\Delta^i}\right)\Phi_{2mn}^{i(l+1)} - \frac{m\pi}{A}\frac{\lambda^i + \mu^i}{2\Delta^i}\Phi_{3mn}^{i(l+1)} = \rho^i\frac{d^2\Phi_{1mn}^{i(l)}}{dt^2},$$

$$\frac{n}{r^{i(l)}}\frac{\lambda^i + \mu^i}{2\Delta^i}\Phi_{1mn}^{i(l-1)} + \mu^i\left(\frac{1}{(\Delta^i)^2} - \frac{1}{r^{i(l)}}\frac{1}{2\Delta^i}\right)\Phi_{2mn}^{i(l-1)} - \frac{n}{(r^{i(l)})^2}(3\mu^i + \lambda^i)\Phi_{1mn}^{i(l)}$$

$$-\left(\mu^i\left(\frac{2}{(\Delta^i)^2} + \frac{m^2\pi^2}{A^2} + \frac{1}{(r^{i(l)})^2}\right) + (2\mu^i + \lambda^i)\frac{n^2}{(r^{i(l)})^2}\right)\Phi_{2mn}^{i(l)}$$

$$+\frac{m\pi}{A}\frac{n}{r^{i(l)}}(\lambda^i + \mu^i)\Phi_{3mn}^{i(l)} - \frac{n}{r^{i(l)}}\frac{\lambda^i + \mu^i}{2\Delta^i}\Phi_{1mn}^{i(l+1)} + \mu^i\left(\frac{1}{(\Delta^i)^2} + \frac{1}{r^{i(l)}}\frac{1}{2\Delta^i}\right)\Phi_{2mn}^{i(l+1)}$$

$$= \rho^i\frac{d^2\Phi_{2mn}^{i(l)}}{dt^2},$$

$$-\frac{m\pi}{A}\frac{\lambda^i + \mu^i}{2\Delta^i}\Phi_{1mn}^{i(l-1)} + \mu^i\left(\frac{1}{(\Delta^i)^2} - \frac{1}{r^{i(l)}}\frac{1}{2\Delta^i}\right)\Phi_{3mn}^{i(l-1)} + \frac{m\pi}{A}\frac{\lambda^i + \mu^i}{r^{i(l)}}\Phi_{1mn}^{i(l)}$$

$$+\frac{m\pi}{A}\frac{n}{r^{i(l)}}(\lambda^i + \mu^i)\Phi_{2mn}^{i(l)} - \left(\mu^i\left(\frac{2}{(\Delta^i)^2} + \frac{n^2}{(r^{i(l)})^2}\right) + \frac{m^2\pi^2}{A^2}(2\mu^i + \lambda^i)\right)\Phi_{3mn}^{i(l)}$$

$$+\frac{m\pi}{A}\frac{\lambda^i + \mu^i}{2\Delta^i}\Phi_{1mn}^{i(l+1)} + \mu^i\left(\frac{1}{(\Delta^i)^2} + \frac{1}{r^{i(l)}}\frac{1}{2\Delta^i}\right)\Phi_{3mn}^{i(l+1)} = \rho^i\frac{d^2\Phi_{3mn}^{i(l)}}{dt^2}.$$

For the same pairs (n, m), conditions (2), (4), and (5) become

$$(14)$$

$$(2\mu^1 + \lambda^1)\frac{\Phi_{1mn}^{1(1)} - \Phi_{1mn}^{1(-1)}}{2\Delta^1} + \lambda^1\left(\frac{1}{r^{1(0)}}\left(n\Phi_{2mn}^{1(0)} + \Phi_{1mn}^{1(0)}\right) - \frac{m\pi}{A}\Phi_{3mn}^{1(0)}\right) = -q_{mn}^-,$$

$$\frac{\Phi_{2mn}^{1(1)} - \Phi_{2mn}^{1(-1)}}{2\Delta^1} + \frac{1}{r^{1(0)}}\left(-n\Phi_{1mn}^{1(0)} - \Phi_{2mn}^{1(0)}\right) = 0, \quad \frac{\Phi_{3mn}^{1(1)} - \Phi_{3mn}^{1(-1)}}{2\Delta^1} + \frac{m\pi}{A}\Phi_{1mn}^{1(0)} = 0,$$

$$(2\mu^I + \lambda^I)\frac{\Phi_{1mn}^{I(L^I+1)} - \Phi_{1mn}^{I(L^I-1)}}{2\Delta^I} + \lambda^I\left(\frac{1}{r^{I(L^I)}}\left(n\Phi_{2mn}^{I(L^I)} + \Phi_{1mn}^{I(L^I)}\right) - \frac{m\pi}{A}\Phi_{3mn}^{I(L^I)}\right)$$

$$= -q_{mn}^+,$$

$$\frac{\Phi_{2mn}^{I(L^I+1)} - \Phi_{2mn}^{I(L^I-1)}}{2\Delta^I} - \frac{1}{r^{I(L^I)}}\left(n\Phi_{1mn}^{I(L^I)} + \Phi_{2mn}^{I(L^I)}\right) = 0, \tag{15}$$

$$\frac{\Phi_{3mn}^{I(L^I+1)} - \Phi_{3mn}^{I(L^I-1)}}{2\Delta^I} + \frac{m\pi}{A}\Phi_{1mn}^{I(L^I)} = 0,$$

$$\Phi_{kmn}^{i(L^i)} = \Phi_{kmn}^{i+1(0)} \quad \text{for } k = 1, 2, 3,$$

$$\frac{(2\mu^i + \lambda^i)}{2\Delta^i}\left(\Phi_{1mn}^{i(L^i+1)} - \Phi_{1mn}^{i(L^i-1)}\right) + \lambda^i\left(\frac{1}{r^{i(L^i)}}\left(n\Phi_{2mn}^{i(L^i)} + \Phi_{1mn}^{i(L^i)}\right) - \frac{m\pi}{A}\Phi_{3mn}^{i(L^i)}\right)$$

$$= \frac{(2\mu^{i+1} + \lambda^{i+1})}{2\Delta^{i+1}}\left(\Phi_{1mn}^{i+1(1)} - \Phi_{1mn}^{i+1(-1)}\right)$$

$$+ \lambda^{i+1}\left(\frac{1}{r^{i+1(0)}}\left(n\Phi_{2mn}^{i+1(0)} + \Phi_{1mn}^{i+1(0)}\right) - \frac{m\pi}{A}\Phi_{3mn}^{i+1(0)}\right),$$

$$\mu^i\left(\frac{\Phi_{2mn}^{i(L^i+1)} - \Phi_{2mn}^{i(L^i-1)}}{2\Delta^i} - \frac{1}{r^{i(L^i)}}\left(n\Phi_{1mn}^{i(L^i)} + \Phi_{2mn}^{i(L^i)}\right)\right)$$

$$= \mu^{i+1}\left(\frac{\Phi_{2mn}^{i+1(1)} - \Phi_{2mn}^{i+1(-1)}}{2\Delta^{i+1}} - \frac{1}{r^{i+1(0)}}\left(n\Phi_{1mn}^{i+1(0)} + \Phi_{2mn}^{i+1(0)}\right)\right),$$

$$\mu^i\left(\frac{\Phi_{3mn}^{i(L^i+1)} - \Phi_{3mn}^{i(L^i-1)}}{2\Delta^i} + \frac{m\pi}{A}\Phi_{1mn}^{i(L^i)}\right) = \mu^{i+1}\left(\frac{\Phi_{3mn}^{i+1(1)} - \Phi_{3mn}^{i+1(-1)}}{2\Delta^{i+1}} + \frac{m\pi}{A}\Phi_{1mn}^{i+1(0)}\right)$$

$$\text{for } i = \overline{1, I-1},$$

with

$$\Phi_{kmn}^{i(s)}(0) = \frac{d\Phi_{kmn}^{i(s)}(0)}{dt} = 0 \quad \text{for } k = 1, 2, 3, \ i = \overline{1, I}.$$

The boundary conditions at the ends of the cylinder (3) are satisfied exactly by selecting the coordinate functions B_{kmn} of (8), which correspond to simply supported conditions.

Conditions (12) allow us to exclude the values $\Phi_{km0}^{i(-1)}$ and $\Phi_{km0}^{i(L^i+1)}$ ($i = \overline{1, I}$, $k = 1, 3$, $n = 0$, $m = \overline{1, M}$) of the sought-for functions in "extra-contour" points from system (10)–(11); and conditions (14)–(15) allow us to exclude the values $\Phi_{kmn}^{i(-1)}$ and $\Phi_{kmn}^{i(L^i+1)}$ ($i = \overline{1, I}$, $k = 1, 2, 3$, $n = \overline{1, N}$, $m = \overline{1, M}$) of the sought-for functions in "extra-contour" points from system (13).

Hence, the solution of problems (1)–(5) on oscillations of a multilayer cylindrical shell subjected to an impulse load is reduced for each pair of values (m, n) to integrating a system of ordinary differential equations with constant coefficients. In this paper, the system obtained is integrated by Taylor expansion [Bakhvalov 1975; Shupikov et al. 2004] as described in the Appendix.

4. Numerical results

We illustrate with examples the method's feasibility and the validity of its results.

Consider an infinite uniform cylinder with $R_0 = 0.08$ m, $h_1 = 0.04$ m, $E_1 = 2.06 \cdot 10^8$ kPa, $\rho_1 = 7.9 \cdot 10^3$ kg/m³, and $\nu_1 = 0.25$, subject to an impulse load applied to the inner surface, the load being a uniformly distributed pressure changing with respect to time according to the law

$$q^-(\theta, z, t) = q_0^- \exp(-t/\tau), \quad q^+(\theta, z, t) = 0,$$

with loading intensity $q_0^- = 1.49 \cdot 10^8$ Pa and load action time $\tau = 14.2 \cdot 10^{-6}$ s.

Figures 2 and 3 show data obtained using the exact solution from [Philippov
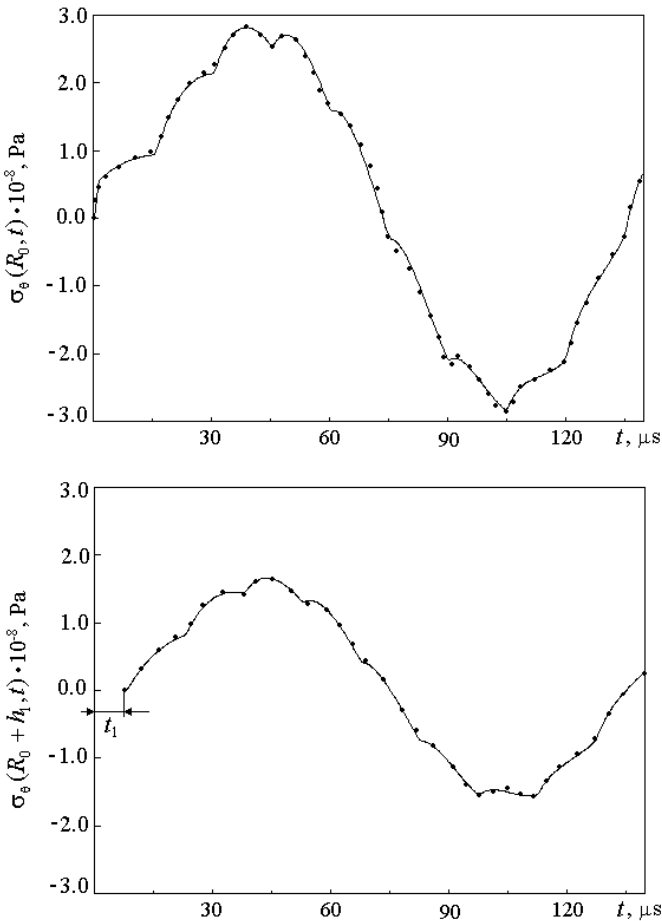
Figure 2. Circumferential stresses of infinite cylinder under impulse loading: dots, exact solution; solid line, present method.

et al. 1978], together with the results obtained with the method of this paper. For
the calculations we took $L^1 = 160$. We observe surges at times of stress $\sigma_\theta^1(R^*, t)$
and $\sigma_r^1(R^*, t)$. For radial stresses $\sigma_r^1(R^*, t)$, such surges are more prominent than
for circumferential ones $\sigma_\theta^1(R^*, t)$, and they have a dramatic impact on both the
absolute values of the stresses $\sigma_r^1(R^*, t)$ and on their change in sign. Stress surges
in time correspond to instances of arrival of waves reflected from the outer surface
($r = R_0 + h_1$) to the surface considered with the coordinate $r = R^* = 0.085$ m.
The interval t_1 corresponds to the time required for the wave to travel the distance
$s_1 = R^* - R_0 = 5$ mm, and it corresponds to the same value obtained from the
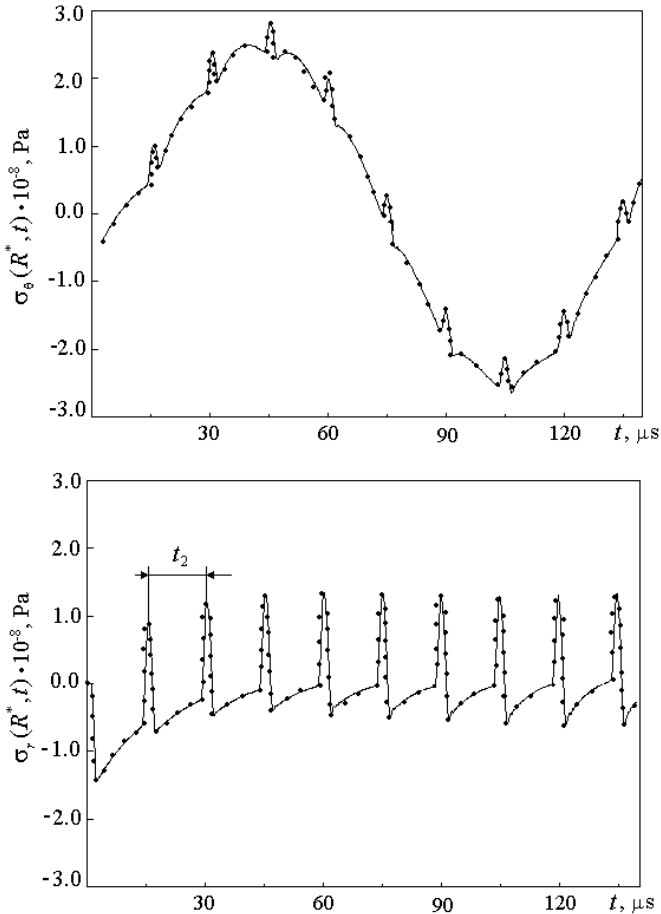
Figure 3. Radial stresses of infinite cylinder under impulse load-
ing: dots, exact solution; solid line, present method.

exact formula

$$t_1 = \frac{s_1}{V} \approx 0.9\,\mu s,$$

where V is the expansion wave [Novatsky 1975]

$$V = \sqrt{\frac{\lambda^1 + \mu^1}{\rho^1}} \approx 5.52 \cdot 10^3 \,\text{m/s}.$$

Interval t_2 corresponds to the time required for the wave to travel the distance $s_2 = 2h_1 - (R^* - R_0) = 7.5\,\text{mm}$, and it corresponds to the same value obtained from the exact formula

$$t_2 = \frac{s_2}{V} \approx 1.36 \cdot 10^{-6} \,\text{s}.$$

In the problem considered, the uniform shell was presented in the form of a one-, two-, and three-ply shell. In all cases, we observed a satisfactory matching of results obtained with the help of the given technique and the analytic solution [Philippov et al. 1978].

For a uniform finite-length cylinder with parameters $A = 0.5\,\text{m}$, $R_0 = 0.095\,\text{m}$, $h^1 = 0.01\,\text{m}$, $E^1 = 6.67 \cdot 10^4 \,\text{MPa}$, $\rho^1 = 2.5 \cdot 10^3 \,\text{kg/m}^3$, and $\nu^1 = 0.3$, subjected to an external radially directed load applied instantaneously to the outer surface, we give a comparison of the results obtained by the analytic method in [Weingarten and Reismann 1974], and those obtained by implementing the given approach.

A load with intensity q_0^+ is applied instantaneously radially outside the cylinder, and distributed over a small area on the outer surface of the shell. The dimensions of the loading areas are ε rad in the direction of axis θ, and λ in the direction of axis z. The centre of the loading area has the coordinates $\theta = 0$, $z = L/2$, i.e., the load is distributed symmetrically with respect to the circumferential coordinate, and it has the following form:

$$q^+(\theta, z, t) = q_0^+ f(\theta) g(z) H(t), \quad q^-(\theta, z, t) = 0,$$

$$f(\theta) = \begin{cases} 0, & |\theta| > \varepsilon/2 \\ 1, & |\theta| \leq \varepsilon/2 \end{cases}, \quad f(\theta + 2\pi) = f(\theta),$$

$$g(z) = \begin{cases} 0, & 0 \leq z < (L-\lambda)/2 \\ 1, & (L-\lambda)/2 \leq z \leq (L+\lambda)/2, \\ 0, & (L+\lambda)/2 < z \leq \lambda \end{cases}$$

$$\lambda = 0.5\,R, \quad R = 0.1\,\text{m}, \quad \varepsilon = 0.5\,\text{rad}.$$

Here q_0^+ is the loading intensity (0.1 MPa); $f(\theta)$ is the load distribution over coordinate θ; $g(z)$ is the load distribution over coordinate z, and $H(t)$ is Heaviside's function.

Figure 4 shows the cylinder's response to dynamic loading. The dots show the analytic solution, and the lines represent the solution obtained by using the given analytic-numerical method. For the calculations we took $L^1 = 150$. The top graph shows the change in radial displacement on the median surface as a function of time. The bottom graph shows the change in circumferential stresses on the outer surface of the shell as a function of time. The abscissa is dimensionless time, which is normed by the value of the time of travel of the shear wave over the shell radius, $\tau^* = tV/R_0$,

$$V = \sqrt{\frac{\mu^1}{\rho^1}} \approx 3.203 \cdot 10^3 \text{ m/s}.$$
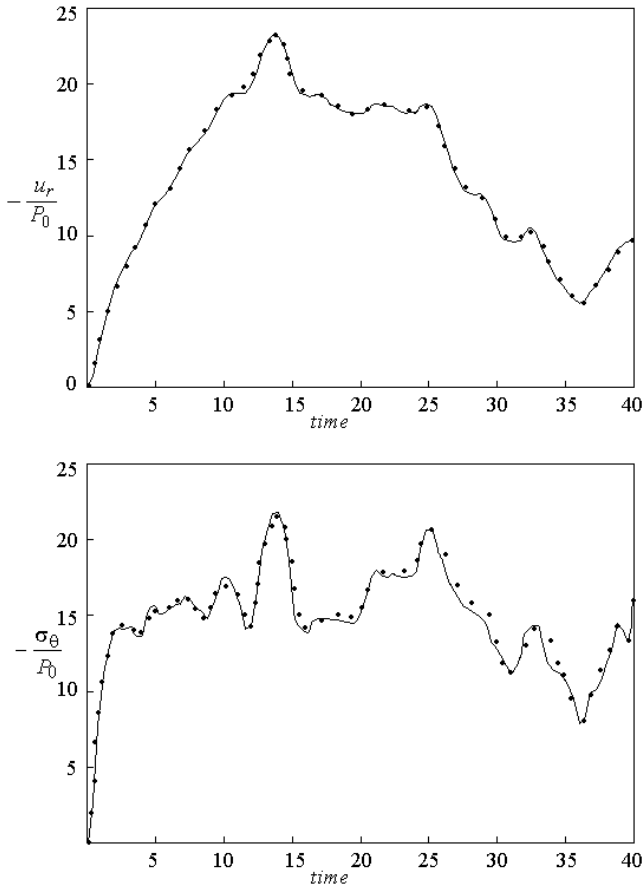
Figure 4. Response of a finite-length cylinder to dynamic loading: dots, analytic solution; solid line, present method.

$R_0(m)$	i	Designation in Fig.	$h^i(m)$	E^i (MPa)	ν^i	ρ^i (kg/m^3)
0.08	1		0.01	$5.59 \cdot 10^3$	0.38	$1.2 \cdot 10^3$
	2		0.02	$6.67 \cdot 10^4$	0.22	$2.5 \cdot 10^3$

Table 1. Parameters of multilayer shell.

These examples show good agreement between the results obtained with the present method and analytic solutions obtained by other authors.

The process of propagation of the disturbance during impulse loading has been investigated. An infinite two-ply cylindrical shell, whose parameters are given in Table 1, is considered.

The load applied to the inner surface is a uniformly distributed pressure that changes with time according to the law

$$q^-(\theta, z, t) = q_0^- H(t), \quad q^+(\theta, z, t) = 0,$$

where $q_0^- = 1.49 \cdot 10^8$ Pa and $H(t)$ is Heaviside's function.

Figure 5 shows the distribution diagrams for stresses $\sigma_r^i(r, t)$ at different times. The symbols

$$\text{and}$$

show the direction of propagation of stress waves. For the calculations we took $L_1 = 70$ and $L_2 = 100$. The figures presented demonstrate the wavelike pattern of the process. Besides, one can see the effect of wave reflection from the boundary between the layers and external surfaces.

Hence, we have shown the possibility of investigating wave processes in thick uniform and multilayer cylindrical shells. Such an approach can be practical for evaluating the area of applicability of two-dimensional theories when it is necessary to investigate the process of propagation of elastic waves, and when the SSS of the object being investigated has an essentially three-dimensional character.

5. Conclusions

The present work suggests an analytic-numerical method of investigating vibrations in a multilayer closed cylindrical shell in terms of the three-dimensional theory of elasticity. The given method allows investigating the behavior of uniform and multilayer cylindrical shells subjected to impulse loading.

The examples given for different kinds of loading (Figures 2–4) show good agreement between the results obtained with the present method and analytic solutions obtained by other authors.

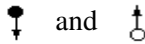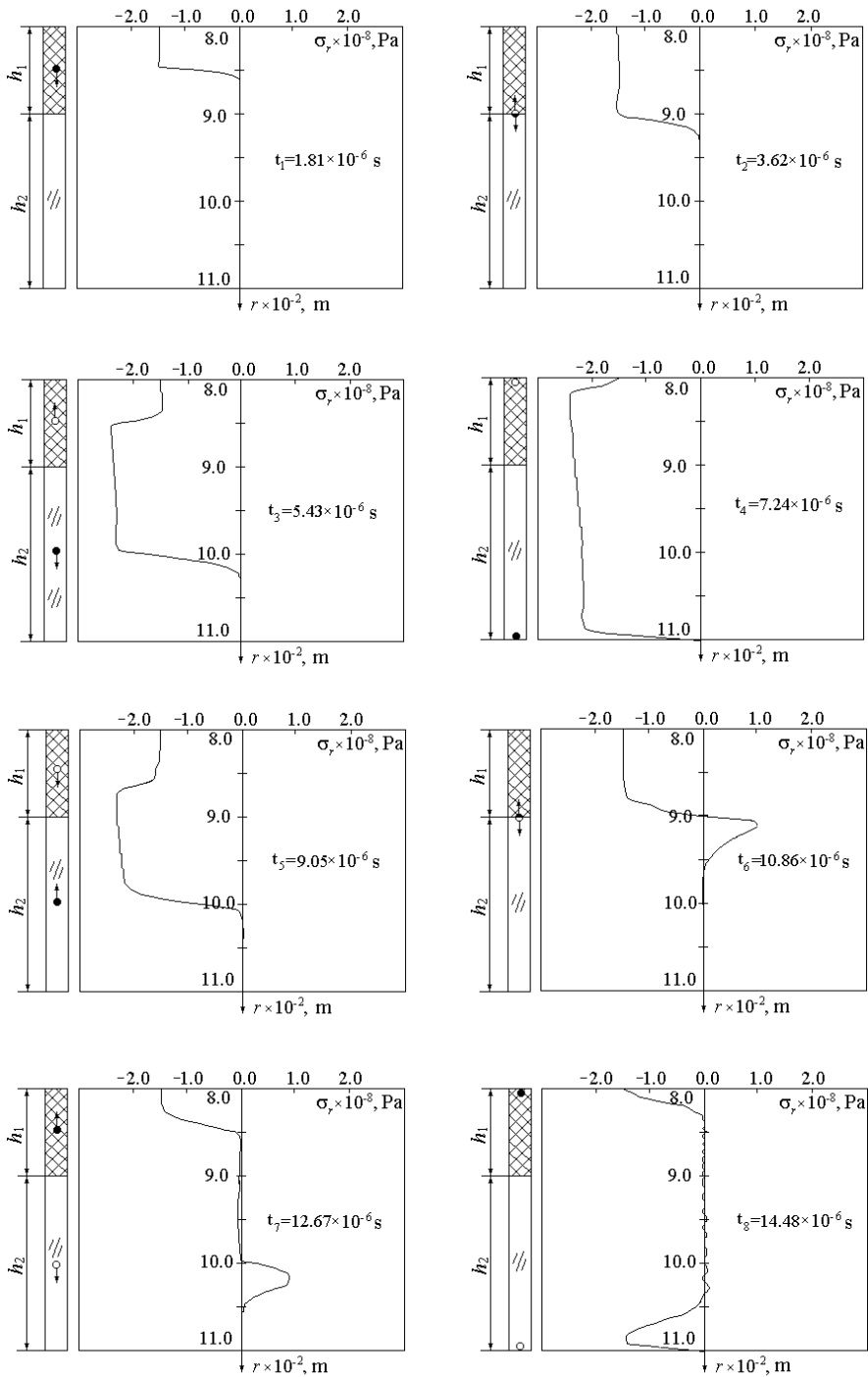**Figure 5.** The propagation of stress waves of a two-layer cylindrical shell.

The plots in Figures 2, 3 and 5 demonstrate the possibility of investigating wave processes in thick uniform and multilayer cylindrical shells.

Hence, the given method can be used for verifying the validity of results based on different two-dimensional theories applied to analyzing vibrations of multilayer cylindrical shells, as well as for investigating wave processes in cylindrically shaped elastic bodies.

Appendix: A modified method of solution expansion into Taylor series

A modified method of solution expansion in Taylor series is applied to integrate a system of ordinary differential equations with constant coefficients.

The initial system of differential equations is written in the form

$$[\Omega^{mn}]\ddot{\overline{\Phi}}_{mn} + [\Lambda^{mn}]\overline{\Phi}_{mn} = \overline{Q}_{mn}.$$

By replacing variables and straightforward transforms, it is reduced to the form

$$\dot{\overline{G}}_{mn} = [R^{mn}]\dot{\overline{G}}_{mn} + \overline{H}_{mn}. \tag{A.1}$$

The integration interval $[0, t]$ is divided into s sections, each with a length of Δt so that $t = s\Delta t$. We denote $\overline{G}_{mn}(s\Delta t) = \overline{G}_{mns}$.

At each integration step Δt, the solution is represented as a Taylor series:

$$\overline{G}_{mns} = \overline{G}_{mns-1} + \frac{\dot{\overline{G}}_{mns-1}}{1!}\Delta t + \frac{\ddot{\overline{G}}_{mns-1}}{2!}\Delta t^2 + \cdots. \tag{A.2}$$

It is assumed that, within the integration step,

$$\overline{H}_{mn}(t) = \overline{H}_{mns}, \qquad (s-1)\Delta t \leq t \leq s\Delta t. \tag{A.3}$$

In this case, given (A.1) and (A.3), the derivatives in series (A.2) can be presented as

$$\overset{(k)}{\overline{G}}_{mn} = [R^{mn}]^k \overline{G}_{mn} + [R^{mn}]^{k-1}\overline{H}_{mn}. \tag{A.4}$$

An example of calculating the derivatives is given here:

$$\ddot{\overline{G}}_{mn} = \frac{d}{dt}\dot{\overline{G}}_{mn} = \frac{d}{dt}\left([R^{mn}]\overline{G}_{mn} + \overline{H}_{mn}\right) = [R^{mn}]\dot{\overline{G}}_{mn} =$$
$$= [R^{mn}]\left([R^{mn}]\overline{G}_{mn} + \overline{H}_{mn}\right) = [R^{mn}]^2\overline{G}_{mn} + [R^{mn}]\overline{H}_{mn}.$$

Substituting the expressions for derivatives (A.4) into (A.2), we obtain the following expressions for solving the system at the s-th step:

$$\overline{G}_{mns} = [K^{mn}]\overline{G}_{mns-1} + [T^{mn}]\overline{H}_{mns}. \tag{A.5}$$

Here

$$[K_{mn}] = [E] + \frac{[R^{mn}]}{1!}\Delta t + \frac{[R^{mn}]^2}{2!}\Delta t^2 + \cdots,$$

$$[T_{mn}] = \frac{[E]}{1!}\Delta t + \frac{[R^{mn}]}{2!}\Delta t^2 + \frac{[R^{mn}]^2}{3!}\Delta t^3 + \cdots,$$

where $[E]$ is the unit matrix.

To refine the solution, the interval $[t_{s-1}, t_s]$ is divided into r sections with the length of $\Delta\tau = \Delta t/r$. In each section, the function $\overline{G}_{mn}$ is calculated using (A.5):

$$\overline{G}_{mn}\left(t_{s-1} + \frac{\Delta t}{r}\right) = [\hat{K}^{mn}]\overline{G}_{mns-1} + [\hat{T}^{mn}]\overline{H}_{mns},$$

$$\vdots$$

$$\overline{G}_{mn}\left(t_{s-1} + \frac{i\,\Delta t}{r}\right) = [\hat{K}^{mn}]^i\,\overline{G}_{mns-1} + \left([\hat{K}^{mn}]^{i-1} + [\hat{K}^{mn}]^{i-2} + \cdots + [E]\right)[\hat{T}^{mn}]\overline{H}_{mns}$$

$$\text{for } i = \overline{1, r}.$$

The matrices $[\hat{K}^{mn}]$ and $[\hat{T}^{mn}]$ are derived from matrices $[K^{mn}]$ and $[T^{mn}]$ by replacing Δt with $\Delta\tau$.

At $i = r$, $\overline{G}_{mn}(t_{s-1} + \Delta t) = \overline{G}_{mns}$, the system solution takes its final form

$$\overline{G}_{mns} = [M^{mn}]\overline{G}_{mns-1} + [J^{mn}]\overline{H}_{mns},$$

where

$$[M^{mn}] = [\hat{K}^{mn}]^r;$$

$$[J^{mn}] = \left([\hat{K}^{mn}]^{r-1} + [\hat{K}^{mn}]^{r-2} + \cdots + [E]\right)[\hat{T}^{mn}].$$

Hence, integrating a system of ordinary differential equations is reduced to calculating the load vector $\overline{H}_{mns}$ and multiplying matrices by vectors. The matrices are calculated once for each pair of values m and n.

In this paper, to ensure stability of the process of numerically integrating a system of ordinary differential equations, the step of integration with respect to time Δt is taken to be equal to the time of strain wave travel between adjacent nodes of the finite-difference grid according to the Courant–Hilbert conditions.

References

[Bakhvalov 1975] N. S. Bakhvalov, *Numerical methods*, Nauka, Moskow, 1975.

[Bhaskar and Ganapathysaran 2002] K. Bhaskar and N. Ganapathysaran, "Three-dimensional analysis of a cross-ply cylindrical shell subjected to a localized circumferential shear force", *Int. J. Pressure Vessels Piping* **79**:7 (2002), 519–524.

[Bhaskar and Ganapathysaran 2003] K. Bhaskar and N. Ganapathysaran, "Elasticity solutions for laminated orthotropic cylindrical shells subjected to localized longitudinal and circumferential moments", *J. Pressure Vessel Technol.(Trans. ASME)* **125**:1 (2003), 26–35.

[Bhaskar and Varadan 1993] K. Bhaskar and T. K. Varadan, "Exact elasticity solution for laminated anisotropic cylindrical shells", *J. Appl. Mech. (Trans. ASME)* **60**:1 (1993), 41–47.

[Bhaskar and Varadan 1994] K. Bhaskar and T. K. Varadan, "Benchmark elasticity solution for locally loaded laminated orthotropic cylindrical shells", *AIAA J.* **32**:3 (1994), 627–632.

[Forsythe and Wasov 1960] G. E. Forsythe and W. R. Wasov, *Finite-difference methods for partial differential equations*, Wiley, New York, 1960.

[Grigoliuk and Kogan 1972] E. I. Grigoliuk and F. A. Kogan, "Modern state of the theory of multilayer shells", *Prikladnaya Mekhanika* **8**:3 (1972), 3–17. In Russian.

[Grigoliuk and Kulikov 1988] E. I. Grigoliuk and G. M. Kulikov, "General direction of development of the theory of multilayer shells", *Mekhanika Kompozitnykh Materialov* **24**:2 (1988), 287–298. In Russian; translation in *Mech. Compos. Materials* **24**:2 (1988), 231–241.

[Kang and Leissa 2000] J.-H. Kang and A. W. Leissa, "Three-dimensional vibrations of thick spherical shell segments with variable thickness", *Int. J. Solids Struct.* **37**:35 (2000), 4811–4823.

[Little 1973] R. W. Little, *Elasticity*, Prentice-Hall, Englewood Cliffs, N. J., 1973.

[Liu 2000] F.-L. Liu, "Static analysis of thick rectangular laminated plates: three-dimensional elasticity solutions via differential quadrature element method", *Int. J. Solids Struct.* **37**:51 (2000), 7671–7688.

[Noor and Burton 1989] A. K. Noor and W. S. Burton, "Assessment of shear deformation for multilayered composite plates", *Appl. Mech. Rev.* **42**:1 (1989), 1–12.

[Noor and Burton 1990a] A. K. Noor and W. S. Burton, "Assessment of computational models for multilayered anisotropic plates", *Compos. Struct.* **14**:3 (1990), 233–265.

[Noor and Burton 1990b] A. K. Noor and W. S. Burton, "Three-dimensional solutions for antisymmetrically laminated anisotropic plates", *J. Appl. Mech. (Trans. ASME)* **57** (1990), 182–188.

[Noor and Rarig 1974] A. K. Noor and P. L. Rarig, "Three-dimensional solutions of laminated cylinders", *Comput. Methods Appl. Mech. Eng.* **3**:3 (1974), 319–334.

[Noor et al. 1996] A. K. Noor, W. S. Burton, and C. W. Bert, "Computational models for sandwich panels and shells", *Appl. Mech. Rev.* **49**:3 (1996), 155–199.

[Novatsky 1975] V. Novatsky, *Theory of elasticity*, Mir, Moskow, 1975.

[Pagano 1969] N. J. Pagano, "Exact solutions for composite laminates in cylindrical bending", *J. Compos. Mater.* **3** (1969), 398–411.

[Pagano 1970a] N. J. Pagano, "Exact solutions for rectangular bidirectional composites and sandwich plates", *J. Compos. Mater.* **4** (1970), 20–34.

[Pagano 1970b] N. J. Pagano, "Influence of shear coupling in cylindrical bending of anisotropic laminates", *J. Compos. Mater.* **4** (1970), 330–343.

[Philippov et al. 1978] A. P. Philippov, S. S. Kohmaniuk, and Y. Yaniutin, *Strain in structural elements subjected to shock and impulse loads*, Naukova Dumka, Kyiv, 1978.

[Reddy 1989] J. N. Reddy, "On the generalization of displacement-based laminate theories", *Appl. Mech. Rev.* **42**:8 (1989), 213–222.

[Reddy 1993] J. N. Reddy, "An evaluation of equivalent-single-layer and layerwise theories of composite laminates", *Compos. Struct.* **25**:1–4 (1993), 21–35.

[Ren 1987] J. G. Ren, "Exact solutions for laminated cylindrical shells in cylindrical bending", *Compos. Sci. Technol.* **29**:3 (1987), 169–187.

[Ren 1989] J. G. Ren, "Analysis of simply-supported laminated circular cylindrical shell roofs", *Compos. Struct.* **11**:4 (1989), 277–292.

[Shupikov and Ugrimov 1997] A. N. Shupikov and S. V. Ugrimov, "High-order theory for multilayer plates", pp. 103–109 in *Current problems of strength and optimization of structures and machines*, Nauka i obrazovaniye, Dnepropetrovsk, 1997.

[Shupikov and Ugrimov 1999] A. N. Shupikov and S. V. Ugrimov, "Vibrations of multilayer plates under the effect of impulse loads, Three-dimensional theory", *Int. J. Solids Struct.* **36**:22 (1999), 3391–3402.

[Shupikov et al. 2004] A. N. Shupikov, N. N. Buzko, N. V. Smetankina, and S. V. Ugrimov, *Non-stationary vibrations of multilayer plates and shells and their optimization*, Publishing House KNEU, Kharkov, 2004.

[Smetankina et al. 1995] N. V. Smetankina, S. Y. Sotrikhin, and A. N. Shupikov, "Theoretical and experimental investigation of vibration of multilayer plates under the action of impulse and impact loads", *Int. J. Solids Struct.* **32**:8-9 (1995), 1247–1258.

[Varadan and Bhaskar 1991] T. K. Varadan and K. Bhaskar, "Bending of laminated orthotropic cylindrical shells: an elasticity approach", *Compos. Struct.* **17**:2 (1991), 141–156.

[Weingarten and Reismann 1974] L. I. Weingarten and H. Reismann, "Forced motion of cylindrical shells: a comparison of shell theory with elasticity theory", *Z. Angew. Math. Mech.* **54** (1974), 181–191.

ALEXANDER SHUPIKOV: shupikov@ipmach.kharkov.ua
Department of Strength of Thin-Walled Structures, Institute for Problems in Machinery,
National Academy of Sciences of the Ukraine, 2/10 Pozharsky Street, Kharkov 61046,
Ukraine

NATALIYA DOLGOPOLOVA: dolna@zeos.net
A. M. Pidgorny Institute for Mechanical Engineering Problems, National Academy of Sciences of the
Ukraine, 2/10 Pozharsky Street, Kharkiv 61046, Ukraine

THE ISOTROPIC MATERIAL CLOSEST TO A GIVEN ANISOTROPIC MATERIAL

ANDREW N. NORRIS

The isotropic elastic moduli closest to a given anisotropic elasticity tensor are defined using three definitions of elastic distance: the standard Frobenius (Euclidean) norm, the Riemannian distance for tensors, and the log-Euclidean norm. The closest moduli are unique for the Riemannian and the log-Euclidean norms, independent of whether the difference in stiffness or compliance is considered. Explicit expressions for the closest bulk and shear moduli are presented for cubic materials, and an algorithm is described for finding them for materials with arbitrary anisotropy. The method is illustrated by application to a variety of materials, which are ranked according to their distance from isotropy.

1. Introduction

The objective here is to answer the question: what is the isotropic material closest to a given anisotropic material? In order to attempt an answer one needs a distance or length function which measures the difference between the elastic moduli of two materials. The Euclidean norm provides a natural definition for distance, and using it one can find the elastic tensor of a given symmetry nearest to an anisotropic elastic tensor [Gazis et al. 1963; Arts et al. 1991; Helbig 1996; Cavallini 1999; Gangi 2000; Browaeys and Chevrot 2004]. The Euclidean distance function is, however, not invariant under inversion, that is, considering compliance instead of stiffness, and as such does not lead to a unique answer to the question posed. To see this, let ΔC_{ijkl} and ΔS_{ijkl} be the elements of the fourth order tensors for the differences in elastic stiffness and compliance, respectively. Define the length of a fourth order tensor with elements T_{ijkl} by $(T_{ijkl}T_{ijkl})^{1/2}$. Then it is clear that the length using ΔC_{ijkl} is not simply related to that of ΔS_{ijkl}.

Recently and separately, Moakher [2006] and Arsigny et al. [2005] (see also [Matthies and Humbert 1995]) introduced two distance functions for elasticity tensors which are unchanged whether one uses stiffness or compliance. The two measures of elastic distances, called the Riemannian distance [Moakher 2006] and the log-Euclidean metric [Arsigny et al. 2005], each provide a means to define

---

Keywords: elastic moduli, anisotropy, Euclidean distance, Riemannian distance.

unambiguously the distance between any two elasticity tensors. The focus here is on finding the isotropic material closest to a given arbitrarily anisotropic material.

The distance functions are first reviewed in Section 2 along with the more common Frobenius or Euclidean norm. The theory is developed in terms of matrices, with obvious application to tensors. Preliminary results for elastic materials are presented in Section 4, where closed-form expressions are derived for the isotropic moduli closest to a given material of cubic symmetry. The general problem for materials of arbitrary anisotropy is solved in Section 5, and applications to sample materials are described in Section 6.

2. Matrix distance functions

We begin with $\mathcal{P}(n)$, the vector space of positive definite symmetric matrices in $\mathbb{M}^{n \times n}$, the space of $n \times n$ real matrices. Recall that a matrix $\boldsymbol{P}$ is symmetric if $\boldsymbol{x}^T \boldsymbol{P} \boldsymbol{y} = \boldsymbol{y}^T \boldsymbol{P} \boldsymbol{x}$ for all $\boldsymbol{x}, \boldsymbol{y}$ in $\mathbb{R}^n$, and positive definite if $\boldsymbol{x}^T \boldsymbol{P} \boldsymbol{x} > 0$ for all nonzero $\boldsymbol{x} \in \mathbb{R}^n$. The spectral decomposition is

$$\boldsymbol{P} = \sum_{i=1}^{n} \lambda_i \, \boldsymbol{v}_i \boldsymbol{v}_i^T \, , \tag{1}$$

where λ_i are the eigenvalues and $\boldsymbol{v}_i \in \mathbb{R}^n$ the eigenvectors, which satisfy $\lambda_i > 0$, $\boldsymbol{v}_i^T \boldsymbol{v}_j = \delta_{ij}$. Functions of $\boldsymbol{P}$ can be readily found based on the diagonalized form; in particular, the logarithm of a matrix is defined as

$$\text{Log } \boldsymbol{P} = \sum_{i=1}^{n} \ln \lambda_i \, \boldsymbol{v}_i \boldsymbol{v}_i^T. \tag{2}$$

Three distinct metrics for positive definite symmetric matrices are considered: the conventional Euclidean or Frobenius metric d_F, the log-Euclidean distance d_L [Arsigny et al. 2005], and the Riemannian distance d_R [Moakher 2006]. Thus, for any pair $\boldsymbol{A}, \boldsymbol{B} \in \mathcal{P}(n)$

$$\text{d}_F(\boldsymbol{A}, \boldsymbol{B}) = \|\boldsymbol{A} - \boldsymbol{B}\| \, , \tag{3}$$

$$\text{d}_L(\boldsymbol{A}, \boldsymbol{B}) = \|\text{Log}(\boldsymbol{A}) - \text{Log}(\boldsymbol{B})\| \, , \tag{4}$$

$$\text{d}_R(\boldsymbol{A}, \boldsymbol{B}) = \left\|\text{Log}(\boldsymbol{A}^{-1/2} \boldsymbol{B} \boldsymbol{A}^{-1/2})\right\| \, , \tag{5}$$

where $\|\boldsymbol{M}\| \equiv [\text{tr}(\boldsymbol{M}^T \boldsymbol{M})]^{1/2}$ for any $\boldsymbol{M} \in \mathbb{M}^{n \times n}$. The distance function d_R is a consequence of the scalar product

$$\langle \boldsymbol{M}_1, \boldsymbol{M}_2 \rangle_{\boldsymbol{P}} \quad \equiv \quad \text{tr}(\boldsymbol{P}^{-1} \boldsymbol{M}_1 \boldsymbol{P}^{-1} \boldsymbol{M}_2), \tag{6}$$

for $\boldsymbol{P} \in \mathcal{P}(n)$ and symmetric $\boldsymbol{M}_1, \boldsymbol{M}_2 \in \mathbb{M}^{n \times n}$, and is also related to the exponential map [Lang 1998; Moakher 2006]. The metric d_L is associated with the Lie

group on $\mathcal{P}(n)$ defined by the following multiplication that preserves symmetry and positive definiteness [Arsigny et al. 2005] :

$$\boldsymbol{P}_1 \odot \boldsymbol{P}_2 \equiv \exp\left(\text{Log}(\boldsymbol{P}_1) + \text{Log}(\boldsymbol{P}_2)\right), \quad \boldsymbol{P}_1, \boldsymbol{P}_2 \in \mathcal{P}(n). \tag{7}$$

The three distance functions possess the properties expected of a distance function d:

(i) it is symmetric with respect to its arguments, $\text{d}(\boldsymbol{A}, \boldsymbol{B}) = \text{d}(\boldsymbol{B}, \boldsymbol{A})$;

(ii) it has nonnegative $\text{d}(\boldsymbol{A}, \boldsymbol{B}) \geq 0$ with equality if and only if $\boldsymbol{A} = \boldsymbol{B}$;

(iii) it is invariant under a change of basis, $\text{d}(\boldsymbol{Q}\boldsymbol{A}\boldsymbol{Q}^T, \boldsymbol{Q}\boldsymbol{B}\boldsymbol{Q}^T) = \text{d}(\boldsymbol{A}, \boldsymbol{B})$ for all orthogonal $\boldsymbol{Q} \in \mathbb{M}^{n \times n}$, $\boldsymbol{Q}\boldsymbol{Q}^T = \boldsymbol{Q}^T\boldsymbol{Q} = \boldsymbol{I}$; and

(iv) it satisfies the triangle inequality $\text{d}(\boldsymbol{A}, \boldsymbol{C}) \leq \text{d}(\boldsymbol{A}, \boldsymbol{B}) + \text{d}(\boldsymbol{B}, \boldsymbol{C})$ for all $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C} \in \mathcal{P}(n)$.

The Riemannian and log-Euclidean distances have additional properties not shared with d_F:

$$\text{d}_{L,R}(a\boldsymbol{A}, a\boldsymbol{B}) = \text{d}_{L,R}(\boldsymbol{A}, \boldsymbol{B}), \quad a \in \mathbb{R}_+, \tag{8}$$

$$\text{d}_{L,R}(\boldsymbol{A}^b, \boldsymbol{B}^b) = |b|\,\text{d}_{L,R}(\boldsymbol{A}, \boldsymbol{B}), \quad b \in \mathbb{R}, \tag{9}$$

where $\text{d}_{L,R}$ signifies either d_L or d_R. Thus d_L and d_R are bi-invariant metrics, that is, distances are invariant under multiplication and inversion. This property makes them consistent and unambiguous metrics for elasticity tensors. Moakher [2006] introduced another bi-invariant distance function, the Kullback–Leibler metric, but it does not satisfy the triangle inequality, and we do not consider it here.

The distance function d_R can be expressed in alternative forms by using the property $\boldsymbol{B}(\text{Log}\ \boldsymbol{A})\boldsymbol{B}^{-1} = \text{Log}\,(\boldsymbol{B}\boldsymbol{A}\boldsymbol{B}^{-1})$, for example,

$$\text{d}_R(\boldsymbol{A}, \boldsymbol{B}) = \left[\,\text{tr}\,\text{Log}^2(\boldsymbol{A}^{-1}\boldsymbol{B})\right]^{1/2} = \left[\,\text{tr}\,\text{Log}^2(\boldsymbol{B}^{-1}\boldsymbol{A})\right]^{1/2}, \tag{10}$$

or in terms of eigenvalues, using Equations (2) and (5),

$$\text{d}_R(\boldsymbol{A}, \boldsymbol{B}) = \left[\sum_{i=1}^{n} (\ln \lambda_i)^2\right]^{1/2}, \tag{11}$$

where λ_i, $i = 1, 2, \ldots, n$ are the eigenvalues of $\boldsymbol{P} = \boldsymbol{A}^{-1/2}\boldsymbol{B}\boldsymbol{A}^{-1/2}$, or equivalently, of the matrices $\boldsymbol{A}^{-1}\boldsymbol{B}$, $\boldsymbol{B}^{-1}\boldsymbol{A}$, $\boldsymbol{A}\boldsymbol{B}^{-1}$, etc. Note that d_R also satisfies

$$\text{d}_R(\boldsymbol{S}\boldsymbol{A}\boldsymbol{S}^T, \boldsymbol{S}\boldsymbol{B}\boldsymbol{S}^T) = \text{d}_R(\boldsymbol{A}, \boldsymbol{B}), \quad \text{for all invertible}\ \ \boldsymbol{S} \in \mathbb{M}^{n \times n}. \tag{12}$$

3. Preliminary examples

The remainder of the paper is concerned with applications to elasticity, with $n = 6$.

3.1. *Definition of elastic moduli.* 6×6 symmetric matrices are used to describe elastic moduli, whether of stiffness or compliance. The matrix representation is based on Kelvin's [Thomson 1856] observation in 1856 that the twenty one coefficients of elasticity define a quadratic form (the energy) in the six strains, and therefore possess six "principal strains". Although Kelvin did not write the elasticity tensor explicitly as a symmetric positive definite matrix, the idea has proved useful and has been developed extensively, notably by Rychlewski [1984] and Mehrabadi and Cowin [1990]. The notation of Mehrabadi and Cowin is employed here. Thus, the matrix $\widehat{C} \in \mathcal{P}(6)$ represents the elastic stiffness, and its inverse is the elastic compliance, $\widehat{S}$, satisfying

$$\widehat{S}\widehat{C} = \widehat{C}\widehat{S} = \widehat{I}, \qquad \text{where } \widehat{I} = \operatorname{diag}(1, 1, 1, 1, 1, 1). \tag{13}$$

The elements of the elastic stiffness matrix are

$$\widehat{C} = \begin{pmatrix} \hat{c}_{11} & \hat{c}_{12} & \hat{c}_{13} & \hat{c}_{14} & \hat{c}_{15} & \hat{c}_{16} \\ \hat{c}_{12} & \hat{c}_{22} & \hat{c}_{23} & \hat{c}_{24} & \hat{c}_{25} & \hat{c}_{26} \\ \hat{c}_{13} & \hat{c}_{23} & \hat{c}_{33} & \hat{c}_{34} & \hat{c}_{35} & \hat{c}_{36} \\ \hat{c}_{14} & \hat{c}_{24} & \hat{c}_{34} & \hat{c}_{44} & \hat{c}_{45} & \hat{c}_{46} \\ \hat{c}_{15} & \hat{c}_{25} & \hat{c}_{35} & \hat{c}_{45} & \hat{c}_{55} & \hat{c}_{56} \\ \hat{c}_{16} & \hat{c}_{26} & \hat{c}_{36} & \hat{c}_{46} & \hat{c}_{56} & \hat{c}_{66} \end{pmatrix} = \begin{pmatrix} c_{11} & c_{12} & c_{13} & \sqrt{2}c_{14} & \sqrt{2}c_{15} & \sqrt{2}c_{16} \\ c_{12} & c_{22} & c_{23} & \sqrt{2}c_{24} & \sqrt{2}c_{25} & \sqrt{2}c_{26} \\ c_{13} & c_{23} & c_{33} & \sqrt{2}c_{34} & \sqrt{2}c_{35} & \sqrt{2}c_{36} \\ \sqrt{2}c_{14} & \sqrt{2}c_{24} & \sqrt{2}c_{34} & 2c_{44} & 2c_{45} & 2c_{46} \\ \sqrt{2}c_{15} & \sqrt{2}c_{25} & \sqrt{2}c_{35} & 2c_{45} & 2c_{55} & 2c_{56} \\ \sqrt{2}c_{16} & \sqrt{2}c_{26} & \sqrt{2}c_{36} & 2c_{46} & 2c_{56} & 2c_{66} \end{pmatrix}, \tag{14}$$

where c_{ij}, $i, j = 1, 2, \ldots 6$ are the coefficients in the Voigt notation.

Before considering materials of arbitrary anisotropy, it is instructive to examine the distance functions for materials possessing the simplest type of anisotropy: cubic symmetry. Materials of cubic symmetry are described by three independent moduli: $c_{11} = c_{22} = c_{33}$, $c_{12} = c_{23} = c_{13}$, $c_{44} = c_{55} = c_{66}$, with the rest equal to zero. The three moduli commonly used are the bulk modulus κ and the two distinct shear moduli μ and η, which are related to the matrix elements by

$$3\kappa = \hat{c}_{11} + 2\hat{c}_{12}, \qquad 2\mu = \hat{c}_{44}, \qquad 2\eta = \hat{c}_{11} - \hat{c}_{12}. \tag{15}$$

Isotropic materials have only two independent moduli, κ, μ, and are of the same form as for cubic materials with the restriction $\hat{c}_{11} - \hat{c}_{12} - \hat{c}_{44} = 0$, or equivalently, $\eta = \mu$.

A concise notation is used for isotropic and cubic matrices, based upon Walpole's [Walpole 1984] general scheme for performing algebra with elasticity tensors. Define the matrices $\widehat{J}$, $\widehat{K}$, $\widehat{L}$ and $\widehat{M}$ by

$$\widehat{K} = \widehat{I} - \widehat{J}, \qquad \widehat{J} = \boldsymbol{u}\boldsymbol{u}^T, \qquad \text{where } \boldsymbol{u} = \left(\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, 0, 0, 0\right)^T, \quad (16)$$

$$\widehat{M} = \widehat{K} - \widehat{L}, \qquad \widehat{L} = \mathrm{diag}\,(0, 0, 0, 1, 1, 1). \qquad (17)$$

Note that $\widehat{I}$ and $\widehat{J}$ correspond, respectively, to the fourth order isotropic symmetric tensors with components $I_{ijkl} = (\delta_{ik}\delta_{jl} + \delta_{il}\delta_{jk})/2$ and $J_{ijkl} = (1/3)\delta_{ij}\delta_{kl}$. Elastic moduli of isotropic and cubic materials are of the generic form

$$\widehat{C}_{\mathrm{iso}}(3\kappa, 2\mu) \equiv 3\kappa\,\widehat{J} + 2\mu\,\widehat{K}, \qquad\qquad\qquad \kappa, \mu > 0, \qquad (18)$$

$$\widehat{C}_{\mathrm{cub}}(3\kappa, 2\mu, 2\eta) \equiv 3\kappa\,\widehat{J} + 2\mu\,\widehat{L} + 2\eta\,\widehat{M}, \qquad \kappa, \mu, \eta > 0. \qquad (19)$$

The isotropic matrices $\{\widehat{J}, \widehat{K}\}$ are idempotent and their matrix product vanishes: $\widehat{J}^2 = \widehat{J}$, $\widehat{K}^2 = \widehat{K}$, $\widehat{J}\widehat{K} = \widehat{K}\widehat{J} = 0$. Similarly, it may be checked that the three basis matrices for cubic materials $\{\widehat{J}, \widehat{L}, \widehat{M}\}$ are idempotent and have zero mutual products. The algebra of matrix multiplication for isotropic and cubic materials follows from these basic multiplication tables:

	$\widehat{J}$	$\widehat{K}$
$\widehat{J}$	$\widehat{J}$	0
$\widehat{K}$	0	$\widehat{K}$

	$\widehat{J}$	$\widehat{L}$	$\widehat{M}$
$\widehat{J}$	$\widehat{J}$	0	0
$\widehat{L}$	0	$\widehat{L}$	0
$\widehat{M}$	0	0	$\widehat{M}$

Thus, the inverses are

$$\widehat{S}_{\mathrm{cub}} = \widehat{C}_{\mathrm{cub}}^{-1} = \widehat{C}_{\mathrm{cub}}\left(\frac{1}{3\kappa}, \frac{1}{2\mu}, \frac{1}{2\eta}\right), \quad \widehat{S}_{\mathrm{iso}} = \widehat{C}_{\mathrm{iso}}\left(\frac{1}{3\kappa}, \frac{1}{2\mu}\right),$$

and the products are

$$\widehat{C}_{\mathrm{iso}}^{-1}(3\kappa_1, 2\mu_1)\,\widehat{C}_{\mathrm{iso}}(3\kappa_2, 2\mu_2) \equiv \frac{\kappa_2}{\kappa_1}\,\widehat{J} + \frac{\mu_2}{\mu_1}\,\widehat{K}, \qquad (20)$$

$$\widehat{C}_{\mathrm{cub}}^{-1}(3\kappa_1, 2\mu_1, 2\eta_1)\,\widehat{C}_{\mathrm{cub}}(3\kappa_2, 2\mu_2, 2\eta_2) \equiv \frac{\kappa_2}{\kappa_1}\,\widehat{J} + \frac{\mu_2}{\mu_1}\,\widehat{L} + \frac{\eta_2}{\eta_1}\,\widehat{M}. \qquad (21)$$

Results for isotropic materials follow from those for cubic with $\eta = \mu$. For the sake of simplicity and brevity we therefore focus on properties for cubic materials in the next subsection.

3.2. *Elastic distance for cubic and isotropic materials.* Consider two cubic materials with moduli $\widehat{C}_1 = \widehat{C}_{\mathrm{cub}}(3\kappa_1, 2\mu_1, 2\eta_1)$ and $\widehat{C}_2 = \widehat{C}_{\mathrm{cub}}(3\kappa_2, 2\mu_2, 2\eta_2)$. The Euclidean distance function of Equation (3) follows from the above properties

and the relations $\operatorname{tr} \widehat{\boldsymbol{J}} = 1$, $\operatorname{tr} \widehat{\boldsymbol{L}} = 3$, $\operatorname{tr} \widehat{\boldsymbol{M}} = 2$. Similarly, the Riemannian and log-Euclidean distances follow from the identities

$$\operatorname{Log}(\widehat{\boldsymbol{C}}_2) - \operatorname{Log}(\widehat{\boldsymbol{C}}_1) = \operatorname{Log}\widehat{\boldsymbol{C}}_1^{-1}\widehat{\boldsymbol{C}}_2 = \ln\left(\frac{\kappa_2}{\kappa_1}\right)\widehat{\boldsymbol{J}} + \ln\left(\frac{\mu_2}{\mu_1}\right)\widehat{\boldsymbol{L}} + \ln\left(\frac{\eta_2}{\eta_1}\right)\widehat{\boldsymbol{M}}. \quad (22)$$

Thus, the distances functions are

$$d_F(\widehat{\boldsymbol{C}}_1, \widehat{\boldsymbol{C}}_2) = \left[(3\kappa_1 - 3\kappa_2)^2 + 3(2\mu_1 - 2\mu_2)^2 + 2(2\eta_1 - 2\eta_2)^2\right]^{1/2}, \quad (23)$$

$$d_{L,R}(\widehat{\boldsymbol{C}}_1, \widehat{\boldsymbol{C}}_2) = \left[\left(\ln\frac{\kappa_2}{\kappa_1}\right)^2 + 3\left(\ln\frac{\mu_2}{\mu_1}\right)^2 + 2\left(\ln\frac{\eta_2}{\eta_1}\right)^2\right]^{1/2}. \quad (24)$$

It is clear that d_L and d_R are invariant under inversion,

$$d_{L,R}(\widehat{\boldsymbol{S}}_1, \widehat{\boldsymbol{S}}_2) = d_{L,R}(\widehat{\boldsymbol{C}}_1, \widehat{\boldsymbol{C}}_2).$$

Note that the first identity in (22) is a consequence of the fact that $\widehat{\boldsymbol{C}}_1$ and $\widehat{\boldsymbol{C}}_2$ commute, which is not true in general for material symmetries lower than cubic.

What is the isotropic material closest to a given cubic material? The answer may be found by considering the distance functions between an arbitrary cubic stiffness $\widehat{\boldsymbol{C}}_{\mathrm{cub}}(3\kappa, 2\mu, 2\eta)$ and the isotropic stiffness $\widehat{\boldsymbol{C}}_{\mathrm{iso}}(3\kappa_*, 2\mu_*)$. The same question will also be considered for the compliances. Minimizing with respect to the isotropic moduli κ_*, μ_* yields

$$\min_{\kappa_*, \mu_*} d_{L,R}(\widehat{\boldsymbol{C}}_{\mathrm{cub}}, \widehat{\boldsymbol{C}}_{\mathrm{iso}}(3\kappa_*, 2\mu_*)) = \min_{\kappa_*, \mu_*} d_{L,R}(\widehat{\boldsymbol{S}}_{\mathrm{cub}}, \widehat{\boldsymbol{S}}_{\mathrm{iso}}) = \sqrt{\frac{6}{5}}\left|\ln\frac{\mu}{\eta}\right|, \quad (25)$$

$$\min_{\kappa_*, \mu_*} d_F(\widehat{\boldsymbol{C}}_{\mathrm{cub}}, \widehat{\boldsymbol{C}}_{\mathrm{iso}}(3\kappa_*, 2\mu_*)) = \sqrt{\frac{6}{5}}\,|2\mu - 2\eta|, \quad (26)$$

$$\min_{\kappa_*, \mu_*} d_F(\widehat{\boldsymbol{C}}_{\mathrm{cub}}^{-1}, \widehat{\boldsymbol{C}}_{\mathrm{iso}}^{-1}(3\kappa_*, 2\mu_*)) = \sqrt{\frac{6}{5}}\left|\frac{1}{2\mu} - \frac{1}{2\eta}\right|. \quad (27)$$

Denote the values of the closest isotropic moduli by (κ_L, μ_L), (κ_R, μ_R) for d_L, d_R, and (κ_A, μ_A) or (κ_H, μ_H) for d_F depending on whether the stiffness (A) or its inverse (H) is used. Thus,

$$\kappa_{L,R,A,H} = \kappa, \quad \mu_{L,R} = (\mu^3\eta^2)^{1/5}, \quad \mu_A = \frac{3}{5}\mu + \frac{2}{5}\eta, \quad \frac{1}{\mu_H} = \frac{3}{5\mu} + \frac{2}{5\eta}. \quad (28)$$

Equations (25) and (28) show clearly that the "closest" isotropic material using the Frobenius metric is ambiguous because it depends on whether one uses stiffness or compliance. Each gives a different isotropic material since $\mu_H < \mu_{L,R} < \mu_A$ for $\mu - \eta \neq 0$. The Riemannian and log-Euclidean metrics give the same unique "closest" isotropic material, regardless of whether the stiffness or the compliance is used. The fact that they agree is particular to the case of cubic symmetry, as noted above, and is not true in general.

In summary, the closest isotropic material to a given cubic material, in the sense of d_R and d_L, is defined by moduli

$$\kappa_R = \kappa_L = \tfrac{1}{3}(\hat{c}_{11} + 2\hat{c}_{12})$$

and

$$\mu_R = \mu_L = \tfrac{1}{2}\left[\hat{c}_{44}^3(\hat{c}_{11} - \hat{c}_{12})^2\right]^{1/5},$$

and the distance from isotropy is

$$d_{L,R} = \sqrt{\frac{6}{5}}\left|\ln\frac{\hat{c}_{11} - \hat{c}_{12}}{\hat{c}_{44}}\right|.$$

These results will be generalized to materials of arbitrary anisotropy next.

4. Closest isotropic moduli

We now turn to the more general question of finding the isotropic material closest to a given anisotropic material characterized by $\widehat{C}$ or its inverse $\widehat{S}$. The solution using the Euclidean metric is relatively simple, and is considered first.

4.1. _Minimum Frobenius distances._ The closest isotropic elastic moduli are assumed to be of general isotropic form $\widehat{C}_{\text{iso}}(3\kappa, 2\mu)$; see Equations (18)–(19). The bulk and shear moduli are found by minimizing $d_F(\widehat{C}_{\text{iso}}, \widehat{C})$, which implies

$$3\kappa \ \text{tr} \ \widehat{J} = \text{tr} \ \widehat{J}\widehat{C}, \qquad 2\mu \ \text{tr} \ \widehat{K} = \text{tr} \ \widehat{K}\widehat{C}. \tag{29}$$

Using suffix A to indicate that the minimization is in the arithmetic sense (in line with [Moakher 2006]),

$$\begin{aligned}
9\kappa_A &= \hat{c}_{11} + \hat{c}_{22} + \hat{c}_{33} + 2(\hat{c}_{23} + \hat{c}_{31} + \hat{c}_{12}), \\
30\mu_A &= 2(\hat{c}_{11} + \hat{c}_{22} + \hat{c}_{33} - \hat{c}_{23} - \hat{c}_{31} - \hat{c}_{12}) + 3(\hat{c}_{44} + \hat{c}_{55} + \hat{c}_{66}),
\end{aligned} \tag{30}$$

which are well known; see, for example, [Fedorov 1968]. Similarly, the closest isotropic elastic compliance can be determined by minimizing

$$d_F(\widehat{C}_{\text{iso}}^{-1}, \widehat{C}^{-1}).$$

Denoting the isotropic moduli with the suffix H for harmonic,

$$\begin{aligned}
1/\kappa_H &= \hat{s}_{11} + \hat{s}_{22} + \hat{s}_{33} + 2(\hat{s}_{23} + \hat{s}_{31} + \hat{s}_{12}), \\
15/(2\mu_H) &= 2(\hat{s}_{11} + \hat{s}_{22} + \hat{s}_{33} - \hat{s}_{23} - \hat{s}_{31} - \hat{s}_{12}) + 3(\hat{s}_{44} + \hat{s}_{55} + \hat{s}_{66}).
\end{aligned} \tag{31}$$

The Euclidean distance does not provide a unique closest isotropic material, although the values in Equations (30) and (31) are sometimes considered as bounds. Equations (29) and (30) also agree with the special case discussed above for cubic materials, Equation (28).

4.2. *Minimum log-Euclidean distance.* The isotropic elasticity $\widehat{C}_{\text{iso}}(3\kappa_L, 2\mu_L)$ is found using the same methods as above by replacing $\widehat{C}_{\text{iso}}$ and $\widehat{C}$ with $\text{Log}(\widehat{C}_{\text{iso}})$ and $\text{Log}(\widehat{C})$, respectively. Thus,

$$\log(3\kappa_L) = \text{tr } \widehat{J} \text{ Log}(\widehat{C}), \qquad 5 \log(2\mu_L) = \text{tr } \widehat{K} \text{ Log}(\widehat{C}). \tag{32}$$

Adding the two equations and using $\widehat{J} + \widehat{K} = \widehat{I}$, implies the identity

$$\det(\widehat{C}_{\text{iso}}) = \det(\widehat{C}). \tag{33}$$

Thus, we have explicit formulae for the closest moduli,

$$\kappa_L = \tfrac{1}{3} \exp\left(\text{tr } \widehat{J} \text{ Log}(\widehat{C})\right), \qquad \mu_L = \tfrac{1}{2} \exp\left(\tfrac{1}{5} \text{ tr } \widehat{K} \text{ Log}(\widehat{C})\right). \tag{34}$$

4.3. *The minimum Riemannian distance.* We look for moduli of the form

$$\widehat{C}_{\text{iso}}(3\kappa_R, 2\mu_R) = 3\kappa_R \widehat{J} + 2\mu_R \widehat{K}, \tag{35}$$

which minimize

$$\text{d}_R^2(\widehat{C}_{\text{iso}}, \widehat{C}) = \text{tr}\left[\text{Log}^2(\widehat{C}_{\text{iso}}^{-1} \widehat{C})\right]. \tag{36}$$

This is achieved using the following result (Proposition 2.1 of [Moakher 2005]) for any invertible matrix $X(t)$ that does not have negative real-valued eigenvalues,

$$\frac{\text{d}}{\text{d}t} \text{tr}\left[\text{Log}^2 X(t)\right] = 2 \text{ tr}\left[\text{Log } X(t)X^{-1}(t)\frac{\text{d}}{\text{d}t} X(t)\right]. \tag{37}$$

Differentiating (36) with respect to κ_R and μ_R separately, implies respectively

$$\text{tr}\left[\widehat{C}_{\text{iso}}^{-1} \widehat{J} \text{ Log}(\widehat{C}_{\text{iso}}^{-1} \widehat{C})\right] = 0, \quad \text{tr}\left[\widehat{C}_{\text{iso}}^{-1} \widehat{K} \text{ Log}(\widehat{C}_{\text{iso}}^{-1} \widehat{C})\right] = 0. \tag{38}$$

Further simplification yields

$$\text{tr}\left[\widehat{J} \text{ Log}(\widehat{C}_{\text{iso}}^{-1} \widehat{C})\right] = 0, \quad \text{tr}\left[\widehat{K} \text{ Log}(\widehat{C}_{\text{iso}}^{-1} \widehat{C})\right] = 0. \tag{39}$$

These conditions, which are necessary for a minimum, can be simplified as follows. Define the eigenvalues and associated eigenvectors by the diagonalization

$$\widehat{C}_{\text{iso}}^{-1/2} \widehat{C} \widehat{C}_{\text{iso}}^{-1/2} = \sum_{i=1}^{n} \lambda_i \, v_i v_i^T. \tag{40}$$

Adding the two conditions (39) using the identity $\widehat{I} = \widehat{J} + \widehat{K}$, along with the expression (2) for the logarithm of a matrix, yields

$$\prod_{i=1}^{n} \lambda_i = 1. \tag{41}$$

A second condition follows by direct substitution from (40) into the first of (39), giving

$$\prod_{i=1}^{n} \lambda_i^{\alpha_i} = 1 , \quad \alpha_i \equiv \boldsymbol{v}_i^T \,\widehat{\boldsymbol{J}} \boldsymbol{v}_i, \quad i = 1, 2, \ldots n. \tag{42}$$

Note that $0 \le \alpha_i \le 1$ and α_i form a partition of unity,

$$\sum_{i=1}^{n} \alpha_i = 1 . \tag{43}$$

This follows from the representation $\widehat{\boldsymbol{J}} = \boldsymbol{u}\boldsymbol{u}^T$ where the unit 6 vector $\boldsymbol{u}$ is defined in Equation (16). Thus, the minimal isotropic moduli are found by satisfying the two simultaneous Equations (41) and (42). We now show how the first of these two conditions can be met, leaving one condition to satisfy.

Let

$$\widehat{\boldsymbol{C}}_{\text{iso}} = 3\kappa_R \left(\widehat{\boldsymbol{J}} + \rho^{-2} \, \widehat{\boldsymbol{K}} \right), \tag{44}$$

where $\rho \ge 0$ is defined by

$$\rho^2 = \frac{3\kappa_R}{2\mu_R} = \frac{1 + \nu_R}{1 - 2\nu_R}, \tag{45}$$

and ν_R is the Poisson's ratio of the minimizer. We choose this form for $\widehat{\boldsymbol{C}}_{\text{iso}}$ so that $\widehat{\boldsymbol{C}}_{\text{iso}}^{-1/2} = (3\kappa_R)^{-1/2} \left(\widehat{\boldsymbol{J}} + \rho \widehat{\boldsymbol{K}} \right)$. Hence, the eigenvalues of (40) are of the form

$$\lambda_i = \frac{\bar{\lambda}_i(\rho)}{3\kappa_R} , \tag{46}$$

where the normalized eigenvectors $\bar{\lambda}_i = \bar{\lambda}_i(\rho)$ and the (unchanged) eigenvectors $\boldsymbol{v}_i$, $i = 1, 2, \ldots, n = 6$ are defined by

$$3\kappa_R \, \widehat{\boldsymbol{C}}_{\text{iso}}^{-1/2} \widehat{\boldsymbol{C}} \, \widehat{\boldsymbol{C}}_{\text{iso}}^{-1/2} = \left(\widehat{\boldsymbol{J}} + \rho \widehat{\boldsymbol{K}} \right) \widehat{\boldsymbol{C}} \left(\widehat{\boldsymbol{J}} + \rho \widehat{\boldsymbol{K}} \right) = \sum_{i=1}^{n} \bar{\lambda}_i \, \boldsymbol{v}_i \boldsymbol{v}_i^T . \tag{47}$$

Turning to the first condition, (41), it is automatically satisfied if the bulk modulus is given by

$$3\kappa_R = \left(\prod_{i=1}^{n} \bar{\lambda}_i \right)^{1/n}. \tag{48}$$

It remains to determine ρ from the second stationary condition, Equation (42), which can be expressed in terms of the modified eigenvalues as

$$\prod_{i=1}^{n} \bar{\lambda}_i^{(\alpha_i - 1/n)} = 1 . \tag{49}$$

Equation (49) involves the eigenvectors v through the inner products α_i. However, α_i vanishes identically for eigenvectors of *deviatoric* form—in fact the definition of a deviatoric eigenvector is $\alpha_i = 0$ [Mehrabadi and Cowin 1990]. Conversely, $\alpha_i = 1$ for purely *dilatational* eigenvectors [Mehrabadi and Cowin 1990], that is, eigenvectors parallel to u of Equation (16).

The solution to Equation (49) may be found numerically by searching for the zero over the permissible range for the Poisson's ratio: $-1 < \nu_R < 1/2$. The minimizing moduli κ_R and μ_R then follow from Equations (48) and (45), or more directly,

$$3\kappa_R = \rho^{5/3} \left(\det \widehat{C} \right)^{1/6}, \qquad 2\mu_R = \rho^{-1/3} \left(\det \widehat{C} \right)^{1/6}, \tag{50}$$

and the minimal distance between $\widehat{C}_{\text{iso}}$ and $\widehat{C}$ is given by

$$d_R(\widehat{C}_{\text{iso}}, \widehat{C}) = \frac{1}{n} \left[\sum_{i=1}^{n} \ln^2 \left((\bar{\lambda}_i)^{-n} \prod_{j=1}^{n} \bar{\lambda}_j \right) \right]^{1/2} \quad (n = 6). \tag{51}$$

We next demonstrate the application of the above procedure to the case of a given elasticity matrix of cubic symmetry.

4.4. *Example: cubic materials.* By substituting the assumed form $\widehat{C} = \widehat{C}_{\text{cub}}$ from Equation (19) into the explicit formulae of Equation (34) for the closest moduli in the log-Euclidean sense, it is a straightforward matter to show that the latter reproduce the results determined directly, in Equation (28). Regarding the closest moduli using the Riemannian distance, the matrix in Equation (47) follows by using the algebra for cubic matrices,

$$\left(\widehat{J} + \rho \widehat{K} \right) \widehat{C} \left(\rho \widehat{J} + \rho \widehat{K} \right) = 3\kappa \widehat{J} + 2\mu\rho^2 \widehat{L} + 2\eta\rho^2 \widehat{M}. \tag{52}$$

Thus, $\bar{\lambda}_1 = 3\kappa$, $\bar{\lambda}_2 = \bar{\lambda}_3 = \bar{\lambda}_4 = 2\mu\rho^2$, $\bar{\lambda}_5 = \bar{\lambda}_6 = 2\eta\rho^2$, and the eigenvectors are either pure dilatational ($\alpha_1 = 1$) or deviatoric ($\alpha_i = 0$, $i = 2, 3, \ldots, 6$). Therefore, Equation (49) becomes

$$(3\kappa)^{5/6} (2\mu)^{-1/2} (2\eta)^{-1/3} \rho^{-5/3} = 1. \tag{53}$$

Solving for the intermediate variable ρ, and evaluating μ_R and κ_R from Equations (48) and (45), respectively, gives $\kappa_R = \kappa$ and $\mu_R = (\mu^3 \eta^2)^{1/5}$, again in agreement with Equation (28).

5. Applications and discussion

Table 1 lists the computed distance from isotropy of various anisotropic materials, using data from Musgrave [2003]. Materials of cubic (cub), hexagonal (hex), tetragonal (tet) and orthotropic (ort) symmetry are considered. In each case the

moduli of the closest isotropic material were found using the algorithm described above. The resulting bulk modulus κ_R and Poisson's ratio ν_R are tabulated.

Table 1 ranks the materials in terms of the Riemannian distance d_R of the original anisotropic moduli from the closest isotropic material. The second column of numbers lists the distance between the closest isotropic materials found using the Riemannian and log-Euclidean distances. That is,

$$
\begin{aligned}
d_{LR} &\equiv d_{L,R}\big(\widehat{\boldsymbol{C}}_{\text{iso}}(3\kappa_R, 2\mu_R), \widehat{\boldsymbol{C}}_{\text{iso}}(3\kappa_L, 2\mu_L)\big) \\
&= \left[\left(\ln\frac{\kappa_L}{\kappa_R}\right)^2 + 5\left(\ln\frac{\mu_L}{\mu_R}\right)^2\right]^{1/2},
\end{aligned}
\tag{54}
$$

which is identically zero for cubic materials. The arithmetic (κ_A, μ_A) and harmonic (κ_H, μ_H) moduli minimizing the Euclidean distances were also computed, and the Riemannian distance between these two is denoted d_{HA}. The distances d_{RA} and d_{RH} are the distances between the closest isotropic material (κ_R, μ_R) and the arithmetic and harmonic isotropic approximants, respectively. All distances listed in Table 1 are based on the Riemannian metric.

Note that the distance between the closest materials using d_R and d_L is less than 0.05 except for the extremely anisotropic spruce. In order to gain some appreciation for the magnitude of the nondimensional distances in Table 1, consider the distance of any $\boldsymbol{P} \in \mathscr{P}(n)$ from a multiple of itself:

$$
d_R(\boldsymbol{P}, a\boldsymbol{P}) = d_L(\boldsymbol{P}, a\boldsymbol{P}) = \sqrt{n}\,|\log a|, \quad a \in \mathbb{R}_+ .
\tag{55}
$$

Small values of the elastic distance can be identified with values of a close to unity, specifically

$$
a = 1 \pm \frac{1}{\sqrt{6}}\,d_{L,R} + O(d_{L,R}^2) \approx 1 \pm 0.4\,d_{L,R} .
\tag{56}
$$

Note that the distance d_{HA} between the arithmetic and harmonic approximations is generally less than the distance from isotropy d_R. This is more so for those materials that are closer to isotropy—at the top of Table 1. As the material gets further from isotropy - the lower half of Table 1—the magnitude of d_{HA} relative to d_R grows as the latter increases. The two distances are of comparable magnitude for the highly anisotropic materials at the very bottom of the table, such as oak and spruce.

As a numerical check on the computations, the triangle inequality

$$
d_{HA} \le d_{RA} + d_{RH}
\tag{57}
$$

Material	Symm	d_R	$100d_{LR}$	d_{RA}	d_{RH}	d_{HA}	ν_R	κ_R
magnesium	hex	0.18	0.00	0.01	0.01	0.02	0.29	3.53
diamond	cub	0.21	0	0.01	0.01	0.02	0.07	44.20
aluminum	cub	0.21	0	0.01	0.01	0.02	0.35	7.69
beryllium	hex	0.22	0.01	0.01	0.01	0.02	0.05	11.44
sodium fluoride	cub	0.29	0	0.02	0.02	0.04	0.24	4.86
ice (H_2O) 257°K	hex	0.31	0.00	0.02	0.02	0.04	0.33	0.89
β-quartz (SiO_2)	hex	0.35	0.02	0.03	0.03	0.05	0.21	5.64
beryllium	hex	0.37	0.23	0.03	0.03	0.06	0.26	14.41
caesium iodide	cub	0.37	0	0.03	0.03	0.06	0.27	1.29
sodium chloride	cub	0.40	0	0.04	0.03	0.07	0.25	2.45
sodium iodide	cub	0.43	0	0.04	0.04	0.08	0.25	1.46
sodium bromide	cub	0.44	0	0.04	0.04	0.09	0.25	1.94
caesium bromide	cub	0.45	0	0.05	0.04	0.09	0.27	1.59
silicon	cub	0.49	0	0.05	0.05	0.11	0.22	9.78
cobalt	hex	0.51	0.00	0.07	0.05	0.12	0.31	19.03
silver bromide	cub	0.52	0	0.06	0.06	0.12	0.40	4.06
germanium	cub	0.56	0	0.07	0.07	0.14	0.21	7.52
caesium chloride	cub	0.58	0	0.08	0.07	0.15	0.27	1.83
gallium antimonide	cub	0.64	0	0.09	0.10	0.18	0.25	5.64
α-uranium	ort	0.68	0.37	0.10	0.10	0.20	0.20	11.28
silver chloride	cub	0.70	0	0.11	0.10	0.22	0.41	4.42
apatite	hex	0.72	0.11	0.10	0.13	0.22	0.21	8.43
indium antimonide	cub	0.75	0	0.12	0.13	0.25	0.29	4.69
potassium fluoride	cub	0.75	0	0.13	0.12	0.25	0.28	3.19
benzophenone	ort	0.85	1.92	0.15	0.14	0.29	0.30	5.14
zircon	tet	0.98	0.74	0.21	0.18	0.39	0.13	1.99
sulphur	ort	0.98	4.13	0.20	0.18	0.39	0.34	1.88
iron	cub	0.99	0	0.20	0.23	0.43	0.30	17.05
nickel	cub	1.01	0	0.21	0.23	0.44	0.29	18.04
cadmium	hex	1.04	3.43	0.20	0.24	0.44	0.30	5.40
rutile (TiO_2)	tet	1.07	0.79	0.21	0.28	0.49	0.27	21.49
potassium chloride	cub	1.08	0	0.27	0.24	0.50	0.28	1.78
barium titanate	tet	1.13	3.20	0.26	0.27	0.52	0.36	17.67
potassium bromide	cub	1.14	0	0.30	0.26	0.56	0.29	1.58
gold	cub	1.16	0	0.27	0.31	0.58	0.42	17.28
Rochelle salt	ort	1.17	0.97	0.24	0.34	0.59	0.31	1.97
zinc	hex	1.18	2.58	0.24	0.34	0.57	0.24	6.61
white tin	tet	1.18	0.04	0.24	0.38	0.62	0.35	5.50
ammon. dihyd. phos.	tet	1.19	0.95	0.36	0.25	0.61	0.33	2.70
silver	cub	1.21	0	0.29	0.33	0.63	0.37	10.36
potassium iodide	cub	1.25	0	0.36	0.31	0.67	0.30	1.20
copper	cub	1.28	0	0.32	0.37	0.70	0.35	13.71
potass. dihyd. phos.	tet	1.34	0.01	0.40	0.38	0.78	0.26	2.67
α-brass	cub	1.46	0	0.41	0.48	0.90	0.34	11.96
indium	tet	1.57	0.01	0.50	0.54	1.04	0.44	4.16
oak	ort	2.30	1.75	0.96	1.09	2.05	0.08	0.17
β-brass	cub	2.34	0	0.94	1.19	2.13	0.36	11.62
spruce	ort	5.66	59.5	7.16	3.33	10.45	0.23	0.09

Table 1. Distance from isotropy for some materials - data from
[Musgrave 2003]. κ_R units 10^{10} N/m^2.

was confirmed for each material in Table 1. Since the three vertices of the triangle are isotropic materials, the inequality may be written, using (24), as

$$
\left[(\ln \frac{\kappa_A}{\kappa_H})^2 + 5(\ln \frac{\mu_A}{\mu_H})^2\right]^{1/2}
$$
$$
\leq \left[(\ln \frac{\kappa_A}{\kappa_R})^2 + 5(\ln \frac{\mu_A}{\mu_R})^2\right]^{1/2} + \left[(\ln \frac{\kappa_R}{\kappa_H})^2 + 5(\ln \frac{\mu_R}{\mu_H})^2\right]^{1/2}. \quad (58)
$$

For cubic materials $\kappa_A = \kappa_H = \kappa_R$, and consequently the triangle is a straight line:

$$
d_{HA} = d_{RA} + d_{RH} \quad \text{for cubic materials only.} \quad (59)
$$

The quantity $(d_{RA} + d_{RH} - d_{HA})/d_{HA}$ was found to be very small for all the cases considered (and numerically zero for the cubic examples), less than 10^{-3} for all materials considered except barium titanate (1.2×10^{-3}) and spruce (2.8×10^{-3}). The "triangle" is almost flat, indicating that the closest moduli (κ_R, μ_R) are in some sense optimally centered between the arithmetic and harmonic approximations. Note however, that κ_R, μ_R are not equal to the Riemannian mean [Moakher 2006] of the arithmetic and harmonic approximations, denoted as κ_{AH}, μ_{AH}. The Riemannian mean of two elasticity matrices $\widehat{C}_1$ and $\widehat{C}_2$ is $\widehat{C}_1(\widehat{C}_1^{-1}\widehat{C}_2)^{1/2}$ [Moakher 2006], and consequently the means of the arithmetic and harmonic moduli are $\kappa_{AH} = (\kappa_A \kappa_H)^{1/2}$, $\mu_{AH} = (\mu_A \mu_H)^{1/2}$. By considering the case of cubic materials, for which all these quantities have explicit expressions, it may be shown that $(\mu_R - \mu_{AH})(\eta - \mu) > 0$ for $\eta - \mu \neq 0$.

6. Conclusions

We have presented a method for finding the isotropic elastic moduli closest to a given material of arbitrary symmetry based on three different metrics. Unlike the Frobenius (Euclidean) distance, the Riemmanian and log-Euclidean metrics provide unique isotropic moduli. The values obtained according to these two metrics are identical if the comparison medium has cubic symmetry, and are otherwise relatively close. The procedures developed here for finding the closest isotropic moduli can be generalized to find the closest material of lower symmetry. The solution for cubic symmetry with the cube axes given is presented in the Appendix, and other, lower symmetries will be considered elsewhere. Another generalization of the present problem is that of determining the closest material of cubic or lower symmetry where the symmetry axes are unrestrained. These and other challenging questions make this an interesting topic for some time to come.

Appendix: The closest cubic material

The cubic stiffness (compliance) closest to $\widehat{C}$ $(\widehat{S})$ in the Euclidean metric d_F has moduli κ_A, μ_A, η_A (κ_H, μ_H and η_H), where κ_A and κ_H are given by Equations (30) and (31), and

$$6\mu_A = \hat{c}_{44} + \hat{c}_{55} + \hat{c}_{66}, \quad 6\eta_A = \hat{c}_{11} + \hat{c}_{22} + \hat{c}_{33} - \hat{c}_{23} - \hat{c}_{31} - \hat{c}_{12}, \tag{A.1}$$

$$\frac{3}{2\mu_H} = \hat{s}_{44} + \hat{s}_{55} + \hat{s}_{66}, \quad \frac{3}{2\eta_H} = \hat{s}_{11} + \hat{s}_{22} + \hat{s}_{33} - \hat{s}_{23} - \hat{s}_{31} - \hat{s}_{12}. \tag{A.2}$$

Using the method for deriving Equation (34), we find the following for the log-Euclidean distance,

$$\begin{aligned}
\kappa_L &= \tfrac{1}{3} \exp\bigl(\operatorname{tr} \widehat{J} \operatorname{Log}(\widehat{C})\bigr), \\
\mu_L &= \tfrac{1}{2} \exp\bigl(\tfrac{1}{3} \operatorname{tr} \widehat{L} \operatorname{Log}(\widehat{C})\bigr), \\
\eta_L &= \tfrac{1}{2} \exp\bigl(\tfrac{1}{2} \operatorname{tr} \widehat{M} \operatorname{Log}(\widehat{C})\bigr).
\end{aligned} \tag{A.3}$$

Note the identity, similar to Equation (33),

$$\det(\widehat{C}_{\text{cub}}) = \det(\widehat{C}). \tag{A.4}$$

For the Riemannian distance d_R we find that the closest cubic material $\widehat{C}_{\text{cub}}$ of the form Equation (19) is determined by three equations:

$$\prod_{i=1}^{n} \lambda_i = 1, \quad \prod_{i=1}^{n} \lambda_i^{\alpha_i} = 1, \quad \prod_{i=1}^{n} \lambda_i^{\beta_i} = 1, \tag{A.5}$$

where

$$\alpha_i \equiv v_i^T \widehat{J} v_i, \quad \beta_i \equiv v_i^T \widehat{L} v_i, \quad i = 1, 2, \ldots n, \tag{A.6}$$

and $\{\lambda_i, v_i\}$ are the eigenvalues and eigenvectors of

$$\widehat{C}_{\text{cub}}^{-1/2} \widehat{C} \widehat{C}_{\text{cub}}^{-1/2}.$$

The parameters α_i satisfy the same properties as before, including the fact that they sum to unity. Since $\{v_i\}$ form an orthonormal basis, it follows that

$$\sum_{i=1}^{n} \beta_i = \dim \widehat{L} = 3.$$

Furthermore, $\beta_i = 0$ if the eigenvector is dilatational. The three equations (A.5) may be reduced to two by assuming the unknown moduli are of the form

$$\widehat{C}_{\text{cub}} = 3\kappa_R \bigl(\widehat{J} + \rho_1^{-2}\widehat{L} + \rho_2^{-2}\widehat{M}\bigr).$$

Define the modified eigenvalues $\bar{\lambda}_i = \bar{\lambda}_i(\rho_1, \rho_2)$ to be the eigenvalues of

$$\left(\widehat{\boldsymbol{J}} + \rho_1 \widehat{\boldsymbol{L}} + \rho_2 \widehat{\boldsymbol{M}}\right) \widehat{\boldsymbol{C}} \left(\widehat{\boldsymbol{J}} + \rho_1 \widehat{\boldsymbol{L}} + \rho_2 \widehat{\boldsymbol{M}}\right),$$

then κ_R is given by the formula (48), while ρ_1, ρ_1 solve the simultaneous equations

$$\prod_{i=1}^{n} \bar{\lambda}_i^{(\alpha_i - 1/n)} = 1, \qquad \prod_{i=1}^{n} \bar{\lambda}_i^{(\beta_i - 1/n)} = 1. \qquad (A.7)$$

Acknowledgements

References

[Arsigny et al. 2005] V. Arsigny, P. Fillard, X. Pennec, and N. Ayache, "Fast and simple calculus on tensors in the Log-Euclidean framework", pp. 115–122 in *8th Int. Conf. on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, edited by J. Duncan and G. Gerig, Lecture Notes in Computer Science **3749**, Springer, 2005.

[Arts et al. 1991] R. J. Arts, K. Helbig, and P. N. J. Rasolofosaon, "General anisotropic elastic tensors in rocks: approximation, invariants and particular directions", pp. 1534–1537 in *61st Annual International Meeting, Society of Exploration Geophysicists*, vol. ST2.4, Tulsa, 1991. Expanded Abstracts.

[Browaeys and Chevrot 2004] J. T. Browaeys and S. Chevrot, "Decomposition of the elastic tensor and geophysical applications", *Geophys. J. Int.* **159** (2004), 667–678.

[Cavallini 1999] F. Cavallini, "The best isotropic approximation of an anisotropic Hooke's law", *Bollettino di Geofisica Teorica e Applicata* **40** (1999), 1–18.

[Fedorov 1968] F. I. Fedorov, *Theory of elastic waves in crystals*, Plenum Press, New York, 1968.

[Gangi 2000] A. F. Gangi, "Fourth-order elastic-moduli tensors by inspection, anisotropy: fractures, converted waves and case studies", pp. 1–10 in *9th International Workshop on Seismic Anisotropy (9IWSA)*, edited by L. Ikelle and A. F. Gangi, Society of Exploration Geophysicists, Tulsa, 2000.

[Gazis et al. 1963] D. C. Gazis, I. Tadjbakhsh, and R. A. Toupin, "The elastic tensor of given symmetry nearest to an anisotropic elastic tensor", *Acta Cryst.* **16** (1963), 917–922.

[Helbig 1996] K. Helbig, "Representation and approximation of elastic tensors", pp. 37–75 in *Seismic anisotropy*, edited by E. Fjaer et al., Society of Exploration Geophysicists, Tulsa, 1996.

[Lang 1998] S. Lang, *Fundamentals of differential geometry*, Springer, New York, 1998.

[Matthies and Humbert 1995] S. Matthies and M. Humbert, "On the principle of a geometric mean of even-rank symmetric tensors for textured polycrystais", *J. Appl. Cryst.* **28** (1995), 254–266.

[Mehrabadi and Cowin 1990] M. M. Mehrabadi and S. C. Cowin, "Eigentensors of linear anisotropic elastic materials", *Q. J. Mech. Appl. Math.* **43** (1990), 15–41.

[Moakher 2005] M. Moakher, "A differential geometric approach to the geometric mean of symmetric positive-definite matrices", *SIAM J. Matrix Anal. Appl.* **26** (2005), 735–747.

[Moakher 2006] M. Moakher, "On the averaging of symmetric positive-definite tensors", 2006. J. Elasticity.

[Musgrave 2003] M. J. P. Musgrave, *Crystal acoustics*, Acoustical Society of America, New York, 2003.

[Rychlewski 1984] J. Rychlewski, "On Hooke's law", *J. Appl. Math. Mech.* **48** (1984), 303–314. Translated from *Prikl. Mat. Mekh.* **18**:3 (1984), 420-435.

[Thomson 1856] W. Thomson, "Elements of a mathematical theory of elasticity", *Phil. Trans. R. Soc. Lond.* **146** (1856), 481–498.

[Walpole 1984] L. J. Walpole, "Fourth rank tensors of the thirty-two crystal classes: multiplication tables", *Proc. R. Soc. Lond. A* **391** (1984), 149–179.

ANDREW N. NORRIS: norris@rutgers.edu
Rutgers University, Mechanical and Aerospace Engineering, 98 Brett Road, Piscataway, NJ 08854-8058, United States
http://mechanical.rutgers.edu/norris

NONLINEAR DYNAMIC CHARACTERISTICS OF A VIBRO-IMPACT SYSTEM UNDER HARMONIC EXCITATION

JIANLIAN CHENG AND HUI XU

Dynamical behaviors of a two-degree-of-freedom (TDOF) vibro-impact system are investigated. The theoretical solution of periodic-one double-impact motion is obtained by differential equations, periodicity and matching conditions, and the Poincaré map is established. The dynamics of the system are studied with special attention to Hopf bifurcations of the impact system in nonresonance, weak resonance, and strong resonance cases. The Hopf bifurcation theory of maps in $\mathbb{R}^2$-strong resonance is applied to reveal the existence of Hopf bifurcations of the system. The theoretical analyses are verified by numerical solutions. The evolution from periodic impacts to chaos in nonresonance, weak resonance, and strong resonance cases, is obtained by numerical simulations. The results show that dynamical behavior of the system in the strong resonance case is more complicated than that of the nonresonance and weak resonance cases.

1. Introduction

An impact damper is basically a small free mass within a main mass with clearances between the moving masses. Impact dampers in various forms have been used successfully for controlling high-amplitude vibration systems in many practical applications, such as in cutting tools, turbine blades and tall flexible structures like chimneys [Ema and Marui 1996; Cheng and Wang 2003; Wang et al. 2003; Dimentberg and Iourtchenko 2004; Chatterjee and Mallik 1995]. If an impact damper is properly designed, the vibration system structure can be effectively simplified and its performance will be less sensitive to the changes of the system parameters, as compared to a conventional dynamic vibration damper. The vibration of the primary system is controlled by the transfer of momentum to a secondary mass through repeated impacts. Impacts occur when the amplitudes of vibration of the system exceed critical values. Investigation of vibro-impact problems is of significance to the optimization design of machinery with clearances or gaps, and to reliability analysis and noise suppression. Since systems

with impact dampers are strongly nonlinear and discontinuous due to the existence
of one or more impact pairs of components, the vibro-systems can show very rich
and complicated dynamic behavior. In recent years, vibro-impact problems have
become a new subject in nonlinear dynamics. Subjects of recent research include
singularity [Chatterjee and Mallik 1996; Whiston 1992]; inelastic vibro-impacts
[Luo et al. 2001]; high codimension bifurcation [Wen 2001; Luo and Xie 2003;
Xie and Ding 2005]; Hopf bifurcations [Padmanabhan and Singh 1995; Luo and
Chen 2005; Ding et al. 2004; Luo 2004a]; and quasiperiodic impacts [Blazejczyk-
Okolewska 2001; Luo 2004b; Cone and Zadoks 1995]; and so on. Dynamics and
bifurcations of a class of single-degree-of-freedom self-excited oscillators with an
impact damper were studied by Chatterjee and Mallik [1995]. Cone and Zadoks
[1995] investigated the nonlinear behavior of an impact oscillator with the addition
of dry friction. The periodic solutions were interpreted by using bifurcation theory
and the nonlinear behavior of this system was identified as a function of both the
excitation amplitude and the excitation frequency for the two levels of dry friction
force. Many nonlinear dynamical behaviors including turning point bifurcations,
symmetry breaking pitchfork bifurcations, period-doubling bifurcation cascades,
and so on, were explained. Asfar and Akour [2005] studied the suppression of
self-excited vibrations with an impact viscous damper and used the optimization
method to determine the design parameters for suppressing self-excited vibrations.

In recent decades, nonsmooth dynamics of mechanical systems with impacts
have been a focus of several investigations; many new results were obtained and
a few new methods have been established. Holmes [1982] found small horse-
shoe maps in a mathematical model for the bouncing ball. The classical pattern
of period-doubling bifurcation cascade was observed numerically by Shaw and
Holmes [1983] and Thompson and Ghaffari [1982]. Recently, a few researchers
have begun to focus on the quasiperiodic and chaotic motions of vibro-impact
systems. Chatterjee and Mallik [1995] studied quasiperiodic vibro-impacts in a
class of single-degree-of-freedom self-excited oscillators with an impact damper.
Budd et al. [1995] studied vibro-impact of a single-degree-of-freedom system con-
tacting a single stop and proved that if the coefficient of restitution is less than
1, quasiperiodic motion cannot occur in the system. Luo and Xie [2003] inves-
tigated codimension-2 bifurcations of a single-degree-of-freedom impact oscilla-
tor and found a Hopf bifurcation of a period-2 two-impact orbit. Xie and Ding
[2005] studied Hopf bifurcations of a two-degree-of-freedom vibro-impact system
in the strong resonance, nonresonance, and weak resonance cases, and analyzed
the routes of quasiperiodic impacts to chaos. Nordmark [1991] investigated the
nonperiodic motion caused by grazing bifurcation. The normal form mapping
for such grazing phenomena was developed in [di Bernardo et al. 2001; 2002].

Therefore, a periodically forced, piecewise, linear system with impacts is of great interest.

In this paper, the impact damper system serves as a two-degree-of-freedom (TDOF) vibro-impact system with a proportional damping property. We focus our attention on Hopf bifurcation of period motions with one impact in the strong resonance, nonresonance or weak resonance cases. Stability and bifurcation conditions for periodic motion will be obtained. A Poincaré section of the vibro-impact system with proportional damping is chosen to establish the Poincaré map, and then periodic motion with one impact and the stability thereof are investigated by analytical methods. Numerical simulations of periodic and chaotic motions will be presented to validate the analytical results.

2. Mechanical model of the vibro-impact system

The mechanical model for a vibro-impact system with masses M_1 and M_2 is shown in Figure 1. The main mass M_1 is connected to the seat with a linear spring with stiffness K and a linear viscous dashpot of damping constant C. In this system, when the impact mass, or *free mass*, collides with the main mass during vibration, an impulsive force acts on both and produces transfer of momentum with loss of energy. Impact damping is used to simulate an inelastic collision with restitution coefficient $R < 1$. We neglect friction and the duration of the impact between the two masses.

The behavior of the system between any two consecutive impacts is considered. For convenience, the time t between any two consecutive impacts is always set to zero directly at the instant when the former impact is over, and the phase angle is used only to make a suitable choice for the origin of time in the calculation. Phase angle, velocities, and displacements of the system at that instant become initial conditions in the subsequent process of the motion. Between impacts, the differential equations of motion of the vibro-impact system are given by
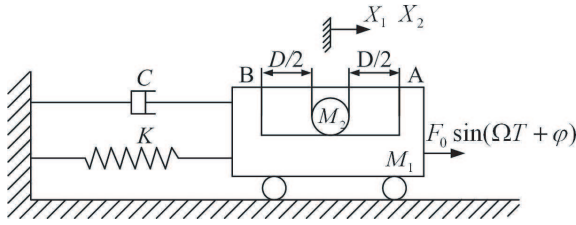
$$M_1 \ddot{X}_1 + C \dot{X}_1 + K X_1 = F_0 \sin(\Omega T + \varphi), \tag{1}$$

$$M_2 \ddot{X}_2 = 0 \quad \text{whenever } |X_1 - X_2| < D/2, \tag{2}$$

where the dot denotes differentiation with respect to time T.

Substituting the nondimensional parameters

$$\omega_n = \sqrt{K/M_1}, \quad \zeta = n/\omega_n,$$
$$n = C/2M_1, \quad \omega = \Omega\sqrt{M_1/K},$$
$$t = T\sqrt{K/M_1}, \quad x_i = K X_i/F_0 \ (i = 1, 2),$$
$$\delta = K D/2F_0, \quad \mu = M_2/M_1$$

Figure 1. Schematic diagram of the impact system.

into Equations (1) and (2) yields

$$\ddot{x}_1 + 2\zeta\omega_n\dot{x}_1 + \omega_n^2 x_1 = \sin(\omega t + \varphi), \tag{3}$$

$$\ddot{x}_2 = 0 \quad \text{whenever } |x_1 - x_2| < \delta, \tag{4}$$

where the dot denotes differentiation with respect to dimensionless time t.

When $|x_1 - x_2| = \delta$, a collision occurs. According to the conservation law of momentum and the definition of the restitution coefficient, we can obtain

$$\dot{x}_{1+} + \mu\dot{x}_{2+} = \dot{x}_{1-} + \mu\dot{x}_{2-}, \tag{5}$$

$$\dot{x}_{2+} - \dot{x}_{1+} = R(\dot{x}_{1-} - \dot{x}_{2-}), \tag{6}$$

where $\dot{x}_{i-}$ and $\dot{x}_{i+}$ ($i = 1, 2$) denotes respectively, the instantaneous velocities before and after impacts. By Equations (5) and (6), the departure velocities of the masses M_1 and M_2 after impact at the instant are given by

$$\begin{aligned}
\dot{x}_{1+} &= \frac{1 - \mu R}{1 + \mu}\dot{x}_{1-} + \frac{\mu(1 + R)}{1 + \mu}\dot{x}_{2-}, \\
\dot{x}_{2+} &= \frac{1 + R}{1 + \mu}\dot{x}_{1-} + \frac{\mu - R}{1 + \mu}\dot{x}_{2-},
\end{aligned} \tag{7}$$

where $\mu = M_2/M_1$.

The general solutions of Equations (3) and (4) are

$$\begin{cases}
x_1(t) = e^{-\zeta\omega_n t}\big(a_{11}\cos(\eta\omega_n t) + b_{11}\sin(\eta\omega_n t)\big) + A\sin(\omega t + \tau), \\
\hphantom{x_1(t) = } 0 \leq t \leq t_1, \\
x_2(t) = a_{12} + b_{12}t,
\end{cases} \tag{8a}$$

$$\begin{cases}
x_1(t) = e^{-\zeta\omega_n(t-t_1)}\big(a_{21}\cos(\eta\omega_n(t - t_1)) + b_{21}\sin(\eta\omega_n(t - t_1))\big) \\
\hphantom{x_1(t) = } + A\sin(\omega t + \tau), \quad t_1 \leq t \leq t_p, \\
x_2(t) = a_{22} + b_{22}(t - t_1)
\end{cases} \tag{8b}$$

where $t_p = t_1 + t_2$, t_1 and t_2 are the traveling time of the impact mass M_2 from $A \to B$ and $B \to A$ in the groove. The integration constants are

$$\eta = \sqrt{1 - \zeta^2},$$
$$\gamma = \omega/\omega_n,$$
$$A = 1/\omega_n^2 \big((1 - \gamma^2)^2 + (2\zeta\gamma)^2\big)^{1/2},$$
$$\tau = \varphi - \psi 1,$$
$$\psi = \tan^{-1}(2\zeta\gamma/(1 - \gamma^2)),$$
$$a_{ki} \text{ and } b_{ki} \quad (k = 1, 2).$$

3. Stability of periodic motion and Poincaré map

We choose a Poincaré section $\sigma \subset \mathbb{R}^4 \times S$, which is given by

$$\sigma = \big\{(x_1, \dot{x}_1, x_2, \dot{x}_2, \theta) \in R^4 \times S, \dot{x}_1 = \dot{x}_{1+}, \dot{x}_2 = \dot{x}_{2+}, x_2 - x_1 = \delta\big\}, \quad (9)$$

to establish a Poincaré map

$$X' = \tilde{f}(v, X), \quad (10)$$

where $\theta = \omega t$, $v \in R^1$ is a real parameter, $X = X^* + \Delta X$, $X' = X^* + \Delta X'$, $\Delta X = (\Delta \dot{x}_{1+}, \Delta x_{10}, \Delta \dot{x}_{2+}, \Delta \tau)^T$, $\Delta X' = (\Delta \dot{x}'_{1+}, \Delta x'_{10}, \Delta \dot{x}'_{2+}, \Delta \tau')^T$, ΔX and $\Delta X'$ are the disturbed vectors of $X^* \cdot X^* = (\dot{x}_{1+}, x_{10}, \dot{x}_{2+}, \tau)^T$ is a fixed point of periodic impacts in Poincaré section, which corresponds to one impact during one forcing cycle.

Under suitable system parameter conditions, the system given in Figure 1 can exhibit 1-1-1 symmetrical periodic motion. We can characterize periodic motions of the vibro-impact system by the symbol n-p-q, where p and q are the number of impacts occurring at the stops A and B, respectively, and n is the number of the forcing cycles. The periodic behavior means that if the dimensionless time t is set to zero directly after an impact, it becomes $2\pi/\omega$ just before the next impact between the masses M_1 and M_2 at point location A, where the mass M_2 come-and-go motion time t_1 and t_2 in the groove are equal to π/ω. After the origin of the θ-coordinate is displaced to an impact point, the determination is based on the fact that they satisfy the following set of periodicity and matching conditions

$$x_i(0) = -x_i(\pi/\omega) = x_{i0},$$
$$\dot{x}_i(2\pi/\omega) = -\dot{x}_{i-}(\pi/\omega) = \dot{x}_{i-},$$
$$\dot{x}(0) = -\dot{x}_{i+}(\pi/\omega) = \dot{x}_{i+},$$
$$x_2(0) - \dot{x}_1(0)i = \delta,$$
$$x_2(\pi/\omega) - x_1(\pi/\omega) = \delta,$$
$$x_2(2\pi/\omega) - x_1(2\pi/\omega) = \delta. \quad (11)$$

Substituting the periodicity boundary condition Equation (11) into the general solution (8), we can solve for the integration constants a_{ki}, b_{ki} $(k, i = 1, 2)$ and the phase angle τ_0. We obtain

$$a_{k1} = (-1)^k m(\delta + A \sin \tau_0) \qquad a_{k2} = (-1)^{k+1}(1+m)(\delta + A \sin \tau_0), \quad (12)$$

$$b_{k1} = (-1)^{k+1} \pi mn(\delta + A \sin \tau_0) \quad b_{k2} = (-1)^k d(1+m)(\delta + A \sin \tau_0), \quad (13)$$

$$\tau_0 = \arccos \frac{hq \pm h\sqrt{h^2 - h^2q^2 + 1}}{1 + h^2}, \tag{14}$$

where

$$m = \frac{s\mu de_1}{(e_1^2 + 2ce_1 + 1)\eta\omega_n - (se_1 d + ce_1 + 1)\mu}, \qquad d = \frac{2\omega}{\pi},$$

$$n = \frac{1 + ce_1}{se_1}, \qquad e_1 = e^{-\pi\zeta/\gamma}, \qquad \gamma = \omega/\omega_n, \qquad s = \sin \eta\pi/\gamma,$$

$$c = \cos \eta\pi/\gamma, \qquad q = \delta/A,$$

$$h = \frac{(\mu - R)(\omega(1 + 2m) + \pi m(1 + 2ce_1)) + \pi mn(1 + \mu)}{\pi\omega(1 + R)}$$
$$+ \frac{2me_1(s\eta - c\zeta) - de_1 mn(s\zeta + c\eta)}{d\gamma}. \tag{15}$$

In Equation (14), "$\pm$" denotes that there may be two different 1-1-1 symmetrical periodic motions under uniform system parameters. Because $|\cos \tau_0| \leq 1$, it should be noted that the existence of periodic impacts meets the condition

$$h^2 - h^2q^2 + 1 \geq 0, \qquad \left| \frac{hq \pm h\sqrt{h^2 - h^2q^2 + 1}}{1 + h^2} \right| \leq 1. \tag{16}$$

For expressing the actual motion of the system, the periodic solution must satisfy simultaneously the conditions of existence and stability. We consider the perturbed motion of 1-1-1 periodic motion to establish its Poincaré map. For simplicity of notations, the origin of the θ-coordinate is displaced to an impact point; the solutions of the perturbed motion are written in the form

$$\begin{cases} \tilde{x}_1(t) = e^{-\zeta\omega_n t}\left(\tilde{a}_{11}\cos(\eta\omega_n t) + \tilde{b}_{11}\sin(\eta\omega_n t)\right) \\ \qquad\qquad\qquad + A\sin(\omega t + \tau_0 + \Delta\tau), \quad 0 \leq t \leq \tilde{t}_1, \quad (17a) \\ \tilde{x}_2(t) = \tilde{a}_{12} + \tilde{b}_{12}t \end{cases}$$

$$\begin{cases} \tilde{x}_1(t) = e^{-\zeta\omega_n(t - t_1)}\left(\tilde{a}_{21}\cos(\eta\omega_n(t - \tilde{t}_1)) + \tilde{b}_{21}\sin(\eta\omega_n(t - \tilde{t}_1))\right) \\ \qquad\qquad\qquad + A\sin(\omega t + \tau_0 + \Delta\tau), \quad \tilde{t}_1 \leq t \leq \tilde{t}_p, \quad (17b) \\ \tilde{x}_2(t) = \tilde{a}_{22} + \tilde{b}_{22}(t - \tilde{t}_1) \end{cases}$$

For the disturbed motion, the dimensionless time is set to zero directly after an impact at point A between the masses M_1 and M_2. It becomes $(2\pi + \Delta\theta)/\omega$ just before the next impact at the same point, and the boundary conditions at two successive impact points are given by

$$\tilde{x}_i(0) = x_{i0} + \Delta x_{i0}, \qquad \dot{\tilde{x}}_i(0) = \dot{x}_{i+} + \Delta\dot{x}_{i+},$$
$$\tilde{x}_i(\tilde{t}_1) = -x_{i0} + \Delta x'_{i0}, \qquad \dot{\tilde{x}}_{i-}(\tilde{t}_1) = -\dot{x}_{i-} + \Delta\dot{x}'_{i-},$$
$$\dot{\tilde{x}}_{i+}(\tilde{t}_1) = -\dot{x}_{i+} + \Delta\dot{x}'_{i+}, \qquad \tilde{x}_i(\tilde{t}_p) = x_{i0} + \Delta x''_{i0},$$
$$\dot{\tilde{x}}_{i-}(\tilde{t}_p) = \dot{x}_{i-} + \Delta\dot{x}''_{i-}, \qquad \dot{\tilde{x}}_{i+}(\tilde{t}_p) = \dot{x}_{i+} + \Delta\dot{x}''_{i+},$$

$$(\dot{x}_{2+} + \Delta\dot{x}_{2+})\tilde{t}_1 + 2x_{10} + \Delta x_{10} - \Delta x'_{10} = -2\delta,$$
$$(-\dot{x}_{2+} + \Delta\dot{x}'_{2+})(\tilde{t}_p - \tilde{t}_1) - 2x_{10} + \Delta x'_{10} - \Delta x''_{10} = 2\delta,$$
$$\Delta\theta = \omega(\Delta t_1 - \Delta t_2), \tag{18}$$

where $\tilde{t}_1 = \pi/\omega + \Delta t_1$, $\tilde{t}_2 = \pi/\omega + \Delta t_2$.

Substituting the boundary condition Equation (12) into Equations (17a) and (17b), we obtain

$$\tilde{a}_{11} = x_{10} + \Delta x_{10} - A\sin(\tau_0 + \Delta\tau),$$
$$\tilde{a}_{21} = -x_{10} + \Delta x'_{10} - A\sin(\omega\tilde{t}_1 + \tau_0 + \Delta\tau),$$
$$\tilde{a}_{12} = x_{20} + \Delta x_{20}, \qquad \tilde{a}_{22} = -x_{20} + \Delta x'_{20}, \qquad \tilde{b}_{12} = \dot{x}_{2+} + \Delta x_{2+},$$
$$\tilde{b}_{22} = -\dot{x}_{2+} + \Delta\dot{x}'_{2+},$$
$$\tilde{b}_{11} = \dot{x}_{1+} + \Delta\dot{x}_{1+} - A\omega\cos(\tau_0 + \Delta\tau) + \zeta\omega_n\sin(\tau_0 + \Delta\tau)$$
$$+ \zeta\omega_n(x_{10} + \Delta x_{10})/\eta\omega_n, \tag{19}$$
$$\tilde{b}_{21} = \left[\begin{array}{c} -\dot{x}_{1+} + \Delta\dot{x}'_{1+} - A\omega\cos(\omega\tilde{t}_1 + \tau_0 + \Delta\tau) \\ -\zeta\omega_n A\sin(\omega\tilde{t}_1 + \tau_0 + \Delta\tau) - \zeta\omega_n(x_{10} - \Delta x'_{10}) \end{array}\right]\bigg/\eta\omega_n.$$

If we substitute the boundary condition Equation (18) into the perturbed solution (Equations (17a) and (17b)) for $t = t_e$, we obtain

$$\Delta x'_{10} = \tilde{x}_1(t_e) - x_{10},$$
$$\Delta\dot{x}'_{1+} = \frac{\mu(1+R)}{1+\mu}\dot{\tilde{x}}_1(t_e) + \frac{1-\mu R}{1+\mu}\dot{\tilde{x}}_{2+}(t_e) - \dot{x}_{1+},$$
$$\Delta\dot{x}'_{2+} = \frac{\mu - R}{1+\mu}\dot{\tilde{x}}_1(t_e) + \frac{1+R}{1+\mu}\dot{\tilde{x}}_2(t_e) - \dot{x}_{2+},$$
$$\Delta\tau' = \Delta\tau + \Delta\theta. \tag{20}$$

Define a function $w(\Delta \dot{x}_{1+}, \Delta x_{10}, \Delta \dot{x}_{2+}, \Delta \tau, \Delta t_1)$ as

$$w(\Delta \dot{x}_{1+}, \Delta x_{10}, \Delta \dot{x}_{2+}, \Delta \tau, \Delta t_1) \stackrel{\text{def}}{=} \tilde{x}_2(\tilde{t}_1) - \tilde{x}_1(\tilde{t}_1) + \delta = 0. \qquad (21)$$

Assuming $\partial w / \partial \Delta t_1|_{(0,0,0,0)} \neq 0$, according to the implicit function theorem, Equation (21) can be solved as

$$\Delta t_1 = \Delta t_1(\Delta \dot{x}_{1+}, \Delta x_{10}, \Delta \dot{x}_{2+}, \Delta \tau).$$

Setting

$$\Delta X = (\Delta \dot{x}_{1+}, \Delta x_{10}, \Delta \dot{x}_{2+}, \Delta \tau)^T = (y_1, y_2, y_3, y_4)^T,$$

we deduce the partial differentials of Δt_1 with respect to $\Delta \dot{x}_{1+}$, Δx_{10}, $\Delta \dot{x}_{2+}$ and $\Delta \tau$, as follows:

$$\frac{\partial \Delta t_1}{\partial y_j} = -\frac{\partial w}{\partial y_j} \bigg/ \frac{\partial w}{\partial \Delta t_1}, \quad j = 1, 2, 3, 4. \qquad (22)$$

In the same way, we define a function

$$h(\Delta \dot{x}_{1+}, \Delta x_{10}, \Delta \dot{x}_{2+}, \Delta \tau, \Delta t_1, \Delta t_2) \stackrel{\text{def}}{=} \tilde{x}_2(t_p) - \tilde{x}_1(t_p) - \delta = 0 \qquad (23)$$

By supposing $\partial h / \partial \Delta t_2|_{(0,0,0,0)} \neq 0$, based on the implicit function theorem, we have

$$\frac{\partial \Delta t_2}{\partial y_j} = -\left(\frac{\partial h}{\partial y_j} + \frac{\partial h}{\partial \Delta t_1} \frac{\partial \Delta t_1}{\partial y_j}\right) \bigg/ \frac{\partial h}{\partial \Delta t_2}. \qquad (24)$$

According to this analysis, we then obtain the Poincaré map, which is given by

$$\Delta \dot{x}'_{1+} = \tilde{f}_1(\Delta \dot{x}_{1+}, \Delta x_{10}, \Delta \dot{x}_{2+}, \Delta \tau, \Delta \theta) - \dot{x}_{1+} \stackrel{\text{def}}{=} f_1(\Delta \dot{x}_{1+}, \Delta x_{10}, \Delta \dot{x}_{2+}, \Delta \tau),$$

$$\Delta \dot{x}'_{10} = \tilde{f}_2(\Delta \dot{x}_{1+}, \Delta x_{10}, \Delta \dot{x}_{2+}, \Delta \tau, \Delta \theta) - x_{10} \stackrel{\text{def}}{=} f_2(\Delta \dot{x}_{1+}, \Delta x_{10}, \Delta \dot{x}_{2+}, \Delta \tau),$$

$$\Delta \dot{x}'_{2+} = \tilde{f}_3(\Delta \dot{x}_{1+}, \Delta x_{10}, \Delta \dot{x}_{2+}, \Delta \tau, \Delta \theta) - \dot{x}_{2+} \stackrel{\text{def}}{=} f_3(\Delta \dot{x}_{1+}, \Delta x_{10}, \Delta \dot{x}_{2+}, \Delta \tau),$$

$$\Delta \tau' = \Delta \tau + \Delta \theta(\Delta \dot{x}_{1+}, \Delta x_{10}, \Delta \dot{x}_{2+}, \Delta \tau) \stackrel{\text{def}}{=} f_4(\Delta \dot{x}_{1+}, \Delta x_{10}, \Delta \dot{x}_{2+}, \Delta \tau). \qquad (25)$$

Letting $v = \omega$, the Poincaré map Equation (25) can be expressed as

$$\Delta X' = \tilde{f}(v, X) - X^* \stackrel{\text{def}}{=} f(v, \Delta X), \qquad (26)$$

in which

$$f(v, \Delta X) = (f_1, f_2, f_3, f_4)^T,$$
$$\Delta X = (\Delta \dot{x}_{1+}, \Delta x_{10}, \Delta \dot{x}_{2+}, \Delta \tau)^T,$$
$$\Delta X' = (\Delta \dot{x}'_{1+}, \Delta x'_{10}, \Delta \dot{x}'_{2+}, \Delta \tau')^T.$$

We expand the function $f(v, \Delta X)$ as a Taylor series in the variables ΔX and v, so that it becomes

$$f(v, \Delta X) = \sum_{p+q \geq 1} F_{pq} v^p \Delta X^q, \tag{27}$$

$$F_{pq} = \frac{1}{p!q!} \frac{\partial^{p+q} f(v, \Delta X)}{\partial v^p \partial X^q} \bigg|_{(v_c, 0)}, \qquad F_{p0} \equiv 0, \ p \geq 1, \tag{28}$$

$$f(v, \Delta X) = F_{01} \Delta X + v F_{11} \Delta X + v^2 F_{21} \Delta X + F_{02}[\Delta X^2] + F_{03}[\Delta X^3] + \cdots, \tag{29}$$

where F_{02} and F_{03} denote the second-order and the third-order terms respectively.

Linearizing the Poincaré map at the fixed point $X^* = (\dot{x}_{1+}, x_{10}, \dot{x}_{2+}, \tau)^T$ results in the matrix

$$Df(v, 0) = \begin{bmatrix} \dfrac{\partial f_1}{\partial \Delta \dot{x}_{1+}} & \dfrac{\partial f_1}{\partial \Delta x_{10}} & \dfrac{\partial f_1}{\partial \Delta \dot{x}_{2+}} & \dfrac{\partial f_1}{\partial \Delta \tau} \\[2mm] \dfrac{\partial f_2}{\partial \Delta \dot{x}_{1+}} & \dfrac{\partial f_2}{\partial \Delta x_{10}} & \dfrac{\partial f_3}{\partial \Delta \dot{x}_{2+}} & \dfrac{\partial f_4}{\partial \Delta \tau} \\[2mm] \dfrac{\partial f_3}{\partial \Delta \dot{x}_{1+}} & \dfrac{\partial f_3}{\partial \Delta x_{10}} & \dfrac{\partial f_3}{\partial \Delta \dot{x}_{2+}} & \dfrac{\partial f_3}{\partial \Delta \tau} \\[2mm] \dfrac{\partial f_4}{\partial \Delta \dot{x}_{1+}} & \dfrac{\partial f_4}{\partial \Delta x_{10}} & \dfrac{\partial f_4}{\partial \Delta \dot{x}_{2+}} & \dfrac{\partial f_4}{\partial \Delta \tau} \end{bmatrix}_{(v,0,0,0,0)}. \tag{30}$$

According to Equations (21) and (23), it is easy to calculate the derivatives in the matrix Equation (30):

$$\frac{\partial f_j}{\partial y_i} = \frac{\partial \tilde{f}_j}{\partial y_i} + \frac{\partial \tilde{f}_j}{\partial \Delta t_1} \frac{\partial \Delta t_1}{\partial y_i} + \frac{\partial \tilde{f}_j}{\partial \Delta t_2} \frac{\partial \Delta t_2}{\partial y_i} \quad i, j = 1, 2, 3, 4. \tag{31}$$

It is possible to determine the stability of periodic impacts by the eigenvalues of $Df(v, 0)$. If all eigenvalues of $Df(v, 0)$ are inside the unit circle, then the periodic solution is stable; otherwise, it is unstable. If some of the eigenvalues of the matrix $Df(v, 0)$ lie on the unit circle in the complex plane when $v = v_c$ (v_c is the bifurcation value), then it is possible for bifurcations to take place. In general, bifurcation occurs in various ways according to the number of the eigenvalues on the unit circle and their position on the circle. When $v = v_c$, $Df(v, 0)$ has a pair of simple complex conjugate eigenvalues $\lambda_1(v_c)$ and $\bar{\lambda}_1(v_c)$ on the unit circle; all other eigenvalues of $Df(v, 0)$ are inside the unit circle. Under this circumstance, 1-1-1 symmetrical periodic motion may lead to Hopf bifurcation. In general, bifurcations from periodic motions to quasiperiodic ones occur under nonresonance or resonance conditions in nonlinear dynamical systems.

4. Hopf bifurcation in nonresonance and weak resonance cases

Consider the Poincaré map

$$\Delta X' = f(v, \Delta X). \tag{32}$$

Let $\Delta X^*(v)$ be a fixed point for the system Equation (32) for v in some neighborhood of a critical value $v = v_c$ at which $Df(v, 0)$ satisfies the following assumptions:

(A1) $Df(v, 0)$ has a pair of complex conjugate eigenvalues

$$\lambda_1 = \lambda_1(v_c) \quad \text{and} \quad \lambda_2 = \bar{\lambda}_2(v_c),$$

and satisfies $|\lambda_1(v_c)| = 1$. The other eigenvalues $\lambda_i(v_c)$ satisfy

$$|\lambda_i(v_c)| < 1, \quad i = 3, 4;$$

(A2) $(d|\lambda_1(v_c)|/dv)|_{v=v_c} > 0$;

(A3) $\lambda_1^m(v_c) \neq 1, \quad m = 1, 2, 3, 4.$

Let k_i denote the eigenvector of $Df(v, 0)$ corresponding to $\lambda_i(v)$, for $i = 1, 2, 3, 4$. If k_3 and k_4 are a complex conjugate pair of nonreal eigenvectors, define

$$H = (\operatorname{Re} k_1, -\operatorname{Im} k_1, \operatorname{Re} k_3, -\operatorname{Im} k_3);$$

otherwise,

$$H = (\operatorname{Re} k_1, -\operatorname{Im} k_1, k_3, k_4).$$

In some neighborhood of v_c, the map Equation (32), under the change of variable

$$\Delta X = HY, \qquad \mu = v - v_c, \tag{33}$$

becomes

$$Y' = F(\mu, Y), \tag{34}$$

where $Y = (y_1, y_2, y_3, y_4)^T$.

For the map Equation (34), there exists a local center manifold $W(z, \bar{z}; \mu)$, on which the local behavior of the map can be reduced to a two-dimensional map $\Phi_\mu(z)$. This map can be presented

$$\Phi_\mu(z) = \lambda(\mu)z + \sum_{i+j=2}^{3} g_{ij}(\mu)\frac{z^i \bar{z}^j}{i!j!} + O(|z|^4), \tag{35}$$

where $\lambda(\mu) = \lambda_1(v_c + \mu), \lambda_0 = \lambda(0), z = y_1 + iy_2, \bar{z} = y_1 - iy_2$.

By center manifold theory, all bifurcation phenomena of $F(\mu, Y)$ take place on a two-dimensional manifold. Local dynamic behavior of the map Equation (34) is equivalent to that of the two-dimensional map Equation (35) for μ in some

neighborhood of a critical value $\mu = 0$, so using the map Equation (35) and applying the following lemma, we can discuss the existence of Hopf bifurcation for map Equation (32) as v passes through v_c.

Lemma 1 [Lanford 1973; Wan 1978]. *Let $\Phi_\mu(z)$ be a one-parameter family of diffeomorphisms on R^2 near $z = 0$, satisfying the following conditions*:

(B1) $\Phi_\mu(0) = 0$ *for all* μ;

(B2) $D\Phi_\mu(0)$ *has two conjugate eigenvalues* $\lambda(\mu)$ *and* $\bar{\lambda}(\mu)$, *with* $|\lambda(0)| = 1$;

(B3) $(d|\lambda(\mu)|/d\mu)|_{\mu=0} > 0$;

(B4) $\lambda^m(0) \neq 1$, $m = 1, 2, 3, 4$.

Subject to assumptions (B1)–(B4), we can make a smooth μ-dependent change of coordinates to put $\Phi_\mu(z)$ into the normal form

$$\Phi_\mu(z) = N\Phi_\mu(z) + O(|Y|^5). \tag{36}$$

In polar co-ordinates,

$$N\Phi_\mu(r, \varphi) = (|\lambda(\mu)|r - f_1(\mu)r^3, \varphi + \theta(\mu) + f_3(\mu)r^3). \tag{37}$$

If $f_1(0) > 0$ (or $f_1(0) < 0$), $\Phi_\mu(z)$ has an attracting (repelling) invariant circle for $\mu > 0$ (or $\mu < 0$). Suppose that the complex form of $\Phi_0(z)$ is

$$\Phi_0(z) = \lambda_0 z + \sum_{i+j=2}^{3} g_{ij}(0)\frac{z^i \bar{z}^j}{i!j!} + O(|z|^4). \tag{38}$$

Then, there is

$$f_1(0) = \mathrm{Re}\left(\frac{(1-2\lambda_0)\bar{\lambda}_0}{2(1-\lambda_0)}g_{20}g_{11}\right) + \tfrac{1}{2}|g_{11}|^2 + \tfrac{1}{4}|g_{02}|^2 - \mathrm{Re}\frac{\bar{\lambda}_0 g_{21}}{2}, \tag{39}$$

where $\lambda_0 = \lambda(0)$, $g_{ij}(\mu)(i + j = 2, 3)$ (see the Appendix).

If the Poincaré map Equation (32) satisfies the conditions (A1)–(A3), then it is easy to show that the map Equation (35) satisfies the conditions (B1)–(B4). If a set of system parameters can be chosen for the vibro-impact system under which the Poincaré map Equation (32) satisfies the conditions (A1)–(A3), then by computing $f_1(0)$, we can conclude the existence of an invariant circle for the map Equation (35) and its stability in terms of the sign of $f_1(0)$. Because on the centre manifold $W(z, \bar{z}; \mu)$ the local behavior of the Poincaré map can be reduced to the two-dimensional behavior of Equation (35), it is certain that if the map Equation (35) has an attracting (repelling) invariant circle for $\mu > 0$ (or $\mu < 0$), a supercritical (subcritical) Hopf bifurcation will take place for the vibro-impact system shown in Figure 1 at $v = v_c$.

5. Numerical simulation of Hopf bifurcation and chaos
for the vibro-impact system

In this section, the analyses developed in the previous section are verified by the numerical results for the impact system in Figure 1. The results of numerical simulations are used to understand the rich dynamical behavior that our model vibro-impact system exhibits. The use of these properties of chaotic systems can offer special advantages in controlling chaotic systems. For instance, small perturbations can lead to large effects, and flexible switching is possible between many different periodic orbits without changing the global configuration of the system. Many feedback control strategies based on this general idea use small perturbations in a control parameter to manipulate the behavior of chaotic systems. These benefits cannot be achieved in nonchaotic systems in which large effects in behavior typically require large changes in the control parameter.

Dynamic behavior of the vibro-impact system is shown in the projected Poincaré sections. The Poincaré section is taken in the form $\Sigma = \{(x_1, \dot{x}_1, x_2, \dot{x}_2, \theta) \in \mathbb{R}^4 \times S, \ \dot{x}_1 = \dot{x}_{1+}, \dot{x}_2 = \dot{x}_{2+}, x_2 - x_1 = \delta\}$, which is four-dimensional. The section is then projected onto the $(x_1, \dot{x}_1)$ plane, which is called the projected Poincaré section. Dynamic behavior of the vibro-impact system near the resonance point can be demonstrated from those projected Poincaré sections.

5.1. *Hopf bifurcation to chaotic motion in nonresonance and weak resonance cases.* Here we briefly analyze Hopf bifurcation to quasiperiodic torus and the break of quasiperiodic torus to chaotic behavior of the model in nonresonance and weak resonance cases. A set of parameters $\mu = 0.1$, $\omega_n = 0.5$, $\zeta = 0.045$, $R = 0.8$, $\delta = 5.0$ are considered. ω is taken as the control parameter, i.e. let $v = \omega$. The eigenvalues of $Df(\omega, 0)$ are computed for $\omega \in [0.5, 1.5]$. A pair of complex conjugate eigenvalues intersects the unit circle and the other eigenvalues are still inside the unit circle as ω passes through $\omega_c = 0.99176$. ω_c is a Hopf bifurcation value, at which $\lambda_{1,2}(\omega_c) = 0.2730961 \pm 0.9619876i$, $\lambda_{3,4}(\omega_c) = 0.3047823 \pm 0.2773642i$, and $|\lambda_{1,2}(\omega_c)| = 1$, $\lambda_{1,2}^m(\omega_c) \neq 1$, $m = 1, 2, 3, 4$. It is apparent that the model with this set of system parameters satisfies a nonresonance or weak resonance condition at the critical point. When $\omega \in (1.015, 1.45)$, the system can exhibit stable symmetry 1-1-1 periodic motion. When $\omega = 1.2$, the impact system exhibits stable periodic 1-1-1 impact motion; see Figure 2(a). Figure 2(b,c) show that the impact system exhibits unstable periodic motion, but in this case invariant circle is not generated. Taking a theoretical fixed point of the system corresponding to $\omega = 1.00857$ as an initial map point, the attracting invariant circle is shown in Figure 2(d,e,f). When the value of ω moves further away from the Hopf bifurcation value, the invariant circle, in projected Poincaré sections, expands markedly; see Figure 2(g,h). With further reduction in the control parameter ω, the attracting

invariant circle is broken; see Figure 2(i,j) and gets locked into a periodic attractor of higher period (higher than one cycle of the forcing). Subsequently, the system becomes unstable and chaotic. The periodic attractors of higher period, via phase locking (that is, the vibration frequencies of the oscillation system are locked in various rational multiples of the forcing frequency), is shown for $\omega = 0.9817$ in Figure 2(k). The chaotic motion of the system, represented by an infinite point set on a nonclosed curve in the projected Poincaré sections, is shown in Figure 2(l).

5.2. *Quasiperiodic and chaotic behavior of the impact system in strong resonance cases.* There exists another route by which Hopf bifurcation leads to chaos in the impact system shown in Figure 1. This route is characterized by a phenomenon in which the system comes into the chaotic motion without a quasi-attracting circle but with a single torus doubling. In order to study such a case, the dynamics of the impact system with system parameters $\mu = 0.04$, $\omega_n = 0.6$, $\zeta = 0.025$, $R = 0.7$, $\delta = 8.0$ is analyzed for $\omega \in [0.93, 0.98]$ by numerical simulation. The dynamic behavior of the system is shown in the projected Poincaré sections; see Figure 4. The eigenvalues of $Df(\omega, 0)$ are computed, and the variation of the eigenvalues is shown in Figure 3(a). When ω decreases through $\omega_c = 0.976327$, a pair of complex conjugate eigenvalues $\lambda_1(\omega_c)$ and $\lambda_2(\omega_c)$ cross the unit circle and all other eigenvalues $\lambda_3(\omega_c)$ and $\lambda_4(\omega_c)$ will still stay inside the unit circle. Then ω_c is a Hopf bifurcation value, and $\lambda_{1,2}(\omega_c) = -0.0000048 \pm 1.000037i$, $\lambda_{1,2}^4(\omega_c) = 1$, $\lambda_{3,4}(\omega_c) = 0.2026327 \pm 0.6311391i$. It is obvious that the model with this set of parameters satisfies the strong resonance condition at the critical point.

Numerical simulation shows that the impact system exhibits an attracting invariant circle in projected Poincaré section for 3561 impacts; see Figure 4(a). It is to be noted that the attracting invariant circle is smooth in nature near the bifurcation point. However, with a further decrease in the control parameter ω, the attracting circle expands and the smoothness of circle is destroyed; see Figure 4(b,c). When $\omega = 0.9454$, the system yields 4-4-4 quasiperiodic impact motion in Figure 4(d). Subsequently, at $\omega = 0.945$, phase locking occurs, and the quasiperiodic motion is locked into the periodic attractors of higher period (than one cycle of the force); see Figure 4(e). With an increase in the control parameter ω, the invariant circle becomes unstable and the system settles into chaotic motion. After the invariant circle loses its stability, no tori doubling or phase locking occurs, and the circle is quasi-attracting (map points inside the circle may be attracted to the circle and map points on or outside the circle may stray from the circle). The chaotic motions of system, represented by "belt-like" attractors in projected Poincaré sections, are shown in Figure 4(f,g), and the width of the belt attractors increases with a decrease in the control parameter. When $\omega = 0.9386$, the chaotic motion of the system grows
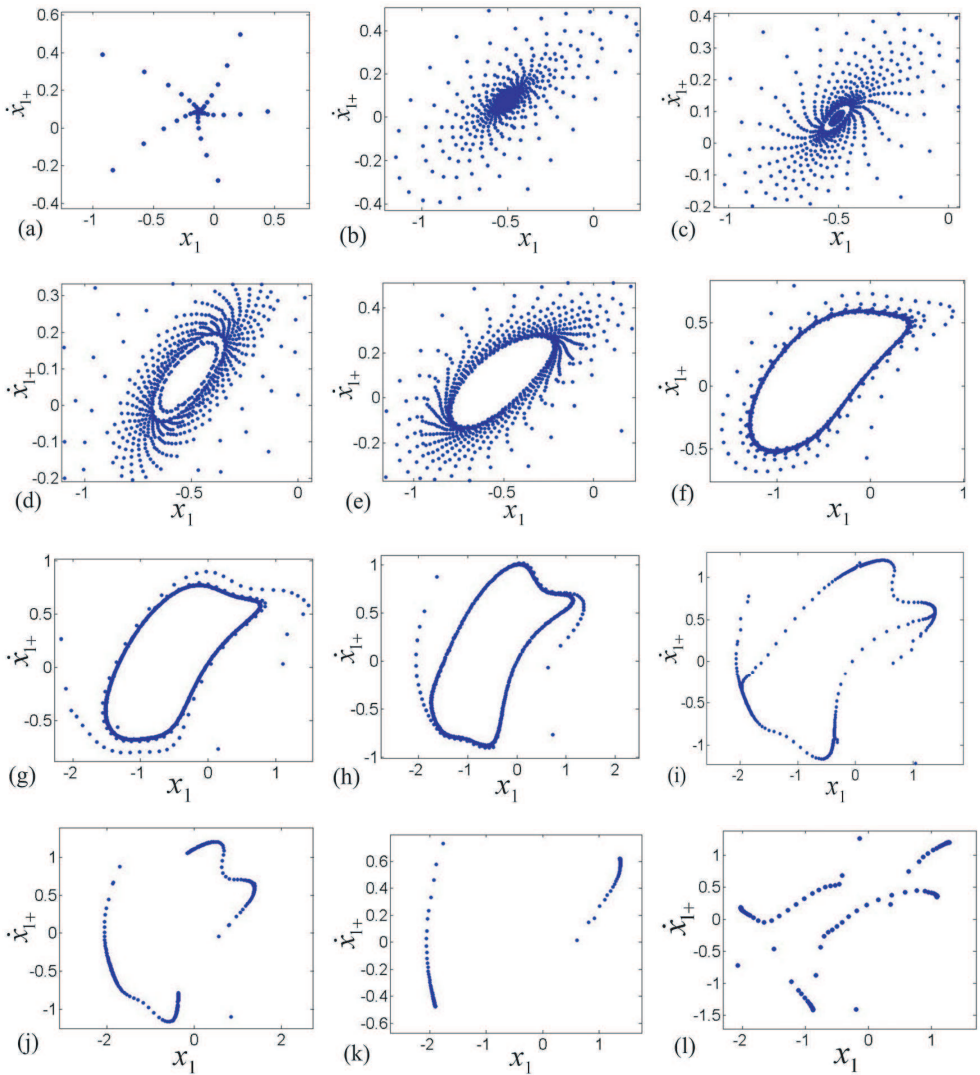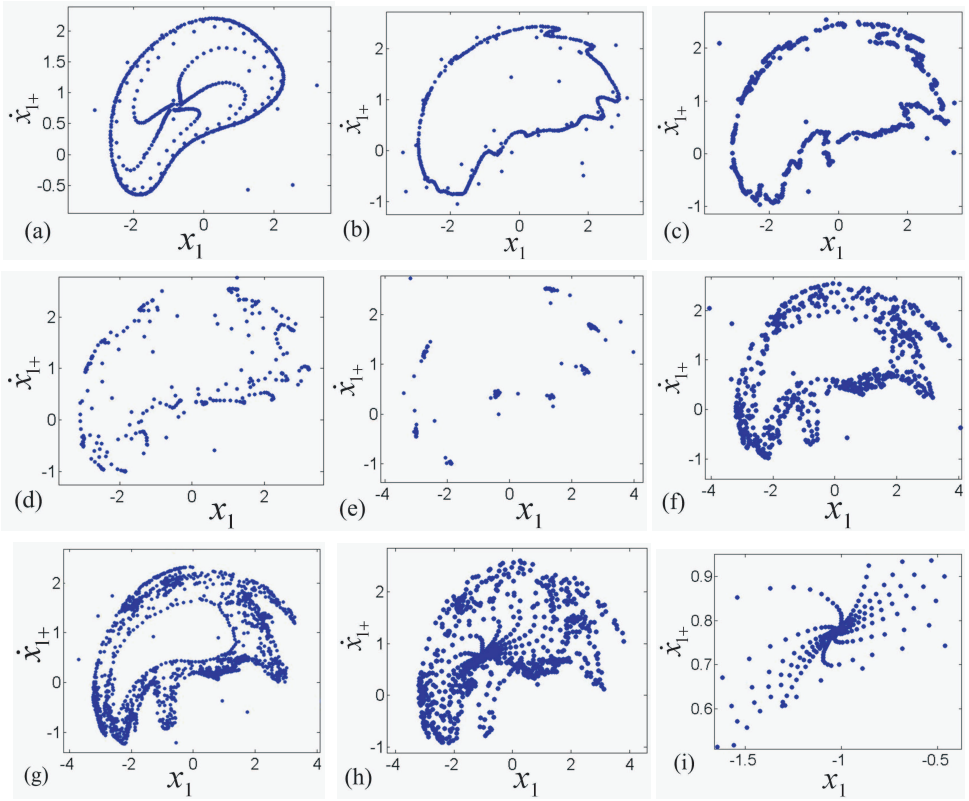
Figure 2. Projected Poincaré mapping section $(x_1, \dot{x}_{1+})$: (a) $\omega = 1.2$, stable 1-1-1 fixed point; (b) $\omega = 1.01962$, stable fixed point; (c) $\omega = 1.0011368$, unstable 1-1-1 fixed point; (d) $\omega = 1.00857$, unstable invariant torus; (e) $\omega = 1.00734$, an attracting invariant circle; (f) $\omega = 1.00314$, quasiperiodic 4-4-4 impacts; (g) $\omega = 0.9958$, quasiperiodic impacts; (h) $\omega = 0.99$, quasiperiodic 10-10-10 impact motions; (i) $\omega = 0.98193$, initial break quasiperiodic invariant circle; (j) $\omega = 0.98185$, break of quasiperiodic circle; (k) $\omega = 0.9817$, phase locked; (l) $\omega = 0.970453$, chaotic motions represented by a discontinuous infinite point set.
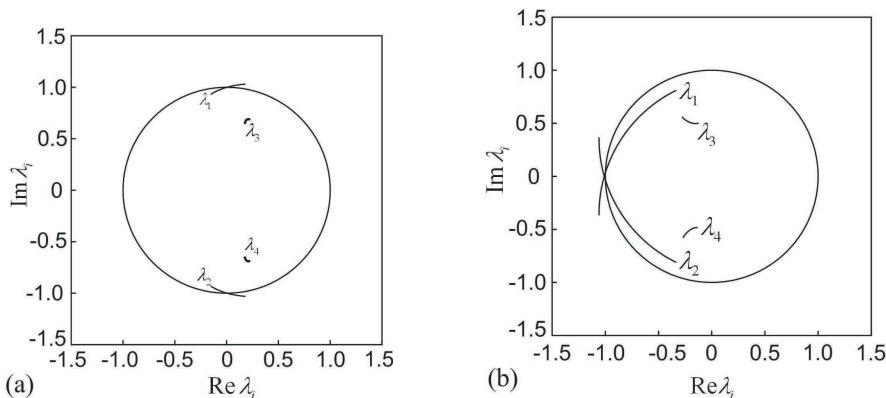
Figure 3. The conjugate pair of eigenvalues intersecting the unit circle for the strong resonance case.

out of the trivial attractor; see Figure 4(h). As the control parameter decreases further, as $\omega = 0.937$, the chaotic motion of the system disappears, and again the system displays a stable 1-1-1 fixed point in projected Poincaré section; see Figure 4(i).

We next study the dynamic behavior of the impact system using numerical simulation, to determine the dynamics near the resonance point $(\lambda_{1,2}^2(\omega_c) = 1)$. The system parameters $\mu = 0.05092$, $\zeta = 0.015$, $\omega_n = 1$, $R = 0.7$, $\delta = 6.0$ are chosen. The eigenvalues of $Df(\omega, 0)$ are computed for $\omega \in [0.415, 0.448]$ and two pairs of complex conjugate eigenvalues $\lambda_{1,2}(\omega)$ and $\lambda_{3,4}(\omega)$ are obtained. The conjugate pair of eigenvalues intersecting the unit circle is shown in Figure 3(b). When the control parameter ω decreases to $\omega_c = 0.43186$, the complex conjugate eigenvalues $\lambda_{1,2}(\omega_c)$ intersect the unit circle through the point $(-1, 0)$, and $\lambda_{3,4}(\omega_c)$ remains inside the unit circle. This ω_c is the Hopf bifurcation value, at which $\lambda_{1,2}(\omega_c) = -1.0000031 \pm 0.00000272i$, $\lambda_{1,2}^2(\omega_c) = 1$, $\lambda_{3,4}(\omega_c) = 0.2054234 \pm 0.4398154i$. Numerical simulation results in the projected Poincaré section are shown in Figure 5. After ω passes through the critical value ω_c, the symmetric 1-1-1 periodic impact motion destabilizes, which subsequently leads to a Hopf bifurcation into quasiperiodic motion, as seen in Figure 5(a). When $\omega_c = 0.418$, the smooth attracting invariant circle is destroyed; see Figure 5(b). With a further decrease in ω, after the invariant circle loses its stability, torus doubling and phase locking occur in succession; see Figure 5(c,d). As ω is decreased further, chaotic motion arises: Figure 5(e,f). The evolution from quasiperiodic motion to chaos is clearly shown in the projected Poincaré map in Figure 5. This kind of route from quasiperiodic
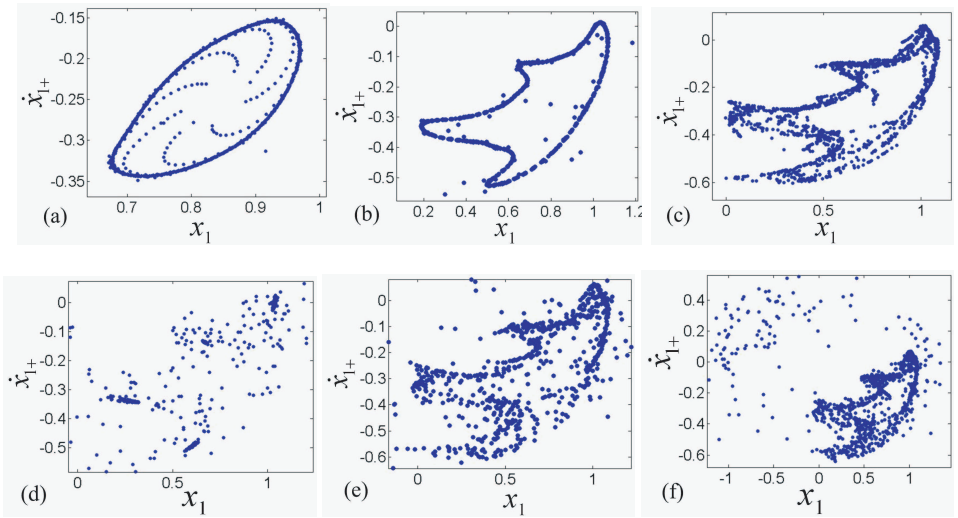
Figure 4. Projected Poincaré section $(x_1, \dot{x}_{1+})$: (a) $\omega = 0.975$, quasiperiodic motion represented by an attracting invariant circle; (b) $\omega = 0.96$, quasiperiodic impacts with wave of oscillation; (c) $\omega = 0.9526$, quasiperiodic impacts; (d) $\omega = 0.9454$, 4-4-4 periodic motion; (e) $\omega = 0.945$, phased locked; (f) $\omega = 0.94$, chaotic motion; (g) $\omega = 0.93998$, chaotic motion; (h) $\omega = 0.9386$, the chaotic motion comes out the trivial attractor; (i) $\omega = 0.937$, stable 1-1-1 fixed point.

impacts to chaos via quasi-attracting invariant circles is often observed in numerical simulations of the dynamics of impact systems.

6. Conclusion

We studied the dynamic behavior of a two-degree-of-freedom impact system via theoretical analysis and numerical simulation. The Poincaré map and fixed point of period 1-1-1 impact are determined analytically. The local dynamical behavior is discussed when the control parameters are changed near the critical point. The

Figure 5. Projected Poincaré section $(x_1, \dot{x}_{1+})$: (a) $\omega = 0.4268$, an attracting invariant circle; (b) $\omega = 0.418$, an attracting invariant circle; (c) $\omega = 0.41754$, the torus doubling; (d) $\omega = 0.41731$, phased locked; (e) $\omega = 0.41721$, chaotic motion; (f) $\omega = 0.41719$, chaotic motion.

dynamic behavior of the system in the strong resonance case is more complicated than in nonresonance and weak resonance cases. In the strong resonance case $\lambda^4_{1,2}(\omega_c) = 1$, the system can exhibit stable 4-4-4 periodic impact motion and quasiperiodic motion, and the route from quasiperiodic motion to chaos is observed by numerical simulation. In the strong resonance cases of $\lambda^2_{1,2}(\omega_c) = 1$, the system transitions from quasiperiodic impacts to chaos in a complicated way: quasiperiodic impacts → torus doubling → various kinds of phase locking → chaos. The method established in the paper can be extended to some other analogous systems. Machines and equipment whose behavior can be attempted to be modeled sy such systems include vibration hammers, gear transmission, shakers, wheel-rail interaction of high-speed coaches, and compacting machinery. However, it is necessary to make further theoretical studies of the routes of bifurcation to chaos.

Appendix

The relational coefficients are shown as follows:

$$g_{11} = \frac{1}{4}\left[\frac{\partial^2 F_1}{\partial y_1^2} + \frac{\partial^2 F_1}{\partial y_2^2} + i\left(\frac{\partial^2 F_2}{\partial y_1^2} + \frac{\partial^2 F_2}{\partial y_2^2}\right)\right] \tag{A1}$$

$$g_{02} = \frac{1}{4}\left[\frac{\partial^2 F_1}{\partial y_1^2} - \frac{\partial^2 F_1}{\partial y_2^2} - 2\frac{\partial^2 F_2}{\partial y_1 \partial y_2} + i\left(\frac{\partial^2 F_2}{\partial y_1^2} - \frac{\partial^2 F_2}{\partial y_2^2} + 2\frac{\partial^2 F_1}{\partial y_1 \partial y_2}\right)\right] \tag{A2}$$

$$g_{20} = \frac{1}{4}\left[\frac{\partial^2 F_1}{\partial y_1^2} - \frac{\partial^2 F_1}{\partial y_2^2} + 2\frac{\partial^2 F_2}{\partial y_1 \partial y_2} + i\left(\frac{\partial^2 F_2}{\partial y_1^2} - \frac{\partial^2 F_2}{\partial y_2^2} - 2\frac{\partial^2 F_1}{\partial y_1 \partial y_2}\right)\right] \tag{A3}$$

$$g_{21} = G_{21} + \sum_{i=1}^{n-2}\left(2G_{10}^s w_{11}^s + G_{01}^s w_{20}^s\right) \tag{A4}$$

$$G_{21} = \frac{1}{8}\left[\frac{\partial^3 F_1}{\partial y_1^3} + \frac{\partial^3 F_1}{\partial y_1 \partial y_2^2} + \frac{\partial^3 F_2}{\partial y_1^2 \partial y_2} + \frac{\partial^3 F_2}{\partial y_2^3}\right.$$
$$\left. + i\left(\frac{\partial^3 F_2}{\partial y_1^3} + \frac{\partial^3 F_2}{\partial y_1 \partial y_2^2} - \frac{\partial^3 F_1}{\partial y_1^2 \partial y_2} - \frac{\partial^3 F_1}{\partial y_2^3}\right)\right] \tag{A5}$$

$$G_{10}^{k-2} = \frac{1}{2}\left[\frac{\partial^2 F_1}{\partial y_1 \partial y_k} + \frac{\partial^3 F_2}{\partial y_2 \partial y_k} + i\left(\frac{\partial^2 F_2}{\partial y_1 \partial y_k} - \frac{\partial^3 F_1}{\partial y_2 \partial y_k}\right)\right](k = 3,\ldots,n) \tag{A6}$$

$$G_{01}^{k-2} = \frac{1}{2}\left[\frac{\partial^2 F_1}{\partial y_1 \partial y_k} - \frac{\partial^3 F_2}{\partial y_2 \partial y_k} + i\left(\frac{\partial^2 F_2}{\partial y_1 \partial y_k} - \frac{\partial^3 F_1}{\partial y_2 \partial y_k}\right)\right](k = 3,\ldots,n) \tag{A7}$$

$$W_{11} = (I - B)^{-1}H_{11}, \qquad W_{02} = \left(\bar{\lambda}^2 I - B\right)^{-1}H_{02}, \tag{A8}$$

where $W_{11} = (w_{11}^1, w_{11}^2, \ldots, w_{11}^{n-2})^T$, $W_{02} = (w_{02}^1, w_{02}^2, \ldots, w_{02}^{n-2})^T$.

$$H_{11}^{k-2} = \frac{1}{4}\left(\frac{\partial^2 F_k}{\partial y_1^2} + \frac{\partial^2 F_k}{\partial y_2^2}\right)(k = 3, 4) \tag{A9}$$

$$H_{02}^{k-2} = \frac{1}{4}\left(\frac{\partial^2 F_k}{\partial y_1^2} - \frac{\partial^2 F_k}{\partial y_2^2} - 2i\frac{\partial^2 F_k}{\partial y_1 \partial y_2}\right)(k = 3, 4) \tag{A10}$$

References

[Asfar and Akour 2005] K. R. Asfar and S. N. Akour, "Optimization analysis of impact viscous damper for controlling self-excited vibrations", *J. Vib. Control* **11**:1 (2005), 103–120.

[di Bernardo et al. 2001] M. di Bernardo, C. J. Budd, and A. R. Champneys, "Normal form maps for grazing bifurcations in n-dimensional piecewise-smooth dynamical systems", *Phys. D: Nonlinear Phenom.* **160**:3-4 (2001), 222–254. MR 2002k:37084

[di Bernardo et al. 2002] M. di Bernardo, P. Kowalczyk, and A. Nordmark, "Bifurcations of dynamical systems with sliding: derivation of normal-form mappings", *Phys. D: Nonlinear Phenom.* **170**:3-4 (2002), 175–205. MR 2003h:37075

[Blazejczyk-Okolewska 2001] B. Blazejczyk-Okolewska, "Analysis of an impact damper of vibrations", *Chaos Solitons Fract.* **12**:11 (2001), 1983–1988.

[Budd et al. 1995] C. Budd, F. Dux, and A. Cliffe, "The effect of frequency and clearance variations on single-degree-of-freedom impact oscillators", *J. Sound Vib.* **184**:3 (1995), 475–502.

[Chatterjee and Mallik 1995] S. Chatterjee and A. Mallik, A. K.and Ghosh, "On impact dampers for non-linear vibrating systems", *J. Sound Vib.* **187**:3 (1995), 403–420.

[Chatterjee and Mallik 1996] S. Chatterjee and A. K. Mallik, "Bifurcations and chaos in autonomous self-excited oscillators with impact damping", *J. Sound Vib.* **191**:4 (1996), 539–562. MR 97f:70049

[Cheng and Wang 2003] C. C. Cheng and J. Y. Wang, "Free vibration analysis of a resilient impact damper", *Int. J. Mech. Sci.* **45**:4 (2003), 589–604.

[Cone and Zadoks 1995] K. M. Cone and R. I. Zadoks, "A numerical study of an impact oscillator with the addition of dry friction", *J. Sound Vib.* **188**:5 (1995), 659–683.

[Dimentberg and Iourtchenko 2004] M. F. Dimentberg and D. V. Iourtchenko, "Random vibrations with impacts: a review", *Nonlinear Dyn.* **36**:2-4 (2004), 229–254. MR 2005c:74038

[Ding et al. 2004] W. C. Ding, J. H. Xie, and Q. S. Sun, "Interaction of Hopf and period doubling bifurcations of a vibro-impact system", *J. Sound Vib.* **275**:1-2 (2004), 27–45.

[Ema and Marui 1996] S. Ema and E. Marui, "Damping characteristics of an impact damper and its application", *Int. J. Mach. Tools Manuf.* **36**:3 (1996), 293–306.

[Holmes 1982] P. J. Holmes, "The dynamics of repeated impacts with a sinusoidally vibrating table", *J. Sound Vib.* **84**:2 (1982), 173–189. Erratum in **88**:2 (1983), 287. MR 676219 (85f:70044a) Zbl 0518.73015

[Lanford 1973] O. E. Lanford, "Bifurcation of periodic solutions into invariant tori: the work of Ruelle and Takens", pp. 159–192 in *Nonlinear problems in the physical sciences and biology* (Seattle, 1972), edited by I. Stakgold et al., Lecture Notes in Mathematics **322**, Springer, Berlin, 1973. MR 0371548 (51 #7766)

[Luo 2004a] A. C. J. Luo, "Period-doubling induced chaotic motion in the LR model of a horizontal impact oscillator", *Chaos Solitons Fract.* **19**:4 (2004), 823–839. MR MR2009668

[Luo 2004b] G. W. Luo, "Period-doubling bifurcations and routes to chaos of the vibratory systems contacting stops", *Phys. Lett. A* **323**:3-4 (2004), 210–217. MR 2050484

[Luo and Chen 2005] A. C. J. Luo and L. Chen, "Periodic motions and grazing in a harmonically forced, piecewise, linear oscillator with impacts", *Chaos Solitons Fract.* **24**:2 (2005), 567–578.

[Luo and Xie 2003] G. W. Luo and J. H. Xie, "Codimension two bifurcation of periodic vibro-impact and chaos of a dual component system", *Phys. Lett. A* **313**:4 (2003), 267–273. MR 1995513

[Luo et al. 2001] G. W. Luo, J. H. Xie, and H. L. Guo, "Periodic motions and global bifurcations of a two-degree-of-freedom system with plastic vibro-impact", *J. Sound Vib.* **240**:5 (2001), 837–858.

[Nordmark 1991] A. B. Nordmark, "Non-periodic motion caused by grazing incidence in an impact oscillator", *J. Sound Vib.* **145**:2 (1991), 279–297.

[Padmanabhan and Singh 1995] C. Padmanabhan and R. Singh, "Analysis of periodically excited non-linear systems by a parametric continuation technique", *J. Sound Vib.* **184**:1 (1995), 35–58.

[Shaw and Holmes 1983] S. W. Shaw and P. Holmes, "Periodically forced linear oscillator with impacts: chaos and long-period motions", *Phys. Rev. Lett.* **51**:8 (1983), 623–626. MR 84k:58171

[Thompson and Ghaffari 1982] J. M. T. Thompson and R. Ghaffari, "Chaos after period-doubling bifurcations in the resonance of an impact oscillator", *Phys. Lett. A* **91**:1 (1982), 5–8. MR 83k:70028

[Wan 1978] Y. H. Wan, "Computation of the stability condition for the Hopf bifurcation of diffeomorphisms on R^2", *SIAM J. Appl. Math.* **34**:1 (1978), 167–175. MR 57 #7684

[Wang et al. 2003] J. F. Wang, C. C. Lin, and B. L. Chen, "Vibration suppression for high-speed railway bridges using tuned mass dampers", *Int. J. Solids Struct.* **40**:2 (2003), 465–491.

[Wen 2001] G. L. Wen, "Codimension-2 Hopf bifurcation of a two-degree-of-freedom vibro-impact system", *J. Sound Vib.* **242**:3 (2001), 475–485.

[Whiston 1992] G. S. Whiston, "Singularities in vibro-impact dynamics", *J. Sound Vib.* **152**:3 (1992), 427–460. MR 92m:58080

[Xie and Ding 2005] J. Xie and W. Ding, "Hopf-Hopf bifurcation and invariant torus $\mathbf{T}^2$ of a vibro-impact system", *Int. J. Non-Linear Mech.* **40**:4 (2005), 531–543. MR 2005j:70052

JIANLIAN CHENG: cheng19690912@163.com
Department of Engineering Mechanics, MOE Key Laboratory for Strength and Vibration, School of Aerospace, Xi'an Jiaotong University, Xi'an 710049, China

HUI XU: xuhui@mail.xjtu.edu.cn
Department of Engineering Mechanics, MOE Key Laboratory for Strength and Vibration, School of Aerospace, Xi'an Jiaotong University, Xi'an 710049, China

VARIATIONAL EIGENSTRAIN ANALYSIS OF SYNCHROTRON DIFFRACTION MEASUREMENTS OF RESIDUAL ELASTIC STRAIN IN A BENT TITANIUM ALLOY BAR

ALEXANDER M. KORSUNSKY

Most procedures for experimental stress evaluation rely on the measurement of elastic strain followed by point-wise calculation of stress based on continuum elasticity assumptions despite the fact that the real purpose of the investigation is to characterise the state of stress everywhere in the object to the greatest possible detail. Using the example of residual elastic strain measurements in a bent titanium alloy bar taken by means of high energy synchrotron X-ray diffraction, an interpretation technique is here introduced based on the variational eigenstrain analysis. An analytical framework is presented for the solution of the direct problem of eigenstrain, that is, the calculation of residual elastic strain distribution within an inelastically bent beam containing a known distribution of eigenstrain. An inverse problem about closest matching between the model and experiment is then cast in a form that allows determination of the underlying eigenstrain distribution from a single noniterative solution of a linear system. Subsequently the complete stress state can be reconstructed everywhere within the object in the form of continuous functions. The value of the approach lies in the fact that subsequent deformation modelling can be carried out with the effects of residual stresses (and their evolution) naturally incorporated. The extension of this approach to more complex geometries within the framework of the finite element method is briefly discussed.

1. Introduction

Residual stresses play an important role in determining the deformation behaviour and fatigue durability of engineering components and assemblies. It is well known, for example, that compressive near surface residual stresses act to inhibit crack initiation and propagation, and thus affect the fatigue life of an object. On the other hand, residual stresses themselves are known to undergo modification during thermal and mechanical loading, through various mechanisms related to time-independent plasticity, creep, phase transformation, etc.

We can pursue two principal avenues in evaluating residual stress states: deformation process modelling and experimental measurement.

Numerical simulation of nonlinear deformation behaviour of solid bodies with complex shapes is usually accomplished with the help of the finite element method. This method allows the introduction of sophisticated constitutive laws that take into account kinematic-isotropic nonlinear hardening, cyclic softening or ratcheting, complex creep and stress relaxation behaviour [Manonukul et al. 2005], coupled nonlocal damage and plasticity [Korsunsky et al. 2005], strain gradient effects [Fleck and Hutchinson 2001], etc. Refined models of deformation processes often involve large numbers of material parameters to be determined from experimental measurements. In order to justify the use of process modelling predictions for undertaking practical design decisions, proper validation procedures must be followed.

An alternative approach to residual stress evaluation is via an experimental procedure. In practice, residual stresses are only ever measured indirectly via observing relaxation or their effect on some other physical quantity, for example, bond vibration frequency as in Raman spectroscopy. Experimental techniques for stress evaluation can be classified into relaxation methods, physical correlation methods and diffraction techniques [Withers and Bhadeshia 2001].

Relaxation methods rely on material removal (slitting, hole drilling, blind hole drilling, layer removal, etc.) accompanied by the measurement of either changes in the object shape (monitored by photogrammetry), or changes in strain measured by means of surface mounted strain gauges, as in hole drilling. Physical correlation methods use various physical effects (thermoelastic, magnetoelastic, ultrasound propagation) to obtain some estimate of a stress state parameter—for example, the hydrostatic stress component.

Diffraction is a highly versatile method for direct measurement of interplanar spacing within the atomic lattice. Consequently residual elastic strain can be calculated on the basis of knowledge of strain-free spacing. Diffraction techniques, particularly those using high flux beams generated at synchrotrons [Korsunsky et al. 2002], can be scaled down to allow micro-diffraction and even nano-diffraction experiments. The use of synchrotron X-ray diffraction to provide the input for the current study is described below.

The objective of the present study is to provide a rational solid mechanics basis for the analysis of residual elastic strain data obtained from state-of-the-art synchrotron X-ray diffraction measurements. It must be pointed out, however, that the method described herein possesses great generality, and can be used with equal success to interpret relaxation method data, such as blind hole drilling or slitting techniques. As a vehicle for the introduction of new concepts, a classical problem of residually bent elastic-plastic beam is deliberately chosen for which an analytical solution of the direct eigenstrain problem is available [Korsunsky 2005]. The size

of the bent beam studied was chosen to be commensurate with the size of residually stressed objects routinely studied in the context of power generation and the aerospace industry. The material of the sample was Ti-6Al-4V aerospace titanium alloy used in the manufacture of fan and compressor blades of jet engines.

The article is organised as follows. In Section 2 the theoretical background for the analysis of residual stress states is presented. In Section 3 a concise presentation is given of the background to diffraction techniques for experimental strain analysis. In Section 4 a solution is presented to the 'direct' problem of determination of residual elastic strain in a beam from known permanent strain (eigenstrain) distributions. In Section 5 a framework is introduced for variational eigenstrain determination from measured residual elastic strain values, by minimising the sum of squares of model-experiment differences [Korsunsky et al. 2004]. In Section 6 the results of interpretation of a particular data set are presented and discussed.

2. Theoretical background

Residual stress states in arbitrarily shaped solid bodies are usually complex, and difficult to describe, since in the general case they must be represented by the six components of the stress tensor varying as a function of three spatial variables. It is virtually impossible to imagine an experimental procedure that would readily and routinely provide this level of detail. At any rate, the interpretation of point-wise data in terms of six independent components is likely to present a serious practical challenge.

Any residual stress state described by the tensor $\boldsymbol{\sigma}$ must, by definition, be self-equilibrating. This requirement in fact establishes a relationship between gradients of different components of the stress tensor, $\boldsymbol{\sigma}$, namely

$$\operatorname{div} \boldsymbol{\sigma} = 0. \tag{1}$$

Furthermore, the stress state deduced within a residually stressed object must satisfy the traction-free boundary conditions, namely,

$$\boldsymbol{\sigma} \cdot \mathbf{n} = 0, \tag{2}$$

where $\mathbf{n}$ denotes the surface normal. However, it is not easy to enforce this requirement on the deduced stress state, or to formulate the constraints that must be imposed on the measured strain data. It is possible to develop a rational analytical approach based on the concept of eigenstrain (stored inelastic strain) that reduces the size of the data array needed to represent a particular residual stress state and at the same time guarantees satisfaction of equations of equilibrium, (1), and traction-free boundary conditions (2).

Eigenstrain modelling is a powerful analytical technique for the representation of residual stress states in solids [Mura 1987]. A practical approach to the use of eigenstrain in residual stress modelling can be developed based on the following fundamental postulates [Korsunsky 1997; 2005]:

(a) In the absence of eigenstrain (stored inelastic strain), any elastic solid is completely free from residual stress. Indeed, the very definition of elastic material response requires that stresses and strains arise in the body upon the application of an external load, and that they vanish completely upon load removal.

(b) Residual stresses within a solid arise in response to the introduction, through some inelastic mechanism (plasticity, creep, cutting and pasting, phase transformation, etc.), of permanent *nonuniform* strains within the body. Note however that the introduction of an entirely spatially uniform permanent strain field does not, in fact, lead to the generation of residual stresses.

(c) Elastic and inelastic strains are additive, that is,

$$\boldsymbol{\varepsilon} = \boldsymbol{\varepsilon}^* + \boldsymbol{e}, \quad \text{or in index notation,} \quad \varepsilon_{ij} = \varepsilon_{ij}^* + e_{ij}, \tag{3}$$

where ε_{ij} denotes the total strain, e_{ij} denotes the elastic strain, and ε_{ij}^* denotes eigenstrain.

(d) Total strain must be compatible, that is, it must satisfy

$$\text{Inc } \boldsymbol{\varepsilon} = \text{rot}((\text{rot } \boldsymbol{\varepsilon})^T) = 0,$$

leading to relationships between strain and components of the type

$$\frac{\partial^2 \varepsilon_{xx}}{\partial y^2} + \frac{\partial^2 \varepsilon_{yy}}{\partial x^2} - \frac{\partial^2 \varepsilon_{xy}}{\partial x \, \partial y} = 0. \tag{4}$$

(e) Eigenstrains (permanent inelastic strains) act as the sources of incompatibility of displacement, and so can be thought of as appearing in the right hand of the Saint-Venant compatibility equations. Indeed, from the compatibility equation (4) one readily obtains the 'incompatibility' equation for the elastic strain, e_{ij}, in the following form:

$$\frac{\partial^2 e_{xx}}{\partial y^2} + \frac{\partial^2 e_{yy}}{\partial x^2} - \frac{\partial^2 e_{xy}}{\partial x \, \partial y} = \Xi, \tag{5}$$

where

$$\Xi = \frac{\partial^2 \varepsilon_{xy}^*}{\partial x \, \partial y} - \frac{\partial^2 \varepsilon_{xx}^*}{\partial y^2} - \frac{\partial^2 \varepsilon_{yy}^*}{\partial x^2}. \tag{6}$$

Note from the above expression that the 'forcing term', Ξ, turns to zero for uniform eigenstrains. In fact, it also vanishes for eigenstrains that depend linearly on one coordinate and not at all on other coordinates, so that bodies

containing such residual strain distributions are also free from residual elastic strain, and hence from residual stress.

(f) The problem of determining the residual elastic fields (residual elastic strain and residual stresses, as well as residual deformations, i.e. distortions) arising from a given eigenstrain, ε_{ij}^* requires the simultaneous solution of equations (5), (1) and (2), together with the elasticity equations (generalised Hooke's law),

$$\boldsymbol{\sigma} = \boldsymbol{C} : \boldsymbol{e}, \quad \text{or in index notation,} \quad \text{sig}_{ij} = \text{C}_{ijkl} \, \text{eps}_{kl} \, . \tag{7}$$

(g) The eigenstrain problem is not in fact in any way different from the well known thermoelastic problem, in which the forcing term Ξ arises from thermal gradients (note that an unconstrained uniformly heated body remains stress-free). In fact, arbitrary eigenstrain distributions can be successfully simulated by means of anisotropic thermal expansion.

Inelastic bending represents one of the most straightforward and well studied processes that leads to the creation of residual stresses. It can be treated very simply within the framework of beam theory whereby only longitudinal elastic strain and stress are considered and differ only by a constant factor that is Young's modulus.

The stresses that arise in inelastic bending can be readily analysed [Gere and Timoshenko 1984] provided the stress-strain behaviour of the sample material under uniaxial tension and compression is known. By modelling numerically the application of a given moment to a beam in bending, the permanent strains induced in the beam by plastic deformation can be readily deduced. Once the externally applied moment is removed, the beam is usually thought to undergo elastic unloading, so the residual stresses and residual elastic strains are easily found.

In the present study a different problem is addressed for which the residual elastic strains after bending are specified in the form of experimental diffraction measurements that may also be subject to some data scatter. The problem is to find the unknown distribution of permanent strains responsible for giving rise to the observed elastic strains. Furthermore, it is also possible to seek to extract approximations to the uniaxial tensile and compressive stress-strain curves from such residual strain data.

3. Experimental

The specimen of Ti64 was machined to the dimensions of $h_y = 50\,\text{mm}$, $h_x = 8.5\,\text{mm}$, $h_z = 4\,\text{mm}$ and bent by applying a bending moment M_z of approximately $100\,\text{N·m}$ using a $100\,\text{kN}$ capacity, screw driven universal testing machine
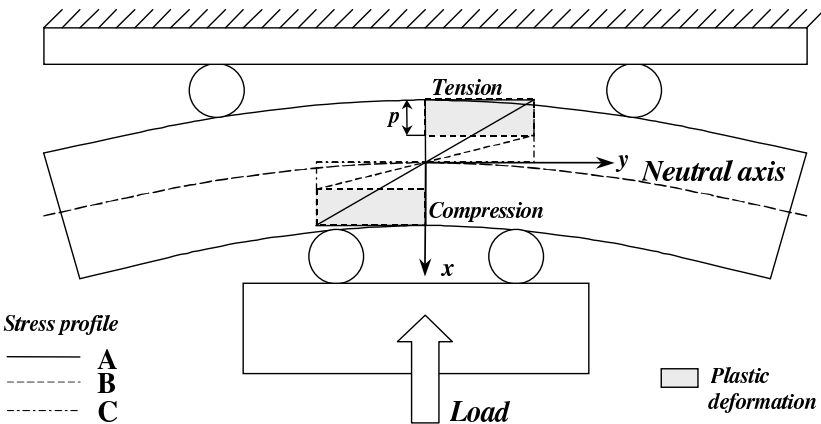
Figure 1. Illustration of plastic deformation of the bar in four point bending.

(Hounsfield Ltd) with a four point bending attachment Figure 1. Under progressively increasing applied bending moment, the lines labelled A, B and C in Figure 1 indicate the profiles of elastic strain, and hence also the longitudinal stress (for the simple case of non-work-hardening material). The elastic strain distribution across the bar remains linear (line A) until the first onset of yielding. Once the applied moment exceeds the yield value (line B), the material undergoes progressive plastic yielding from the surface. Note that the *total* strain remains linear across the beam, but a proportion of it is now accommodated plastically. With the increased applied moment (line C) the tensile and compressive plastic zones expand progressively inwards from two surfaces. In the course of analysis carried out in this paper we pay particular attention to the extent of the plastic zones, the deduction of the distribution of permanent inelastic strains within these zones, and the relationship between these distributions and the macroscopic and microscopic residual stresses.

The two principal methods for extracting residual elastic strain information within objects with the help of X-ray diffraction are the angle-dispersive (monochromatic beam) technique and the energy-dispersive (white-beam) technique.

In obtaining the data used in the present study a single bounce bent Laue mode monochromator was employed [Laundy et al. 2004] as shown in Figure 2. Bending the Laue monochromator crystal induces tensile and compressive strains on different sides of the crystal, thus increasing the band pass and hence the flux incident on the sample. In many engineering applications this is in fact an advantage, since peak broadening is dominated by sample effects (strain spread within the gauge
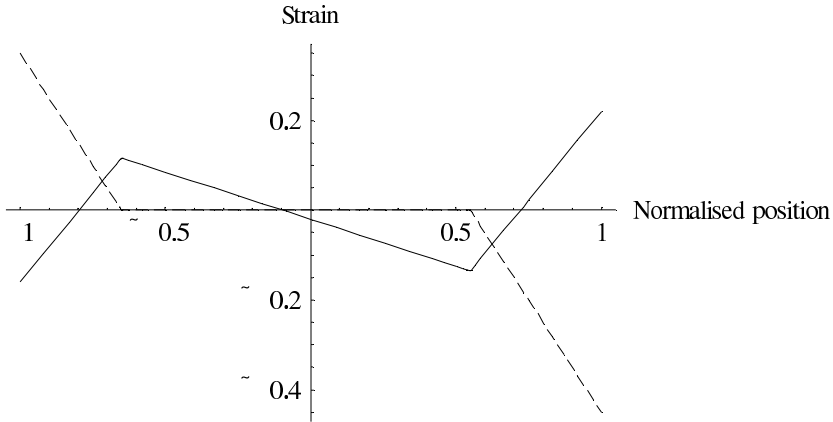
Figure 2. Schematic illustration of the relationship between linear segments of tensile and compressive eigenstrain induced by bending (dashed lines), and the residual elastic strain distribution (solid line) arising by the process of elastic equilibration.

volume), and no advantages are obtained by using extremely fine monochromation.

Diffraction patterns were collected by employing a detector scanning the scattering angle 2θ, or a position sensitive detector capable of registering total photon flux simultaneously at several positions along a line or over a two-dimensional surface. This mode allows accurate determination of diffraction peak intensity, shape and position. However, it usually requires significantly longer counting times in comparison with the white beam mode in order to collect the data from comparable sections of the diffraction pattern, primarily due to the reduction of flux by monochromation, but also due to the necessity of scanning the detector.

Energy dispersive setup allows multiple diffraction peaks to be collected simultaneously, thus achieving particularly efficient counting statistics at energies above 30 keV [Korsunsky et al. 2002]. The accuracy of determination of individual peak position and shape resolution in the white-beam mode is usually related to the resolution of the energy-dispersive detector, but can in fact be several orders of magnitude better. The accuracy of interpretation in terms of lattice parameters and hence strain can be significantly improved by using multiple peak analysis or whole pattern fitting [Liu et al. 2005].

The material used in the present study was a Ti-6Al-4V alloy widely used to manufacture components located at the front of the aeroengine, such as fan and

compressor blades and disks; it also has many other structural applications. Production comprises vacuum arc melting above the β phase transus followed by plate rolling in order to minimise crystallographic texture in the material which results in anisotropy of both elastic and inelastic properties. The structure achieved by the manufacturing process is obtained by solution treating in the middle of the $\alpha + \beta$ phase field and air cooling. This gives a mixture of primary α and a transformation product which comprises α and β phases. The α phase is a hexagonal close-packed structure (hcp) which is typical of titanium at room temperature and transforms to a body centred cubic structure, β phase, at 883° C. The size of the grains was approximately 5μm.

Bragg's law

$$2d \sin \theta = \lambda = \frac{hc}{E} \tag{8}$$

is used to determine a lattice spacing parameter d that can be related to any particular phase and may correspond to the phase average, if pattern refinement is used, or to the average taken over crystallites of a certain orientation within a particular phase.

The residual elastic strain is computed using the formula

$$e = \frac{d - d_0}{d_0}. \tag{9}$$

For each position of the gauge volume within the sample, peak centre positions were determined for the reflections (00.2), (10.1), (10.2) and (11.0) from the hcp α phase of the titanium alloy, as well as the macroscopic average computed by Pawley refinement of a section of the diffraction profile containing multiple peaks. Unstrained lattice spacing values d_0 for each of these reflections, are also needed to calculate strain using (9). To this end the data were collected by performing a similar measurement for the gauge volume located at the very corner of the sample. This choice of reference was based on the argument that the sampling volume in such a position must be free from tractions (and hence stresses) in the x and y directions. The gauge volume should be free from *macroscopic average* stress, and hence free from macroscopic average residual elastic strain. It should be noted that the above argument does not quite prevent the 'corner' gauge volume from containing microscopic residual stresses that exist between grains of different orientations.

In the sequel an interpretation is developed for the analysis of macroscopic residual elastic strains by employing the eigenstrain formalism. Since this development is undertaken within the framework of continuum elastic theory, the most appropriate residual elastic strain value that should be used for the purpose is that of macroscopic average.

4. Direct problem: determination of residual elastic strain from given eigenstrain

Consider an elastic beam occupying the region $x_L < x < x_R$, $-\infty < y < \infty$ and containing a distribution of eigenstrain $\varepsilon^*_{yy} = \varepsilon^*(x)$. The basic framework for evaluating the residual elastic strain (r.e.s.) distribution that arises in the beam has been presented in [Korsunsky 2005], and will only be reproduced here in brief to introduce some modifications to the previously published results.

The following statements provide the basis for the analysis:

(i) Total strain in the beam is given by the sum of the elastic and inelastic strain (eigenstrain).

(ii) Following Kirchhoff's hypothesis of straight normals, it is assumed that material points originally lying on a line perpendicular to the beam axis remain on a straight line, that is, any normal to the beam axis undergoes only rotation without distortion.

(iii) Hence displacements, and therefore *total* strain must vary linearly through the plate thickness, that is, they must be be given by

$$\varepsilon = e + \varepsilon^* = a + bx/h, \tag{10}$$

where $h = x_R - x_L$ is the beam thickness. Here the parameter a characterises the amount of axial straining experienced by the beam, and the term b characterises the intensity of bending.

(iv) In the absence of external loading being applied, elastic strain e presents an example of macroscopic residual elastic strain, such as that measured in a diffraction experiment.

(v) From equation (10), residual elastic strain is given by

$$e = a + bx/h - \varepsilon^*(x), \tag{11}$$

If the dependence of parameters a and b on the eigenstrain distribution $\varepsilon^*(x)$ is known, then the relationship between the residual elastic strain e and the eigenstrain is established.

(vi) It will be shown (below) that parameters a and b depend solely on two integral parameters, namely, the zeroth and first moments of the eigenstrain distribution given by

$$\Gamma = \frac{1}{h} \int_{x_L}^{x_R} \varepsilon^*(x)\, dx, \quad \Gamma_1 = \frac{1}{h^2} \int_{x_L}^{x_R} \varepsilon^*(x)\, dx. \tag{12}$$

The relationship between parameters a and b, on the one hand, and Γ and Γ_1, on the other, is established using the requirements of force and moment balance across

the beam, given by

$$F = \int_{x_L}^{x_R} [a + bx/h - \varepsilon^*(x)]\, dx = 0, \tag{13}$$

$$M = \int_{x_L}^{x_R} [a + bx/h - \varepsilon^*(x)]\, x\, dx = 0. \tag{14}$$

leading to the following relationships:

$$(x_R + x_L)\, b/2 + a(x_R - x_L) - (x_R - x_L) = 0, \tag{15}$$

$$(x_R^2 + x_R x_L + X_L^2)\, b/3 + (x_R^2 - x_L^2)\, a/2 - (x_R - x_L)^2 \Gamma_1 = 0. \tag{16}$$

Expressions are given explicitly in terms of the beam boundaries x_L and x_R for the purposes of generality, e.g., to allow the consideration of effects of surface layer removal.

The solution of the linear system for parameters a and b has the form

$$a = \frac{6\Gamma_1(X_R^2 - x_L^2) - 4\Gamma(x_R^2 + x_L + X_L^2)}{(x_R - x_L)^2}, \tag{17}$$

$$b = \frac{12\Gamma_1(x_R - x_L) - 6\Gamma(x_R - x_L)}{(x_R - x_L)}. \tag{18}$$

Noting that the since bending component of strain in terms of beam bending radius R and the beam curvature K is given by

$$e = \frac{x}{R} = xK, \tag{19}$$

then from equation (8) the curvature of the bent beam is found as

$$K = \frac{b}{h} = \frac{12\Gamma_1(x_R - x_L) - 6\Gamma(x_R - x_L)}{(x_R - x_L)^2}, \tag{20}$$

Equation (20) contains an expression that is useful for the analysis of beam curvature as an function of the eigenstrain distribution $\varepsilon^*(x)$.

Substituting equations (17) and (18) back into equation (11) gives the resulting prediction for the residual elastic strain distribution in the form

$$e(x) =$$

$$\frac{1}{(x_R - x_L)^2}\big(6\Gamma_1(x_R - x_L)(2x - x_R - x_L) + 2\Gamma((x_R^2 + x_R x_L + x_L^2) - 3x(x_R + x_L))\big)$$

$$- \varepsilon^*(x). \tag{21}$$

Equation (21) establishes the solution of the direct problem about the determination of residual elastic strain for arbitrary given distribution of eigenstrain.

Figure 1 gives an example of the above solution and shows the relationship between the eigenstrain distribution, $\varepsilon^*(x)$, and residual elastic strain e. This result should be seen as the simplest illustration of the relationship between eigenstrain and the residual elastic strain. Although in the present treatment this relationship is established analytically for a rather trivial case, the method is not restricted to such situations. In fact, arbitrary eigenstrain distributions can be readily incorporated into the finite element framework through the use of virtual anisotropic thermal expansion [Korsunsky et al. 2005; Hill 1996].

5. Inverse problem: determination of eigenstrain distribution from measured residual elastic strain

The problem that we wish to address in the present study stands in an inverse relationship to the one solved in the previous section. In practice it is the residual elastic strain distribution that may be known, for example, from diffraction measurement. Alternatively, changes in the elastic strain values can be monitored, say using strain gauges, in the course of material removal; and the underlying eigenstrain distribution then needs to be determined.

In practice the residual elastic strain, or its increments, can only be measured at a finite number of points. We are therefore seeking to *reconstruct* an unknown functional distribution, that is, an object with infinite number of degrees of freedom, using a finite data set. Several difficulties may arise in this situation, e.g. whether the problem described in the previous section can be inverted; whether the inverse problem is regular, i.e. varies in a smooth fashion depending on the data; and whether the obtained solution is unique. In the present study we do not attempt to answer these questions. Instead, we offer an efficient inversion procedure, leaving the evaluation of its uniqueness and regularity for future consideration.

Consider a set of experimental data consisting of the values of residual elastic strain (r.e.s.) y_j collected at positions x_j, $j = 1, \ldots, m$. In the present study we assume that the data was collected from a one-dimensional scan in coordinate x. It is worth noting, however, that the approach presented below is not in any way limited to one-dimensional problems, and can be readily generalised to two- and three-dimensional cases.

Denote by $e(x)$, as in the previous section, the *predicted*, or *modelled* residual elastic strain distribution. Evaluating $e(x)$ at each of the measurement points gives the *predicted values* $e_j = e(x_j)$. In order to measure the goodness of the prediction we form a functional J given by the sum of squares of differences between actual

measurements and the predicted values, with weights:

$$J = \sum_{j=1}^{m} w_j (y_j - e_j)^2. \tag{22}$$

The choice of weights w_j is left to the modeller; for example, they could be chosen based on the accuracy of measurements being interpreted.

Minimisation of functional J provides a rational variational basis for selecting the most suitable model to match the measurements, in terms of the overall goodness of fit.

Let us now assume that the unknown eigenstrain distribution, yet to be determined, is given by a truncated series of *basis* distributions,

$$e^*(x) = \sum_{i=1}^{N} c_i \xi_i(x). \tag{23}$$

Here N is the total number of *basis* distributions used in the prediction.

The results of the previous section contain the analytical procedure for the solution of the direct problem, that is, the determination of the residual elastic strain distribution that arises in response to an arbitrary eigenstrain distribution $e^*(x)$. This procedure can now be applied to each of the N basis distributions $\xi_i(x)$ in turn. As a result, a family of residual elastic strain solutions $E_i(x)$ is obtained.

Due to the linearity of the direct problem, the predicted values of e_j the residual elastic strain arising from the eigenstrain distribution $\varepsilon^*(x)$ of equation (23) can themselves be written in the form of a superposition of responses to the *basis* eigenstrain distributions,

$$e_j = \sum_{i=1}^{N} c_i E_i(x_j) = \sum_{i=l}^{N} c_i e_{ij}, \tag{24}$$

with the same coefficients c_i as in equation (23).

The inverse problem of determining the unknown eigenstrain distribution $\varepsilon^*(x)$ has now been reduced to the problem of determination of N unknown coefficients c_i that deliver a minimum to the functional J in equation (22), which may now be rewritten as

$$J = \sum_{j=1}^{m} w_j \left(\sum_{i=1}^{N} c_i e_{ij} - y_j \right)^2. \tag{25}$$

The above expression is quadratic and positive definite in the unknown coefficients c_i. It follows that the functional has a unique minimum that is found by

satisfying the condition

$$\nabla_c J = 0, \quad \text{or} \quad \frac{\partial J}{\partial c_i} = 0, \quad i = 1, \ldots, N. \tag{26}$$

Due to the quadratic nature of the functional in equation (25), the system of equations in equation (26) is linear. Therefore, the solution for the unknown coefficients c_i can be readily found *without iteration* by inverting the linear system arising in equation (26). This system is written out explicitly below.

The partial derivative of J with respect to the coefficient c_i can be written explicitly as

$$\frac{\partial J}{\partial c_i} = 2 \sum_{j=1}^{m} w_j e_{ij} \left(\sum_{k=1}^{N} c_k e_{kj} - y_j \right) = 2 \left(\sum_{k=1}^{N} c_k \sum_{j=1}^{m} w_j e_{ij} e_{kj} - \sum_{j=1}^{m} w_j e_{ij} y_j \right) = 0. \tag{27}$$

For purposes of illustration, let us now assume that the weights are equal to unity, so that equation (27) simplifies to:

$$\frac{\partial J}{\partial c_i} = 2 \left(\sum_{k=1}^{N} c_k \sum_{j=1}^{m} e_{ij} e_{kj} - \sum_{j=1}^{m} e_{ij} y_j \right) = 0. \tag{28}$$

We introduce the following matrix and vector notation

$$\mathbf{E} = \{e_{ij}\}, \quad \mathbf{y} = \{y_j\}, \quad \mathbf{c} = \{c_i\}. \tag{29}$$

Noting that notation e_{kj} corresponds to the transpose of matrix E, the entities appearing in (28) can be written in matrix form as:

$$\mathbf{A} = \sum_{j=1}^{m} e_{ij} e_{kj} = \mathbf{E}\mathbf{E}^T, \quad \mathbf{b} = \sum_{j=1}^{m} e_{ij} y_j = \mathbf{E}\mathbf{y}. \tag{30}$$

Hence equation (28) assumes the form

$$\nabla_c J = 2(\mathbf{A}\mathbf{c} - \mathbf{b}) = 0. \tag{31}$$

The solution of the inverse problem has thus been reduced to the solution of the linear system

$$\mathbf{A}\mathbf{c} = \mathbf{b} \tag{32}$$

for the unknown vector of coefficients $\mathbf{c} = \{c_i\}$.

Whenever the solution of an inverse problem is sought, questions arise concerning the existence and uniqueness of the solution, and also concerning the well-posedness of the problem, that is, the continuity of the dependence of the solution on the problem parameters, the choice of the basis functions, the number of terms N in the truncated series, etc.

Within the present regularised formulation of the problem, for an *arbitrary* choice of the family of basis functions and an *arbitrary* number of basis functions N, a unique solution is guaranteed to exist. This is a consequence of the positive definiteness of the quadratic functional J. Furthermore, it is clear that increasing the number of terms N is guaranteed to deliver a sequence of monotonically non-increasing values of J, in other words, the goodness of approximation will not be diminished.

An interesting question concerns the convergence of the solution in terms of eigenstrain distribution $\varepsilon^*(x)$, to the 'true' solution in the limit $N \to \infty$. Similarly, the continuity in the behaviour of the solution with the choice of basis functions deserves to be discussed. While it must be emphasised that these questions are clearly fundamental and ought to be addressed, the focus is currently placed on the development of a practical tool for residual strain analysis. In so far as this is the aim of the present study, the proposed framework offers an efficient 'one shot' approach to the solution of an inverse problem. Furthermore, the choice of moderate values N, compared to the number of measurements, m, also offers a rational procedure for smoothing the data.

Figure 2 illustrates the relationship between the simple kind of eigenstrain distribution that may be introduced by inelastic bending (shown by the dashed lines) in tension and compression on the opposites sides of the sample, and the residual elastic strain (shown by the solid line) that arises by the process of elastic equilibration in response. For simplicity, the eigenstrain distributions are assumed to be linear in both tension and compression. This assumption corresponds to the case of elastic-ideally plastic material. Note, however, that the depths of the plastic zones on the two sides of the sample are allowed to be different. As a result the residual elastic strain state that arises in the bent beam is asymmetric, and illustrates how asymmetry of residual stress distribution is connected with the asymmetry of material response (yielding) in tension and compression.

6. Results and discussion

The variation procedures for eigenstrain determination described in the previous section were applied to the experimental data obtained from synchrotron diffraction measurements. As noted earlier, the diffraction strain estimate that is obtained by whole pattern refinement provides the most reliable estimate of the average macroscopic residual elastic strain. These data were used in the present analysis.

The unknown eigenstrain distributions were represented by the following series:

$$\varepsilon^{*T} = \sum_{i=1}^{N} c_i (x - d)^i, \quad \varepsilon^{*C}(x) = -\sum_{i=1}^{N} c_i' (x - d')^i, \tag{33}$$
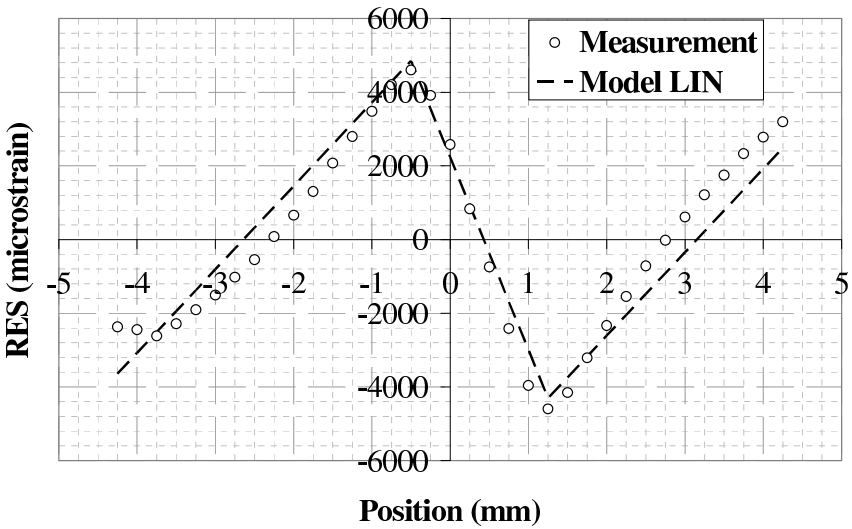
Figure 3. The measured profile of residual elastic strains in the bent Ti-6Al-4V bar (markers) compared with the prediction of the linear eigenstrain model (dashed line).

where superscripts refer to the tensile and compressive eigenstrains, and parameters d and d' denote the positions of the tensile and compressive plastic zone boundaries, respectively.

Several versions of the variational interpretation were investigated. In the first version a very simple interpretation was used of the type illustrated in Figure 2, that is, with linear assumed eigenstrain profiles in both tension and compression. Nevertheless, even with such simple assumptions it was possible to capture the salient features of the residual elastic strain distribution. The comparison between the model and experiment is illustrated in Figure 3, where the experimental measurement points are shown by the markers, while the continuous line shows the model prediction obtained using only the linear terms in the eigenstrain distribution. Note that the eigenstrain interpretation provides a 'balanced' approximation in the least squares sense; the model provides a 'smoothing' of the data at the chosen level of detail in the description of eigenstrain distribution.

Figure 4 shows the improvement to the model prediction afforded by allowing higher order of eigenstrain distribution functions (up to order 6): the agreement between the model and experimental measurements shown by markers is clearly improved. However, in the model used for this reconstruction the tensile and compressive eigenstrain distributions remained linked, in that the same coefficients were used in the expressions for the tensile and compressive eigenstrains in equation (33), that is, $c_i' = c_i$.
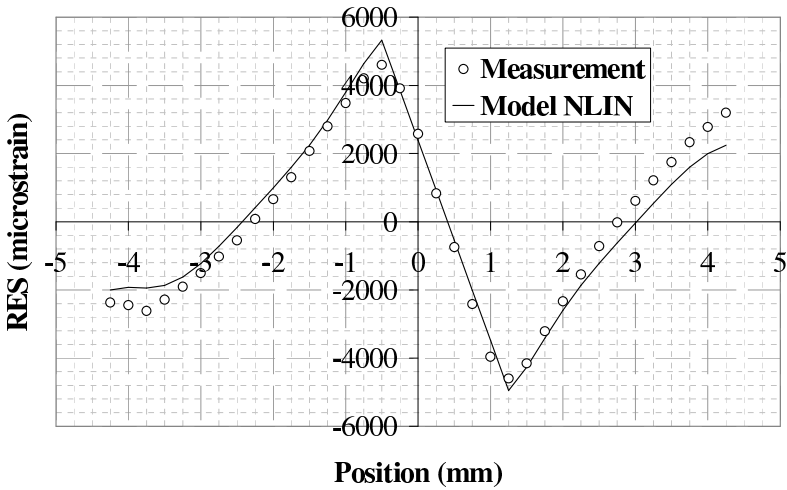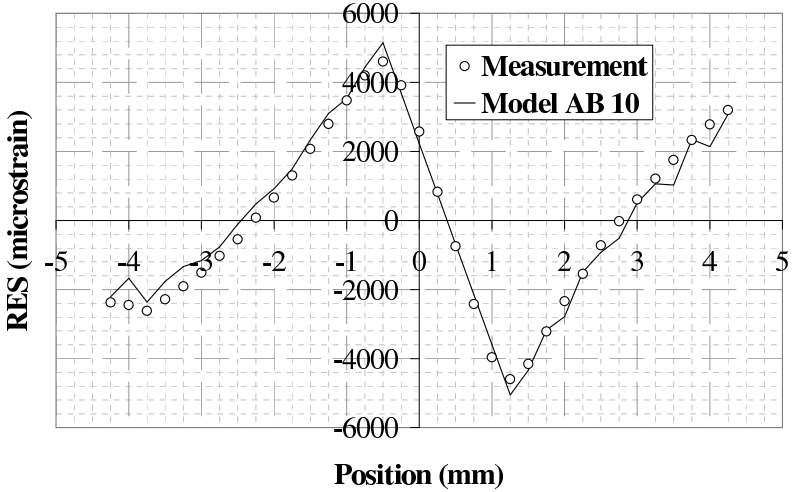
Figure 4. The measured profile of residual elastic strains in the bent Ti-6Al-4V bar (markers) compared with the predictions of the higher order eigenstrain model (continuous curve).

Figure 5 shows the result of interpretation, with the coefficients for the tensile and compressive eigenstrain distributions allowed to vary independently, that is, $c_i' \neq c_i$. Some small improvement can be detected, although it is not thought to be particularly significant.

Finally, in Figure 6 the order of approximation was increased to $N = 10$. This clearly delivers an improvement in the apparent quality of fit, but also leads to some oscillatory behaviour of the prediction curve. This situation might perhaps be expected for any approximation that involves higher order polynomial representation of an unknown distribution. The problem of this type could be overcome by representing the unknown distribution by a set of smooth radial basis functions with bounded support.

7. Conclusions

The purpose of the present paper was to introduce a self-contained framework that can serve as a convenient vehicle for introducing the fundamental ideas for residual stress reconstruction using the concept of continuous distributions of eigenstrain. Kirchhoff bending theory allows a simple analytical formulation to be developed for the prediction of the residual elastic strain (and hence bending stress) within inelastically bent bars due to distributions of tensile and compressive eigenstrains.
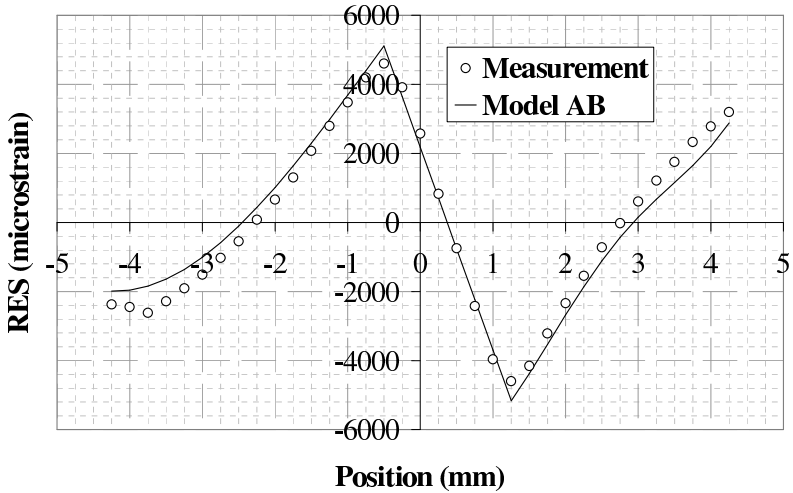
Figure 5. The measured profile of residual elastic strains in the bent Ti-6Al-4V bar (markers) compared with the predictions of the higher order eigenstrain model (continuous curve) with separate description of the tensile and compressive eigenstrain distributions.

Once these analytical formulae are established, they are used as the direct eigenstrain problem solver within the inverse framework for variational determination of unknown eigenstrains.

The solutions are obtained for residual elastic strain profiles measured by high energy synchrotron X-ray diffraction. The stability of the solutions is investigated by way of numerical experiments involving different formulation of the functional basis and different orders of approximation. It is found that the solutions display good stability, although the use of higher order polynomial approximations leads to some evidence of oscillatory behaviour of the solution. An approach using radial basis functions may be able to overcome this difficulty.

The findings of this paper are particularly relevant to the task of modelling the effects of residual stresses on subsequent deformation behaviour of engineering components. Assuming the residual elastic strain distribution can somehow be measured, for example, by diffraction, the underlying eigenstrain distribution can them be determined via an implementation of the variational approach presented here. When once such distribution is found, it can be used to continue deformation simulation onwards from the corresponding instant in the component's history. It then becomes possible not only to account accurately for the effects of residual

Figure 6. The measured profile of residual elastic strains in the bent Ti-6Al-4V bar (markers) compared with the predictions of the higher order eigenstrain model (continuous curve) with high order ($N = 10$) separate description of the tensile and compressive eigenstrain distributions.

stresses on subsequent deformation, but also *vice versa* to observe the evolution of the residual stress state (or, perhaps even more appropriately, of the underlying eigenstrain distribution) under deformation.

References

[Fleck and Hutchinson 2001] N. A. Fleck and J. W. Hutchinson, "A reformulation of strain gradient plasticity", *J. Mech. Phys. Solids* **49** (2001), 2245–2271.

[Gere and Timoshenko 1984] J. M. Gere and S. P. Timoshenko, *Mechanics of materials*, Chapman and Hall, London, 1984.

[Hill 1996] M. R. Hill, *Determination of residual stress based on the estimation of eigenstrain*, Ph.D. Thesis, Stanford University, 1996.

[Korsunsky 1997] A. M. Korsunsky, "An analysis of residual stresses and strains in shot peening", in *5th Intl. Conf. Residual Stresses*, edited by T. Ericsson, Linköping, Sweden, 1997.

[Korsunsky 2005] A. Korsunsky, "On the modelling of residual stresses due to surface peening using eigenstrain distributions", *J. Strain Anal.* **40** (2005), 817–824.

[Korsunsky et al. 2002] A. M. Korsunsky, S. P. Collins, R. A. Owen, M. R. Daymond, S. Achtioui, and K. James, "Fast residual stress mapping using energy-dispersive synchrotron X-ray diffraction on station 16.3 at the SR S", *J. Synch. Rad.* **9** (2002), 77–81.

[Korsunsky et al. 2004] A. M. Korsunsky, K. E. James, and M. R. Daymond, "Intergranular stresses in polycrystalline fatigue: diffraction measurement and self–consistent modelling", *Eng. Fract. Mech.* **71** (2004), 805–812.

[Korsunsky et al. 2005] A. M. Korsunsky, G. D. Nguyen, and G. T. Houlsby, "Analysis of essential work of rupture using non-local damage-plasticity modelling", *Intl. J. Fract.* **135** (2005), L19–26.

[Laundy et al. 2004] D. Laundy, A. Lennie, M. Golshan, D. Taylor, M. Roberts, G. Bushnell-Wye, R. Cernik, J. Flaherty, and I. Burrows, "A focusing Laue monochromator optimised for diamond anvil cell diffraction experiments", pp. 683–686 in *Synchrotron radiation instrumentation*, vol. 705, San Francisco, 2004.

[Liu et al. 2005] J. Liu, K. Kim, M. Golshan, D. Laundy, and A. M. Korsunsky, "Energy calibration and full-pattern refinement for strain analysis using energy-dispersive and monochromatic x-ray diffraction", *J. Appl. Cryst.* **38** (2005), 661–667.

[Manonukul et al. 2005] A. Manonukul, F. P. E. Dunne, D. Knowles, and S. Williams, "Multiaxial creep and cyclic plasticity in nickel-base superalloy C263", *Intl. J. Plasticity* **21** (2005), 1–20.

[Mura 1987] T. Mura, *Micromechanics of defects in solids*, Nijhoff, Dordrecht, 1987.

[Withers and Bhadeshia 2001] P. J. Withers and H. K. D. H. Bhadeshia, "Residual stress, 1: Measurement techniques", *Matls. Sci. Technol.* **17** (2001), 355–365.

ALEXANDER M. KORSUNSKY: *Department of Engineering Science, University of Oxford, Parks Road, Oxford OX1 3PJ, UK*
alexander.korsunsky@eng.ox.ac.uk
http://www.eng.ox.ac.uk/~ftgamk

EFFECT OF AIR GAPS ON THE BALLISTIC RESISTANCE OF DUCTILE SHIELDS PERFORATED BY NONCONICAL IMPACTORS

GABI BEN-DOR, ANATOLY DUBINSKY AND TOV ELPERIN

In *Int. J. Solids Struct.* **35**:23 (1998), pp. 3097–3103, we proved that localized interaction model (LIM) for shield-impactor interaction implies independence of the ballistic limit velocity (BLV) of spaced shield on air gaps and their widths for conical striker. In this study, the effect of deviation from a conical shape on ballistic properties of spaced shield is investigated using two-term LIM. It is found that this effect is insignificant and it causes small changes (of the order of few percent) in the magnitude of the BLV and energy absorbed by a shield. These theoretical predictions are in agreement with the available experimental results.

1. Introduction

Currently there is no consensus on the effect of layering and spacing on the ballistic properties of shields, although interest in this topic has existed for a long time.

Hurlich [1950] noted that the earliest study he found on the modern use of spaced armor was performed in 1913 for armor of naval vessels. He presented some qualitative arguments in favor of spaced armor (mostly for tanks), a number of tables with experimental results, some references and curious historical information. Honda et al. [1930] investigated experimentally the impact of steel plates by conical-nosed projectiles. It was found that a spaced shield with thicknesses of the plates equal to the half-thickness of a monolithic shield performed better than a monolithic shield. Marom and Bodner [1979] conducted a combined analytical and experimental comparative study of monolithic, layered and spaced thin aluminum shields. They found that the ballistic resistance of a monolithic shield is higher than that of a multilayered shield with the plates in contact and lower than the ballistic resistance of a spaced shield. Radin's study [1988] was also based on semi-empirical models and experimental investigations. They found a monolithic aluminum shield to be superior to a layered shield with the same total thickness for conical-nose and blunt projectiles, while spaced shields were less effective. Zukas [1996] performed calculations with thick plates impacted by long rod projectiles

---

Keywords: localized interaction model, ballistic limit velocity, spaced shield, nonconical impactor.

moving at 1500 m/s. The calculations showed that air gaps of one and four projectile diameters between plates involved the increase of projectile residual velocity when compared to their monoblock equivalent and a shield consisting of plates in contact. The sizes of the gaps play a minor role in determining the residual velocity. Using experimental results obtained for aluminum and steel plates and armor-piercing projectiles, Gupta and Madhu [1997] found that for the same impact velocity the residual velocity for the spaced shield was larger than for the plates in contact. Corran [1983a; 1983b], using experimental results on penetration of mild steel plates by impactors having "increasingly rounded nose shape", presented some data on perforation energy of spaced shields. Almohandes et al. [1996] conducted a comprehensive experimental study on the perforation of mild steel by standard 7.62 mm bullets. They investigated layered in contact, spaced and monolithic shields with total thickness in the range 8–14 mm. The efficiencies of shields were assessed by comparing their residual velocities for the same magnitude of the impact velocity. Almohandes' experimental results were used by Liang et al. [2005] for validating their penetration model. Applying this model for comparative analysis of shields with different structures, they concluded that an air gap slightly influenced the resistance to perforation in multilayered shields. Elek et al. [2005] developed a simple model to describe the perforation of monolithic and multilayered thin metallic plates by a flat-ended cylindrical impactor and used their model for the analysis of the ballistic properties of multilayered spaced shields. The main results of this study may be summarized as follows. The suggested model predicted that the monolithic shield will have larger resistance than any other multilayered shield with standoff distance between layers and equivalent total mass. The analysis of penetration in a two-layered shield showed that the maximum resistance could be obtained for very low or very high first-layer thickness (less than 20% or more than 80% of total thickness). The increase of the number of spaced layers of a multilayered shield, at constant total mass, caused a further decrease of the ballistic resistance.

In [Ben-Dor et al. 1998b; 1998a; 1999], we studied analytically the influence of air gaps between the plates on the ballistic limit velocity (BLV) of a multilayered shield. Using the general localized interaction model (LIM). In the first of these articles we found that the ballistic performance of the shield against 3D conical-nosed impactors is independent of the widths of the air gaps and of the sequence of plates in the shield and that it is determined only by the total thickness of the plates if the plates are manufactured from the same material. The influence of air gaps on the BLV of a shield that consisted of two plates manufactured from different materials was studied in [Ben-Dor et al. 1999] using the two-term localized interaction model. They found the criterion (depending on mechanical properties of the materials of the plates) that governs the decrease or the increase of the BLV
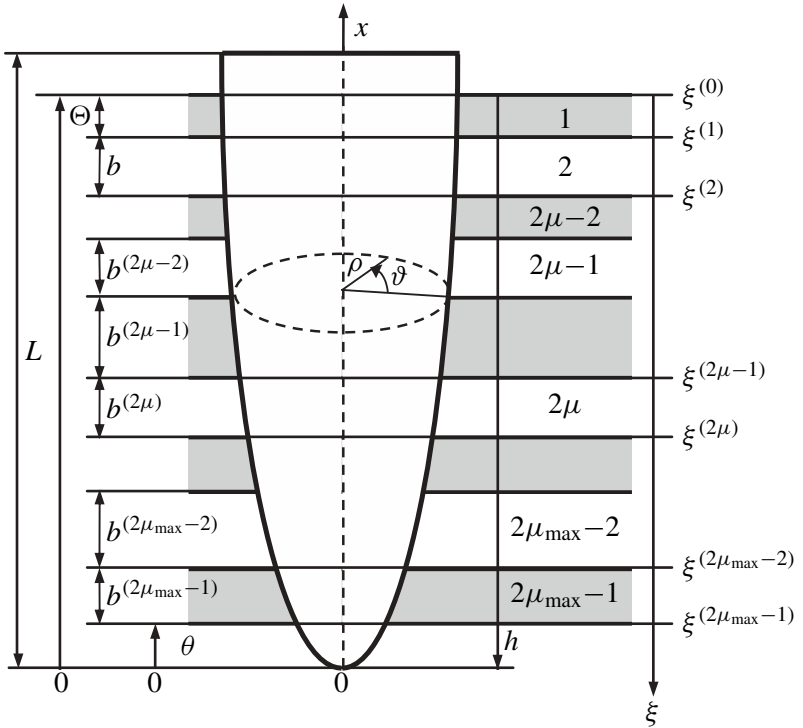
Figure 1. Notations.

of the shield with increasing the air gap thickness. Using the cylindrical cavity expansion model, we studied in [Ben-Dor et al. 1998a] the effect of air gaps on the ballistic performance of a spaced shield comprising plates manufactured from the same material and found that the BLV of the shield slightly increased with the increase of the widths of air gaps. In this study we investigate the effect of air gaps on ballistic properties of shields against nonconical impactors.

2. Formulation of the problem

Consider a high speed normal penetration of a rigid sharp striker (a body of revolution) into a ductile spaced shield with a finite thickness. We assume that the conditions of penetration are determined mainly by the "ductile hole enlargement perforation mechanism" ([Backman and Goldsmith 1978]). The basic notations are shown in Figure 1, and it is assumed that only the nose of the cylindrical impactor can interact with the shield. The coordinate h, the instantaneous depth of penetration, is defined as the distance between the leading edge of the nose of the impactor and the rear surface of the shield. The coordinate ξ is associated with the shield. In cylindrical coordinates (x, ρ, ϑ) associated with the impactor, the

surface of the nose is described by the equation

$$\rho = \phi(x), \quad 0 \le x \le L, \ 0 \le \vartheta \le 2\pi,$$

where L is the length of the impactor's nose and $\Phi(x)$ is an increasing convex function. Assume that the shield consists of $2\mu_{\max} - 1$ layers including $\mu_{\max}$ plates with the thicknesses $b^{(1)}, b^{(3)}, \ldots, b^{(2\mu_{\max}-1)}$ and air gaps between the plates with the thicknesses $b^{(2)}, b^{(4)}, \ldots, b^{(2\mu_{\max}-2)}$. The plate with number $2\mu - 1$ is located between the cross-sections $\xi = \xi^{(2\mu-2)}$ and $\xi = \xi^{(2\mu-1)}$, where $\mu = 1, 2, \ldots, \mu_{\max}$ and $\xi^{(0)} = 0$. The total thickness of the shield (the sum of the thicknesses of all layers including the air gaps) and the sum of the thickness of all plates are denoted b and b_{sum}, respectively. It is assumed that the plates are manufactured from the same material. The part of the lateral surface of the impactor between the cross-sections $x = \theta(h)$ and $x = \Theta(h)$ (see Figure 1) interacts with some layers of the shield or is in contact with some air gaps (see Figure 2, top):

$$\theta(h) = \begin{cases} 0 & \text{if } 0 \le h \le b, \\ h - b & \text{if } b \le h \le b + L, \end{cases} \qquad \Theta(h) = \begin{cases} h & \text{if } 0 \le h \le L, \\ L & \text{if } h \ge L. \end{cases}$$

The equation of motion of the impactor, $m(d^2h/dt^2) = -D$, can be rewritten as

$$mv(dv/dh) = -D, \tag{1}$$

where the velocity of the impactor v is considered to be a function of h, m is the mass of the impactor, and D is the resistance force. We consider the range of impact velocities v_{imp} whereby the projectile perforates the shield. Perforation occurs when the position of the striker is $h = b + L$. The ballistic limit velocity v_{bl} is defined as the impact velocity of the impactor required to emerge from the shield with zero residual velocity, $v_{\mathrm{res}} = 0$.

We assume that the impactor-target interaction at a given location at the surface of the impactor that is in contact with a plate can be represented as

$$d\vec{F} = \left(\gamma(-\vec{v}^0 \cdot \vec{n}^0)^2 v^2 + \sigma\right)\vec{n}^0 dS, \tag{2}$$

where $d\vec{F}$ is the force acting at the surface element dS of the impactor, $\vec{n}^0$ is the inner normal unit vector at a given location on the impactor's surface, $\vec{v}^0$ the unit vector of the impactor's velocity, the parameters γ and σ depend on the properties of the material of the shield. Equation (2) comprises most of the widely used phenomenological models for homogenous shields (see [Recht 1990; Ben-Dor et al. 2005] for details). In particular, in the model proposed and validated in comprehensive experimental studies in [Vitman and Stepanov 1959], σ and γ are the "dynamical hardness" and material density of the shield, respectively. The values of these parameters for some materials are given in [Vitman and Ioffe 1948,
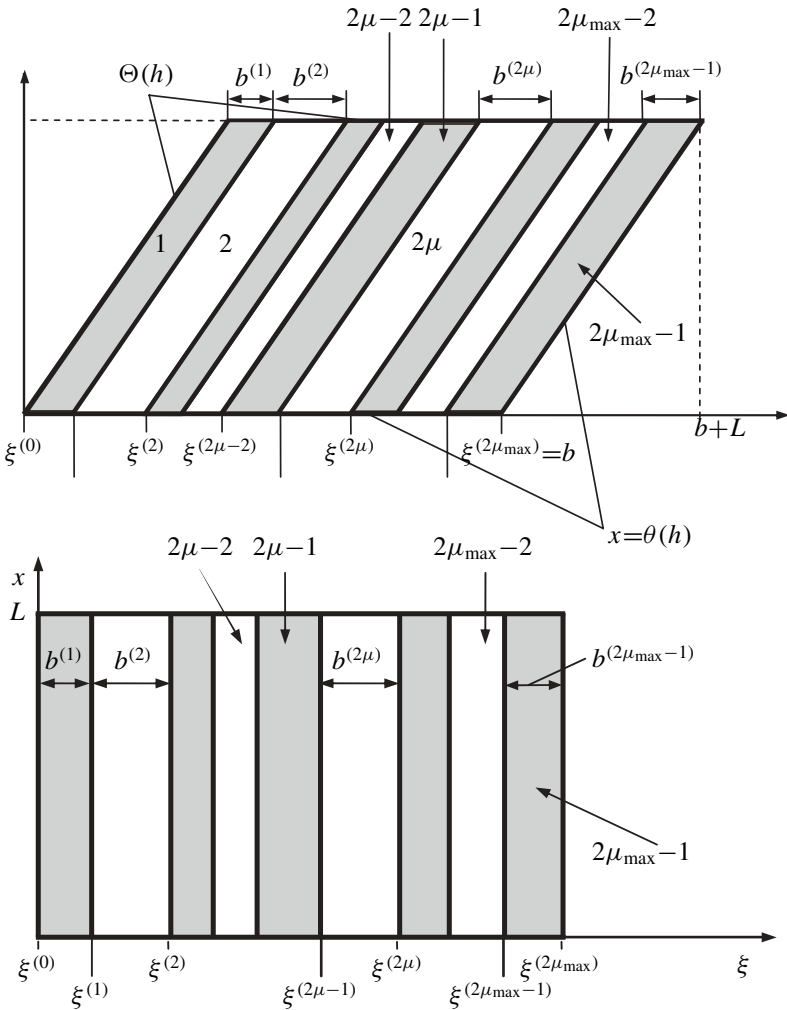
Figure 2. Model of the spaced shield, before and after a change of variables.

Table 1]. The remark by Recht [1990] concerning similar semi-empirical models is confirmed in this case: the parameter σ is significantly larger than the compressive yield strength.

In order to adapt (2) for a spaced shield we define a function $\varepsilon(\xi)$, which is equal to 1 if the point with the coordinate ξ (Figure 1) is located in any plate and is equal to 0 if this point is located in an air gap:

$$\varepsilon(\xi) = \varepsilon^{(j)} \text{ if } \xi^{(j-1)} \leq \xi \leq \xi^{(j)}, \quad j = 1, 2, \ldots, \mu_{\max} - 1,$$
$$\varepsilon^{(1)} = \cdots = \varepsilon^{(2\mu-1)} = \cdots = \varepsilon^{(2\mu_{\max}-1)} = 1,$$
$$\varepsilon^{(2)} = \cdots = \varepsilon^{(2\mu)} = \cdots = \varepsilon^{(2\mu_{\max}-2)} = 0.$$

Then (2) can be rewritten as

$$d\vec{F} = \varepsilon(\xi)\big(\gamma(-\vec{v}^0 \cdot \vec{n}^0)^2 v^2 + \sigma\big)\vec{n}^0 dS.$$

The total force $\vec{F}$ for some location of the impactor inside the shield is found by integrating the local force over the impactor-shield contact surface area, formally including the air gaps, that is, over the portion of the impactor's surface S determined by the inequalities $0 \leq \vartheta \leq 2\pi$ and $\theta(h) \leq x \leq \Theta(h)$. Taking into account the identity

$$\xi = h - x,$$

using the differential geometry formulas

$$-\vec{v}^0 \cdot \vec{n}^0 = \Phi'/\sqrt{\Phi'^2 + 1}, \quad dS = \sqrt{\Phi'^2 + 1}\, dx\, d\upsilon, \quad \Phi' = d\Phi/dx,$$

we obtain for the drag force D the expression

$$D = \vec{F} \cdot (-\vec{v}^0) = \iint_S \varepsilon(\xi)\big(\gamma(-\vec{v}^0 \cdot \vec{n}^0)^2 v^2 + \sigma\big)(-\vec{v}^0 \cdot \vec{n}^0) dS$$
$$= \frac{m}{2}\big(f_2(h)v^2 + f_0(h)\big), \tag{3}$$

where

$$f_2(h) = \frac{4\pi\gamma}{m} \int_{\theta(h)}^{\Theta(h)} \frac{\varepsilon(h-x)\Phi\Phi'^3}{\Phi'^2 + 1} dx, \quad f_0(h) = \frac{4\pi\sigma}{m} \int_{\theta(h)}^{\Theta(h)} \varepsilon(h-x)\Phi\Phi' dx.$$

Substituting D from (3) into (1) we obtain, after some algebra, an ordinary linear differential equation with respect to v^2:

$$dv^2/dh + f_2(h)v^2 + f_0(h) = 0.$$

The solution of this equation with the initial condition $v(0) = v_{\mathrm{imp}}$, which corresponds to the beginning of the motion of the impactor with the impact velocity v_{imp} reads

$$v^2(h) = \frac{1}{q(h)}\big(v_{\mathrm{imp}}^2 - g(h)\big), \tag{4}$$

(see [Kamke 1959]), where

$$q(h) = \exp\left(\int_0^h f_2(\eta)\, d\eta\right), \quad g(h) = \int_0^h f_0(H)q(H)dH.$$

Equation (4) yields formulas for the residual velocity, $v_{\mathrm{res}} = v(b+L)$, and the ballistic limit velocity, v_{bl}:

$$v_{\mathrm{res}}^2 = \frac{1}{q(b+L)}\big(v_{\mathrm{imp}}^2 - g(b+L)\big), \quad v_{\mathrm{bl}}^2 = g(b+L). \tag{5}$$

We now prove that $q(b + L)$ does not depend on the widths of the air gaps. To this end, we change the variables (Figure 2, bottom), $x \to x$, $h \to x + \xi$, in the integral in the expression for $q(b + L)$:

$$\frac{m}{4\pi\gamma} \ln(q(b+L)) = \int_0^{b+L} dh \int_{\theta(h)}^{\Theta(h)} \varepsilon(h-x)\psi(x)\,dx$$

$$= \int_0^L \psi(x)\,dx \int_0^b \varepsilon(\xi)\,d\xi = b_{\text{sum}} \int_0^L \psi(x)\,dx,$$

where $\psi(x) = \Phi\Phi'^3/(\Phi'^2 + 1)$.

For further analysis it is convenient to rewrite the expression for v_{bl} in dimensionless variables where L is chosen as a characteristic length:

$$v_{\text{bl}} = \sqrt{\chi}\,\Psi(\alpha), \tag{6}$$

where function Ψ depends also on structure of the shield and the shape of the impactor, and

$$\Psi(\alpha) = \sqrt{\alpha \int_0^{\bar{b}+1} Q(\bar{h})\,d\bar{h} \int_{\bar{\theta}(\bar{h})}^{\bar{\Theta}(\bar{h})} \bar{\varepsilon}(\bar{h}-\bar{x})\bar{\Phi}\bar{\Phi}'d\bar{x}},$$

$$\alpha = \frac{4\pi L^3}{m}\gamma, \quad \bar{x} = \frac{x}{L}, \quad \bar{\Phi} = \frac{\Phi}{L}, \quad \bar{\Phi}' = \frac{d\bar{\Phi}}{d\bar{x}}, \quad \bar{h} = \frac{h}{L},$$

$$Q(\bar{h}) = \exp\left(\alpha \int_0^{\bar{h}} d\bar{H} \int_{\bar{\theta}(\bar{H})}^{\bar{\Theta}(\bar{H})} \frac{\bar{\varepsilon}(\bar{H}-\bar{x})\overline{\bar{\Phi}\,\bar{\Phi}'}^3}{\bar{\Phi}'^2 + 1}d\bar{x}\right), \quad \bar{\varepsilon}(\bar{\xi}) = \varepsilon(L\bar{\xi}),$$

$$\bar{\theta}(\bar{h}) = \begin{cases} 0 & \text{if } 0 \le \bar{h} \le \bar{b}, \\ \bar{h}-\bar{b} & \text{if } \bar{b} \le \bar{h} \le \bar{b}+1, \end{cases} \qquad \bar{\Theta}(\bar{h}) = \begin{cases} \bar{h} & \text{if } 0 \le \bar{h} \le 1, \\ 1 & \text{if } \bar{h} \ge 1. \end{cases}$$

Let v_{bl}^0 be the BLV of the shield that consists of the plates in contact with the same total thickness b_{sum}. Clearly, the latter structure is equivalent to a monolithic shield with the total thickness b_{sum}. Equation (6) implies that the ratio $\delta = v_{\text{bl}}/v_{\text{bl}}^0$ depends on the dimensionless parameter α as well as on the shape of the impactor and the dimensionless thicknesses of the layers (plates and air gaps). In our previous investigation it was found that spacing does not affect ballistic properties of shields against conical impactors [Ben-Dor et al. 1998b]. In the next section, using the obtained expression for the BLV we study numerically the effect of spacing on the BLV of the shields for nonconical impactors.
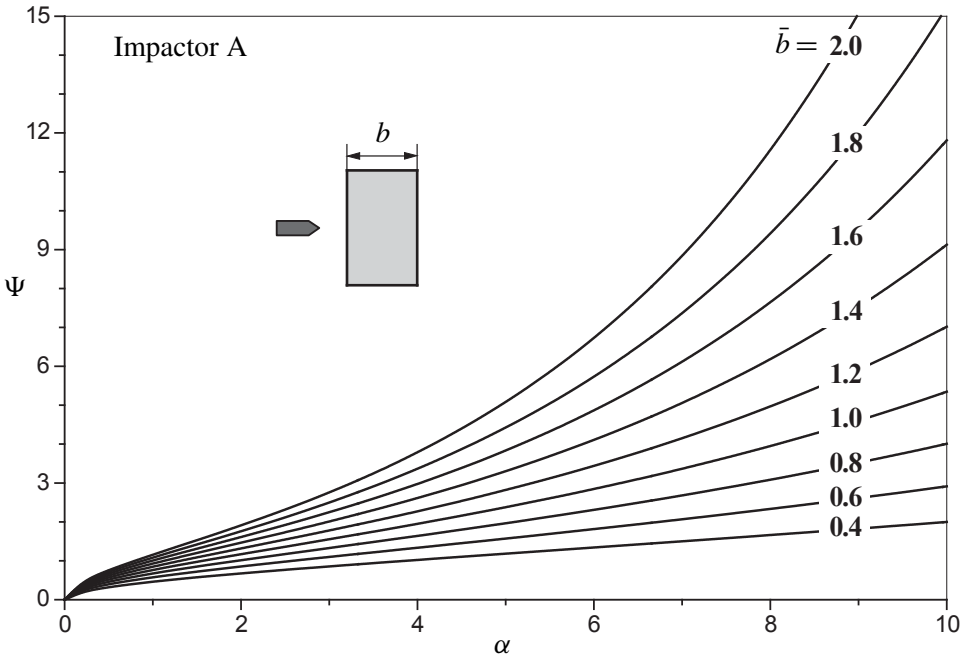
Figure 3. Function $\Psi(\alpha)$ for different values of $\bar{b}$.

3. Result of numerical calculations and discussion

We performed calculations for two impactors (cylindrical bodies of revolution with different nose shapes). The first impactor (Impactor A) is the cylinder with a semi-spherical nose. The generatrix of the nose of the second impactor (Impactor B) is the broken line consisting of segments between three points with the dimension-less coordinates $(\bar{x}, \bar{\rho})$: $(0, 0)$; $(0.3, 0.7)$; $(1, 1)$. The validity of this model for blunt, thick impactors against shield with a finite thickness is questionable, and, consequently, Impactor A can be considered only as a model.

In the following analysis, we use dimensionless variables. We show in Figure 3 plots of function Ψ for a monolithic plate. Using (6), one can determine the ranges of variation of parameters $\bar{b}_{\text{sum}}$ and α that correspond to a given range of BLVs for a shield manufactured from a given material (for some materials one can use Table 1). The plots in Figure 3 correspond to Impactor A, and they differ only insignificantly from those obtained for Impactor B.

The top halves of Figures 4–10 correspond to Impactor A, while the bottom parts correspond to Impactor B.

Figure 4 shows the dependence of the parameter $\delta = v_{\text{bl}}/v_{\text{bl}}^0$ versus α for different widths of the gap (case with two plates). Inspection of these figures shows that ballistic efficiency of the shield decreases when the width of the gap between the
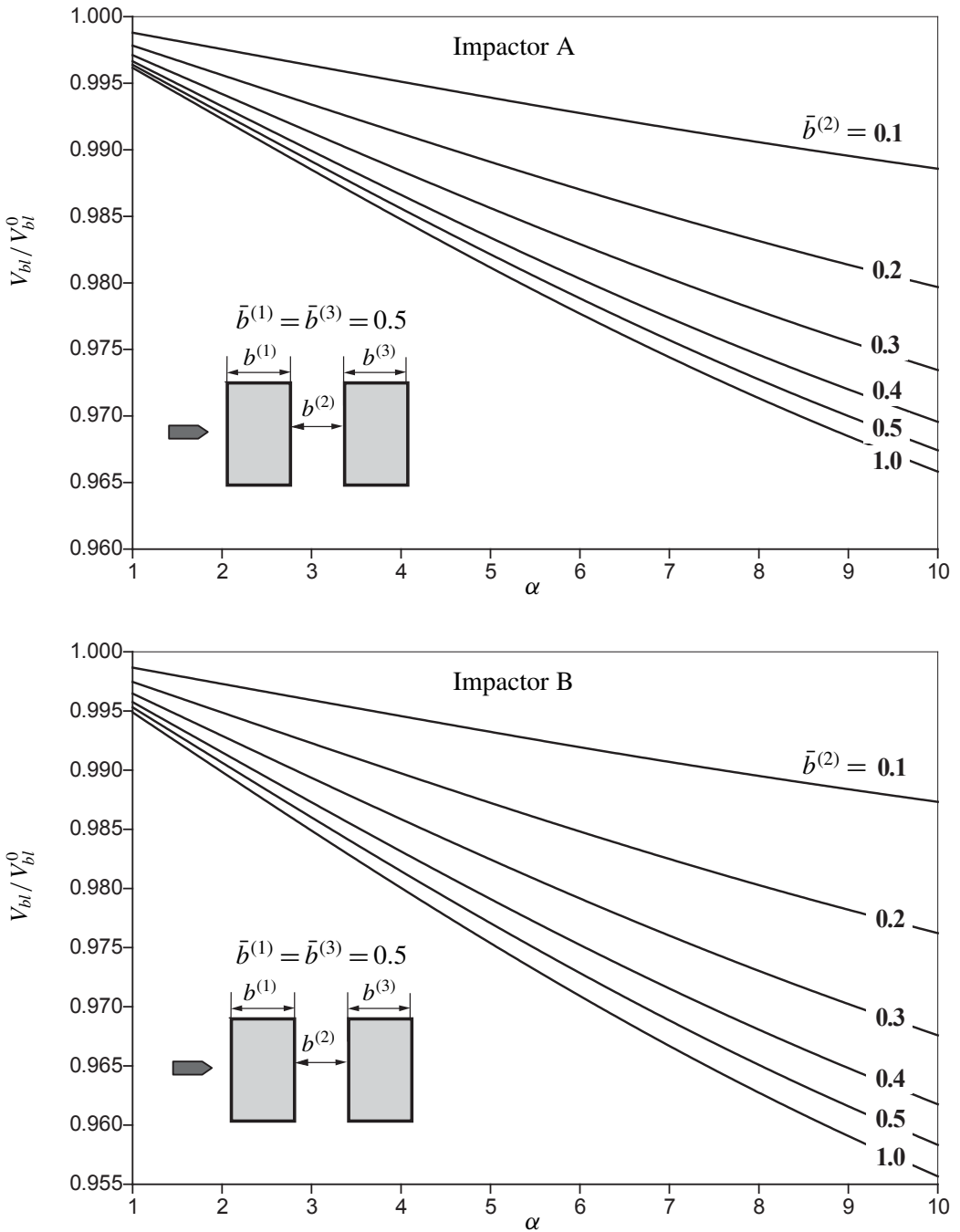
Figure 4. Effect of the thickness of the air gap on ballistic proper-
ties of a two-layer shield.

Number	Material	Dynamical hardness σ (N/m^2)	Density γ (kg/m^3)	$\chi = \sigma/\gamma$ (m^2/s^2)
1	Aluminum	$350 \cdot 10^6$	2765	$0.127 \cdot 10^6$
2	Soft steel	$1850 \cdot 10^6$	7830	$0.236 \cdot 10^6$
3	Copper	$910 \cdot 10^6$	8930	$0.102 \cdot 10^6$
4	Duraluminum	$1330 \cdot 10^6$	2765	$0.481 \cdot 10^6$

Table 1. Parameters of the model. (Based on [Vitman and Ioffe 1948]).

plates is increased (the larger is parameter $\bar{b}^{(2)}$, the lower is the curve). The rate of change of parameter δ that characterizes ballistic efficiency sharply reduces with increase of $\bar{b}^{(2)}$. Thus, for instance, inspection of Figure 4 shows that increase of $\bar{b}^{(2)}$ from 0.4 to 0.5 and from 0.5 to 1.0 is accompanied by approximately the same change of δ. The effect of gaps on δ becomes more pronounced as the parameter α increases. Analysis of Figure 5 allows us to arrive at similar conclusions. In this figure we showed the results obtained for a shield consisting of three plates whereby the widths of air gaps between the first and the second plate and between the second and the third plate are varied but remain equal. Therefore the maximum negative effect of spacing (which is the most interesting) occurs for large air gaps equal to the length of the impactor's nose since in the framework of the used model, further increase of the air gap width does not change the BLV of the shield. Our further analysis is performed exactly for this width of air gap ($\bar{b}^{(2)} \geq 1$).

In Figure 6, we show the plots of function $\delta(\alpha)$ for the case of two identical plates and for different magnitudes of the total width of the shield, $\bar{b}_{\text{sum}}$. Inspection of these plots shows that the effect of air gaps is more pronounced for large α. It must be noted that some curves in Figure 6 intersect, that is, for two shields with different total thicknesses parameter δ can be larger for the first shield than for a second one for one magnitude of the parameter α, and it can become smaller than for the second shield for a different value of parameter α. The same dependencies are observed in the case when a plate is separated into three identical plates with equal widths of the gaps between the plates (Figure 7).

In Figures 8 and 9, we show the plots of $\delta(\lambda)$ for $\bar{b}_{\text{sum}} = 1$ and $\bar{b}_{\text{sum}} = 2$, respectively, and for different values of parameter α (λ is the ratio of the width of the first plate to the total width, $\lambda = b^{(1)}/b_{\text{sum}}$). Inspection of these figures suggests that the shape of the impactor affects the plots only weakly, and that for every value of parameter α the curves $\delta(\lambda)$ are concave; that is, initially the increase in relative thickness of the first plate causes an increase in the negative effect of spacing, until some maximum value is attained, and then it starts to decrease; the magnitude of λ
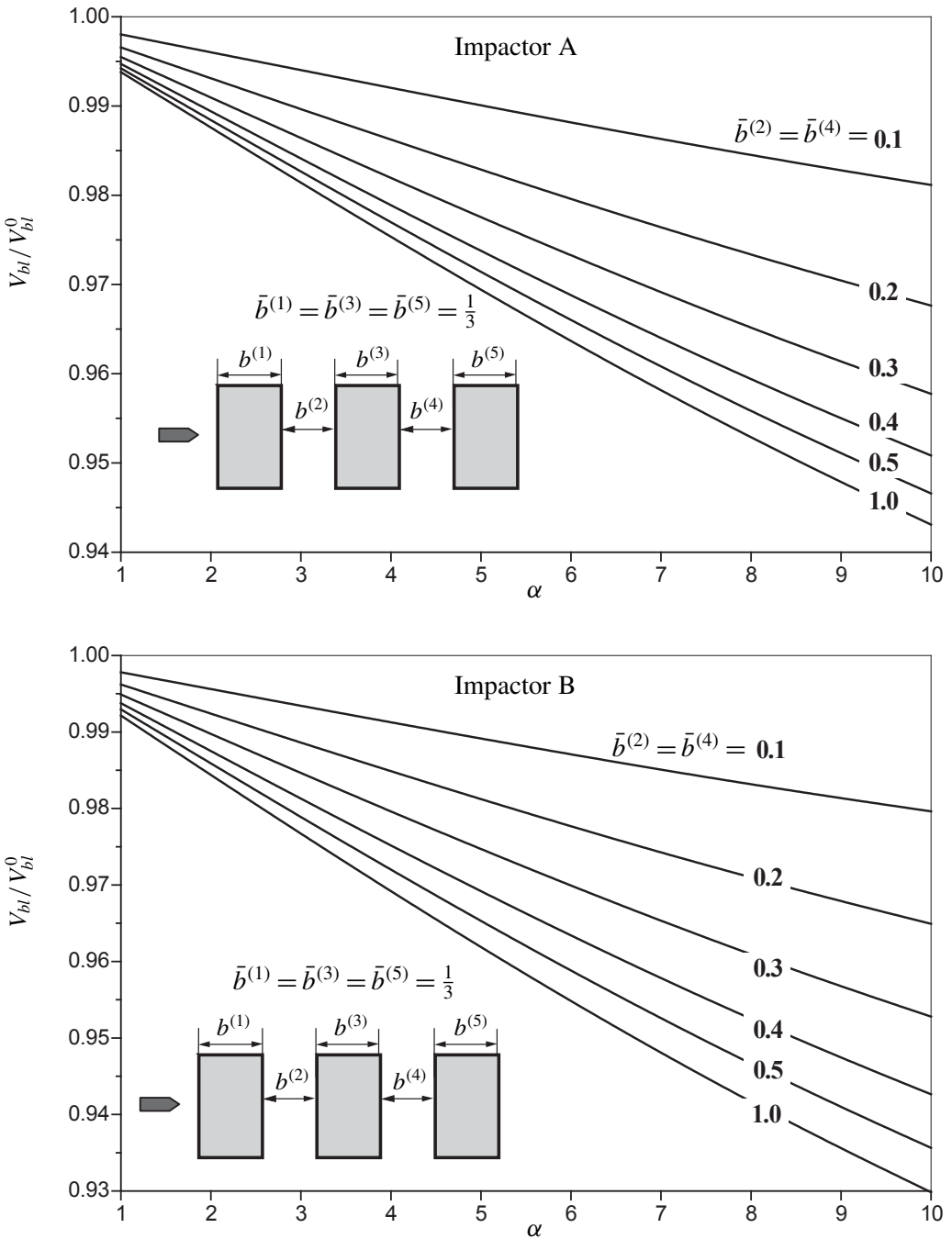
Figure 5. Effect of the thickness of the air gap on ballistic properties of a three-layer shield.
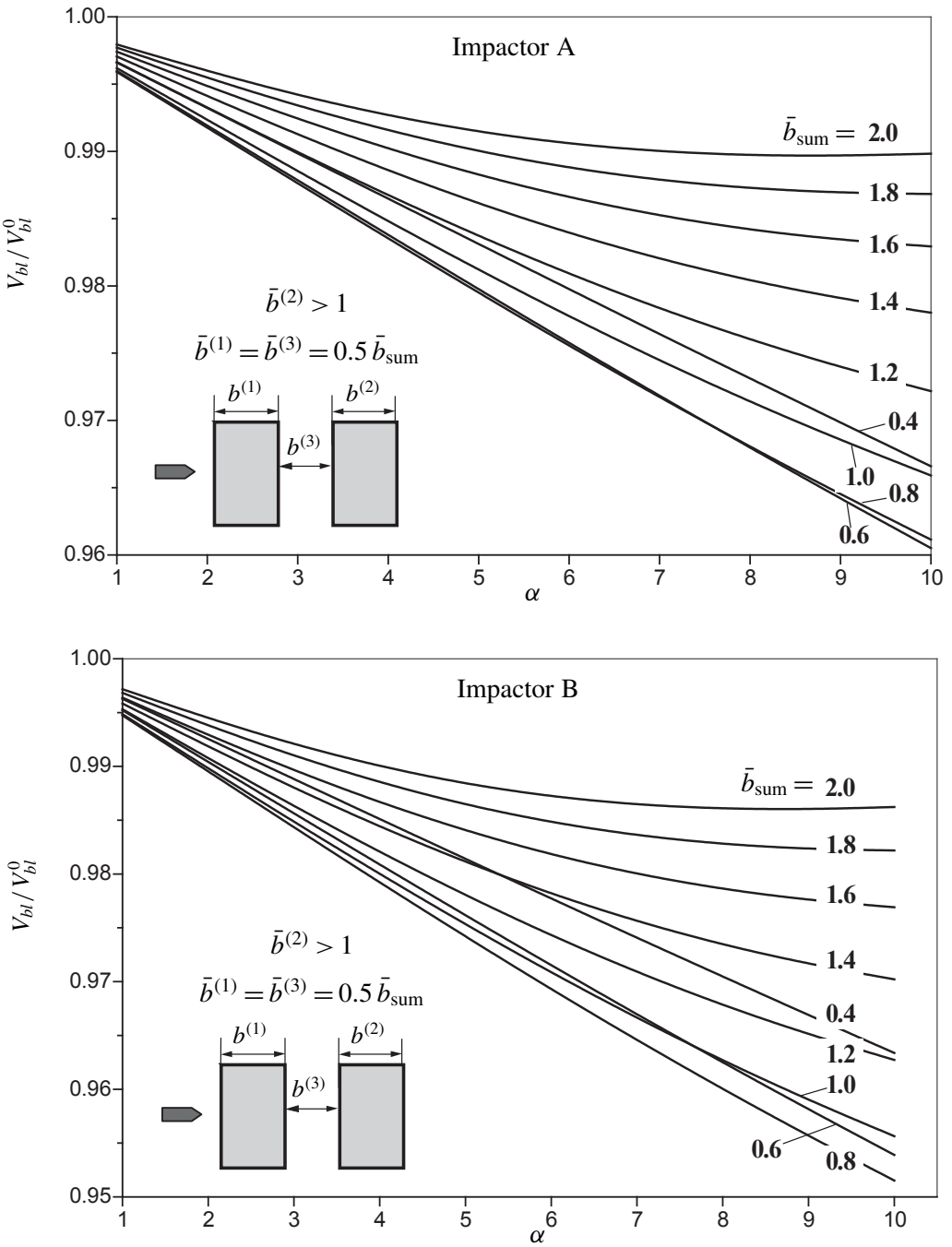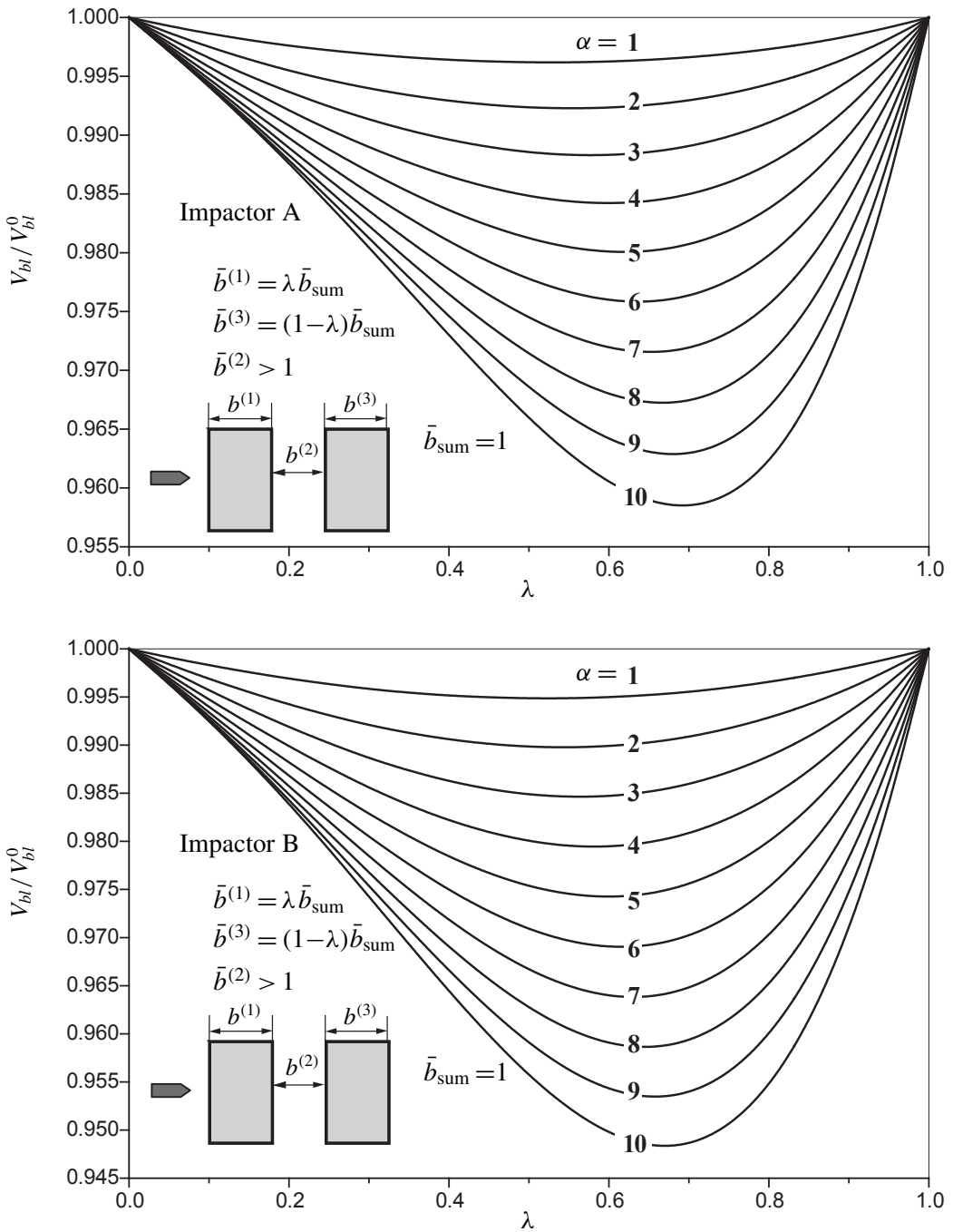
Figure 6. Effect of the parameters α and $\bar{b}_{\text{sum}}$ on ballistic properties of two-layer shield with large air gap.
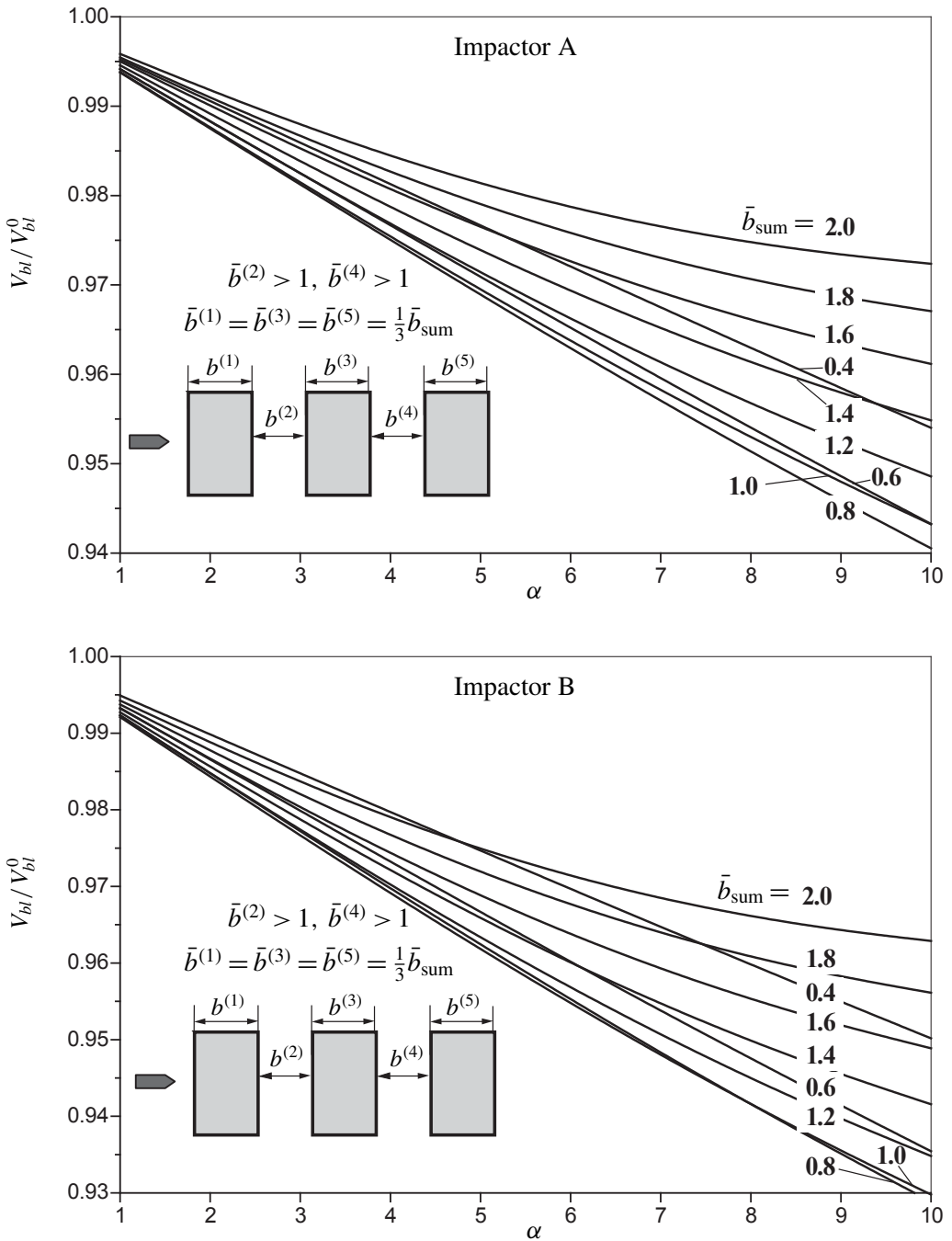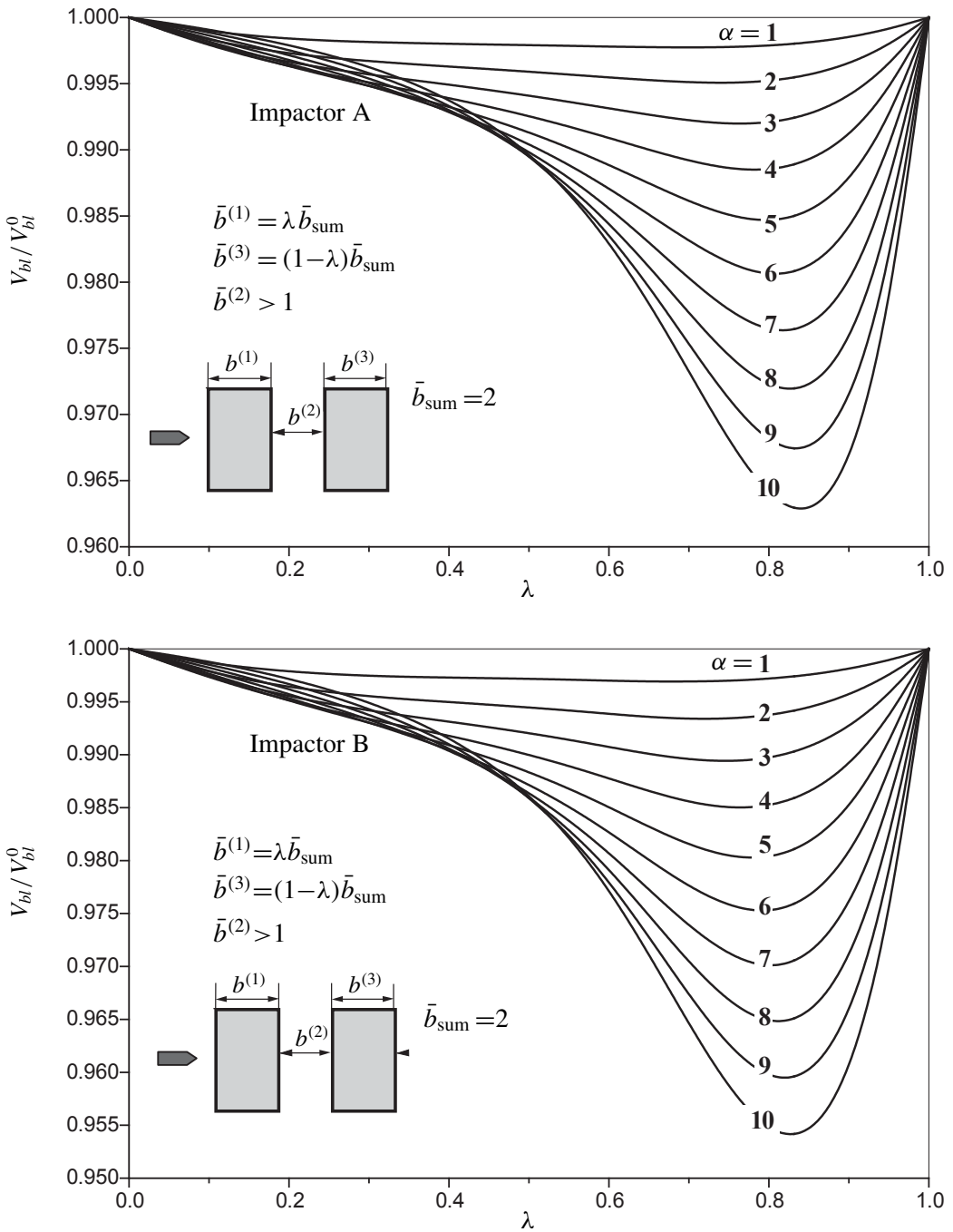
Figure 7. Effect of parameters α and $\bar{b}_{\text{sum}}$ on ballistic properties of three-layer shield with large air gaps.
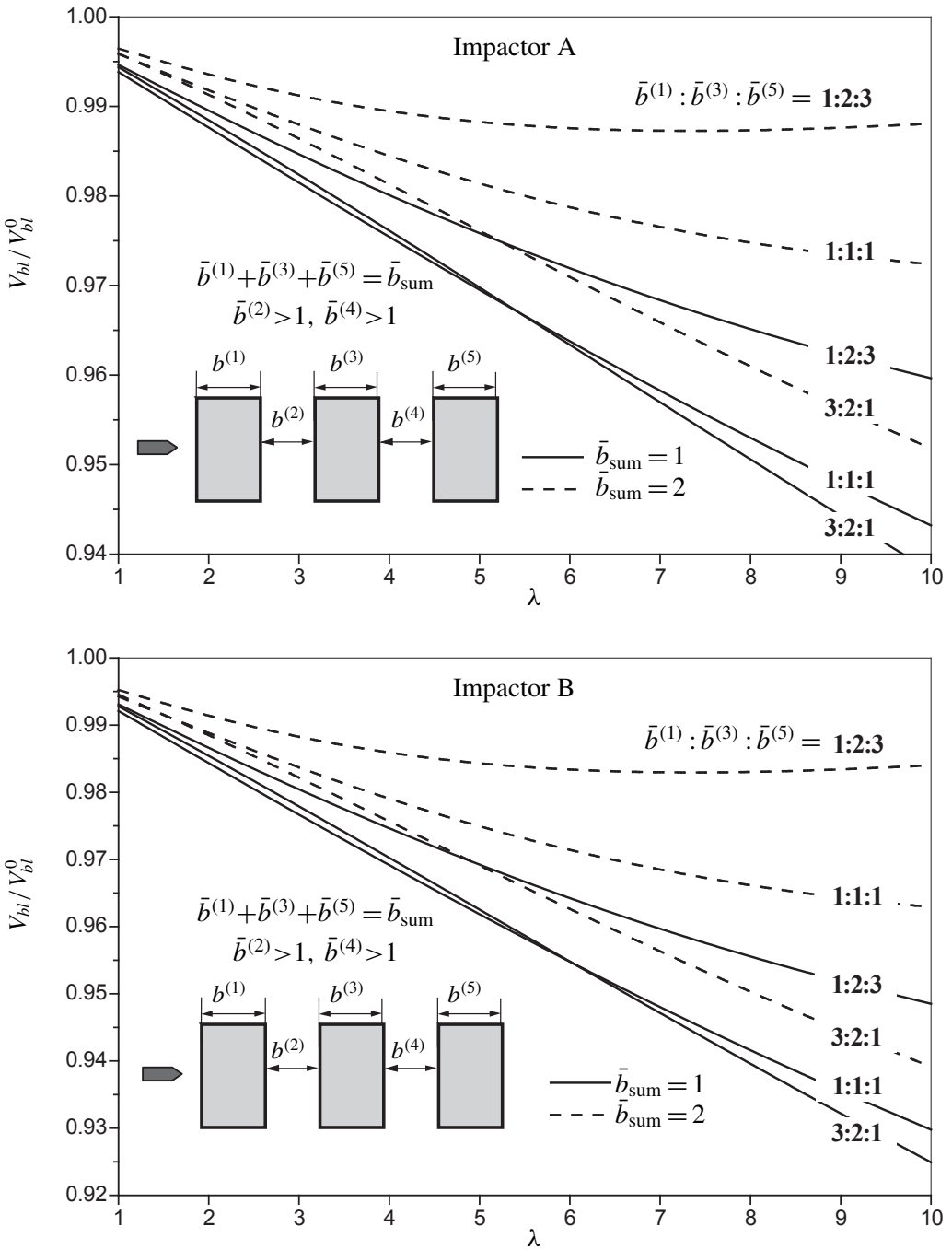
Figure 8. Effect of the thicknesses of plates on ballistic properties of two-layer shield with large air gap and $\bar{b}_{\text{sum}} = 1$.

Figure 9. Effect of the thicknesses of plates on ballistic properties of two-layer shield with large air gap and $\bar{b}_{\text{sum}} = 2$.

Figure 10. Effect of the thicknesses of plates on ballistic properties of three-layer shield with large air gaps.

where the negative effect of spacing is maximum depends upon the total thickness, and it increases with the increase of parameter α.

Figure 10 compares the results obtained for a spaced shield consisting of three plates with different relative thicknesses of the plates with a constant total thickness, $\bar{b}_{\text{sum}}$, for $\bar{b}_{\text{sum}} = 1$ and $\bar{b}_{\text{sum}} = 2$. Calculations were performed for three sets of the relative thicknesses of the plates, $b^{(1)} : b^{(2)} : b^{(3)}$, namely, $1:2:3$, $1:1:1$ and $3:2:1$. The results show that the negative effect of spacing is minimum for configuration $1:2:3$, and it depends on α for two other configurations.

Therefore, in the framework of the employed model, the effect of spacing on the BLV of the nonconical impactors is of the order of several percent. The results of the calculations showed that for slender projectiles this effect becomes even smaller.

4. Discussion of experimental results

Experimental data that allow us to compare directly the BLV of spaced shields and shields with plates in contact are not available. However we can use the results of the experiments performed with spaced and nonspaced shields for the same magnitudes of the impact velocity, v_{imp}. These data were published in [Almohandes et al. 1996] for 7.62 mm bullets perforating a mild steel shield. First, using the model, we analyze the connection between the value of energy absorbed and the value of BLV of the spaced and nonspaced shields for the same magnitude v_{imp}.

Using relationships for the impact energy E_{imp} and residual energy E_{res} of the impactor,

$$E_{\text{imp}} = \tfrac{1}{2}mv_{\text{imp}}^2, \qquad E_{\text{res}} = \tfrac{1}{2}mv_{\text{res}}^2,$$

and Equation (5) rewritten as

$$v_{\text{res}}^2 = \frac{1}{q_*}\big[v_{\text{imp}}^2 - v_{\text{bl}}^2\big], \qquad q_* = q(b+L) = Q(\bar{b}+1),$$

we obtain for the relative energy absorbed by the spaced shield:

$$e_{\text{abs}} = \frac{E_{\text{imp}} - E_{\text{res}}}{E_{\text{imp}}} = \frac{q_* - 1}{q_*} + \frac{1}{q_*}\left(\frac{v_{\text{bl}}}{v_{\text{imp}}}\right)^2. \tag{7}$$

Since q_* is the same for the spaced and nonspaced shield (see below), we may write an equation similar to (7) for the spaced shield:

$$e_{\text{abs}}^0 = \frac{E_{\text{imp}} - E_{\text{res}}^0}{E_{\text{imp}}} = \frac{q_* - 1}{q_*} + \frac{1}{q_*}\left(\frac{v_{\text{bl}}^0}{v_{\text{imp}}}\right)^2.$$

Then

$$e_{\text{abs}} - e_{\text{abs}}^0 = \frac{1}{q_*}\left(\frac{v_{\text{bl}}^0}{v_{\text{imp}}}\right)^2\left[\left(\frac{v_{\text{bl}}}{v_{\text{bl}}^0}\right)^2 - 1\right]. \tag{8}$$
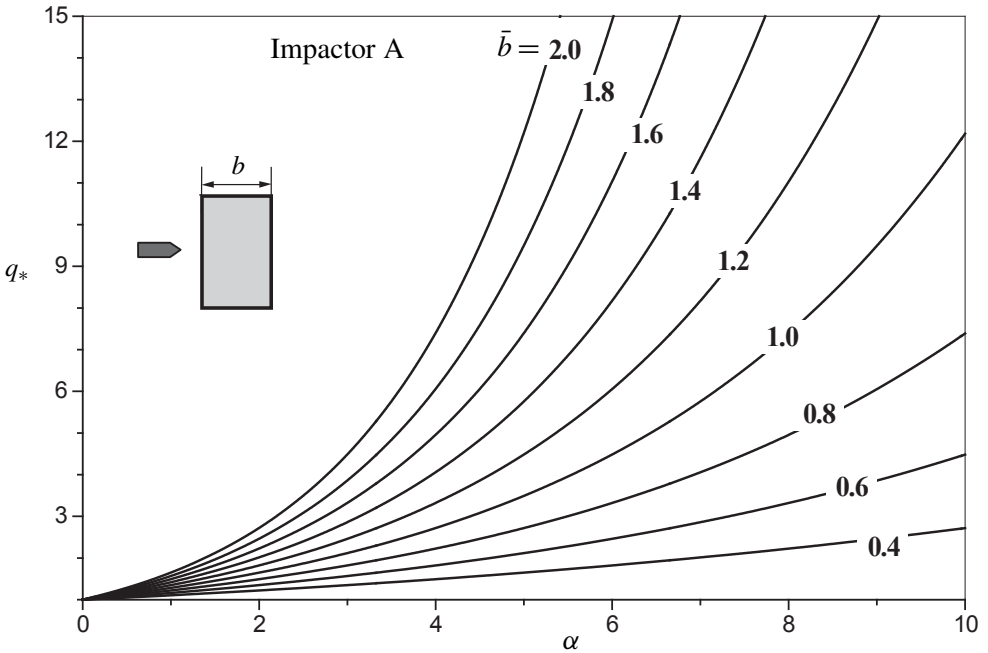
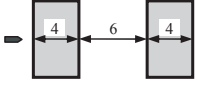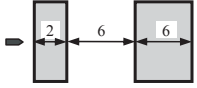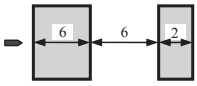**Figure 11.** Function $q_*(\alpha)$ for different values of $\bar{b}$.

Taking into account that for $|v_{bl}/v_{bl}^0 - 1| = |\delta - 1| \ll 1$,

$$\left(\frac{v_{bl}}{v_{bl}^0}\right)^2 - 1 \equiv (\delta - 1)^2 + 2(\delta - 1) \approx 2(\delta - 1),$$
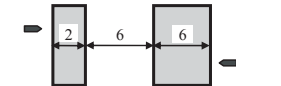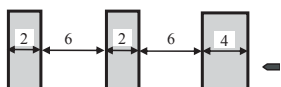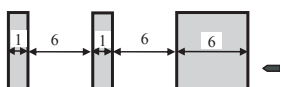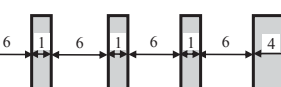
we can rewrite (8) as follows:

$$e_{abs} - e_{abs}^0 \approx k(1 - \delta), \quad k = \frac{2}{q_*}\left(\frac{v_{bl}^0}{v_{imp}}\right)^2.$$

Since $v_{bl}^0 \leq v_{imp}$ and $q_* > 1$, then $k < 2$. Moreover, when α increases, the magnitude of q_* increases (see Figure 11), which compensates for a certain increase in δ discussed above. Therefore, the model predicts insignificant change of the absorbed energy in the whole practical range of variation of parameter α. Since the predicted magnitude of the change lies within the range of experimental error and model uncertainty, it is conceivable that experimentally observed v_{bl}^0 and v_{bl} will be different by several percent, and that there will be situations when $v_{bl}^0 < v_{bl}$, and when $v_{bl}^0 > v_{bl}$. The latter conclusion is supported by the data in Table 2, compiled from experimental results of Almohandes et al. [1996]. The table shows, for each shield configuration and each impact velocity, the percent impact energy absorbed by a spaced shield, and the corresponding percentage for a shield consisting of the

#	Structure (sizes in mm)	$v_{imp} =$ 706.0	754.5	775.4	804.5	826.2
1	▶ [4] 6 [4]	52.2	49.6	47.4	35.9	34.5
		52.5	50.0	46.5	36.1	34.8
2	▶ [2] 6 [6]	54.9	52.2	50.7	42.1	37.4
		54.3	53.6	49.8	42.7	36.5
3	▶ [6] 6 [2]	54.4	52.0	49.7	41.5	35.1
		55.5	52.1	51.7	41.4	37.3

Table 2. Relative energy absorption, in percent. For each structure, the top row indicates the percent absorption for the spaced shield, and the second, for a shield with plates in contact. Based on experimental results from [Almohandes et al. 1996].

#	Structure (sizes in mm)	$v_{imp} =$ 706.0	754.5	775.4	804.5	826.2
1	▶ [2] 6 [6] ◀	54.9	52.2	50.7	42.1	37.4
		54.4	52.0	49.7	41.5	35.1
2	▶ [2] 6 [2] 6 [4] ◀	51.3	49.0	44.0	33.8	32.4
		50.9	48.4	44.3	36.0	32.7
3	▶ [1] 6 [1] 6 [6] ◀	53.0	49.1	46.8	35.9	34.7
		54.7	54.3	52.1	37.2	36.4
4	▶ [1] 6 [1] 6 [1] 6 [1] 6 [4] ◀	44.2	42.3	38.4	34.0	31.5
		46.0	46.3	37.5	32.6	33.2

Table 3. Comparison of the relative energy absorption, in percent, caused by the interchange of the order of the plates. For each structure the first row describes entry from the left and the second, entry from the right. Based on experimental results from [Almohandes et al. 1996].

same plates in the same order but without air gaps. Table 3 shows that reversing the order of the plates in a spaced shield does not cause significant changes in the

absorbed energy. In the table one also observes cases where $v_{bl}^0 < v_{bl}$ and where $v_{bl}^0 > v_{bl}$.

5. Concluding remarks

Using an approximate model that takes into account the plastic deformation of the shield during perforation, we analyzed the effect of air gaps upon ballistic properties of the shield against nonconical rigid impactors. It was found that nonconical shape of the impactor causes insignificant change (in the range of several percent) of BLV and energy absorbed by a shield. The obtained results are supported by available experimental data.

References

[Almohandes et al. 1996] A. A. Almohandes, M. S. Abdel-Kader, and A. M. Eleiche, "Experimental investigation of the ballistic resistance of steel-fiberglass reinforced polyester laminated plates", *Compos. B: Eng.* **27**:5 (1996), 447–458.

[Backman and Goldsmith 1978] M. E. Backman and W. Goldsmith, "The mechanics of penetration of projectiles into targets", *Int. J. Eng. Sci.* **16**:1 (1978), 1–99.

[Ben-Dor et al. 1998a] G. Ben-Dor, A. Dubinsky, and T. Elperin, "Effect of air gaps on ballistic resistance of targets for conical impactors", *Theor. Appl. Fract. Mech.* **30**:3 (1998), 243–249.

[Ben-Dor et al. 1998b] G. Ben-Dor, A. Dubinsky, and T. Elperin, "On the ballistic resistance of multi-layered targets with air gaps", *Int. J. Solids Struct.* **35**:23 (1998), 3097–3103.

[Ben-Dor et al. 1999] G. Ben-Dor, A. Dubinsky, and T. Elperin, "Effect of air gap and order of plates on ballistic resistance of two layered armor", *Theor. Appl. Fract. Mech.* **31**:3 (1999), 233–241.

[Ben-Dor et al. 2005] G. Ben-Dor, A. Dubinsky, and T. Elperin, "Ballistic impact: recent advances in analytical modeling of plate penetration dynamics, a review", *Appl. Mech. Rev.* **58**:6 (2005), 355–371.

[Corran et al. 1983a] R. S. J. Corran, C. Ruiz, and P. J. Shadbolt, "On the design of containment shields", *Comput. Struct.* **16**:1-4 (1983), 563–572.

[Corran et al. 1983b] R. S. J. Corran, P. J. Shadbolt, and C. Ruiz, "Impact loading of plates-an experimental investigation", *Int. J. Impact Eng.* **1**:1 (1983), 3–22.

[Elek et al. 2005] P. Elek, S. Jaramaz, and D. Mickovic, "Modeling of perforation of plates and multi-layered metallic targets", *Int. J. Solids Struct.* **42**:3–4 (2005), 1209–1224.

[Gupta and Madhu 1997] N. K. Gupta and V. Madhu, "An experimental study of normal and oblique impact of hard-core projectile on single and layered plates", *Int. J. Impact Eng.* **19**:5–6 (1997), 395–414.

[Honda et al. 1930] K. Honda, G. Takamae, and T. Watanabe, "On the measurement of the resistance of shield plates to penetration by a rifle bullet", *Tohoku Imperial Univ. 1st Ser.* **19** (1930), 703–725.

[Hurlich 1950] A. Hurlich, "Spaced armor", Report WAL-710/930-1, Watertown Arsenal Lab., MA, 1950.

[Kamke 1959] E. Kamke, *Differentialgleichungen: Lösungsmethoden und Lösungen — gewönliche Differentialgleichungen*, Geest und Portig, Leipzig, 1959.

[Liang et al. 2005] C.-C. Liang, M.-F. Yang, P.-W. Wu, and T.-L. Teng, "Resistant performance of perforation of multi-layered targets using an estimation procedure with marine application", *Ocean Eng.* **32**:3–4 (2005), 441–468.

[Marom and Bodner 1979] I. Marom and S. R. Bodner, "Projectile perforation of multi-layered beams", *Int. J. Mech. Sci.* **21**:8 (1979), 489–504.

[Radin and Goldsmith 1988] J. Radin and W. Goldsmith, "Normal projectile penetration and perforation of layered targets", *Int. J. Impact Eng.* **7**:2 (1988), 229–259.

[Recht 1990] R. F. Recht, "High velocity impact dynamics: analytical modeling of plate penetration dynamics", pp. 443–513 in *High velocity impact dynamics*, edited by J. A. Zukas, Wiley, New York, 1990.

[Vitman and Ioffe 1948] F. F. Vitman and B. S. Ioffe, "A simple method of determining the dynamical hardness of metals using a double cone", *Zavodskaja Laboratorija* **14**:6 (1948), 727–732. In Russian.

[Vitman and Stepanov 1959] F. F. Vitman and V. A. Stepanov, "Effect of the strain rate on the resistance of metals to deformation at impact velocities of 100–1000 m/s", pp. 207–221 in *Nekotoryie problemy prochnosti tvyordogo tela*, USSR Acad. of Sci., Moscow and Leningrad, 1959. In Russian.

[Zukas 1996] J. A. Zukas, "Effect of lamination and spacing on finite thickness plate perforation", pp. 103–115 in *Structures under shock and impact, IV*, edited by C. A. B. N. Jones and J. A. Watson, Comput. Mech. Publ., Southampton, 1996.

GABI BEN-DOR: bendorg@bgu.ac.il
Department of Mechanical Engineering, Ben-Gurion University of the Negev, POB 653, Beer-Sheva 84105, Israel

ANATOLY DUBINSKY: dubin@bgu.ac.il
Department of Mechanical Engineering, Ben-Gurion University of the Negev, POB 653, Beer-Sheva 84105, Israel

TOV ELPERIN: elperin@bgu.ac.il
Department of Mechanical Engineering, Ben-Gurion University of the Negev, POB 653, Beer-Sheva 84105, Israel

OBSERVATIONS OF ANISOTROPY EVOLUTION
AND IDENTIFICATION OF PLASTIC SPIN PARAMETERS
BY UNIAXIAL TENSILE TESTS

YANGWOOK CHOI, MARK E. WALTER, JUNE K. LEE AND CHUNG-SOUK HAN

Micromechanical effects such as the development of crystallographic texture and of dislocation structures lead to evolution of material anisotropy during plastic deformation. The anisotropy of sheet metals is commonly quantified by its R-values. The R-value is defined as the ratio of the transverse strain to the thickness strain at a certain longitudinal strain, and it changes if the anisotropy changes. Conventional hardening models do not account for the evolution of anisotropy along an arbitrary orientation. Therefore, although R-values are measured from experiments, predictions of hardening behavior based on R-values using conventional hardening models do not reproduce the experiments for arbitrary orientation. The R-value evolution for large strains can be observed in simple uniaxial tension tests by measuring the transverse and longitudinal strains continuously up to large strains. A digital image correlation (DIC) method is introduced as superior to strain gages for measuring large strains. To model the experimental response, a rotational-isotropic-kinematic (RIK) hardening model is investigated. Because of this model's ability to represent the rotational evolution of the anisotropy, it can predict the hardening behavior for non-RD and non-TD directions. Methods to identify the plastic spin and kinematic hardening parameters are also discussed.

1. Introduction

In sheet metal forming and springback processes, incorporating evolution of material anisotropy is important in order to obtain accurate predictions. Cold rolled sheet metals are known to have initial anisotropy, incurred during the rolling process. The anisotropy causes different flow stresses with respect to the orientation that is measured from the rolling direction (RD). The different flow stresses have significant effects on the forming process: different earing of edges, punch force-displacements, and springback, etc. To address the material anisotropy, anisotropic yield functions such as those of [Hill 1948; 1990] and [Barlat et al. 1991; 1997; 2003] have been developed. Hill's quadratic yield function [Hill 1948] is known to

be appropriate for BCC materials such as mild steel, while Barlat's yield function [Barlat et al. 1991; 1997; 2003] is good for FCC materials such as aluminum alloys. However, these yield functions do not account for the *evolution* of material anisotropy during the plastic deformation.

The evolution of anisotropy has been observed by both micromechanical and macromechanical experiments [Boehler and Koss 1991; Bunge and Nielsen 1997; Kim and Yin 1997; Peeters et al. 2001]. At a micromechanics level, the anisotropy evolution is generally considered to be the result of the crystallographic texture development, which changes the preferred orientation of grain aggregates, and the development of substructures [Peeters et al. 2001]. Using micromechanical approaches, the evolution of anisotropy has been observed and modeled by many investigators [Agnew and Weertman 1998; Asaro 1983; Beaudoin et al. 1994; Kocks et al. 1998; Nakamachi and Doug 1997]. On the other hand, the anisotropy evolution has been observed and modeled by using macromechanical approaches that rely on phenomenological descriptions [Boehler and Koss 1991; Kim and Yin 1997; Kuroda 1997; Dafalias 2000; Han et al. 2002].

For sheet metals, R-values in several directions are used to represent the planar anisotropy and to compare the anisotropy between different materials. The R-value is defined as the ratio of transverse strain to thickness strain at a certain longitudinal strain. Since R-values are different for different longitudinal strains, R-values are usually provided along with the longitudinal strain at which the transverse strains were measured.

Even though the R-values are used to define the anisotropic relationship of the hardening curve between a certain direction and the RD, the computed hardening curve may not properly predict the experiment results for orientations not in the RD or TD (transverse direction) when conventional hardening models are used. The computed hardening curves in three directions are compared with the experimental results in Figure 1. The computed curves were generated by ABAQUS/Standard using Hill's quadratic yield function and an isotropic hardening model. The material is deep drawing quality (DDQ) mild steel which is used for the NUMISHEET 2002 benchmark problem [Yang et al. 2002]. The curves of the RD and TD correlate quite well with the experimental results. However, the computed 45° orientation curve overestimates the experiment when constant R-values are used: $R_0 = 2.64$, $R_{45} = 1.57$, and $R_{90} = 2.17$.

The use of constant R-values is effective for characterizing the flow stresses in RD and TD. However, it is not capable of modeling the flow stress in the 45° orientation. It is assumed that the unexpected higher estimation of the hardening at 45° in Figure 1 can be explained by the evolution of anisotropy and the subsequent R-value evolution for the orientation. In addition, Hill's yield criterion [1948] has
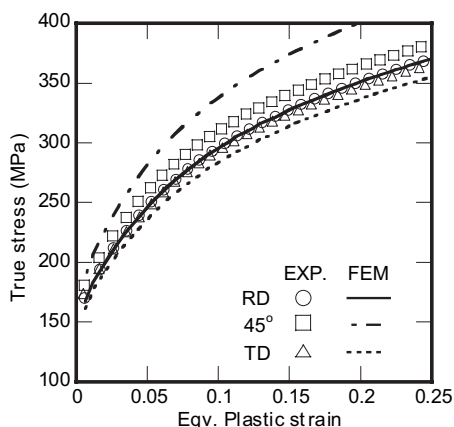
Figure 1. Flow stresses in three orientations, compared to measured data from NUMISHEET 2002 [Yang et al. 2002]. ABAQUS/ Standard was used to compute the curves by using constant *R*-values. The flow stress at 45° overestimates the experiments, while the others correlate with experiments relatively well.

major drawbacks in predicting flow stress curves that depend on the loading orientation. In other words, the dependence of the yield stress on orientation is poorly predicted by the conventional theory [Banabic et al. 2000]. These shortcomings are usually interpreted to be the result of not incorporating the evolution of anisotropy during the plastic deformation.

Understanding the evolution of anisotropy is critical for predicting material behavior in large strain deformation. It is also essential to be able to represent the anisotropy evolution in multiaxial and multipath loading. To model the multiaxial and multipath elastoplastic deformation of planar anisotropic materials, an RIK (rotational-isotropic-kinematic) hardening model [Choi et al. 2006a] is proposed to incorporate rotation of the yield surface with the isotropic combined kinematic hardening model. To measure anisotropy evolution, tensile tests for RD, TD, and 45° orientations were performed.

A simple method to measure the rotation of the symmetry axes is suggested with the following assumptions: the anisotropy of the yield surface shape can be described by the rotation of the orthogonal symmetry axes, and the symmetry axes will not rotate if the loading is along the symmetry axes. The method requires tensile experiments with continuous measurement of *R*-values for specimens cut at 45° to the RD. According to the theory, the symmetry axes do not rotate for the deformation along the RD or TD. To measure the strains continuously up to large strains, we used the DIC (digital image correlation) method developed by

Sutton et al. [1983]. DIC is a technique that compares digital images of a specimen surface before and after deformation to deduce its two-dimensional surface displacement field and strains; see, for example, [Vendroux and Knauss 1998]. In spite of its low accuracy at small strains, DIC can be used at large strains and is simple to implement. Although large strains were measured using digital images of grid marks on uniaxial sheet metal specimens by Rao and Mohan [2001], the grid deformation was not used to determine R-values. The authors are not aware of any other investigations that use the DIC method to determine R-values.

Here we present a DIC-based experimental method for determining anisotropy evolution and demonstrates how the material parameters for the RIK hardening model can be determined from the simple tensile experiments.

2. Definition of the R-value

The definition of the R-value for a planar anisotropic material is the ratio of transverse strain (ε_w) to thickness strain (ε_t) and is therefore given by Hill [1950] as

$$R = \frac{\varepsilon_w}{\varepsilon_t}. \tag{1}$$

Because the specimen thickness is generally small compared to the other dimensions, an accurate measurement of the strain in the thickness direction for sheet metals is difficult. By applying the constant volume condition for plastic deformation, Equation (1) can be reformulated using the longitudinal strain (ε_l) as

$$R = \frac{\varepsilon_w}{\varepsilon_t} = -\frac{\varepsilon_w/\varepsilon_l}{(1 + \varepsilon_w/\varepsilon_l)},$$

which is widely used to measure R-values. With these formulas the R-values are determined by fitting the slope ($\varepsilon_w/\varepsilon_l$) of the $\varepsilon_w - \varepsilon_l$ (transverse strain versus longitudinal strain) curve of experiments. An example for DDQ material used in NUMISHEET 2002 benchmark problem [Yang et al. 2002] is shown in Figure 2 (left). For DDQ, the TD and RD lines are very relatively straight. It should be noted that although the 45° response appears to be linear, it has measurable nonlinearity. Given the linear and nonlinear responses, the anisotropy in the RD and TD expressed through R-values remains constant, as seen in Figure 2 (right), while the R-value evolves for tensile tests in the 45° orientation.

This is consistent with the assumptions stated in the introduction that no rotation of the anisotropy axes will occur for the loading along the RD or TD. As shown in Figure 3, Stout and Kocks have obtained similar R-value evolution on a cube of copper with rolling texture [Kocks et al. 1998].

It is evident that R_{45} evolves with the strain. However, the current industry practice is to measure an R-value at a certain longitudinal strain (15–20%) and
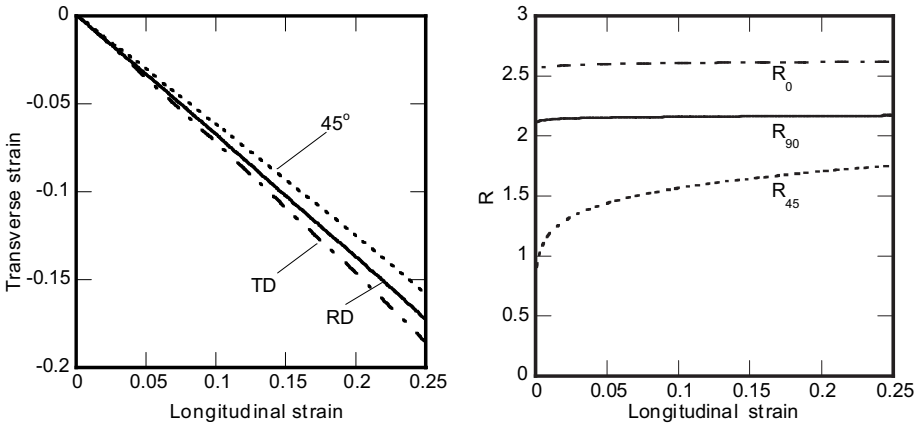
Figure 2. Transverse strain versus longitudinal strain of DDQ (left). Fitted R-values for each orientation (right). The R_{45} value is evolving with the strain and documents the evolution of anisotropy in the 45° orientation.
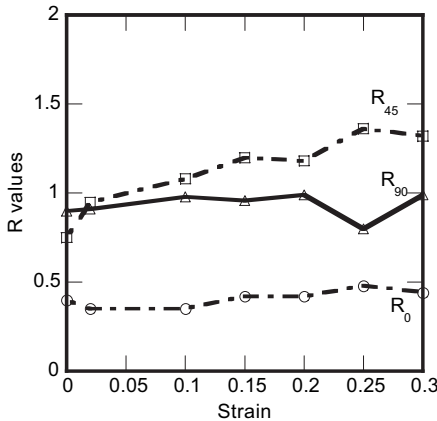

Figure 3. R-value measurements for rolled copper in compression experiments. Particularly R_{45} shows significant changes. >From [Kocks et al. 1998].

use it as a constant. A suitable modeling method for the anisotropy evolution is addressed by Choi et al. [2006a].

3. Experimental setup

Experiments were performed on an MTS Model 810 servohydraulic load frame with an Instron 8500 digital controller. Figure 4 shows the experimental setup. The controller was operated in displacement control mode. Load was measured
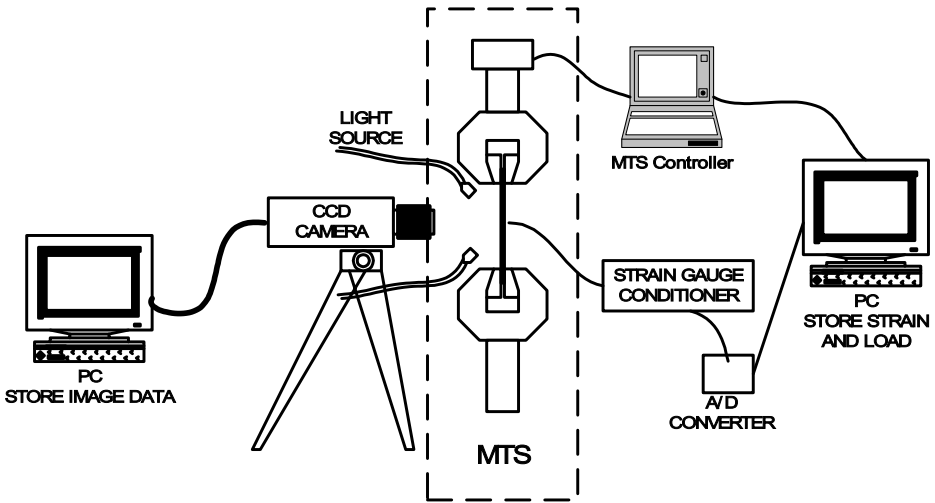
Figure 4. Schematic diagram of the experimental setup.

from a 25 MT load cell and was subsequently converted to true stress. For some experiments, strain gages were used to verify the accuracy of the DIC strain measurement.

Large strains were measured using DIC, the required images being acquired by a Pulnix TMC-9701 digital CCD camera with a resolution of 768×472 pixels at a continuous rate of 4 frames per second. To minimize intensity change during the experiment, a Schott KL1500 directed light source was used to illuminate the specimen.

Specimens used in the experiments were cut from sheet metal at 0°, 45°, and 90° to the RD direction. Specimens were 1 inch wide and 0.039 inch (1 mm) thick. Since the view of the CCD camera is fixed and only the lower grip of the load frame moves, the region of interest has significant downward displacement. Therefore, the size of the specimen must be carefully chosen to ensure that a significant portion of the originally undeformed image remains inside the camera's view for all subsequent deformation. For this reason, 1.5 inch gage length specimens were used. FEM simulations were performed to confirm that the region where strains were measured was far enough away from the grips to avoid grip effects. The results of these FEM simulations are shown in Figure 5(a). By comparing transverse strain versus longitudinal strain curves for four locations along the specimen centerline — see Figure 5(a) — no significant differences were observed in the curves shown in Figure 5(b). Hence, end effects do not influence the region of interest.

The DIC method works best when the starting image contains a random pattern that can carry the specimen deformation exactly. To create such a random pattern
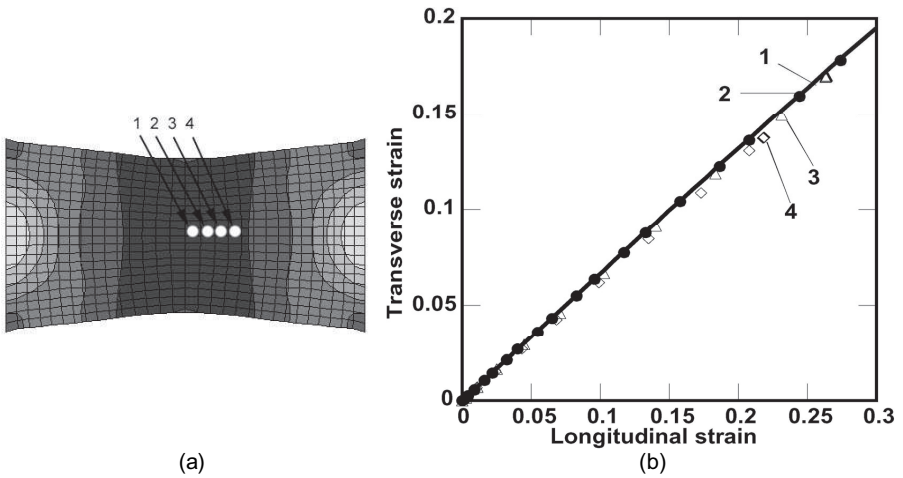
Figure 5. FEM validation of specimen dimensions to avoid grip effects: (a) strain measurement locations; (b) computed strains.

on the current specimens, spray paint was used. First white paint provided a bright background and then black spray paint gave a high contrast speckle pattern. The black paint was applied by spraying parallel to the specimen surface and letting paint droplets fall randomly on the specimen. Examples of the random pattern before and after deformation are shown in Figure 6.

The DIC program was set to measure incremental transverse and longitudinal strain between subsequent images. For random patterns such as the one shown in Figure 6, the best subset size was 100×100 pixels, corresponding to 4.2×4.2 mm.
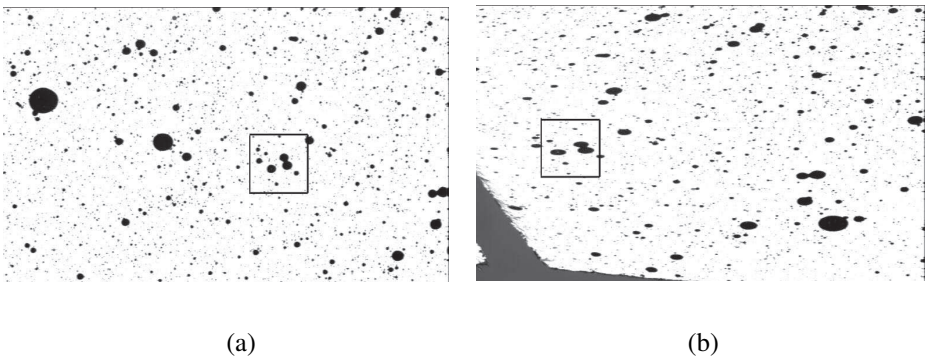
Figure 6. The random pattern sprayed on the specimen and captured by the CCD camera (a) before and (b) after deformation. The small rectangles highlight a 100×100 pixel subset of the image that has been deformed.
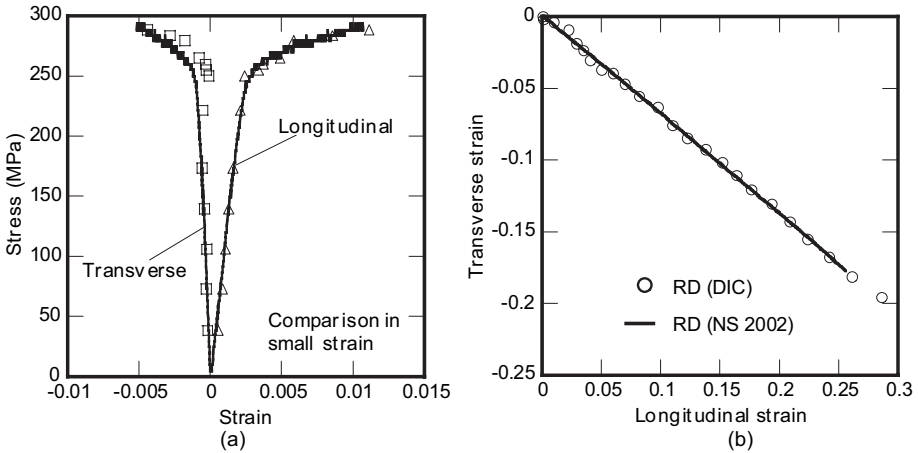
Figure 7. (a) Comparison of stress-strain measurement by DIC (symbols) and strain gauge (lines) for tensile test on RD specimens; (b) comparison of strain measurement by DIC and NUMISHEET 2002 data.

The displacement was measured at 25 equally spaced points in the subset. Since the subset used for correlation at the beginning of the loading has undergone significant translation by the end the deformation, it was necessary to devise an algorithm to track the approximate position of each subset used for correlation. Total strain was obtained by adding the incremental strains. In order to reduce measurement error, the correlation was performed on all 25 points in the subset, and then strain values that were beyond one standard deviation were eliminated. The average strain was then determined from the remaining strain values. The original DIC program was written in FORTRAN and C by Vendroux and Knauss [1998]. The modifications to track the overall displacement of the subsets and to treat the strain measurements statistically were done by the authors. The accuracy of the DIC method for strain measurements is demonstrated by comparison with a strain gage for smaller strains and with the data provided by NUMISHEET 2002 [Yang et al. 2002] for larger strains. The results are shown in Figure 7. The strain measurement with DIC has poor resolution at lower strains but shows quite good correlation with the NUMISHEET 2002 data at higher strains. Usually it is difficult to measure large strains with a strain gage, but DIC can measure 25–30% strains with relative ease. The limitation at strains greater than 30% is related to the loss of adherence of the applied speckle pattern to the specimen surface.

4. Hardening model

The procedure for modeling isotropic, kinematic and rotational hardening is rather complex. Since this article is mainly concerned with experimental findings, the theory is only briefly introduced in this section. For a more detailed description of the applied RIK (rotational-isotropic-kinematic) hardening model the interested reader is referred to [Choi et al. 2006a].

For the isotropic hardening, we chose a description based on [Chaboche 1989], modified to improve the agreement between experiments and predictions for tensile stress-strain curves [Chun et al. 2002]. The magnitude of the yield stress corresponding to isotropic hardening is described as

$$\bar{\sigma}_y = \bar{\sigma}_0 + K(1 - e^{Ns}) - \frac{c_1}{b_1}(1 - e^{b_1 s}) - c_2 s,$$

where s is the effective plastic strain, $\bar{\sigma}_0$ is the initial yield stress, K and N are fit parameters of the hardening curve in monotonic loading in RD, and c_1, c_2, b_1 are parameters associated with the kinematic hardening and coupling the isotropic to the kinematic hardening.

A related kinematic hardening model with permanent softening, similar to that of [Armstrong and O. 1966], is defined with different backstress terms whose evolution equations are given by

$$\boldsymbol{\alpha}_1^\nabla = \frac{c_1}{\beta}(\boldsymbol{\tau} - \boldsymbol{\alpha})\dot{s} - b_1 \dot{s} \boldsymbol{\alpha}_1, \quad \boldsymbol{\alpha}_2^\nabla = k\frac{c_2}{\beta}(\boldsymbol{\tau} - \boldsymbol{\alpha})\dot{s}, \qquad (2)$$

where $\boldsymbol{\tau}$ denotes the Kirchhoff stress, $\boldsymbol{\alpha} = \boldsymbol{\alpha}_1 + \boldsymbol{\alpha}_2$ is total backstress of the backstress components $\boldsymbol{\alpha}_1$, $\boldsymbol{\alpha}_2$, and the variable β is defined by

$$\beta = \sqrt{\frac{2}{3}} \left\| \frac{\partial \phi}{\partial \boldsymbol{\tau}} \right\|.$$

The Oldroyd rate $(\cdot)^\nabla$ is applied for the stress and backstress evolution equations as suggested in [Haupt and Tsakmakis 1986] and [Han et al. 2003].

The rotational hardening is expressed by the rotation of the symmetry axes of anisotropy, e_i^ϕ, defining the orientation of the anisotropic yield function ϕ. The rotation of the symmetry axes is described by

$$\dot{e}_i^\phi = \boldsymbol{\theta}^\phi e_i^\phi,$$

where the constitutive spin $\boldsymbol{\theta}^\phi$ is implicitly determined by the material spin $\boldsymbol{w}$ describing rigid body rotations and the plastic spin $\boldsymbol{\omega}_p$, in the following manner:

$$\boldsymbol{\theta}^\phi = \boldsymbol{w} - \boldsymbol{\omega}_p^\phi.$$

With the introduction of the plastic spin into the model, the anisotropy axes are allowed to rotate relative to the rigid body rotations. Several expressions have been suggested in the literature to describe plastic spin, for example, [Dafalias 1993; 2000; 2001]. Here we consider the expressions suggested in [Han et al. 2002], which describe the experimentally determined rotations of [Kim and Yin 1997] fairly well with

$$\boldsymbol{\omega}_p^\phi = \frac{a}{\bar{\sigma}_y} \tan(\vartheta)(\boldsymbol{\tau}\dot{\boldsymbol{\varepsilon}}_p - \dot{\boldsymbol{\varepsilon}}_p\boldsymbol{\tau}), \tag{3}$$

where a is the only material parameter needed to describe the plastic spin. This parameter can be identified with the evolution of R_{45}, the R-value for specimens at $45°$ to the RD.

Assuming that the anisotropy axes remain orthogonal during deformation, all these components can be incorporated into the yield function

$$\Phi = \phi - \bar{\sigma}_y^2,$$

where $\phi = (\boldsymbol{\tau} - \boldsymbol{\alpha}) \cdot \boldsymbol{K}_\phi(\boldsymbol{\tau} - \boldsymbol{\alpha})$ characterizes a Hill-type yield surface with the fourth order tensor $\boldsymbol{K}_\phi$ reflecting the anisotropy and the total backstress is obtained as $\boldsymbol{\alpha} = \boldsymbol{\alpha}_1 + \boldsymbol{\alpha}_2$. For plane stress problems the initial matrix form of the fourth order tensor $\boldsymbol{K}_\phi$ can be given in vector notation as

$$\boldsymbol{P}^0 = \begin{bmatrix} 1 & -\beta_{12} & 0 \\ -\beta_{12} & \beta_{22} & 0 \\ 0 & 0 & \beta_{66} \end{bmatrix}, \qquad \boldsymbol{\tau} = \begin{bmatrix} \tau_{11} \\ \tau_{22} \\ \tau_{12} \end{bmatrix}, \qquad \boldsymbol{\alpha} = \begin{bmatrix} \alpha_{11} \\ \alpha_{22} \\ \alpha_{12} \end{bmatrix},$$

where the components are related to R-values through

$$\beta_{12} = \frac{R_0}{1 + R_0}, \qquad \beta_{22} = \frac{R_0(1 + R_{90})}{R_{90}(1 + R_0)}, \qquad \beta_{66} = \frac{(R_0 + R_{90})(1 + 2R_{45})}{R_{90}(1 + R_0)};$$

see [Valliappan et al. 1976].

The mechanical tests and texture analysis illustrated in [Boehler and Koss 1991] indicate that the rotation of the anisotropy axes can be related to texture development and rotation of grains. The grain rotation in turn will also affect the backstresses and kinematic hardening as back stresses are usually related to dislocation substructures. The mechanisms of backstresses are however not very well understood. One difficulty may be seen in the different length scales backstresses are associated with; for example, Barton et al. [1999] pointed out that Bauschinger effects can in part be explained by texture development but it is also well known that grain boundaries, dislocation cells [Hughes et al. 2003] and dislocation structures within grains and cells [Feaugas 1999] can have significant influence on the kinematic hardening.

Here the rotation of anisotropy axes is assumed to have an influence on the backstress and an interrelation is suggested here by assuming that the function k in equation (2) depends on the loading directions and anisotropy axes; specifically, k is suggested to have the form

$$k(\varphi, \vartheta) = \begin{cases} 1 + \kappa\vartheta & \text{for initial loading,} \quad \varphi \leq \pm 90°, \\ \kappa\vartheta & \text{for reversal loading,} \quad \varphi > \pm 90°, \end{cases} \tag{4}$$

where ϑ is the angle between the direction of the symmetry axes and direction of straining as shown in Figure 8 and φ is the angle between the directions of the previous and current loadings. The angle $\vartheta \in (0, \pi/4)$ is defined by

$$\vartheta = \min\left(\cos^{-1} \frac{e_i^{\phi} \cdot n}{|e_i^{\phi}||n|}\right),$$

where e_i^{ϕ} is the direction vector of the anisotropy axes and n is one of the eigenvectors of the strain tensor closer to e_i^{ϕ}, and angle φ is defined by

$$\varphi = \cos^{-1} \frac{\dot{\varepsilon}_p^* \cdot \dot{\varepsilon}_p}{|\dot{\varepsilon}_p^*||\dot{\varepsilon}_p|},$$

where $\dot{\varepsilon}_p^*$ is the plastic strain increment of the previous load step and $\dot{\varepsilon}_p$ is the plastic strain increment of the current loading step. Other approaches where the interaction between kinematic and rotational hardening has been considered are found in [Dafalias 1993; 1998; Tsakmakis 2004].
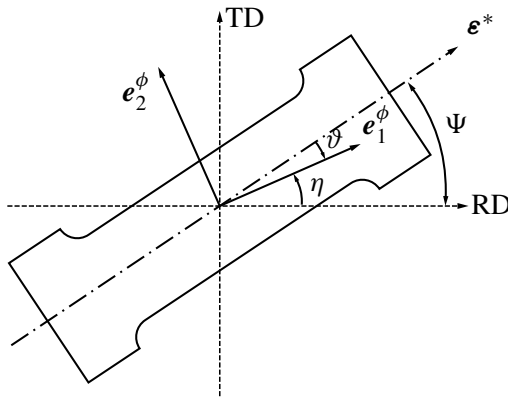
Figure 8. Definition of the angles used in formulation: Ψ is the initial orientation to the RD, η the rotation angle of the symmetry axes of anisotropy, and $\vartheta = \Psi - \eta$ the difference between the major strain direction and the symmetry axes.

The nonlinear kinematic hardening parameters c_1, b_1, and c_2 in equations (2) are identified by using an inverse method, which optimizes the parameters using the results of three-point bend experiments [Zhao and Lee 1999; 2000]. If rotational hardening is ignored, a and κ can be set to zero. Then the RIK hardening model reduces to Chun's ANK model [Chun et al. 2002]. The hardening model reduces to Chaboche's model [1989] if c_2 is set to zero, and to a regular isotropic hardening model if c_1 and b_1 are set to zero. It should be noted that the description of k in Equation (4) could theoretically result in a discontinuous response. In numerical tests however such discontinuities have never occurred — neither in implicit nor explicit finite element simulations [Choi et al. 2006a; 2006b].

5. Experimental results and material parameters

The experimental results, the determination of the plastic spin parameters, and the corresponding predictions are discussed in this section. All experimental results presented in this paper are averages over three different tensile tests. The investigated materials are mild steels for deep drawing quality (DDQ) and drawing quality (DQ) and a high strength steel (HSS). The results for DDQ and DQ are shown in Figure 9. The transverse strain values are fit to the longitudinal strain by $\varepsilon_w = \{a \ln(\varepsilon_l) + b\}\varepsilon_l$. The procedure to determine the plastic spin parameters will be described for the DDQ steel material in the following. The determination of the plastic spin parameters involves the results of experiments described in the previous sections or, alternatively, experimental data given in the proceedings of
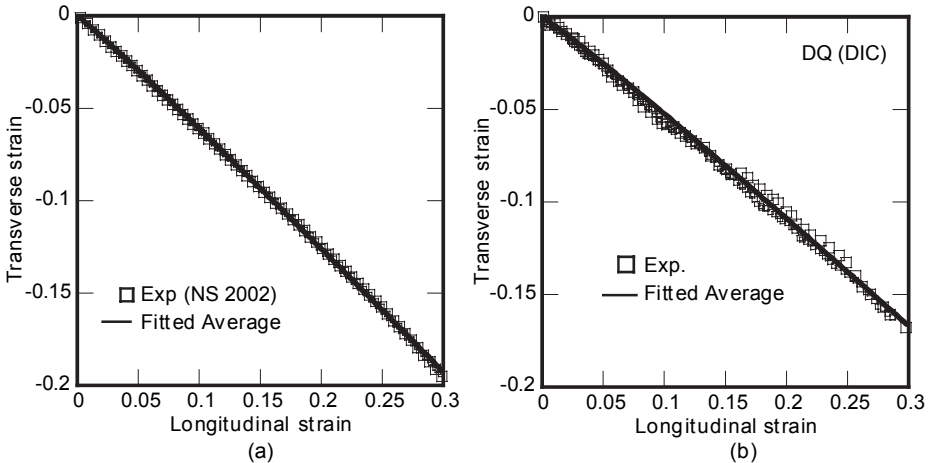
Figure 9. Average fitted curves of the transverse and longitudinal strain measurement for 45° orientation tensile tests: (a) DDQ; (b) DQ.

Material	Elastic	Anisotropy	Isotropic	Kinematic	Rotational
DDQ (Exp.)	$E = 210\,\text{GPa}$ $v = 0.3$	$R_0 = 2.137$ $R_{45} = 0.93$ $R_{90} = 1.508$	$\sigma_0 = 152.22\,\text{MPa}$ $K = 222.01\,\text{MPa}$ $N = -7.87$	$c_1 = 3.0\,\text{GPa}$ $b_1 = 300$ $c_2 = 70.0\,\text{MPa}$	$a = -155$ $\kappa = 50.0$
DDQ (NS2002)	$E = 210\,\text{GPa}$ $v = 0.3$	$R_0 = 2.722$ $R_{45} = 1.474$ $R_{90} = 2.169$	$\sigma_0 = 152.0\,\text{MPa}$ $K = 235.81\,\text{MPa}$ $N = 9.23$	$c_1 = 3.0\,\text{GPa}$ $b_1 = 300$ $c_2 = 70.0\,\text{MPa}$	$a = -57$ $\kappa = 1.8$
DQ	$E = 180\,\text{GPa}$ $v = 0.3$	$R_0 = 1.60$ $R_{45} = 1.010$ $R_{90} = 1.46$	$\sigma_0 = 198.0\,\text{MPa}$ $K = 242.47\,\text{MPa}$ $N = 9.95$	$c_1 = 3.3\,\text{GPa}$ $b_1 = 220$ $c_2 = 103\,\text{MPa}$	$a = -100$ $\kappa = 1.5$
HSS	$E = 210\,\text{GPa}$ $v = 0.3$	$R_0 = 0.832$ $R_{45} = 1.185$ $R_{90} = 0.560$	$\sigma_0 = 332.22\,\text{MPa}$ $K = 430.55\,\text{MPa}$ $N = 1.7$	$c_1 = 3.0\,\text{GPa}$ $b_1 = 150$ $c_2 = 80\,\text{MPa}$	$a = 0$ $\kappa = 0$

Table 1. Material properties of RIK hardening model for DDQ, DQ and HSS.

NUMISHEET 2002. The experimental results for the other materials are shown with fitted curves, and the parameters for all materials are summarized in Table 1.

DDQ (Experiment with DIC). The DDQ sheet metal tested in the experiments is the same material as in the NUMISHEET 2002 benchmark problems [Yang et al. 2002]. The measured transverse strain versus longitudinal strain curves for the RD, TD and 45° to RD direction are shown in Figure 10, together with the corresponding flow curves. Also shown in Figure 10(b) are simulation results of conventional models without rotational hardening. When the rotational evolution of the anisotropy is not considered and R_{45} is fixed at a value measured between 0% and 5% strain, the strain-strain relation for the 45° orientation prediction is higher than the experimental results. The estimated flow curve for the 45° orientation differs from the experimental results after approximately 5% strain. The ratio of stress evolution between 0% and 5% strain correlates well with the experiment. For larger deformations, better results can be obtained when an averaged R_{45} value is applied which is computed from the linear fit of the strain-strain relation from 0% to 20% strain.

Comparisons of experiments and predictions of a conventional hardening model using the averaged R_{45} value are given in Figure 11. Even though the predicted stress level is much closer to the experimental result than the prediction using initial anisotropy values in Figure 10(b), the rate of stress evolution does not correlate with the experimental results for the whole strain range.

With conventional hardening models, the anisotropy of the yield function remains unchanged during the deformation process. Therefore, the evolution of the anisotropy in 45° to the RD cannot be predicted with qualitatively or quantitatively satisfying accuracy. As can be seen in Figure 12(a), this evolution of R_{45} value is present for DDQ material and can be captured with the appropriate rotational hardening parameter. The corresponding flow stresses of tensile tests in the 45° to RD evolve with an abrupt change in stress as the symmetry axes rotate, obeying the
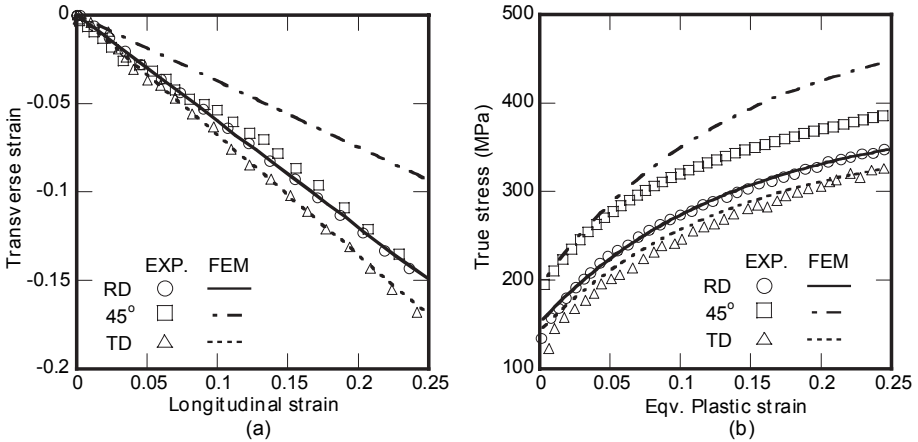
Figure 10. Comparing simulation and experimental results of the *initial R*-value for 45° orientation: (a) in transverse strain versus longitudinal strain; (b) in the flow stresses.
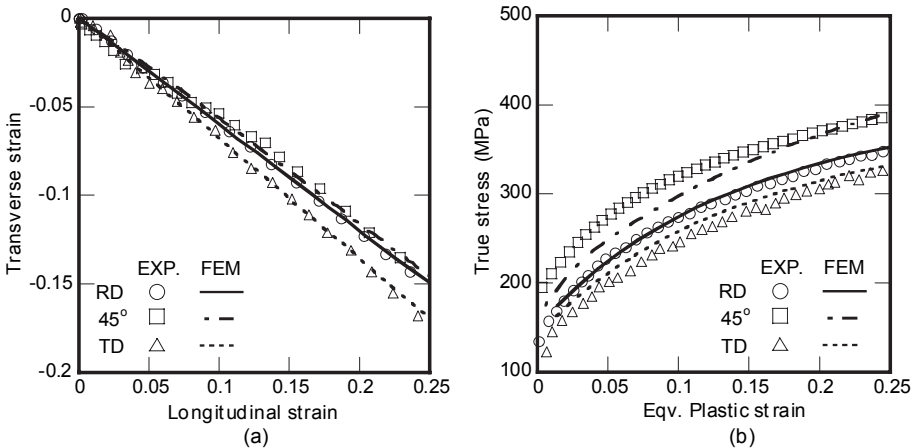
Figure 11. Comparing simulation and experimental results of the *averaged R*-value for 45° orientation: (a) Transverse strain versus longitudinal strain; (b) flow stresses in each orientation.
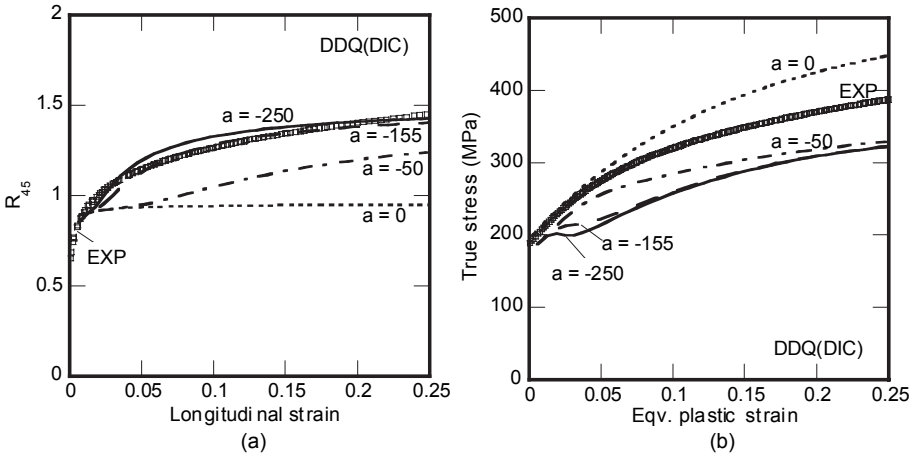
Figure 12. Sensitivity of plastic spin parameter: (a) R_{45} evolution; (b) flow stresses of 45° orientation. The softening due to the rotational hardening will be compensated for by backstress.

plastic spin description for the rotation of anisotropy axes. The corresponding flow stress results of experiments and predictions with various values for the parameter a are shown in Figure 12(b).
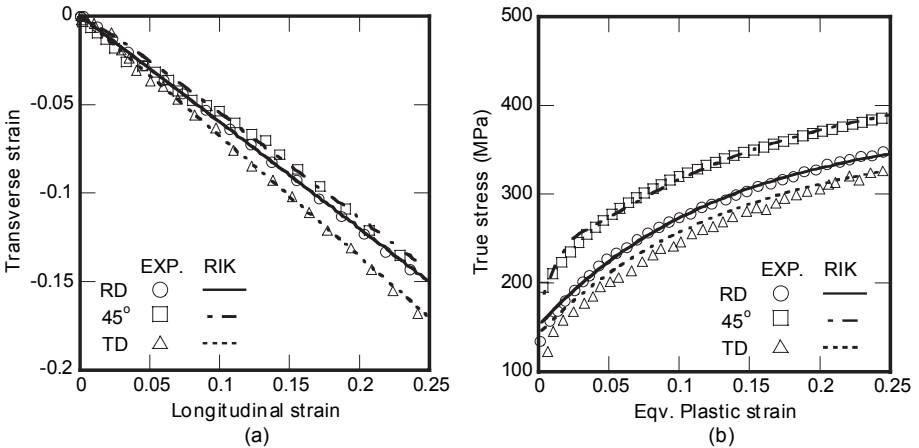
Figure 13. Simulation with the RIK hardening model and comparison with experimental results: (a) transverse strain versus longitudinal strain, (b) flow stresses in each orientation. For the 45° orientation the flow stress loss due to the rotational hardening is recovered by the backstress.

The diagrams related to stretch tests in RD and TD are omitted because there is no rotation in these two directions and the predictions in these directions are identical to those without rotational hardening. A plastic spin parameter of $a = -155$ is found to be the best fit for the experimental R_{45} data in Figure 12(a). The differences between experimental and computed flow stress curves in Figure 12(b) are then corrected by adding kinematic hardening through equation (4). As shown in Figure 13, kinematic hardening improves the fit to the experimental flow stress curves without changing the slope of the strain-strain curves.

DDQ (NUMISHEET 2002 Data). Figure 1 showed the flow stress versus strain curves of the DDQ material from NUMISHEET 2002, and Figure 2 the transverse strain versus longitudinal strain curves. As discussed on page 304, conventional
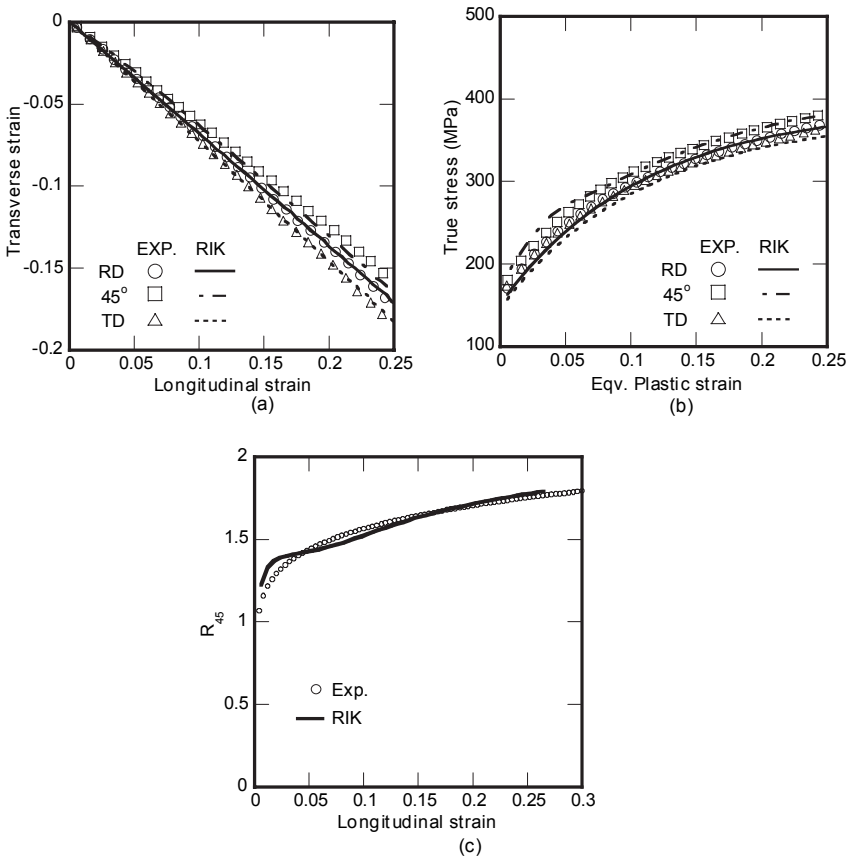
Figure 14. Results for DDQ mild steel using data from NU-MISHEET 2002: (a) transverse strain versus longitudinal strain; (b) flow stress curves in each orientation; (c) R_{45} evolution.

models without rotational hardening overestimate the flow stress by applying the average R-value for the 45° orientation stretch tests.

Figure 14 plots the predictions obtained with the RIK hardening model together with the experimental results. Parts (a) and (b) show there is good agreement in the case of transverse strain versus longitudinal strain and stress versus strain. By applying a plastic spin parameter, $a = -155$, the change of R_{45} is reasonably well represented, as shown in Figure 14(c).

DQ mild steel. In addition to DDQ mild steel, drawing quality (DQ) mild steel was also examined. The same procedure applied to DDQ mild steel was performed for DQ mild steel. The strain-strain, strain-stress at 45° to RD and strain-R_{45} results
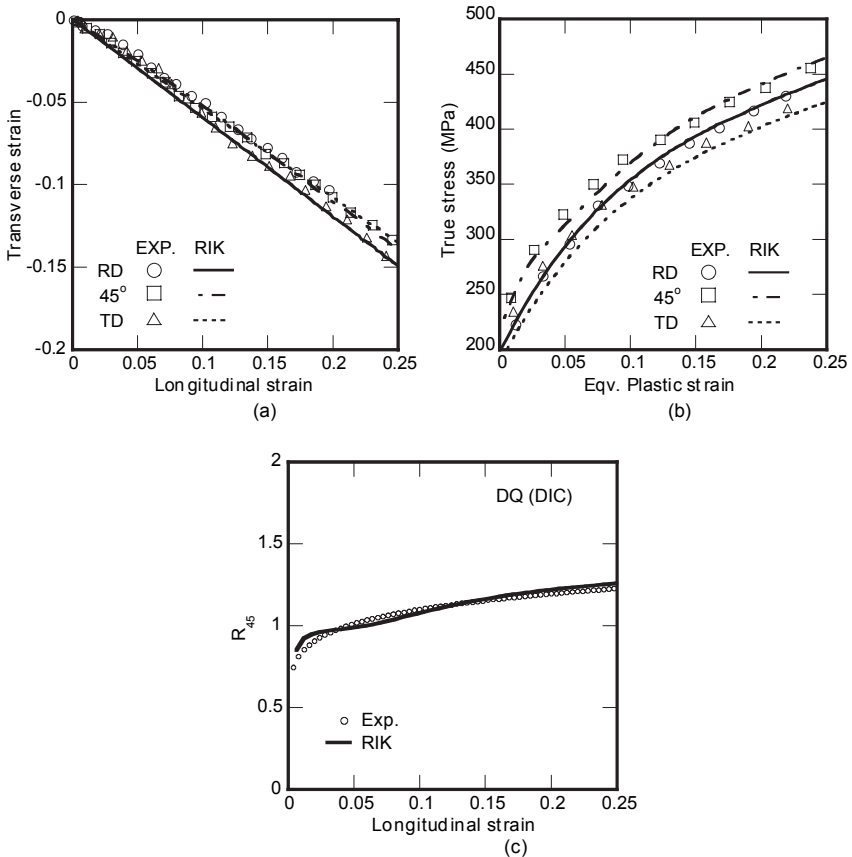
Figure 15. Results for DQ mild steel: (a) Transverse strain versus longitudinal strain; (b) Flow stresses of the RD, 45° and the TD; (c) R_{45} evolution.

from experiments and simulations are illustrated in Figure 15. Experiments and predictions show excellent agreement for a wide range of applied strains.

HSS. High strength steel (HSS) is a material widely used in the automotive industry and, like the mild steels discussed above, it is subject to large strain in the forming processes. This material was also tested for rotational hardening. In contrast to the mild steels described earlier, HSS does not show much rotation, since its R_{45}-values do not significantly change during deformation. Consequently, the spin parameter is set to zero for this material, which corresponds to anisotropy axes aligned with the material axes. Experimental and computational results are compared in Figure 16.

6. Discussion

By comparing the R-values of the different materials, it can be concluded that only the HSS material has a R_{45}-value higher than the R-values in RD and TD. Thus, the HSS material in the 45° orientation does not exhibit rotation of the anisotropy axes toward the straining direction as in the DDQ and DQ materials. Furthermore, for HSS, the flow curves are actually modeled with good agreement by applying initial R-values determined from the transverse strain versus longitudinal strain curves. Therefore, it may be assumed that the rotational hardening in HSS is very small. For these tests, the anisotropy of HSS hardly evolves with the deformation. This may be due to a pronounced texture that is not severely affected by the deformation of the tensile tests. To verify this interpretation, we recall the results of Boehler and Koss [1991] and Bunge and Nielsen [1997] who determined symmetry axes
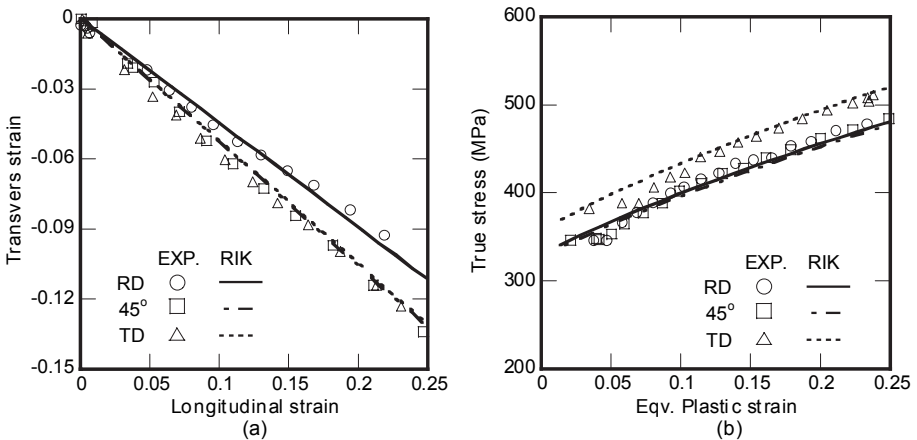
Figure 16. Results for HSS: (a) transverse strain versus longitudinal strain; (b) flow stresses of the RD, 45°, and the TD.

by texture analysis via orientation distribution function (ODF). In the first of these references, significant rotations of the symmetry axes for mild steel were observed in pole figures, whereas in the second, the rotation of the symmetry axes is quite small for cold rolled aluminum, where the texture is quite pronounced. The same may be true for HSS; this would have to be confirmed by monitoring the texture.

The R-values are used to compare the anisotropy of materials in a quantitative manner. The R-values measured in tensile experiments are useful for predicting the rotational evolution of anisotropy. On the other hand, $\bar{R}$ and ΔR, which are defined as

$$\bar{R} = \frac{R_0 + 2R_{45} + R_{90}}{4}, \quad \Delta R = \frac{R_0 - 2R_{45} + R_{90}}{2},$$

are also used to judge the formability ($\bar{R}$) and the earing pattern (ΔR) in applied problems like the circular cup drawing. Formability increases as $\bar{R}$ increases. For mild steel with rotational hardening, $\bar{R}$ changes with respect to R_{45} during deformation, in agreement with experimental results. Therefore, it may be concluded that models with rotational hardening can describe the formability better than non-rotational hardening models.

The plastic spin in sheet metals implicitly defines the planar rotation of the symmetry axes of anisotropy, and ΔR is known to be related to the planar anisotropy. Therefore, there must be a relationship between the parameters of the plastic spin and the planar anisotropy ΔR. The plastic spin and planar anisotropy values for the tested materials are shown in Table 1. As shown in Figure 17, there is evidence that the plastic spin may be linearly related to ΔR. Detailed experimental determination of the rotation of the symmetry axes of anisotropy would help to verify these results. In Figure 17 the comparison between the plastic spin parameter
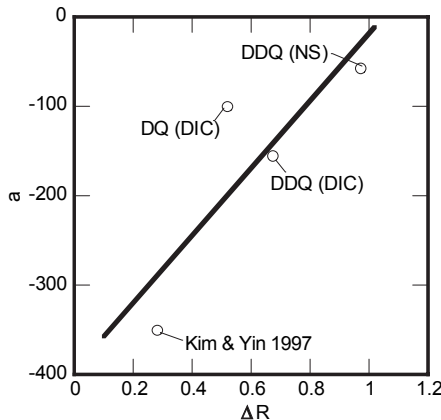
Figure 17. Plastic spin parameter a versus planar anisotropy ΔR.

using the measured data of the rotation of the symmetry axes [Kim and Yin 1997] shows good correlation. Therefore, it may be possible to determine the unknown plastic spin parameter from the R-values (ΔR) for mild steel. More experiments are required to verify this methodology for other alloys and crystal structures.

7. Concluding remarks

For two mild steels and a high strength steel sheet metal, experimental results of conventional flow stress curves and transverse versus longitudinal strain curves of two orientations relative to the RD have been presented. Assuming that rotation of the anisotropy axes can describe the evolution of anisotropy qualitatively, the material parameters for the plastic spin and for correcting the kinematic harden- ing description can be identified from tensile tests. Using the DIC method, it is relatively easy to measure large strains. The DIC method was applied to generate experimental transverse strain versus longitudinal strain for simple tensile tests. The evolution of anisotropy is then correlated with the slope of the transverse strain versus longitudinal strain curve. Only the plastic spin parameter allows a change to the slope of the computed transverse strain versus longitudinal strain curve. The flow stress results can further be improved by incorporating the kinematic hard- ening. The parameters for the RIK model can be determined by three-point bend tests [Zhao and Lee 1999; 2001] and tensile tests for three directions: RD, TD, and 45°. With the aid of the rotational evolution of the anisotropy axes, the RIK model can essentially capture anisotropic plastic hardening behavior with Hill's quadratic yield criteria. For mild steel, a linear relationship between the planar anisotropy ΔR and the plastic spin parameter a has been determined.

Acknowledgment

References

[Agnew and Weertman 1998] S. R. Agnew and J. R. Weertman, "The influence of texture on the elastic properties of ultrafine-grain copper", *Mater. Sci. Eng. A* **242**:1 (1998), 174–180.

[Armstrong and O. 1966] P. J. Armstrong and F. C. O., "A mathematical representation of the multi- axial Bauschinger effect", Report rd/b/bn 731, Central Electricity Generating Board, 1966.

[Asaro 1983] R. J. Asaro, "Micromechanics of crystals and polycrystals", *Adv. Appl. Mech.* **23** (1983), 1–115.

[Banabic et al. 2000] D. Banabic, H. J. Bunge, K. Pohlandt, and A. E. Tekkaya, *Formability of metallic materials*, Springer, 2000.

[Barlat et al. 1991] F. Barlat, D. J. Lege, and J. C. Brem, "A six-component yield function for aniso- tropic materials", *Int. J. Plast.* **7**:7 (1991), 693–712.

[Barlat et al. 1997] F. Barlat, Y. Maedat, K. Chung, M. Yanagawa, J. C. Brem, Y. Hayashida, D. J. Lege, K. Matsui, S. J. Murtha, S. Hattori, R. C. Becker, and S. Makosey, "Yield function development for aluminum alloy sheets", *J. Mech. Phys. Solids* **45**:11 (1997), 1727–1763.

[Barlat et al. 2003] F. Barlat, J. C. Brem, J. W. Yoon, K. Chung, R. E. Dick, D. J. Lege, F. Pourboghrat, S. H. Choi, and E. Chu, "Plane stress yield function for aluminum alloy sheets, I: theory", *Int. J. Plast.* **19**:9 (2003), 1297–1319.

[Barton et al. 1999] N. Barton, P. R. Dawson, and M. Miller, "Yield strength asymmetry predictions from polycrystal elastoplasticity", *J. Eng. Mater. Technol. (ASME)* **121**:2 (1999), 230–239.

[Beaudoin et al. 1994] A. J. Beaudoin, P. R. Dawson, K. K. Mathur, U. F. Kocks, and D. A. Korzekwa, "Application of polycrystal plasticity to sheet forming", *Comput. Methods Appl. Mech. Eng.* **117**:1–2 (1994), 49–70.

[Boehler and Koss 1991] J. P. Boehler and S. Koss, "Evolution of anisotropy in sheet-steels submitted to off-axes large deformations", pp. 143–158 in *Advances in continuum mechanics*, edited by O. e. a. Brueller, Springer, 1991.

[Bunge and Nielsen 1997] H. J. Bunge and I. Nielsen, "Experimental determination of plastic spin in polycrystalline materials", *Int. J. Plast.* **13**:5 (1997), 435–446.

[Chaboche 1989] J. L. Chaboche, "Constitutive equations for cyclic plasticity and cyclic viscoplasticity", *Int. J. Plast.* **5**:3 (1989), 247–302.

[Choi et al. 2006a] Y. Choi, C. S. Han, J. K. Lee, and R. H. Wagoner, "Modeling multi-axial elastoplastic deformation of planar anisotropic materials, I: Theory", *Int. J. Plast.* **22**:9 (2006), 1745–1764.

[Choi et al. 2006b] Y. Choi, C. S. Han, J. K. Lee, and R. H. Wagoner, "Modeling multi-axial elastoplastic deformation of planar anisotropic materials, II: Applications", *Int. J. Plast.* **22**:9 (2006), 1765–1783.

[Chun et al. 2002] B. K. Chun, J. T. Jinn, and J. K. Lee, "Modeling of the Bauschinger effect for sheet metals, I: Theory", *Int. J. Plast.* **18**:5 (2002), 571–595.

[Dafalias 1993] Y. F. Dafalias, "On multiple spins and texture development, Case study: kinematic and orthotropic hardening", *Acta Mech.* **100**:3–4 (1993), 171–194.

[Dafalias 1998] Y. F. Dafalias, "Plastic spin: Necessity or redundancy?", *Int. J. Plast.* **14**:9 (1998), 909–931.

[Dafalias 2000] Y. F. Dafalias, "Orientational evolution of plastic orthotropy in sheet metals", *J. Mech. Phys. Solids* **48**:11 (2000), 2231–2255.

[Dafalias 2001] Y. F. Dafalias, "Plasticity in large deformations", pp. 247–254 (Section 4.8) in *Handbook of materials behavior models*, edited by J. Lemaitre, Academic Press, 2001.

[Feaugas 1999] X. Feaugas, "On the origin of the tensile flow stress in the stainless steel AISI 316L at 300 K: back stress and effective stress", *Acta Mater.* **47**:13 (1999), 3617–3632.

[Han et al. 2002] C. S. Han, Y. Choi, J. K. Lee, and R. H. Wagoner, "A FE formulation for elastoplastic materials with planar anisotropic yield functions and plastic spin", *Int. J. Solids Struct.* **39**:20 (2002), 5123–5141.

[Han et al. 2003] C. S. Han, M. G. Lee, K. Chung, and R. H. Wagoner, "Integration algorithms for planar anisotropic shells with isotropic and kinematic hardening at finite strains", *Commun. Numer. Methods Eng.* **19**:6 (2003), 473–490.

[Haupt and Tsakmakis 1986] P. Haupt and C. Tsakmakis, "On kinematic hardening and large plastic deformation", *Int. J. Plast.* **2**:3 (1986), 279–293.

[Hill 1948] R. Hill, "A theory of the yielding and plastic flow of anisotropic metals", *P. Roy. Soc. Lond. A Mat.* **193**:1033 (1948), 281–297.

[Hill 1950] R. Hill, *The mathematical theory of plasticity*, Oxford University Press, 1950.

[Hill 1990] R. Hill, "Constitutive modeling of orthotropic plasticity in sheet metals", *J. Mech. Phys. Solids* **38**:3 (1990), 405–417.

[Hughes et al. 2003] D. A. Hughes, N. Hansen, and D. J. Bammann, "Geometrically necessary boundaries, incidental dislocation boundaries and geometrically necessary dislocations", *Scr. Mater.* **48**:2 (2003), 147–153.

[Kim and Yin 1997] K. H. Kim and J. J. Yin, "Evolution of anisotropy under plane stress", *J. Mech. Phys. Solids* **45**:5 (1997), 841–851.

[Kocks et al. 1998] U. F. Kocks, C. N. Tome, and H. R. Wenk, *Texture and anisotropy*, Cambridge University Press, 1998.

[Kuroda 1997] M. Kuroda, "Interpretation of the behavior of metals under large plastic shear deformations: a macroscopic approach", *Int. J. Plast.* **13**:4 (1997), 359–383.

[Nakamachi and Doug 1997] E. Nakamachi and X. Doug, "Study of texture effect on sheet failure in a limit dome height test by using elastic/crystalline viscoplastic finite element analysis", *J. Appl. Mech.* (*ASME*) **64**:3 (1997), 519–524.

[Peeters et al. 2001] B. Peeters, M. Seefeldt, C. Tedodosiu, S. R. Kalidindi, P. Van Houtte, and E. Aernoudt, "Work-hardening/softening behavior of B.C.C. polycrystals during changing strain paths: I. An integrated model based on substructure and texture evolution, and its prediction of the stress-strain behavior of an IF steel during two-stage strain paths", *Acta Mater.* **49**:9 (2001), 1607–1629.

[Rao and Mohan 2001] K. P. Rao and E. V. R. Mohan, "A vision-integrated tension test for use in sheet-metal formability studies", *J. Mater. Process. Technol.* **118**:1–3 (2001), 238–245.

[Sutton et al. 1983] M. A. Sutton, W. J. Wolters, W. H. Peters, W. F. Ranson, and S. R. McNeill, "Determination of displacements using an improved digital correlation method", *Image Vision Comput.* **1**:3 (1983), 133–139.

[Tsakmakis 2004] C. Tsakmakis, "Description of plastic anisotropy effects at large deformations, I: restrictions imposed by the second law and the postulate of Il'iushin", *Int. J. Plast.* **20**:2 (2004), 167–198.

[Valliappan et al. 1976] S. Valliappan, P. Boonlaulohr, and I. K. Lee, "Non-linear analysis for anisotropic materials", *Int. J. Numer. Meth. Eng.* **10**:3 (1976), 597–606.

[Vendroux and Knauss 1998] G. Vendroux and W. G. Knauss, "Submicron deformation field measurements, 2: Improved digital image correlation", *Exp. Mech.* **38**:2 (1998), 86–92.

[Yang et al. 2002] D. Y. Yang, S. I. Oh, H. Huh, and Y. H. Kim, "NUMISHEET 2002", Proceedings -Verification of Simulations with Experiments, 2002.

[Zhao and Lee 1999] K. Zhao and J. K. Lee, "On simulation of bending/reverse bending of sheet metals", *J. Manuf. Sci. Eng.* (*ASME*) **10** (1999), 929–933.

[Zhao and Lee 2000] K. Zhao and J. K. Lee, "Generation of cyclic stress-strain curves for sheet metals", *ASME MED* **11** (2000), 667–674.

[Zhao and Lee 2001] K. Zhao and J. K. Lee, "Generation of cyclic stress-strain curves for sheet metals", *J. Eng. Mater. Technol.* (*ASME*) **123**:4 (2001), 391–397.

YANGWOOK CHOI: yangwook.choi@abaqus.com

Dr. Yangwook Choi, Department of Mechanical Engineering, The Ohio State University, 650 Ackerman Road, Columbus, Ohio 43210, United States

MARK E. WALTER: walter.80@osu.edu
Dr. Mark E. Walter, Department of Mechanical Engineering, The Ohio State University, 650 Ackerman Road, Columbus, Ohio 43210, United States

JUNE K. LEE: lee.71@osu.edu
Dr. June K. Lee, Department of Mechanical Engineering, The Ohio State University, 650 Ackerman Road, Columbus, Ohio 43210, United States

CHUNG-SOUK HAN: chung-souk.han@ndsu.edu
Dr. Chung-Souk Han, Department of Civil Engineering, North Dakota State University, 1410 North 14th Avenue, Fargo, ND 58105, United States

THE ESHELBY TENSOR IN NONLOCAL ELASTICITY AND IN NONLOCAL MICROPOLAR ELASTICITY

MARKUS LAZAR AND HELMUT O. K. KIRCHNER

The Eshelby tensor is formulated for anisotropic linear nonlocal elasticity and nonlocal micropolar elasticity in a nonhomogeneous medium. The divergence of this tensor gives the configurational forces on geometric and physical defects in such a medium. Some examples of the Peach–Koehler force and the Mathison–Papapetrou force between dislocations and/or disclinations are given.

1. Introduction

We consider anisotropic nonlocal elasticity and anisotropic nonlocal micropolar elasticity for a medium of arbitrary inhomogeneity. Such nonlocal theories can predict dispersion relations in the entire Brillouin zone; they suppress nonphysical singularities: crack tip singularities do not occur, and the stresses of dislocations are finite [Eringen 2002]. These results are features of linear nonlocal theories which cannot be obtained in linear elasticity and linear micropolar elasticity. They agree very well with those predicted by atomistic theories and experiments.

The aim of this paper is to derive the Eshelby tensor [Eshelby 1951;1975] in the theories of nonlocal elasticity and nonlocal micropolar elasticity. This represents a vast generalization of this tensor written for local, linear elasticity by [Morse and Feshbach 1953]. The Eshelby tensor, which is the static energy-momentum tensor, is of fundamental importance in any field theory, and in particular in the field theory of generalized elasticity. The divergence of the Eshelby tensor gives the configurational forces on the sources of the field. Few results are known about the Peach–Koehler force and conservation laws in nonlocal elasticity [Kovács and Vörös 1979; Vukobrat and Kuzmanović 1992; Lazar 2005]. This is one motivation for the investigations in the present paper. We will derive all configurational forces felt by topological defects (dislocations and disclinations), physical sources (body force, body moment) and all others due to inhomogeneities in nonlocal elasticity and nonlocal micropolar elasticity. We calculate the J-integral for these nonlocal theories, relevant in fracture mechanics of nonlocal materials. In addition, we will present some examples of interaction forces between dislocations as well as disclinations in nonlocal theories.

Keywords: Eshelby tensor, J-integral, nonlocal elasticity.

2. Nonlocal elasticity

The goal of this section is the construction of the Eshelby tensor and the related configurational forces for nonlocal elasticity. In nonlocal elasticity the elastic energy is given by [Kröner and Datta 1966]

$$W = \frac{1}{2} \int \int C_{ijkl}(\boldsymbol{x}, \boldsymbol{x}') \beta_{ij}(\boldsymbol{x}) \beta_{kl}(\boldsymbol{x}') \, \mathrm{d}^3 \boldsymbol{x} \, \mathrm{d}^3 \boldsymbol{x}', \tag{2.1}$$

where $C_{ijkl}(\boldsymbol{x}, \boldsymbol{x}')$ is the tensor of nonlocal elastic constants and $\beta_{ij}(\boldsymbol{x})$ denotes the elastic distortion. For simplicity, we assume a linear relationship. The nonlocal constitutive law for full anisotropy reads:

$$t_{ij}(\boldsymbol{x}) = \int C_{ijkl}(\boldsymbol{x}, \boldsymbol{x}') \beta_{kl}(\boldsymbol{x}') \, \mathrm{d}^3 \boldsymbol{x}'. \tag{2.2}$$

The tensor of nonlocal elastic constants possesses the symmetry

$$C_{ijkl}(\boldsymbol{x}, \boldsymbol{x}') = C_{klij}(\boldsymbol{x}', \boldsymbol{x}). \tag{2.3}$$

The equilibrium condition is given by

$$\partial_j t_{ij}(\boldsymbol{x}) + f_i(\boldsymbol{x}) = 0, \tag{2.4}$$

where $f(\boldsymbol{x})$ denotes the body force in nonlocal elasticity. The incompatibility condition reads

$$\epsilon_{jkl} \partial_k \beta_{il}(\boldsymbol{x}) = \alpha_{ij}(\boldsymbol{x}). \tag{2.5}$$

Here α_{ij} is the dislocation density tensor, divergence free in the second index. The field (2.4) and the incompatibility condition (2.5) have the same form as in local elasticity; the generalization to nonlocal elasticity occurs through Hooke's law (2.2). By multiplying Equation (2.5) with ϵ_{mnj} one finds for the elastic distortion

$$\partial_m \beta_{in}(\boldsymbol{x}) - \partial_n \beta_{im}(\boldsymbol{x}) = \epsilon_{mnj} \alpha_{ij}(\boldsymbol{x}). \tag{2.6}$$

If no dislocations are present, the elastic distortion is just the gradient of a displacement $u_i(\boldsymbol{x})$: $\beta_{ij}(\boldsymbol{x}) = \partial_j u_i(\boldsymbol{x})$.

Following the procedure of [Kirchner 1999], we construct the Eshelby (or static energy-momentum) tensor for nonhomogeneous nonlocal elasticity. Let us take an arbitrary infinitesimal functional derivative δW of the elastic energy density. From Equation (2.1) we get

$$\delta W = \frac{1}{2} \int \int \big\{ C_{ijkl}(\boldsymbol{x}, \boldsymbol{x}') [\delta \beta_{ij}(\boldsymbol{x})] \beta_{kl}(\boldsymbol{x}') + C_{ijkl}(\boldsymbol{x}, \boldsymbol{x}') \beta_{ij}(\boldsymbol{x}) [\delta \beta_{kl}(\boldsymbol{x}')]$$

$$+ [\delta C_{ijkl}(\boldsymbol{x}, \boldsymbol{x}')] \beta_{ij}(\boldsymbol{x}) \beta_{kl}(\boldsymbol{x}') \big\} \, \mathrm{d}^3 \boldsymbol{x} \, \mathrm{d}^3 \boldsymbol{x}'. \tag{2.7}$$

Using the symmetry (2.3) and Hooke's law (2.2) for nonlocality, there remains

$$\delta W = \int t_{ij}(\boldsymbol{x})[\delta\beta_{ij}(\boldsymbol{x})]\,\mathrm{d}^3\boldsymbol{x} + \frac{1}{2}\iint \beta_{ij}(\boldsymbol{x})[\delta C_{ijkl}(\boldsymbol{x},\boldsymbol{x}')]\beta_{kl}(\boldsymbol{x}')\,\mathrm{d}^3\boldsymbol{x}\,\mathrm{d}^3\boldsymbol{x}'. \quad (2.8)$$

Since we want to obtain configurational forces, we specify the functional derivative to be translational:

$$\delta = (\delta x_k)\partial_k. \quad (2.9)$$

On the left hand side of Equation (2.7) we write

$$\begin{aligned}
\delta W &= \int \delta w(\boldsymbol{x})\,\mathrm{d}^3\boldsymbol{x} \\
&= \int [\partial_k w(\boldsymbol{x})](\delta x_k)\,\mathrm{d}^3\boldsymbol{x} \\
&= \int \partial_i[w(\boldsymbol{x})\delta_{ik}](\delta x_k)\,\mathrm{d}^3\boldsymbol{x},
\end{aligned} \quad (2.10)$$

with the energy density

$$w(\boldsymbol{x}) = \tfrac{1}{2}\,t_{ij}(\boldsymbol{x})\beta_{ij}(\boldsymbol{x}). \quad (2.11)$$

On the right hand side of Equation (2.7) we obtain with (2.3)

$$\begin{aligned}
\delta W = \int \Big\{ t_{ij}(\boldsymbol{x})[\partial_k\beta_{ij}(\boldsymbol{x}) - \partial_j\beta_{ik}(\boldsymbol{x})] + t_{ij}(\boldsymbol{x})[\partial_j\beta_{ik}(\boldsymbol{x})]\Big\}(\delta x_k)\,\mathrm{d}^3\boldsymbol{x} \\
+ \frac{1}{2}\iint \beta_{ij}(\boldsymbol{x})[\partial_k C_{ijmn}(\boldsymbol{x},\boldsymbol{x}')]\beta_{mn}(\boldsymbol{x}')(\delta x_k)\,\mathrm{d}^3\boldsymbol{x}\,\mathrm{d}^3\boldsymbol{x}', \quad (2.12)
\end{aligned}$$

where the second and third terms have been subtracted and added. The purpose is to obtain the square bracket with the meaning of Equation (2.6). The third term may be written with (2.4) as

$$\begin{aligned}
t_{ij}(\boldsymbol{x})[\partial_j\beta_{ik}(\boldsymbol{x})] &= \partial_j[t_{ij}(\boldsymbol{x})\beta_{ik}(\boldsymbol{x})] - [\partial_j t_{ij}(\boldsymbol{x})]\beta_{ik}(\boldsymbol{x}) \\
&= \partial_j[t_{ij}(\boldsymbol{x})\beta_{ik}(\boldsymbol{x})] + f_i(\boldsymbol{x})\beta_{ik}(\boldsymbol{x}). \quad (2.13)
\end{aligned}$$

By equating (2.10) and (2.12), using Equations (2.11) and (2.13) we obtain the expression

$$\begin{aligned}
\int \partial_i\big(w(\boldsymbol{x})\delta_{ik} - t_{li}(\boldsymbol{x})\beta_{lk}(\boldsymbol{x})\big)\mathrm{d}^3\boldsymbol{x} = \int \Big(\epsilon_{kjl}t_{ij}(\boldsymbol{x})\alpha_{il}(\boldsymbol{x}) + f_i(\boldsymbol{x})\beta_{ik}(\boldsymbol{x}) \\
+ \frac{1}{2}\int \beta_{ij}(\boldsymbol{x})[\partial_k C_{ijmn}(\boldsymbol{x},\boldsymbol{x}')]\beta_{mn}(\boldsymbol{x}')\mathrm{d}^3\boldsymbol{x}'\Big)\mathrm{d}^3\boldsymbol{x} \\
= J_k. \quad (2.14)
\end{aligned}$$

The second integral contains the sources of the elastic fields: the dislocation density, the body force and the inhomogeneity of the material. The integrand of the first integral in Equation (2.14) is the divergence of the Eshelby tensor of nonlocal elasticity

$$P_{ki}(\boldsymbol{x}) = \big[w(\boldsymbol{x})\delta_{ik} - t_{li}(\boldsymbol{x})\beta_{lk}(\boldsymbol{x})\big]. \tag{2.15}$$

It may be transformed into a surface integral

$$J_k = \int P_{ki}(\boldsymbol{x})n_i \, \mathrm{d}^2\boldsymbol{x}. \tag{2.16}$$

Equation (2.16) is the J-integral in nonlocal elasticity. Notice that in terms of energy, stresses and distortions, it is of the same form as in local elasticity. This is because the field (2.4) and the incompatibility condition (2.5) have the same form in local and nonlocal elasticity. The configurational force density is the divergence of the Eshelby tensor

$$\partial_i P_{ki} = F_k \tag{2.17}$$

with

$$F_k = \epsilon_{kjl}t_{ij}(\boldsymbol{x})\alpha_{il}(\boldsymbol{x}) + f_i(\boldsymbol{x})\beta_{ik}(\boldsymbol{x})$$
$$+ \frac{1}{2}\int \beta_{ij}(\boldsymbol{x})[\partial_k C_{ijmn}(\boldsymbol{x}, \boldsymbol{x}')]\beta_{mn}(\boldsymbol{x}') \, \mathrm{d}^3\boldsymbol{x}'. \tag{2.18}$$

The first term is the configurational force on a dislocation density like the Peach–Koehler force in local elasticity [Peach and Koehler 1950]. We have obtained the Peach–Koehler force generalized to nonlocal elasticity. The second term is the configurational force on a body force $f_i(\boldsymbol{x})$ in presence of an elastic distortion $\beta_{ik}(\boldsymbol{x})$—it is the nonlocal generalization of the Cherepanov force [Cherepanov 1981]. The third term is the material force on the inhomogeneity $\partial_k C_{ijmn}(\boldsymbol{x}, \boldsymbol{x}')$ in nonlocal elasticity—the nonlocal generalization of the Eshelby force [Eshelby 1951].

For a homogeneous defect-free and source-free material the Eshelby tensor (2.15) reduces to

$$P_{ki}(\boldsymbol{x}) = \big[w(\boldsymbol{x})\delta_{ik} - t_{li}(\boldsymbol{x})\partial_k u_l(\boldsymbol{x})\big], \tag{2.19}$$

which is divergenceless. Then the J-integral (2.16) is zero.

On the other hand, if we use the dislocation density of a single straight dislocation

$$\alpha_{ij} = b_i' n_j \, \delta(x - x')\delta(y - y'), \tag{2.20}$$

we obtain the expression for the Peach–Koehler force in nonlocal elasticity as follows:

$$F_k^{\text{PK}} = \epsilon_{kjl} b_i' n_l t_{ij}. \tag{2.21}$$

Here b_i is the Burgers vector and n_j the tangent line element of the dislocation, in agreement with the formula given by [Kovács and Vörös 1979]. These authors did not use the concept of the Eshelby tensor and configurational force in their calculation. They gave just a formal derivation of Equation (2.21).

From invariance arguments it follows that in an isotropic nonlocal medium the tensor of nonlocal elastic moduli must be of the form

$$C_{ijkl}(\boldsymbol{x}, \boldsymbol{x}') = \left\{ \lambda \delta_{ij} \delta_{kl} + \mu (\delta_{ik} \delta_{jl} + \delta_{il} \delta_{jk}) \right\} G(|\boldsymbol{x} - \boldsymbol{x}'|), \tag{2.22}$$

where $G(|\boldsymbol{x} - \boldsymbol{x}'|)$ is called the nonlocal kernel [Eringen 2002].

In the following, we choose the two-dimensional nonlocal kernel (see, for example, [Eringen 2002])

$$G(|\boldsymbol{x} - \boldsymbol{x}'|) = \frac{1}{2\pi \varepsilon^2} K_0 \left(\frac{\sqrt{(x - x')^2 + (y - y')^2}}{\varepsilon} \right), \quad \varepsilon \geq 0, \tag{2.23}$$

which is the Green function of the two-dimensional Helmholtz-equation and ε is the parameter of nonlocality. Here K_n denotes the modified Bessel function of the second kind and n is the order of this function.

With Equation (2.22) we obtain for the Peach–Koehler force of two parallel screw dislocations ($n_z = 1$)

$$F_r^{\text{PK}} = b_z' t_{z\varphi}, \tag{2.24}$$

where

$$t_{z\varphi} = \frac{\mu b_z}{2\pi} \frac{1}{r} \{1 - \frac{r}{\varepsilon} K_1(r/\varepsilon)\}, \tag{2.25}$$

with $r = \sqrt{x^2 + y^2}$. Due to the nonlocal theory the $1/r$-singularity has disappeared. This force is zero at $r = 0$. It has an extremum value of $0.399 \mu b_z' b_z / [2\pi \varepsilon]$ at $r \simeq 1.114\varepsilon$ (see Figure 1).

The Peach–Koehler force between two parallel edge dislocations has been found by [Lazar 2005] in the framework of nonlocal elasticity. Unlike in classical elasticity, both for screw and edge dislocations the Peach–Koehler forces are finite and nonsingular in nonlocal elasticity.

3. Nonlocal micropolar elasticity

The aim of this section is to derive the Eshelby tensor and the corresponding configurational forces for nonlocal micropolar elasticity, another generalization of
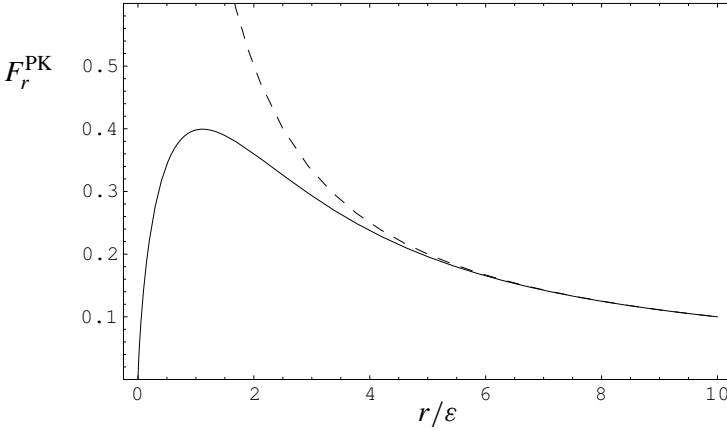
Figure 1. Peach–Koehler force F_r^{PK} between two screw disloca-tions. F_r^{PK} is given in units of $\mu b'_z b_z/[2\pi\varepsilon]$. The dashed curve represents the classical result.

classical elasticity. For linear anisotropic nonlocal micropolar elasticity, the strain energy density is given as follows [Eringen 2002]

$$W = \frac{1}{2} \int\!\!\int \left\{ \gamma_{ij}(\boldsymbol{x})\, C_{ijkl}(\boldsymbol{x}, \boldsymbol{x}')\gamma_{kl}(\boldsymbol{x}') + \kappa_{ij}(\boldsymbol{x})\, A_{ijkl}(\boldsymbol{x}, \boldsymbol{x}')\kappa_{kl}(\boldsymbol{x}') \right.$$
$$\left. + 2\gamma_{ij}(\boldsymbol{x})\, B_{ijkl}(\boldsymbol{x}, \boldsymbol{x}')\kappa_{kl}(\boldsymbol{x}') \right\} \mathrm{d}^3\boldsymbol{x}\, \mathrm{d}^3\boldsymbol{x}', \quad (3.1)$$

where $\gamma_{ij}(\boldsymbol{x})$ and $\kappa_{ij}(\boldsymbol{x})$ denote the relative distortion tensor and the wryness tensor, respectively. The nonlocal constitutive moduli possess the symmetries

$$A_{ijkl}(\boldsymbol{x}, \boldsymbol{x}') = A_{klij}(\boldsymbol{x}', \boldsymbol{x}), \qquad C_{ijkl}(\boldsymbol{x}, \boldsymbol{x}') = C_{klij}(\boldsymbol{x}', \boldsymbol{x}). \qquad (3.2)$$

In nonlocal micropolar elasticity, the force stress tensor $t_{ij}(\boldsymbol{x})$ and the couple stress tensor $m_{ij}(\boldsymbol{x})$ are given in integral form by the nonlocal constitutive relations:

$$t_{ij}(\boldsymbol{x}) = \int \left\{ C_{ijkl}(\boldsymbol{x}, \boldsymbol{x}')\gamma_{kl}(\boldsymbol{x}') + B_{ijkl}(\boldsymbol{x}, \boldsymbol{x}')\kappa_{kl}(\boldsymbol{x}') \right\} \mathrm{d}^3\boldsymbol{x}', \qquad (3.3)$$

$$m_{ij}(\boldsymbol{x}) = \int \left\{ B_{klij}(\boldsymbol{x}, \boldsymbol{x}')\gamma_{kl}(\boldsymbol{x}') + A_{ijkl}(\boldsymbol{x}, \boldsymbol{x}')\kappa_{kl}(\boldsymbol{x}') \right\} \mathrm{d}^3\boldsymbol{x}'. \qquad (3.4)$$

The force and the moment equilibrium conditions read

$$\partial_j t_{ij}(\boldsymbol{x}) + f_i(\boldsymbol{x}) = 0, \qquad (3.5)$$

$$\partial_j m_{ij}(\boldsymbol{x}) - \epsilon_{ijk} t_{jk}(\boldsymbol{x}) + l_i(\boldsymbol{x}) = 0, \qquad (3.6)$$

where $f_i(x)$ and $l_i(x)$ are the body force and the body couple, respectively. The incompatibility conditions in micropolar elasticity [Eringen 1999] are the definitions for the dislocation density tensor $\alpha_{ij}(x)$ and the disclination density tensor $\Theta_{ij}(x)$:

$$\epsilon_{jkl}\left[\partial_k\gamma_{il}(x)+\epsilon_{ikm}\kappa_{ml}(x)\right]=\alpha_{ij}(x), \tag{3.7}$$

$$\epsilon_{jkl}\partial_k\kappa_{il}(x)=\Theta_{ij}(x). \tag{3.8}$$

Again, the form of Equations (3.5)–(3.8) is the same as in local micropolar elasticity. If no dislocations and disclinations are present, the micropolar strain quantities are of the form: $\gamma_{ij}(x)=\partial_j u_i(x)+\epsilon_{ijk}\varphi_k(x)$ and $\kappa_{ij}(x)=\partial_j\varphi_i(x)$. Here $\varphi_k(x)$ denotes the micro-rotation.

Using the same procedure for the calculation of the Eshelby tensor in nonlocal micropolar elasticity as in Section 2 for the Eshelby tensor in nonlocal elasticity, we obtain

$$P_{ki}(x)=\left[w(x)\delta_{ik}-t_{li}(x)\bar{\gamma}_{lk}(x)-m_{li}(x)\kappa_{lk}(x)\right], \tag{3.9}$$

where $\bar{\gamma}_{lk}=\gamma_{lk}-\epsilon_{lkm}\varphi_m$ and

$$w(x)=\tfrac{1}{2}t_{ij}(x)\gamma_{ij}(x)+\tfrac{1}{2}m_{ij}(x)\kappa_{ij}(x). \tag{3.10}$$

Equation (3.9) is the Eshelby tensor for nonlocal micropolar elasticity. Using the Noether theorem, in fact the translational invariance, it is the generalization of the Eshelby tensor for micropolar elasticity given by [Kluge 1969] to nonlocality and, on the other hand, it is the generalization of the Eshelby tensor for nonlocal elasticity derived in Section 2 to micropolarity. With Equation (3.9) we obtain a surface integral

$$J_k=\int P_{ki}(x)n_i\,\mathrm{d}^2x. \tag{3.11}$$

Equation (2.16) is the J-integral in nonlocal micropolar elasticity. The divergence of the Eshelby tensor (3.9) gives the configurational force density:

$$F_k(x)=\partial_i P_{ki}(x), \tag{3.12}$$

with

$$\begin{aligned}
F_k(x)={}&\epsilon_{kjl}t_{ij}(x)\alpha_{il}(x)+\epsilon_{kjl}m_{ij}(x)\Theta_{il}(x)\\
&-\epsilon_{kjl}t_{ji}(x)\kappa_{li}^{\mathrm{P}}(x)+f_i(x)\bar{\gamma}_{ik}(x)+l_i(x)\kappa_{ik}(x)\\
&+\tfrac{1}{2}\int\left\{\gamma_{ij}(x)[\partial_k C_{ijmn}(x,x')]\gamma_{mn}(x')+\kappa_{ij}(x)[\partial_k A_{ijmn}(x,x')]\kappa_{mn}(x')\right.\\
&\left.+2\gamma_{ij}(x)[\partial_k B_{ijmn}(x,x')]\kappa_{mn}(x')\right\}\mathrm{d}^3x'.
\end{aligned} \tag{3.13}$$

It can be seen that Equation (3.13) is a sum of several configurational force densities in nonlocal micropolar elasticity:

(i) the Peach–Koehler force density on a dislocation density $\alpha_{il}(\boldsymbol{x})$ in the presence of the force stress $t_{ij}(\boldsymbol{x})$ [Kluge 1969];

(ii) the force density on a disclination density $\Theta_{il}(\boldsymbol{x})$ in the presence of the couple stress $m_{ij}(\boldsymbol{x})$, which is called a generalized Mathisson–Papapetrou type force density [Gairola 1981; Maugin 1993; Hehl et al. 1995];

(iii) a Cherepanov force density on a body force $f_i(\boldsymbol{x})$ in the presence of a distortion $\bar{\gamma}_{ik}(\boldsymbol{x})$;

(iv) a force density on a body couple $l_i(\boldsymbol{x})$ in presence of the elastic wryness $\kappa_{ik}(\boldsymbol{x})$;

(v) a force density on the force stress $t_{ji}(\boldsymbol{x})$ in presence of the plastic wryness $\kappa_{li}^{\mathrm{P}}(\boldsymbol{x})$;

(vi) three force densities on inhomogeneities: $\partial_k C_{ijmn}(\boldsymbol{x}, \boldsymbol{x}')$, $\partial_k A_{ijmn}(\boldsymbol{x}, \boldsymbol{x}')$ and $\partial_k B_{ijmn}(\boldsymbol{x}, \boldsymbol{x}')$.

For a homogeneous defect-free and source-free micropolar material, the Eshelby tensor (3.9) simplifies to

$$P_{ki}(\boldsymbol{x}) = \left[w(\boldsymbol{x})\delta_{ik} - t_{li}(\boldsymbol{x})\partial_k u_l(\boldsymbol{x}) - m_{li}(\boldsymbol{x})\partial_k \varphi_l(\boldsymbol{x}) \right], \qquad (3.14)$$

which is divergenceless. Then the J-integral (3.11) is zero. The formula (3.14) is the nonlocal generalization of the Eshelby tensor for micropolar elasticity given by [Lubarda and Markenscoff 2003]. The corresponding Eshelby tensor for finite local polar elasticity has been given by [Maugin 1998].

If we use the dislocation density tensor of a straight dislocation and the disclination density tensor of a straight disclination

$$\alpha_{ij} = b_i' n_j \, \delta(x - x')\delta(y - y'), \qquad (3.15)$$

$$\Theta_{ij} = \Omega_i' n_j \, \delta(x - x')\delta(y - y'), \qquad (3.16)$$

we obtain for the Peach–Koehler force and the Mathisson–Papapetrou force, respectively,

$$F_k^{\mathrm{PK}} = \epsilon_{kjl} b_i' n_l t_{ij}, \qquad (3.17)$$

$$F_k^{\mathrm{MP}} = \epsilon_{kjl} \Omega_i' n_l m_{ij}. \qquad (3.18)$$

Here Ω_i denotes the Frank vector (the topological charge of a disclination).

For isotropic nonlocal micropolar elasticity the nonlocal elastic moduli must be of the form

$$C_{ijkl}(\boldsymbol{x}, \boldsymbol{x}') = \{\lambda\delta_{ij}\delta_{kl} + \mu(\delta_{ik}\delta_{jl} + \delta_{il}\delta_{jk}) + \mu_c(\delta_{ik}\delta_{jl} - \delta_{il}\delta_{jk})\}G(|\boldsymbol{x}-\boldsymbol{x}'|), \quad (3.19)$$

$$A_{ijkl}(\boldsymbol{x}, \boldsymbol{x}') = \{\alpha\delta_{ij}\delta_{kl} + \beta(\delta_{ik}\delta_{jl} + \delta_{il}\delta_{jk}) + \gamma(\delta_{ik}\delta_{jl} - \delta_{il}\delta_{jk})\}G(|\boldsymbol{x}-\boldsymbol{x}'|), \quad (3.20)$$

$$B_{ijkl}(\boldsymbol{x}, \boldsymbol{x}') = 0, \quad (3.21)$$

in terms of six material constants λ, μ, μ_c, α, β, γ, characteristic for the medium under consideration. Again, $G(|\boldsymbol{x} - \boldsymbol{x}'|)$ is the nonlocal kernel. If we use the six material constants of micropolar elasticity, two characteristic lengths l and h can be defined by [Nowacki 1986]

$$l^2 = \frac{(\mu + \mu_c)(\beta + \gamma)}{4\mu\,\mu_c}, \qquad h^2 = \frac{\alpha + 2\beta}{4\mu_c}. \quad (3.22)$$

In the following, we use the two-dimensional nonlocal kernel (2.23). Then the Peach–Koehler force for two parallel screw dislocations in a micropolar medium is

$$F_r^{\mathrm{PK}} = b_z' t_{z\varphi}, \quad (3.23)$$

with [Lazar et al. 2005]

$$t_{z\varphi} = \frac{b_z}{2\pi} \frac{1}{r} \left\{ \mu\left[1 - \frac{r}{\varepsilon} K_1(r/\varepsilon)\right] + \frac{\mu_c h^2}{h^2 - \varepsilon^2}\left[\frac{r}{h} K_1(r/h) - \frac{r}{\varepsilon} K_1(r/\varepsilon)\right] \right\}. \quad (3.24)$$

The force (3.23) is nonsingular. It is zero at $r = 0$ and has an extremum value which depends on the coefficients ε and h (see Figure 2). In addition, it can be seen that the Peach–Koehler force between two screw dislocations in nonlocal micropolar elasticity is slightly different from the force in nonlocal elasticity (2.24).

Another interesting situation is the interaction of two parallel wedge disclinations. The Mathisson–Papapetrou force for two parallel wedge disclinations in a micropolar medium is

$$F_r^{\mathrm{MK}} = \Omega_z' m_{z\varphi}, \quad (3.25)$$

with [Lazar and Maugin 2004]

$$m_{z\varphi} = \frac{(\beta + \mu_c)\Omega_z}{2\pi} \frac{1}{r} \left\{1 - \frac{r}{\varepsilon} K_1(r/\varepsilon)\right\}. \quad (3.26)$$

It is zero at $r = 0$ and has an extremum value of $0.399(\beta + \mu_c)\Omega_z'\Omega_z/[2\pi\varepsilon]$ at $r \simeq 1.114\varepsilon$. It is similar in form to the Peach–Koehler force between two screw dislocations (2.24).
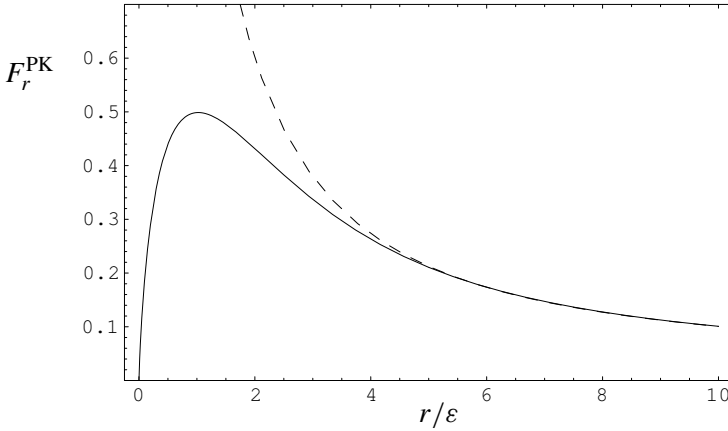
Figure 2. Peach–Koehler force F_r^{PK} between two screw disloca-
tions in nonlocal micropolar elasticity. F_r is given in units of
$\mu b_z/[2\pi\varepsilon]$ with $h = 2\varepsilon$ and $\mu = 3\eta$. The dashed curve represents
the micropolar result.

The Peach–Koehler force between an edge dislocation and the force stress pro-
duced by a wedge disclination is given by

$$F_x^{\mathrm{PK}} = b'_x t_{xy}, \tag{3.27}$$

$$F_y^{\mathrm{PK}} = -b'_x t_{xx}, \tag{3.28}$$

where [Lazar and Maugin 2004]

$$t_{xx} = \frac{\mu\Omega_z}{2\pi(1-\nu)}\left\{\ln r + \frac{y^2}{r^2} + K_0(r/\varepsilon) + \frac{(x^2-y^2)\varepsilon^2}{r^4}\left(2 - \frac{r^2}{\varepsilon^2}K_2(r/\varepsilon)\right)\right\}, \tag{3.29}$$

$$t_{xy} = -\frac{\mu\Omega_z}{2\pi(1-\nu)}\frac{xy}{r^2}\left\{1 - \frac{2\varepsilon^2}{r^2}\left(2 - \frac{r^2}{\varepsilon^2}K_2(r/\varepsilon^2)\right)\right\}. \tag{3.30}$$

F_x^{PK} is zero at $x = 0$ and $y = 0$ and has extremum values at $x = y$. On the other
hand F_y^{PK} has a finite extremum at $r = 0$ (see Figure 3). F_x^{PK} is the glide force
and F_y^{PK} is the climb force for the edge dislocation caused by the stress field of the
wedge disclination.

The Mathisson–Papapetrou force between a wedge disclination and the couple
stress produced by an edge dislocation reads

$$F_x^{\mathrm{MP}} = \Omega'_z m_{zy}, \tag{3.31}$$

$$F_y^{\mathrm{MP}} = -\Omega'_z m_{zx}, \tag{3.32}$$
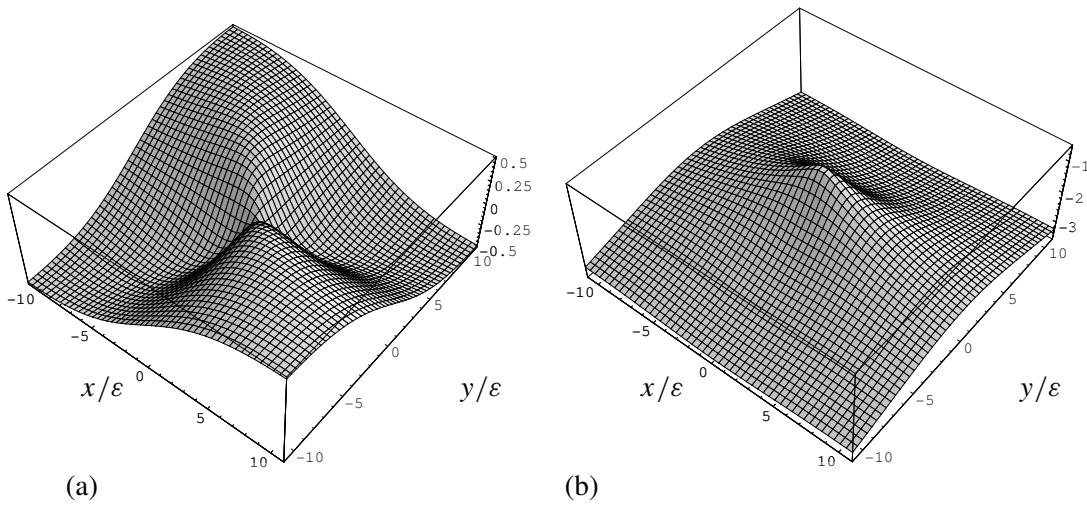
Figure 3. Peach–Koehler force between an edge dislocation and a wedge disclination: (a) F_x^{PK} and (b) F_y^{PK} are given in units of $\mu b_x' \Omega_z / [2\pi(1-\nu)]$.
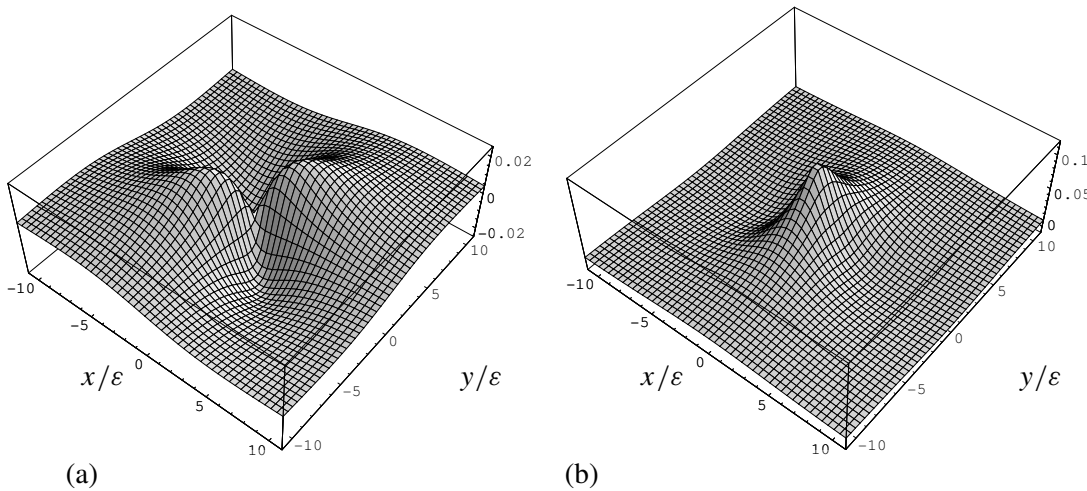
Figure 4. Mathisson–Papapetrou force between a wedge disclination and an edge dislocation: (a) F_x^{MP} and (b) F_y^{MP} are given in units of $(\beta+\gamma)\Omega_z b / [2\pi]$ with $l = 2\varepsilon$.

with the couple stress [Lazar and Maugin 2004]

$$
m_{zx} = \frac{(\beta+\gamma)b_x}{2\pi}\left\{\frac{x^2-y^2}{r^4}\left(1-\frac{1}{l^2-\varepsilon^2}\Big[lrK_1(r/l)-\varepsilon rK_1(r/\varepsilon)\Big]\right)\right.
$$
$$
\left. -\frac{x^2}{r^2}\frac{1}{l^2-\varepsilon^2}\Big[K_0(r/l)-K_0(r/\varepsilon)\Big]\right\}, \quad (3.33)
$$

$$m_{zy} = \frac{(\beta + \gamma)b_x}{2\pi} \frac{xy}{r^4} \left\{ 2\left(1 - \frac{1}{l^2 - \varepsilon^2}\Big[lr K_1(r/l) - \varepsilon r K_1(r/\varepsilon)\Big]\right) \right.$$
$$\left. - \frac{r^2}{l^2 - \varepsilon^2}\Big[K_0(r/l) - K_0(r/\varepsilon)\Big] \right\}. \quad (3.34)$$

F_x^{MP} is zero at $x = 0$ and $y = 0$ and has extremum values at $x = y$. F_y^{MP} has a finite extremum at $r = 0$ (see Figure 4).

The main feature in nonlocal micropolar elasticity is that the Peach–Koehler and the Mathisson–Papapetrou forces are nonsingular and they have finite extremum values unlike the results obtained in micropolar elasticity.

Acknowledgement

References

[Cherepanov 1981] G. P. Cherepanov, "Invariant Γ integrals", *Eng. Fract. Mech.* **14**:1 (1981), 39–58.

[Eringen 1999] A. C. Eringen, *Microcontinuum field theories I: foundations and solids*, Springer, New York, 1999.

[Eringen 2002] A. C. Eringen, *Nonlocal continuum field theories*, Springer, New York, 2002.

[Eshelby 1951] J. D. Eshelby, "The force on an elastic singularity", *Philos. Tr. R. Soc. S. A* **A244**:877 (1951), 87–112. London.

[Eshelby 1975] J. D. Eshelby, "The elastic energy-momentum tensor", *J. Elasticity* **5**:3–4 (1975), 321–335.

[Gairola 1981] B. K. D. Gairola, "Gauge invariant formulation of continuum theory of defects", pp. 55–65 in *Continuum models and discrete systems 4th Int. Symp.*, edited by O. Brulin and R. K. T. Hsieh, North-Holland, Amsterdam, 1981.

[Hehl et al. 1995] F. W. Hehl, J. D. McCrea, E. W. Mielke, and Y. Ne'eman, "Metric-affine gauge theory of gravity: field equations, Noether identities, world spinors, and breaking of dilation invariance", *Phys. Rep.* **258**:1–2 (1995), 1–171.

[Kirchner 1999] H. O. K. Kirchner, "The force on an elastic singularity in a nonhomogenous medium", *J. Mech. Phys. Solids* **47**:4 (1999), 993–998.

[Kluge 1969] G. Kluge, "Über den Zusammenhang der allgemeinen Versetzungstheorie mit dem Cosserat-Kontinuum", *Wissenschaftliche Zeitschrift der Technischen Hochschule Otto von Guericke Magdeburg* **13** (1969), 377–380.

[Kovács and Vörös 1979] I. Kovács and G. Vörös, "Lattice defects in nonlocal elasticity", *Physica B* **96**:1 (1979), 111–115.

[Kröner and Datta 1966] E. Kröner and B. K. Datta, "Nichtlokale elastostatik: ableitung aus der gittertheorie", *Z. Phys. A Hadron. Nucl.* **196**:3 (1966), 203–211.

[Lazar 2005] M. Lazar, "Peach-Koehler forces within the theory of nonlocal elasticity", pp. 149–158 in *Mechanics of material forces*, edited by P. Steinmann and G. A. Maugin, Springer, Berlin, 2005.

[Lazar and Maugin 2004] M. Lazar and G. A. Maugin, "Defects in gradient micropolar elasticity: II. Edge dislocation and wedge disclination", *J. Mech. Phys. Solids* **52**:10 (2004), 2285–2307.

[Lazar et al. 2005] M. Lazar, G. A. Maugin, and E. C. Aifantis, "An overview of dislocations in a special class of generalized elasticity", *Phys. Status Solidi B* **242**:12 (2005), 2365–2390.

[Lubarda and Markenscoff 2003] V. A. Lubarda and X. Markenscoff, "On conservation integrals in micropolar elasticity", *Philos. Mag.* **83**:11 (2003), 1365–1377.

[Maugin 1993] G. A. Maugin, *Material inhomogeneities in elasticity*, Chapman and Hall, London, 1993.

[Maugin 1998] G. A. Maugin, "On the structure of the theory of polar elasticity", *Philosophical Transactions: Mathematical, Physical and Engineering Sciences* **356**:1741 (1998), 1367–1395.

[Morse and Feshbach 1953] P. M. Morse and H. Feshbach, *Methods of theoretical physics*, McGraw-Hill, New York, 1953.

[Nowacki 1986] W. Nowacki, *Theory of asymmetric elasticity*, Pergamon Press, Oxford, 1986.

[Peach and Koehler 1950] M. O. Peach and J. S. Koehler, "Forces exerted on dislocations and the stress field produced by them", *Phys. Rev.* **80**:3 (1950), 436–439.

[Vukobrat and Kuzmanović 1992] M. Vukobrat and D. Kuzmanović, "Conservation laws in nonlocal elasticity", *Acta Mech.* **92**:1–4 (1992), 1–8.

MARKUS LAZAR: lazar@fkp.tu-darmstadt.de
Emmy Noether Research Group, Institute of Condensed Matter Physics, Darmstadt University of Technology, Hochschulstr. 6, D-64289 Darmstadt, Germany

HELMUT O. K. KIRCHNER: kirchnerhok@hotmail.com
Institut des Sciences des Materiaux, Université Paris-Sud, F-91405 Orsay, France

TEMPERATURE DEPENDENCE OF A NiTi SHAPE MEMORY ALLOY'S SUPERELASTIC BEHAVIOR AT A HIGH STRAIN RATE

WEINONG CHEN AND BO SONG

The temperature dependence of the dynamic compressive stress-strain behavior of a NiTi shape memory alloy (SMA) has been determined at a strain rate of 4.35×10^2 s^{-1} with a split Hopkinson pressure bar (SHPB) modified for obtaining the dynamic stress-strain loops at constant strain rates. The environmental temperature was varied from 0 to 50° C, where the SMA exhibits superelastic behavior through stress-induced martensite (SIM). Experimental results show that both the loading and unloading portions of the stress-strain loop are significantly temperature dependent. Shape memory effect below the austenite finish transition temperature, A_f, and superelastic behavior above A_f are also observed at the high strain rate.

1. Introduction

Engineering and medical applications of shape memory alloys (SMA) have attracted research efforts to reveal the mechanical responses of these materials. Recent research in the field is extensively reviewed, for example, by [Birman 1997], [Otsuka and Wayman 1988], [James and Hane 2000], and [Bhattacharya and James 2005]. Reversible martensitic phase transformations provide SMA's with the capabilities of shape memory and superelastic deformation. An SMA possesses an austenite phase at high temperature and a martensitic phase at low temperature. There is a range of transition temperatures over which temperature-induced phase transformation occurs. An object that appears to be permanently deformed at low temperature in its martensitic phase will return to its original shape in austenite phase when heated above the transition temperature, because of a solid phase transformation from martensitic back to austenite. Two-way shape memory has also been proposed where the alloy remembers its shapes at both low and high temperatures [Wayman 1993]. It has been proposed that the shape memory effects may be utilized to build small-scale machines [Bhattacharya and James 2005].

In addition to the temperature-induced phase transformations, martensitic transformation can be induced by mechanical stress in a certain temperature range. The

range starts just above the temperature-induced phase transformation region (typically marked by the austenite finish transformation temperature A_f) and ends at a temperature above which stress-induced martensitic transformation is no longer possible, M_d [Stoeckel and Yu 1991; Wayman 1993; Duerig and Pelton 1994]. The alloy is austenite in this temperature range when stress-free. When the applied stress exceeds a threshold level, called the on-set stress for stress-induced martensite (SIM), the austenite crystal structure will transform into a martensitic phase. Therefore, martensitic transformation in the material can occur at much higher temperatures when assisted by mechanical stress. This SIM-induced deformation is achieved by detwinning [Bhattacharya 2003; Liu et al. 2002], which requires much less energy than to deform the austenite by conventional metal deformation mechanisms. It takes a stress level an order of magnitude higher to deform the austenite material than the martensite at the same strain [Wayman 1993]. The detwinning in the martensitic phase can accumulate up to 10% strain, which can be recovered completely by the reverse transformations back to austenite when the applied stress is removed. This large but reversible deformation is named superelasticity, a distinct property of SMA's.

When the temperature is in the superelasticity range, $A_f < T < M_d$, the typical stress-strain curve of an SMA obtained under quasistatic loading conditions at engineering strain rate, $\dot{\varepsilon} = 1.2 \times 10^{-4}/s$ is shown in Figure 1 [Song and Chen 2004].
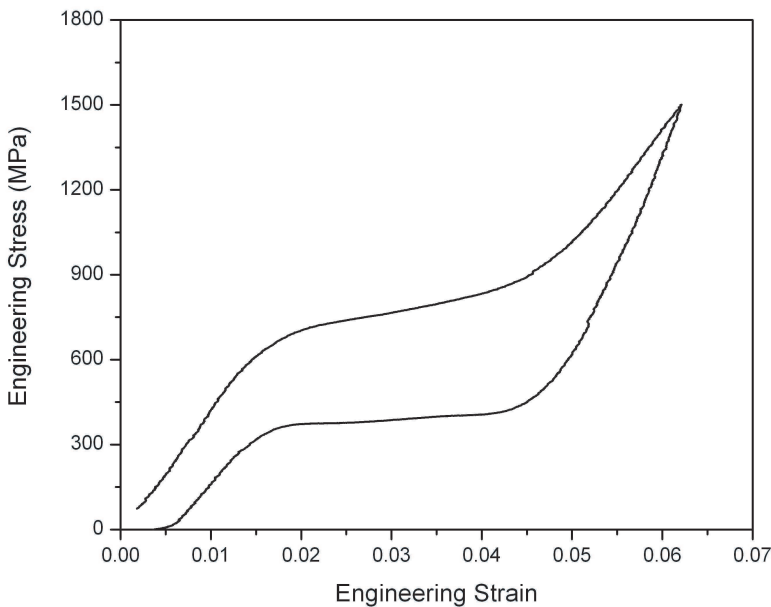
Figure 1. A quasistatic stress-strain curve of a NiTi SMA.

The compressive stress-strain curve contains two plateaus. The upper plateau corresponds to the loading portion where mechanical stress drives the austenite to transform into martensite and where detwinning occurs in the newly formed SIM. The lower plateau represents the stress-strain behavior of the SMA during the unloading process where the SIM is transforming back to austenite. Thermal energy is needed to assist this reverse transformation. This stress-strain curve indicates that, even though the mechanical behavior is termed superelastic because there is hardly any permanent deformation upon unloading, part of the mechanical energy used to deform the alloy is lost during unloading. Therefore, the loading/unloading cycle on the SMA leaves no permanent deformation, but dissipates mechanical energy. The volume density of the lost energy in a loading/unloading cycle can be computed from the area within the hysteresis loop on the stress-strain curve shown in Figure 1. Such an energy loss over a loading/unloading cycle makes it possible for the shape memory alloy to be used as a shock/vibration absorption medium. In fact, SMA's have been described as "quiet" alloys [Schetky and Perkins 1978]. In order to properly use the SMA's as an energy-absorbing member in applications subjected to high-rate loading, it is essential to quantitatively determine and understand the constitutive behavior of these alloys under dynamic loading conditions.

Considerable research efforts have been focused on the dynamic mechanical responses of SMA's under high rates of loading. Ogawa [1988] used a conventional split Hopkinson pressure bar (SHPB) and characterized a shape memory alloy over a temperature range of 201 K to 363 K. Lin et al. [1996] and Wayman [1993] studied the rate effects on the mechanical behavior of a NiTi alloy within the quasistatic range. Chen et al. [2001] determined the compressive behavior of a NiTi SMA over the strain rate range of $0.001\text{--}750 \text{ s}^{-1}$. A modified SHPB was used to control the loading profile such that the loading portion of the stress-strain curve was obtained at a constant strain rate. It was found that the on-set stress for SIM depends on the strain rate, and that the strain history in the specimen lags behind the stress history. Similar high-rate experimental results have also been reported recently in compression by Nemat-Nasser et al. [2005] and Nakayama et al. [2005], and in tension by Liu et al. [2002]. By using a smaller specimen to achieve higher strain rates in their dynamic compression experiments, Nemat-Nasser et al. [2005] found that the stress-strain relation of their NiTi SMA differs from that achieved at a moderately high strain rate. By comparing the shape memory effects produced by different rates of loading, Belyaev et al. [2002] discovered that an increase in strain rate can lead to an increase in shape memory effect. At even higher strain rates, Millett et al. [2002] studied the shock response of NiTi SMA. Dai et al. [2004] investigated the propagation of macroscopic phase boundaries under impact loading analytically and numerically, and found that the thermal effect should be taken into account due to the temperature coupling nature of the phase transition process. The effects of

loading conditions other than one-dimensional were explored recently [McNaney et al. 2003]. Besides the experimental research, analytical modeling is also under intensive development; examples include [Lagoudas et al. 2003], [Iadicola and Shaw 2004], and [Lovey et al. 2004].

Although the mechanical behavior of SMA's has been experimentally explored recently at high strain rates, most dynamic experiments have been performed in a controlled manner only during the loading portions of the stress-strain curves. The unloading portions have been left uncontrolled due to difficulties in dynamic experimentation. However, since both of the loading and unloading portions of stress-strain curves for an SMA are important in determining the energy absorbing capacity of the material at high rates, it is desirable that the loading conditions during both loading and unloading stages in a dynamic experiment be precisely controlled to produce dynamic loading and unloading stress-strain loops at constant strain rates. Furthermore, most of the dynamic experiments are performed at room temperature. Since the superelastic behavior of SMA's only exists within a certain temperature range, it is important to determine the high-rate behavior over this temperature range. In this paper, we employ a recently developed SHPB technique, which produces dynamic compression experimental results for a stress-strain loop at a common constant strain rate over both loading and unloading portions [Song and Chen 2004], to conduct the dynamic compressive experiments on a NiTi SMA at a high strain rate over a temperature range where superelastic behavior is clearly observed.

A valid SHPB experiment requires that the specimen undergo homogeneous deformation under dynamic stress equilibrium [Gray 2000; Song et al. 2003]. In addition, a constant strain rate is convenient for the purpose of developing more accurate constitutive models based on experimental results. To ensure that these conditions were satisfied when dynamically testing the SMA specimens using a SHPB in this study, pulse shaping techniques were employed to control both the loading and unloading portions of the incident pulse so that the specimen was deformed under valid dynamic testing conditions. The dynamic stress equilibrium in the specimen was monitored using 1-wave/2-wave analysis [Gray 2000] on the nearly nondispersive waves created by the pulse-shaping technique. Experiments were conducted on Nitinol SE508 SMA. The following sections briefly describe the experiments and then present the results for the NiTi alloy.

2. Experimental setup and specimen preparation

2.1. *Experimental setup of the modified SHPB.* To study the dynamic compressive loading/unloading behavior of the SMA, a SHPB with pulse shaping on both loading and unloading stages [Song and Chen 2004] was used to conduct dynamic
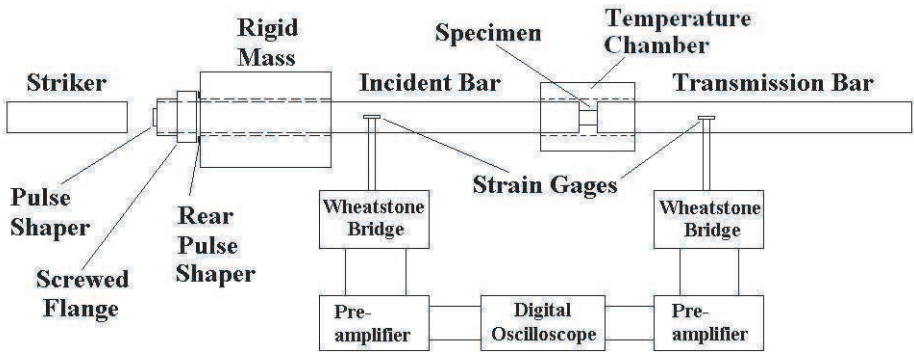
Figure 2. A schematic of the modified split Hopkinson pressure bar.

compressive experiments. A temperature chamber covered the testing section of the SHPB. A schematic of the modified SHPB facility is shown in Figure 2. In addition to the striker, incident and transmission bars included in a standard SHPB, the modified SHPB apparatus consists of front and rear pulse shapers in association with a momentum trapping system to facilitate constant strain-rate loading and unloading on the specimen. The front pulse-shaper is attached at the impact end of the incident bar to control the loading profile in the incident pulse; the rear pulse-shapers are placed on the surface of the rigid mass. A gap between the flange, which is screwed on the impact end of the incident bar, and the rigid mass, through which the incident bar passes, is precisely preset. Upon the impact of the striker during an experiment, the front pulse shaper is extensively compressed, generating a desired incident loading profile to deform the specimen at a constant strain rate under dynamic stress equilibrium over the loading phase of the experiment. In the meantime, a compression wave is also generated and then propagates in the striker. This compression wave is reflected back as an unloading wave (tension wave) when it arrives at the free end of the striker. The unloading pulse from the striker travels into and then passes through the incident bar to unload the specimen. In this momentum trap design, the gap between the flange and the rigid mass is precisely controlled to close before the unloading pulse travels into the incident bar. The rear pulse shapers on the surface of the rigid mass are compressed by the flange, generating a desired unloading profile in the incident bar. The unloading pulse is well designed by varying the dimensions of the rear pulse shapers to ensure that the specimen recovers at the same constant strain rate under dynamic stress equilibrium during unloading. Thus, a dynamic stress-strain loop is obtained at the same constant loading and unloading strain rate.

In the modified SHPB, a heating/cooling chamber is placed between the incident bar and the transmission bar (around the specimen) to create environmental temperatures that are monitored with an embedded thermocouple inside the heater/cooler chamber. Heating is automatically controlled by a temperature controller, manufactured by WATLOW (96A0-DAAA-00RG), while cooling is manually controlled by pouring liquid nitrogen into a surrounded 6.0 mm diameter brass tube inside the heater/cooler. The environmental temperature inside the chamber is controlled through the flow rate control of liquid nitrogen inside the brass tube when testing below room temperature. This temperature chamber was used in SHPB experiments on other materials and its details were described in a previous paper [Song et al. 2005]. Furthermore, besides the environmental temperature, the adiabatic temperature rising and falling during the forward and reverse SIM in the specimen is monitored by a small thermocouple embedded in one of the specimens. The specimen temperature measuring technique was also previously employed and a detailed description of it can be found in a previous publication [Song et al. 2003].

The lengths of the VascoMax maraging steel bars used for the experiments were 1830, 762, and 305 mm for the incident, transmission, and striker bars, respectively, with a common diameter of 12.3 mm. The strain signals sensed by strain gages from the incident and transmission bar surfaces were recorded using a high-speed digital storage oscilloscope. Since the temperature variation range (0–50° C) in the chamber is not significant enough to affect the stress-wave propagation through the temperature gradient along the bar length, the temperature gradient effects on the wave propagation in the bars are neglected in data reduction.

2.2. *Materials and specimens.* The SMA investigated in this research is NDC (Nitinol Devices & Components, Fremont, CA) SE508, nominally 55.8% nickel by weight and the balance titanium. The specified density is 6.5 g/cm^3, with an austenite finish transition temperature A_f of 5–18° C and a melting point of 1310° C, as provided by the manufacturer. The temperature above which stress-induced martensitic transformation is no longer possible, M_d, for SE508 is approximately 150° C. The temperature range for this alloy to exhibit superelasticity (the capability of returning to its original shape upon unloading after a substantial deformation) is 15–150° C. In a previous study [Chen et al. 2001], it was found that the strain far lagged behind the stress during the dynamic unloading stages in experiments performed at room temperature (close to A_f), indicating that there may exist unique dynamic deformation phenomena in the alloy during high-rate unloading near room temperature. However, the unloading paths were uncontrolled in those experiments. Therefore, the choice of the temperature range in this study reflects the effort to focus on the behavior near A_f (0–50° C) under much better controlled experimental conditions.
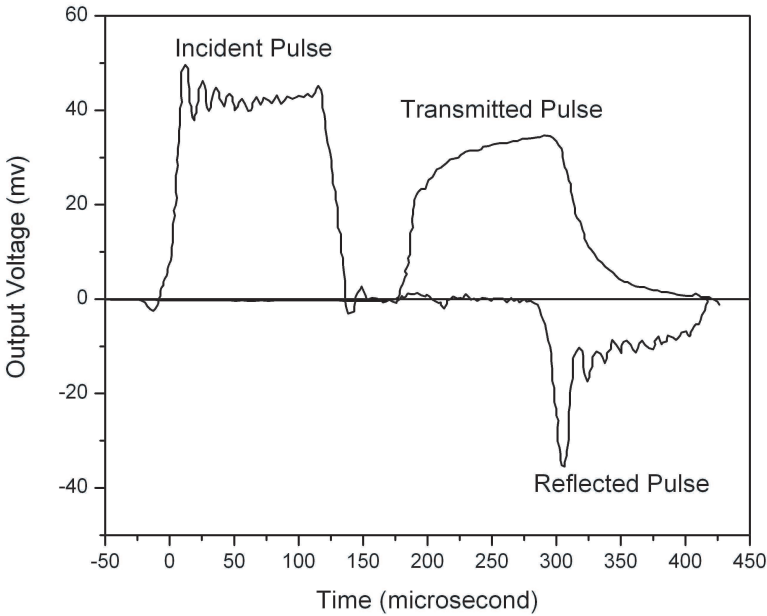
Figure 3. Typical oscilloscope records of a conventional SHPB experiment on an SMA.

Cylindrical specimens of 5.94 mm diameter by 6.10 mm long were sliced from a heat-treated Nitinol rod with a water-jet cooled abrasive saw. Chilled water was continuously sprayed on the contact area of the cutting blade and the Nitinol bar to keep the temperature low in the NiTi alloy during machining.

2.3. *Dynamic experimental technique for the SMA.* When the SHPB is used in its conventional configuration, there are high-frequency oscillations riding on the base incident pulse as shown in Figure 3. These oscillations in the loading pulse profile, which are the results of wave dispersion due to radial inertia effects when the wave propagates in the bars, cause complications in the loading conditions in the specimen. The nature of stress-induced phase transformation makes the specimen sensitive to alternations of loading and unloading. Figure 3 shows a typical oscilloscope record of an SMA undergoing dynamic loading by a conventional SHPB. There is no plateau in the reflected pulse shown in Figure 3, which indicates that the specimen does not deform at a constant strain rate. The oscillations in the reflected pulse also indicate that the specimen experiences a complicated deformation process. The point-wise dynamic material properties extracted by averaging the response over the specimen's volume cannot be expected to be accurate and reliable when the specimen is undergoing such a complicated deformation.

In order to obtain reliable dynamic material behavior from an SHPB experiment for the SMA, the rapid oscillations in the incident pulse must be avoided. Furthermore, efforts must be made to ensure that a dynamic equilibrium stress state is reached in the specimen during the experiment and to maintain a constant strain rate in the specimen. A homogeneously-deforming specimen under a dynamic equilibrium state of stress makes the volume-average of the specimen's behavior representative of the point-wise material properties. A constant strain rate during an experiment makes it convenient to report the stress-strain response as a function of strain rates, which is necessary to fit material constants in constitutive models. To obtain the stress-strain loop of an SMA at a constant strain rate, it is also desirable to maintain the same strain rate during both loading and unloading phases of the experiment. These requirements pose significant challenges in the dynamic experiment design.

To eliminate the oscillations in the loading pulse and to ensure that the specimen deforms at a constant strain rate in dynamic equilibrium, a pulse shaper was placed on the impact end of the incident bar to control the shape of the loading pulse (Figure 2). There are a variety of pulse-shaping devices. For example, Duffy et al. [1971] used a pulse-shaper in the form of a concentric tube to smooth pulses generated by explosive loading in a torsional Hopkinson bar. Ravichandran and Chen [1991] and Frew et al. [2002] used copper pulse shapers to achieve ramp loading profiles when testing ceramics and rocks using SHPB. Togami et al. [1996] used a pulse shaper in a modified SHPB to control the loading pulse shape and to filter out high-frequency components in the incident pulse for accelerometer calibrations up to 200,000 g. In the present research, annealed C11000 copper disks were used to control the shape of the loading pulse during both loading and unloading phases. The diameters and the thickness of the pulse shapers necessary to control the strain rate in the SMA specimen at a desired and constant level were difficult to determine since the dynamic behavior of the SMA was unknown before the experiments. Trial experiments were conducted to select a proper pulse shaper for each combination of desired strain rate, environmental temperature, and maximum strain.

3. Experiments and Results

3.1. *Dynamic experiments with the modified SHPB.* A typical set of incident, reflected, and transmitted signals recorded with the oscilloscope for an SHPB experiment on the NiTi SMA at 23° C with shaped pulses is shown in Figure 4. The dashed line is the transmitted pulse, which records the stress history in the specimen. The first pulse in the solid line is the incident pulse, whereas the second and third are the reflected pulses associated with the deformations during loading
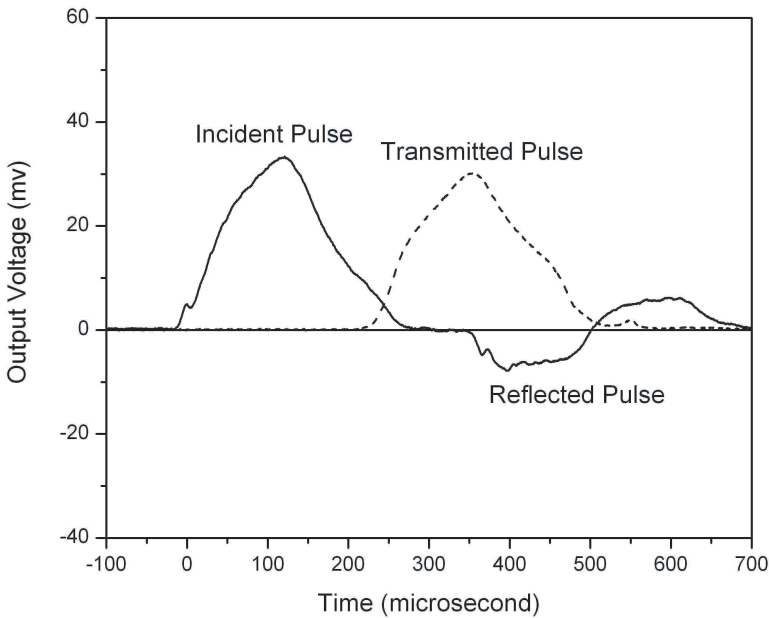
Figure 4. Oscilloscope records of a pulse-shaping SHPB experiment on a NiTi SMA.

and unloading in the specimen. If the specimen's mechanical impedance ($\rho c A$, where ρ is the mass density, c the bar wave velocity, and A the cross-sectional area) is less than that of the bar, the incident and first reflected pulses are opposite in sign, as seen in Figure 4. Furthermore, the second reflected pulse indicates the strain rate of the recovery deformation in the specimen and has an opposite sign from the first reflected pulse. Although the first and second reflected pulses have opposite signs, the magnitudes are nearly the same, which indicates that the strain rates at both the loading and unloading stages are nearly the same. We also note that the magnitudes of the reflected pulses are nearly constant in each of the reflected pulses, which indicates that the loading and unloading strain rates are nearly the same constant. This is a necessary condition to obtain a dynamic stress-strain loop or hysteretic loop at a certain strain rate. A comparison between the incident and reflected pulses of Figures 3 and 4 indicates that the shape of the incident pulse in the pulse-shaped experiment shown in Figure 4 is very different from that obtained in a conventional SHPB experiment (Figure 3). The high-frequency oscillations on the incident pulse shown in Figure 3 are eliminated after pulse shaping, as shown in Figure 4. The nearly flat and equal plateaus on the first and second reflected pulses shown in Figure 4 indicate that the specimen deformed at a nearly constant strain rate during both loading and unloading on the NiTi SMA specimen in a pulse-shaped SHPB experiment. The transmitted pulses were used to calculate stress
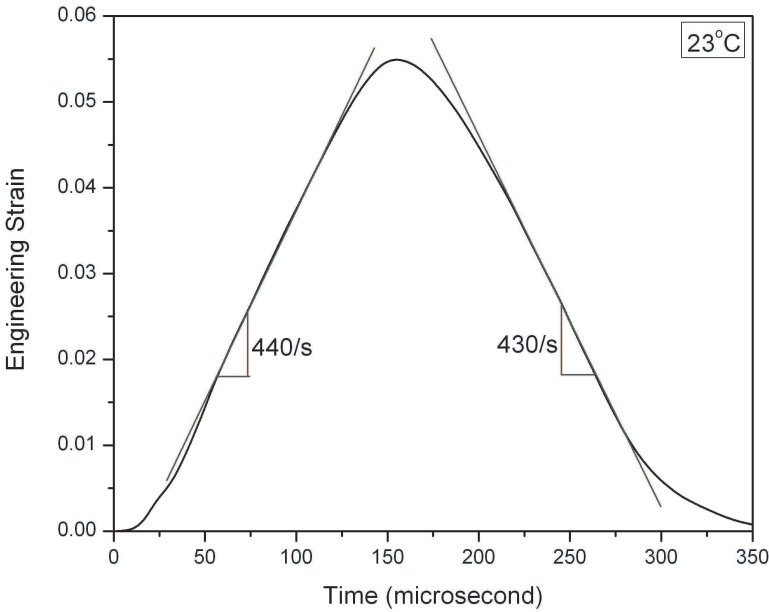
Figure 5. Dynamic loading and unloading strain histories in the NiTi Specimen.

history in the NiTi specimen according to conventional SHPB theory based on 1-D wave propagation theory [Kolsky 1949; Lindholm 1964; Follansbee 1985; Gray 2000]. The strain rate in the experiment was determined by the average magnitude of the plateau in the reflected pulses.

The dynamic strain history in the specimen is shown in Figure 5. An examination of Figure 5 shows that both the loading (rising) and unloading (descending) portions of the strain histories are nearly straight lines, indicating constant rates of deformation. The slope of the loading portion, that is, the loading strain rate, is observed to be 440/s, whereas that for the unloading portion is 430/s. Therefore, the strain rates during both loading and unloading stages may be considered as the same constant. It should be noted that the dynamic strain history in a NiTi SMA lags behind the dynamic stress [Chen et al. 2001; Song and Chen 2004], which needs to be taken into account in the pulse shaper design.

In SHPB experiments, dynamic stress equilibrium in the specimen is a fundamental requirement for valid data processing because equilibrium is one of the basic assumptions upon which SHPB theory is built [Meyers 1994; Gray 2000]. When dynamic equilibrium is impossible to reach, the constitutive behavior may be still found using a hybrid approach assisted by computer simulation if the form of the stress-strain relation is known. However, in the case of the NiTi SMA, the dynamic stress-strain relations need to be determined by the SHPB experiments.
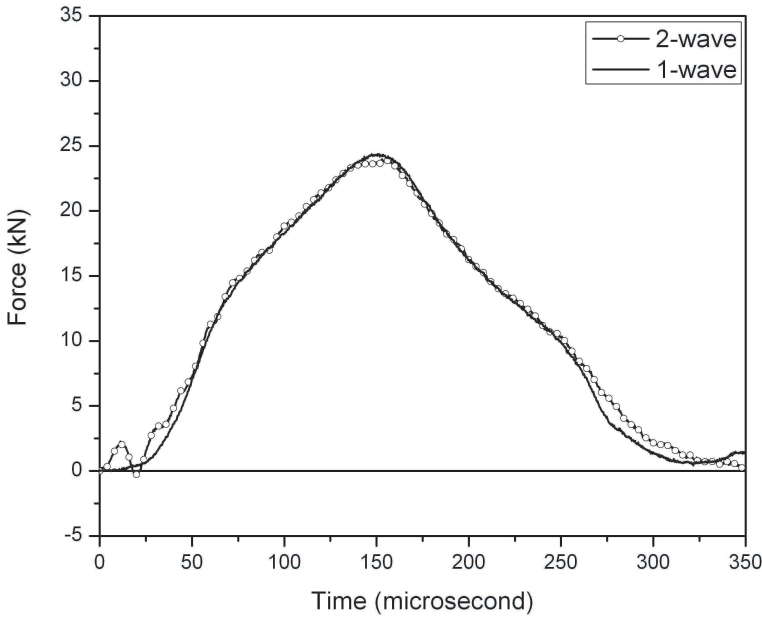
Figure 6. Dynamic equilibrium analysis by 1-wave/2-wave method.

Dynamic equilibrium must be achieved such that the volume average of the stress-strain behavior over the entire specimen may be used as the point-wise material responses at a certain strain rate. It is therefore critical to ensure that the dynamic equilibrium conditions are satisfied in order to obtain valid experimental results.

In our experiments, we checked stress equilibrium by comparing the transmitted signal (1-wave) with the difference between the incident and reflected signals (2-wave) [Gray et al. 1997; Wu and Gorham 1997; Gray 2000]. The incident waves created in our experiments by pulse shaping are nearly nondispersive since there are few high-frequency components associated with the main loading pulses. The results on stress equilibrium from 1-wave, 2-wave analysis are thus considered to be reliable. Figure 6 shows the results of such an analysis on the stress pulses of Figure 4. There are two nearly overlapping curves shown in Figure 6. One is the transmitted pulse profile (1-wave), and the other is the difference between the incident and the reflected pulses (2-wave). The 2-wave curve represents the axial force history on the face of the specimen that is in contact with the incident bar [Gray et al. 1997; Wu and Gorham 1997]. The 1-wave curve, which is used to calculate the stress history in SHPB data reduction, is the axial force history on the face of the SMA specimen that is in contact with the transmission bar during an experiment. Dynamic stress equilibrium in the specimen requires that the 1-wave and the 2-wave are the same over the duration of the SHPB experiment. Figure 6
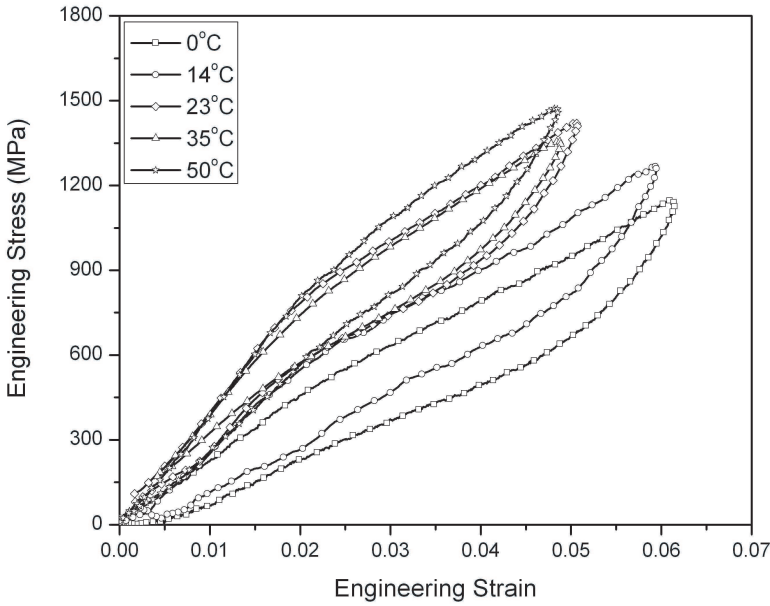
Figure 7. Dynamic compressive stress-strain curves of the NiTi SMA.

shows that the 1-wave curve nearly overlaps the 2-wave curve in this pulse-shaped experiment, which indicates that dynamic equilibrium stress state in the specimen has been achieved. It is noted that the fluctuations are observed in the 2-wave curve in Figure 6 within the first 20 μs. The slight wave dispersion, even though the pulse shaping technique was employed, may result in fluctuations in the subtraction of the reflected pulse from the incident pulse. However, the fluctuations do not affect the examination of the stress equilibration process in the specimen, as shown in Figure 6. The amplitude of the reflected pulses is nearly constant as shown in Figure 5, which indicates a constant strain rate over the duration of the experiment, including both loading and unloading phases. Therefore the SHPB experiment recorded in Figure 4 is a valid dynamic experiment. Such analysis was performed on every experiment presented in this paper.

3.2. *Dynamic stress-strain loops at various temperatures for the SMA.* Figure 7 summarizes the dynamic compressive stress-strain curves at a common dynamic compression strain rate of 430/s for the Nitinol SE508 SMA at environmental temperatures of 0°, 14°, 23°, 35°, and 50°. The stress-strain curves are plotted in the units of engineering stress and strain, which are nearly identical to true stress and strain curves, since the maximum strain ($\sim$ 6%) experienced by the specimens is small. As shown in Figure 7, the shape of the dynamic stress-strain curves has similarities and differences as compared to those obtained under quasistatic

loading conditions. The quasistatic compressive stress-strain curve obtained at a temperature well above A_f is typically in the form presented in Figure 1 (23° in this case). After a linearly elastic region, the stress-strain curve bends into a plateau-like region caused by stress induced martensite starting at a strain of $\sim$ 1.5–2%. If the deformation is too large (for example, > 4.5% strain), significant work-hardening in this SMA drives a steep rise in the stress-strain curve. The reverse transformation from martensite to austenite upon unloading takes place at a smaller stress, which is schematically displayed by the lower plateau in the stress-strain loop. At the dynamic strain rate of 430/s, we can examine one of the stress-strain curves (for example, at 23°) to illustrate its similarities and differences from its quasistatic behavior. Similar to the quasistatic case, the dynamic stress-strain curve also exhibits a linear initial portion up to a strain of $\sim$ 1.7–2%, where a transition in the stress-strain curve occurs (on-set stress for SIM). However, instead of following a plateau-like stress-induced martensite transformation region, the dynamic stress-strain curve displays a work-hardening behavior after the SIM on-set stress. A similar behavior is observed on the unloading branch of the stress-strain loop. There is no plateau observed for the reverse phase transformation. Instead, a delayed unloading curve that is nearly parallel to the loading curve is observed. This clear difference from quasistatic behavior may be an indication that SIM phenomena are rate dependent. Adiabatic temperature change in the specimen under high-rate deformation is considered to be one of the parameters that depicts this shape change in stress-strain curve from quasistatic to dynamic loading conditions. The discussion on this point will be continued after results on the adiabatic temperature change are presented.

The results shown in Figure 7 also clearly illustrate the effects of temperature on the high-rate mechanical response of the SMA. Although the environmental temperature was varied only within a rather narrow range, the resultant stress-strain responses show clear temperature dependence. As the environmental temperature decreases from 50° to 0° C, the slopes of both the loading and unloading portions of the stress-strain loops decrease. When the environmental temperature is below room temperature, i.e., 14° C and 0° C, the strain does not return to zero when the specimen is completely unloaded, as indicated by the residual strain in the unloading stress-strain curves at these temperatures. The specimen eventually recovers all strains at room temperature. This phenomenon indicates that, under high-rate deformation, the reverse phase transformation, which needs thermal energy input, may not be as fast as the forward SIM, which is driven by mechanical stress wave loading. Furthermore, when the environmental temperature is close to A_f of the alloy, the drawing of heat from the specimen to assist the reverse transformation may be so much that the temperature in the specimen is actually below A_f during the unloading, which stops the reverse transformation until more heat is drawn from
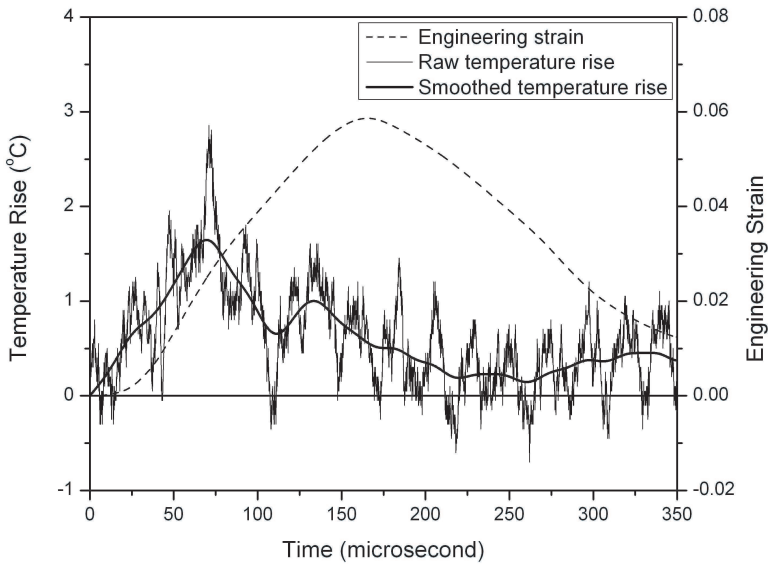
Figure 8. Local temperature variations in the NiTi SMA during
dynamic deformation.

the environment around the specimen to drive the specimen temperature back above
A_f. It is also observed that, although $0°$ C is below A_f of the SMA, superelastic-
ity is still reached. This indicates that the forward SIM driven by stress actually
releases heat into the specimen, resulting in the actual specimen temperature above
A_f during the superelastic deformation in the specimen. A similar phenomenon
was also observed in a previous study, although the testing conditions were not
controlled during the unloading stages of the dynamic experiments [Chen et al.
2001].

3.3. *Temperature variation in specimen during dynamic tests.* To better under-
stand the temperature effects on the mechanical response of the NiTi SMA, be-
sides the variations in environmental temperature, the temperature change in the
specimen during dynamic deformation needs to be monitored. To accurately mon-
itor the local temperature change during high-rate deformation, a small hole was
drilled along the centerline of the cylindrical specimen from one end, and a small
thermocouple was placed inside [Song et al. 2003]. The temperature variation
during one experiment from the base room temperature is shown in Figure 8. The
results in Figure 8 indicate that the specimen temperature rises with increasing
strain in the specimen during early stages of loading. The forward SIM driven by
mechanical stress waves clearly releases heat as evidenced by the temperature rise.
Partly due to the small superelastic strain achieved in the experiments, the ampli-
tude of the temperature rise is rather small. A distinctive and interesting feature

recorded in the temperature history is that the temperature starts to decrease before the maximum strain is reached. In the records shown in Figure 8, the specimen temperature recorded by the small thermocouple starts to decrease at only half of the maximum strain. This repeatable phenomenon in experiments indicates that the deformation in the specimen may not be uniform. With a small hole drilled to accommodate the thermocouple, there are inevitable stress concentrations around the hole, where the SIM driven by the locally concentrated stresses occurs much earlier than the rest of the specimen, resulting in locally higher temperature around the hole. At half of the average maximum strain experienced by the entire specimen, the SIM near the hole is completed. Without further thermal energy released from the SIM, the heat in the specimen actually flows from near the hole to the rest of the specimen, resulting in a decrease in local temperature even though the entire specimen is still under compression. Iadicola and Shaw [2004] showed analytically that SIM occurs at localized transformation fronts even in samples without stress concentrations. The results in Figure 8 also clearly show the existence of specimen temperature changes during dynamic deformation of the SMA. This fluctuation in temperature may drive the specimen material above A_f even though the starting temperature is below A_f. On the other hand, during the reverse transformation, the specimen temperature may drop below A_f to stop the reverse transformation. These results indicate that, depending on the environmental temperature, there is a frequency range where the SMA may be used for dynamic energy dissipation. If the loading/deformation frequency in the application is too high, the transformations— in particular, the reverse transformation in the alloy—may not be fast enough to satisfy the impact/vibration energy dissipation requirements. Furthermore, the temperature variation in the specimen causes a changing temperature behind a specific dynamic stress-strain curve. This nonisothermal testing condition during stress-wave loading causes the shape of the resultant stress-strain curves to deviate from the typical curves obtained under quasistatic loading conditions.

4. Conclusions

A new SHPB technique was employed to determine the dynamic compressive stress-strain behavior of a Nitinol SE508 SMA in its superelasticity phase at a dynamic strain rate of 430/s over an environmental temperature range of 0° to 50° C. The experimental technique had pulse shapers to control both the dynamic loading and unloading profiles to control the loading on the specimen so that the specimen deformed under dynamic stress equilibrium at a constant strain rate during both the loading and unloading stages of the experiments. Local temperature in the specimen was also monitored during dynamic deformation to record the effects of adiabatic heating and cooling associated with the high-rate deformation of the

specimen. Valid dynamic testing conditions were checked for each experiment to ensure that the experiments were valid and the results accurate.

Experimental results show that the compressive stress-strain behavior of the SMA is dependent on both temperature and strain rate. The higher the environmental temperature within the range explored in this study, the stiffer the stress-strain behavior. The shape of the dynamic stress-strain loops obtained over the entire environmental temperature range differs from those obtained quasistatically. Under dynamic loading conditions, the forward and reverse SIM plateaus are inclined, mainly due to temperature variations in the specimen caused by adiabatic heating and cooling during dynamic loading and unloading. The unloading portions of the dynamic compressive stress-strain curves obtained below room temperature exhibit residual strains after the specimen is completely unloaded. This is because the reverse transformation may be stopped when the specimen temperature drops below A_f during unloading.

References

[Belyaev et al. 2002] S. P. Belyaev, N. F. Morozov, A. I. Razov, A. E. Volkov, L. Wang, S. Shi, S. Can, J. Chen, and X. Dong, "Shape-memory effect in titanium-nickel after preliminary dynamic deformation", *Mater. Sci. Forum* **394–395** (2002), 337–340.

[Bhattacharya 2003] K. Bhattacharya, *Microstructure of Martensite: why it forms and how it gives rise to shape memory effort*, Oxford Univ. Press, Oxford, 2003.

[Bhattacharya and James 2005] K. Bhattacharya and R. D. James, "The material is the machine", *Science* **307**:5706 (2005), 53–54.

[Birman 1997] V. Birman, "Review of mechanics of shape memory alloy structures", *Appl. Mech. Rev.* **50**:11 (1997), 629–645.

[Chen et al. 2001] W. W. Chen, Q. Wu, J. H. Kang, and N. A. Winfree, "Compressive superelastic behavior of a NiTi shape memory alloy at strain rates of 0.001-750 s^{-1}", *Int. J. Solids Struct.* **38**:50–51 (2001), 8989–8998.

[Dai et al. 2004] X. Dai, Z. P. Tang, S. Xu, Y. Guo, and W. Wang, "Propagation of macroscopic phase boundaries under impact loading", *Int. J. Impact Eng.* **30**:4 (2004), 385–401.

[Duerig and Pelton 1994] T. W. Duerig and A. R. Pelton, "Ti-Ni shape memory alloys", pp. 1035–1048 in *Materials properties handbook, titanium alloys*, ASM International, Materials Park, Ohio, 1994.

[Duffy et al. 1971] J. Duffy, J. D. Campbell, and R. H. Hawley, "On the use of torsional Hopkinson bar to study rate effects in 1100-0 aluminum", *J. Appl. Mech. (ASME)* **38** (1971), 83–91.

[Follansbee 1985] P. S. Follansbee, "The Hopkinson bar", pp. 198–217 in *Mechanical testing, metals handbook*, 9th ed., vol. 8, American Society for Metals, Materials Park, Ohio, 1985.

[Frew et al. 2002] D. J. Frew, M. J. Forrestal, and W. Chen, "Pulse-shaping techniques for testing brittle materials with a split Hopkinson pressure bar", *Exp. Mech.* **42**:1 (2002), 93–106.

[Gray 2000] G. T. Gray, III, "Classic split-Hopkinson pressure bar testing", pp. 462–476 in *ASM handbook: Mechanical testing and evaluation*, vol. 8, edited by H. Kuhn and D. Medlin, ASM International, Materials Park, Ohio, 2000.

[Gray et al. 1997] G. T. Gray, III, W. R. Blumenthal, C. P. Trujillo, and R. W. Carpenter, II, "Influence of temperature and strain rate on the mechanical behavior of Adiprene L-100", *J. Phys. (France) IV* **7**:C3 (1997), 523–528. International Conference on Mechanical and Physical Behaviour of Materials under Dynamic Loading (EURODYMAT'97), Toledo (Spain), 22-26 Sep 1997.

[Iadicola and Shaw 2004] M. A. Iadicola and J. A. Shaw, "Rate and thermal sensitivities of unstable transformation behavior in a shape memory alloy", *Int. J. Plast.* **20**:4–5 (2004), 577–605.

[James and Hane 2000] R. D. James and K. F. Hane, "Martensitic transformations and shape-memory materials", *Acta Mater.* **48**:1 (2000), 197–222.

[Kolsky 1949] H. Kolsky, "An investigation of mechanical properties of materials at very high rates of loading", *P. Phys. Soc. Lond. B* **62**:11 (1949), 676–700.

[Lagoudas et al. 2003] D. C. Lagoudas, K. Ravi-Chandar, K. and. Sarh, and P. Popov, "Dynamic loading of polycrystalline shape memory alloy rods", *Mech. Mater.* **35**:10.1016/S0167-6636(02)00199-0 (2003), 689–716.

[Lin et al. 1996] P. H. Lin, H. Tobushi, K. Tanaka, T. Hattori, and A. Ikai, "Influence of strain rate on deformation properties of TiNi shape memory alloy", *JSME Int. J. A Mech. M.* **39**:1 (1996), 117–123.

[Lindholm 1964] U. S. Lindholm, "Some experiments with the split Hopkinson pressure bar", *J. Mech. Phys. Solids* **12**:5 (1964), 317–335.

[Liu et al. 2002] Y. Liu, Y. Li, and K. T. Ramesh, "Rate dependence of deformation mechanisms in a shape memory alloy", *Philos. Mag. A* **82**:12 (2002), 2461–2473.

[Lovey et al. 2004] F. C. Lovey, A. M. Condó, and V. Torra, "A model for the interaction of martensitic transformation with dislocations in shape memory alloys", *Int. J. Plast.* **20**:2 (2004), 309–321.

[McNaney et al. 2003] J. M. McNaney, V. Imbeni, Y. Jung, P. Papadopoulos, and R. O. Ritchie, "An experimental study of the superelastic effort in a shape-memory nitinol alloy under biaxial loading", *Mech. Mater.* **35**:10 (2003), 969–986.

[Meyers 1994] M. A. Meyers, *Dynamic behavior of materials*, Wiley-Interscience, New York, 1994. 305–310.

[Millett et al. 2002] J. C. F. Millett, N. K. Bourne, and G. T. Gray, "Behavior of the shape memory alloy NiTi during one-dimensional shock loading", *J. Appl. Phys.* **92**:6 (2002), 3107–3110.

[Nakayama et al. 2005] H. Nakayama, Y. Zhao, M. Taya, W. W. Chen, Y. Urushiyama, and S. Suzuki, "Strain rate effects of TiNi and TiNiCu shape memory alloys", pp. 355–363 in *Proceedings of SPIE, Smart structures and materials 2005: Active materials: Behaviors and mechanics*, vol. 5761, edited by W. D. Armstrong, San Diego, CA, 2005.

[Nemat-Nasser et al. 2005] S. Nemat-Nasser, J. Y. Choi, W. G. Guo, J. B. Isaacs, and M. Taya, "High strain-rate, small strain response of a NiTi shape-memory alloy", *J. Eng. Mater. Technol.* (*ASME*) **127**:1 (2005), 83–89.

[Ogawa 1988] K. Ogawa, "Characteristics of shape memory alloy at high strain rate", *J. Phys.* **49**:C3 (1988), 115–120.

[Otsuka and Wayman 1988] K. Otsuka and C. M. Wayman, *Shape memory materials*, Cambridge Univ. Press, 1988.

[Ravichandran and Chen 1991] G. Ravichandran and W. Chen, "Dynamic behavior of brittle materials under uniaxial compression", pp. 85–90 in *Experiments in micromechanics of fracture resistant materials*, edited by K. S. Kim, AMD-130, ASME, New York, 1991.

[Schetky and Perkins 1978] L. M. Schetky and J. Perkins, "The 'quiet' alloys", *Machine Design* **50**:8 (April 6 1978), 202–206.

[Song and Chen 2004] B. Song and W. Chen, "Loading and unloading SHPB pulse shaping tech-
niques for dynamic hysteretic loops", *Exp. Mech.* **44**:6 (2004), 622–627.

[Song et al. 2003] B. Song, W. Chen, and T. Weerasooriya, "Quasi-static and dynamic compressive
behaviors of a S-2 glass/SC15 composite", *J. Compos. Mater.* **37**:19 (2003), 1723–1743.

[Song et al. 2005] B. Song, W. Chen, T. Yanagita, and D. J. Frew, "Temperature effects on the
dynamic compressive and failure behaviors of an epoxy syntactic foam", *Compos. Struct.* **67**:3
(2005), 289–298.

[Stoeckel and Yu 1991] D. Stoeckel and W. Yu, "Superelastic Ni-Ti wire", *Wire J. Inter.* **24**:3 (March
1991), 45–50.

[Togami et al. 1996] T. C. Togami, W. E. Baker, and M. J. Forrestal, "A split Hopkinson bar tech-
nique to evaluate the performance of accelerometers", *J. Appl. Mech.* (*ASME*)) **63** (1996), 353–356.

[Wayman 1993] C. M. Wayman, "Shape memory alloys", MRS Bulletin, April 1993. 49–56.

[Wu and Gorham 1997] X. J. Wu and D. A. Gorham, "Stress equilibrium in the split Hopkinson pres-
sure bar test", *J. Phys. (France) IV* **7**:C3 (1997), 91–96. International Conference on Mechanical
and Physical Behaviour of Materials under Dynamic Loading (EURODYMAT'97), Toledo (Spain),
22-26 Sep 1997.

WEINONG CHEN: wchen@purdue.edu
Schools of Aeronautics and Astronautics and Materials Engineering, Purdue University,
West Lafayette, IN 47907-2023, United States

BO SONG: songb@purdue.edu
School of Aeronautics and Astronautics, Purdue University, West Lafayette, IN 47907-2023,
United States

RANDOM FIELD AND HOMOGENIZATION FOR MASONRY WITH NONPERIODIC MICROSTRUCTURE

VITTORIO GUSELLA AND FEDERICO CLUNI

The purpose of this paper is to illustrate a method for homogenizing masonry with a nonperiodic microstructure. The proposed approach is based on the concept of the representative volume element and on the finite-size test-window method. First, the peculiarities of masonry as a composite continuum are highlighted. Then, the heterogeneity of the microstructure (elements and texture) is modeled by statistical descriptors. To improve the classical test-window method a probabilistic convergence criterion is coupled with the well-known mechanical convergence criterion. Both criteria must be met in order to check the convergence of the material window with the statistically equivalent representative volume element. An application shows the effectiveness of the proposed approach.

1. Introduction

Within the framework of micromechanics theory, masonry is modeled as a heterogeneous material composed of bricks or stones in a matrix of mortar. In dealing with a heterogeneous continuum, homogenization techniques allow one to define an equivalent body in order to study linear and nonlinear behavior [Christensen 1980; Suquet 1987].

The application of this approach to masonry was proposed by Pande et al. [1989], Pietruszczak and Niu [1992], and Maier et al. [1991]. On the basis of asymptotic analysis [Bensoussan et al. 1978; Sanchez-Palencia 1980], a rigorous application of the homogenization theory to periodic media was developed by Anthoine [1995]. The effect of rigid or elastic blocks was analyzed by Cecchi and Sab [2002], while failure analysis, ultimate strength, and damage models were considered in [Alpa and Monetto 1994; De Buhan and De Felice 1997; Luciano and Sacco 1997], respectively.

The hypothesis of a "periodic microstructure" in masonry has been adopted in all previous papers. This means that the bricks and the mortar joints are assumed to be of equal dimensions and characteristics. Moreover, these elements must be arranged in a periodic pattern.

Keywords: masonry, random heterogeneous material, homogenization.

However, the most interesting aspect of masonry structure analysis is related to the maintenance and restoration of historical and monumental buildings. In these cases the assumption of a periodic microstructure in stone masonry or brickwork would be mistaken. In order to apply the homogenization theory to old masonry it is necessary to use a different approach.

This aspect was considered in [Cluni and Gusella 2004], where the representative volume element of the masonry wall was determined by employing a formulation based on finite-size test-windows. In analyzing a masonry wall, the homogenized medium elastic stiffness tensor was obtained by considering the hierarchy of estimates relative to essential and natural boundary conditions. An adequate linear model for masonry is very important, because it permits one to analyze very large structures (monumental and historical buildings), and indicates those parts that bear the greatest stresses, where one must consider a more sophisticated nonlinear analysis.

A different procedure was proposed by Šejnoha et al. [2004] to analyze masonry structures with irregular geometry. This methodology, based on [Povirk 1995] and further developed in [Zeman and Šejnoha 2001; Šejnoha and Zeman 2002], introduces a periodic unit cell that possesses statistical proprieties similar to the original material and can therefore be considered a reasonable approximation.

In [Cluni and Gusella 2004], the material window is considered an adequate estimation of the representative volume element when the difference between natural and essential elastic moduli is limited. However, an estimation of the representative volume element based solely on mechanical convergence may be inaccurate. It is necessary to check that this convergence does not reflect the conditions in a specific portion of the wall in question and that the test-window is sufficiently representative of the masonry in terms of its constituent elements (stones and mortar joints) and its texture.

For particular composites, this aspect was highlighted by Bochenek and Pyrz [2004], who introduced statistical and geometrical measures and constraints to reconstruct families of plane and spatial dispersion of inclusions resembling reference patterns and to predict overall properties. Introducing the concept of periodization, the combination of mechanical and statistical criteria was also proposed in [Sab and B. 2005].

This paper proposes an improvement of the finite-size test-window method in order to overcome these limitations of the representative volume element estimation based solely on the mechanical convergence criterion. First, the peculiarities of masonry, which are related to the construction technique employed, are highlighted. Nonperiodic masonry is then analyzed within the framework of random heterogeneous material theory, and statistical descriptors are introduced in order to describe the random field modeling of masonry. Mechanical and probabilistic

Figure 1. Portion of the medieval defense walls of an Italian town.
Note that the hypothesis of periodic continuum is not applicable.

criteria are used to check the convergence of the material window to the statistical
equivalent volume element. These criteria are illustrated by applying them to an
actual masonry wall.

2. Masonry as a peculiar heterogeneous material

Masonry can be considered a heterogeneous material composed of stones or bricks
in a matrix of mortar. However it is a very particular composite: taken together,
the inclusions (bricks or stones) have a much larger surface area than the matrix
(in the case of dry masonry the matrix disappears); the mortar could in fact be
regarded as merely joining the inclusions; the constituent blocks and mortar joints
have different dimensions.

Leaving aside very chaotic typologies, some regularity is imposed by the build-
ing procedure even in stone masonry (see, for example, Figure 1):

– the masonry is built with courses of blocks connected by head and bed mortar
 joints;
– the bed mortar joints are continuous and more or less horizontal;
– the bed thickness is fairly constant for each course;

– the head joints are almost always interrupted vertically by inclusions;

– the inclusions are roughly rectangular;

– blocks belonging to the same course have similar dimensions, especially the heights, inasmuch as the bed joints are horizontal;

– there is no relation between consecutive courses, in fact their heights are purposely differentiated so as to obtain good quality masonry (in practice courses with larger stones are not grouped together but more or less evenly distributed vertically).

These features give the *masonry pattern* (or *masonry texture*) which characterize the mechanical behavior of the wall.

Given that the hypothesis of periodic continuum is not applicable, an important issue, in order to define a homogeneous continuum equivalent to the masonry, is the minimum size of the representative volume element [Drugan and Willis 1996].

Based on these observations, acceptance of a volume as being representative requires a set of conditions which are both mechanical and probabilistic. Thus the masonry wall must be analyzed within the framework of random field theory.

3. Masonry random field

Consider the masonry as a random heterogeneous material or simply a random medium [Torquato 2001]. A masonry wall is a realization of this two-phase random medium and occupies a region $D \subseteq \mathbb{R}^2$ that is partitioned into two disjoint random sets or phases (we consider here the mechanical and probabilistic problems in 2-dimensional terms; however, the proposed approach can be extended to the third dimension).

Let D_1 be the region relative to stone and D_2 the region relative to mortar. Since D_1 and D_2 are the complements of one another, then $D_1 \cup D_2 = D$ and $D_1 \cap D_2 = \varnothing$.

For a given realization, the indicator function (or characteristic function) $\chi^i(x)$ for the phase $i = 1, 2$, given $x \in D$, is defined by

$$\chi^i(x) = \begin{cases} 1 & \text{if } x \in D_i, \\ 0 & \text{otherwise,} \end{cases} \quad \text{with } \chi^1(x) + \chi^2(x) = 1. \tag{1}$$

The probabilistic descriptor of $\chi^i(x)$ is given by the n-point probability function for phase i [Torquato and Stell 1982]

$$S_n^i(x_1, x_2, \ldots x_n) = P\{\chi^i(x_1) = 1, \chi^i(x_2) = 1, \ldots \chi^i(x_n) = 1\}, \tag{2}$$

which gives the probability that n points at positions $x_1, x_2, \ldots, x_n$ are found in phase i.

The random medium is strictly statistically homogeneous if there is not a preferred origin in the system so that, for all $n \geq 1$ and all $y \in D$

$$S_n^i(x_1, x_2, \ldots, x_n) = S_n^i(x_1 + y, x_2 + y, \ldots, x_n + y). \tag{3}$$

In particular, the one-point probability function is constant and equal to the volume fraction ϕ_i of the phase $i : S_1^i(x_1) = \phi_i$.

From this point of view, the masonry has a locally heterogeneous microstructure, but it can be considered homogeneous overall when a sufficiently large portion of it is taken into account. In the following, statistical descriptors up to the second order will be considered so that the masonry wall will be assumed to be a weakly homogeneous random field.

If the random medium is ergodic then the result of averaging over all the realizations of the ensemble is equivalent to that of averaging over the surface for one realization in the infinite-surface limit. In this way probabilistic information can be obtained from a single realization of the infinite medium.

The masonry will be assumed, in the following, to be an ergodic medium. Moreover, it is assumed that the probabilistic characteristics of the medium can be estimated by analyzing a portion of the wall. This portion must be sufficiently large to make it adequately representative of the medium; furthermore this condition permits one to replace the averaging in the infinite-surface limit with the numerical approximation derived from the averaging over the surface of the portion.

3.1. *Statistical descriptors of stones and mortar joints.* Masonry is, as we noted above, a very distinctive random medium, and we shall therefore introduce statistical descriptors to represent this peculiarity. These descriptors will refer to the type of masonry shown in Figure 1, but they can be applied to a very large number of masonry typologies.

The portion of masonry shown in Figure 1 has been taken as a representative sample of the structure as a whole. Considering, in particular, the construction of the wall in superimposed courses, the random field of masonry was described using a grid that was set up as follows (Figure 2):

- The rows of the grid are horizontal lines drawn in correspondence with the center points of the stones; rows are indicated by R_j with $j = 1, \ldots, N_{Rj}$ ($N_{Rj} = 23$).

- The columns are vertical lines starting from the center points of the stones of R_1; columns are indicated by C_j with $j = 1, \ldots, N_{Cj}$ ($N_{Cj} = 18$).

- Along the rows, for each course $j = 1, \ldots, N_{Rj}$, the widths of the stones $B_{st,k}^j$ $k = 1, 2, \ldots$ and the sizes of the mortar joints $B_{m,k}^j$ $k = 1, 2, \ldots$ were measured.
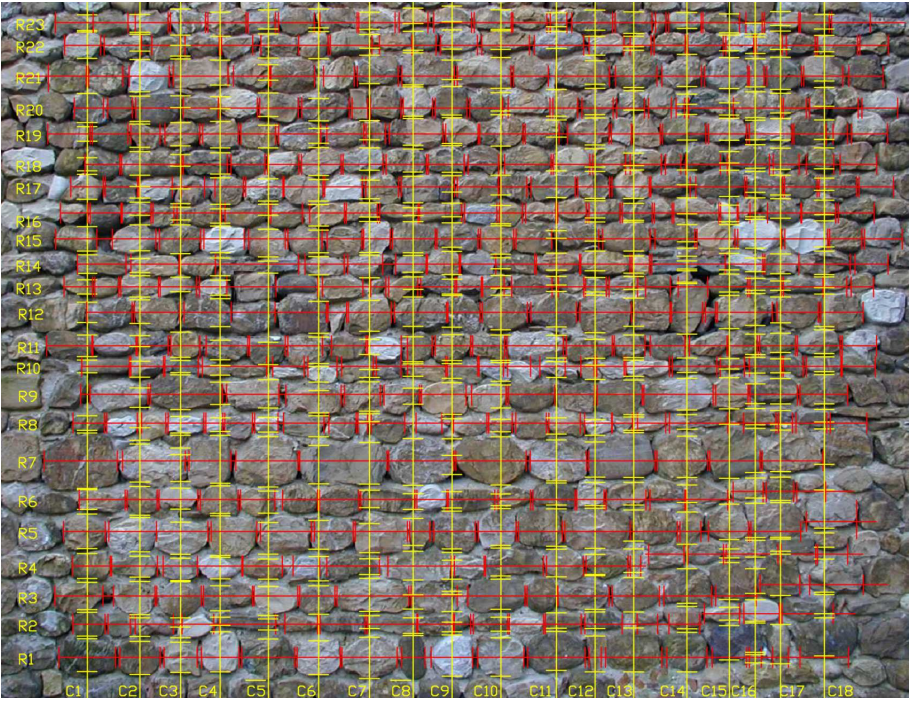
Figure 2. Masonry wall with superimposed orthogonal grid; rows R_j ($j = 1, \ldots, N_{Rj}$; $N_{Rj} = 23$) and columns C_j ($j = 1, \ldots, N_{Cj}$; $N_{Cj} = 18$).

– The heights of the stones $H^j_{st,k}$ $k = 1, 2, \ldots$ and the thicknesses of the mortar joints $H^j_{m,k}$ $k = 1, 2, \ldots$ were measured along each column $j = 1, \ldots, N_{Cj}$ (note that $H^j_{m,k}$ represents the thickness of the bed joints, but it can, in some cases, indicate the height of the head joint between rows).

A grid rather than image analysis with small pixels was preferred because the former reflects the intrinsically horizontal/vertical structure of the masonry and because the irregularities of the surface would make it extremely difficult to use the pixels in distinguishing mortar from stone.

Considering the previous values the following samples were obtained: $\{B_{st}\}$ and $\{H_{st}\}$ for the widths and heights of the stones, and $\{B_m\}$ and $\{H_m\}$ for the sizes of bed and head mortar joints.

The portion of the wall considered, shown in Figure 2, was sufficiently large (432×341 cm) that the previous samples were assumed to be statistically representative of the geometric characteristics of the inclusions and of the matrix. These samples were therefore taken to be the statistical descriptors of the masonry continuum.
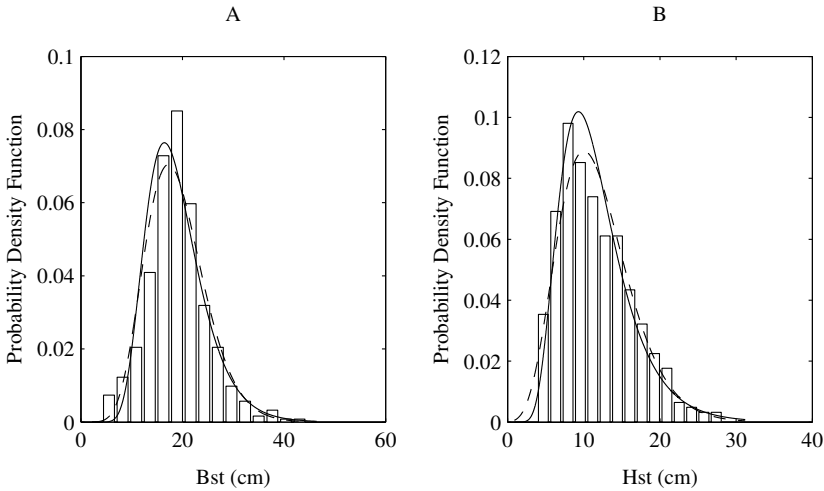
Figure 3. Comparison among the probability density functions of the width of the stones P_{Bst}, the heights of the stones P_{Hst}, the gamma (dashed line), and the log-normal curve (continuous line).

These samples were analyzed to determine their statistical moments up to the second order:

(a) the mean values of the width and height of the stones: E_{Bst} and E_{Hst}, respectively;

(b) the mean values of the mortar joint dimensions: E_{Bm} and E_{Hm}, respectively;

(c) the standard deviation values for the same samples: Σ_{Bst}, Σ_{Hst}, Σ_{Bm} and Σ_{Hm}.

These values are reported in Table 3.

The comparison among the probability density functions of the width of the stones P_{Bst}, the gamma curve, and the log-normal curve are shown in Figure 3 (A). The same comparison for the probability density functions of the height of the stones P_{Hst} is reported in Figure 3 (B). Note that both gamma and log-normal laws are not rejected by the chi-square test.

The probability density functions of the characteristics of the mortar joints P_{Bm} and P_{Hm} are reported in Figure 4 (A) and (B). In these figures the probability density functions are compared only with the log-normal curve, since the gamma law was rejected.

3.2. *Statistical descriptors of the masonry wall texture.* The previous statistical descriptors give information only about stones and mortar joints. Conversely, the mechanical behavior of a masonry wall is influenced by its texture.

We use the following approach to describe this feature.

A

B
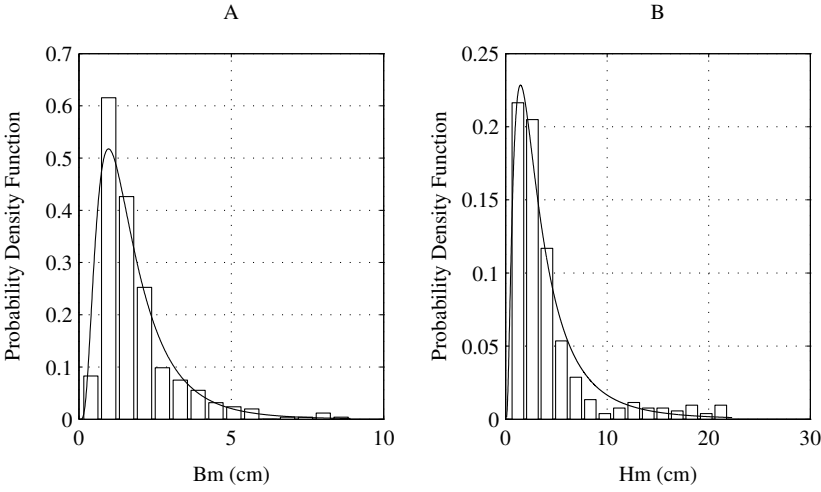
Figure 4. Comparison between the probability density functions of the characteristics of the mortar joints: P_{Bm} and P_{Hm}, and the log-normal curve.

Let $R_j(x)$ be the characteristic function relative to the row R_j $j = 1, \ldots, N_{Rj}$: if the point, with abscissa x, belongs to the "stone phase", the function assumes the value 1; if the point belongs to the "mortar phase", the function value is 0:

$$R_j(x) = \begin{cases} 1 & x \in \text{ stone phase}, \\ 0 & x \in \text{ mortar phase}, \end{cases} \tag{4}$$

where $0 \le x \le L_{Rj}$ and $L_{Rj} = \sum_k (B_{st,k}^j + B_{m,k}^j)$ is the total length of the row R_j.

Let $C_j(y)$ be the characteristic function relative to the column C_j $j = 1, \ldots, N_{Cj}$:

$$C_j(y) = \begin{cases} 1 & y \in \text{ stone phase}, \\ 0 & y \in \text{ mortar phase}, \end{cases} \tag{5}$$

where $0 \le y \le L_{Cj}$ and $L_{Cj} = \sum_k (H_{st,k}^j + H_{m,k}^j)$ is the total length of the column C_j.

In order to estimate the second-order characteristics of $R_j(x)$, the "shifted-area function" (SAF) $A_{Rj,Rj}(\xi)$ was introduced. Given $\xi \in \mathbf{R}$, this function $A_{Rj,Rj}(\xi)$ corresponds to the mean square value of the area below the curve expressing the difference between the shifted function $R_j(x + \xi)$ and the function $R_j(x)$:

$$A_{Rj,Rj}(\xi) = \lim_{\Delta x \to \infty} \frac{1}{\Delta x} \int_0^{\Delta x} \left[R_j(x + \xi) - R_j(x)\right]^2 dx. \tag{6}$$

In the application the integral is replaced by a summation extended on the row.

This function gives immediate information about periodic texture (Figure 5) because it assumes null value for $\xi = 0$ and $\xi = nX$, where $X \in R$ is the period and $n = 1, 2, \ldots$ (see Figure 5, where the shifted-area function is normalized to its maximum value).
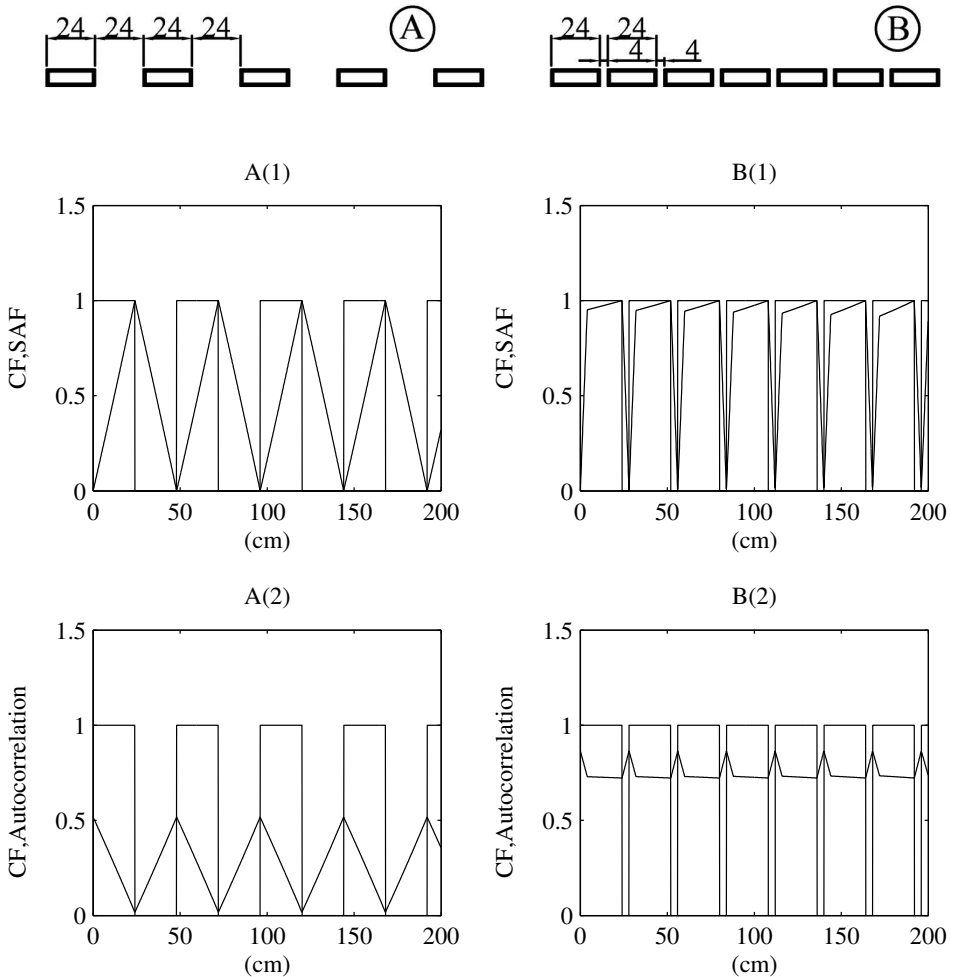
Figure 5. Courses in masonry with different periodic textures (measured in cm). Characteristic functions (CF), normalized shifted-area functions (SAF) (top) and autocorrelation functions (bottom).

Given the ergodicity hypothesis, the shifted-area function $A_{Rj,Rj}(\xi)$ is related to the autocorrelation function $AC_{Rj,Rj}(\xi)$ of $R_j(x)$ by

$$A_{Rj,Rj}(\xi) = \lim_{\Delta x \to \infty} \frac{1}{\Delta x} \int_0^{\Delta x} [R_j(x+\xi)]^2 dx$$

$$-2 \lim_{\Delta x \to \infty} \frac{1}{\Delta x} \int_0^{\Delta x} R_j(x+\xi) R_j(x)\, dx + \lim_{\Delta x \to \infty} \frac{1}{\Delta x} \int_0^{\Delta x} [R_j(x)]^2\, dx$$

$$= 2 \lim_{\Delta x \to \infty} \frac{1}{\Delta x} \int_0^{\Delta x} [R_j(x)]^2\, dx - 2AC_{Rj,Rj}(\xi). \qquad (7)$$

We then have

$$AC_{Rj,Rj}(\xi) = \lim_{\Delta x \to \infty} \frac{1}{\Delta x} \int_0^{\Delta x} R_j(x+\xi) R_j(x)\, dx = \frac{2E_{Rj}^2 - A_{Rj,Rj}(\xi)}{2}, \qquad (8)$$
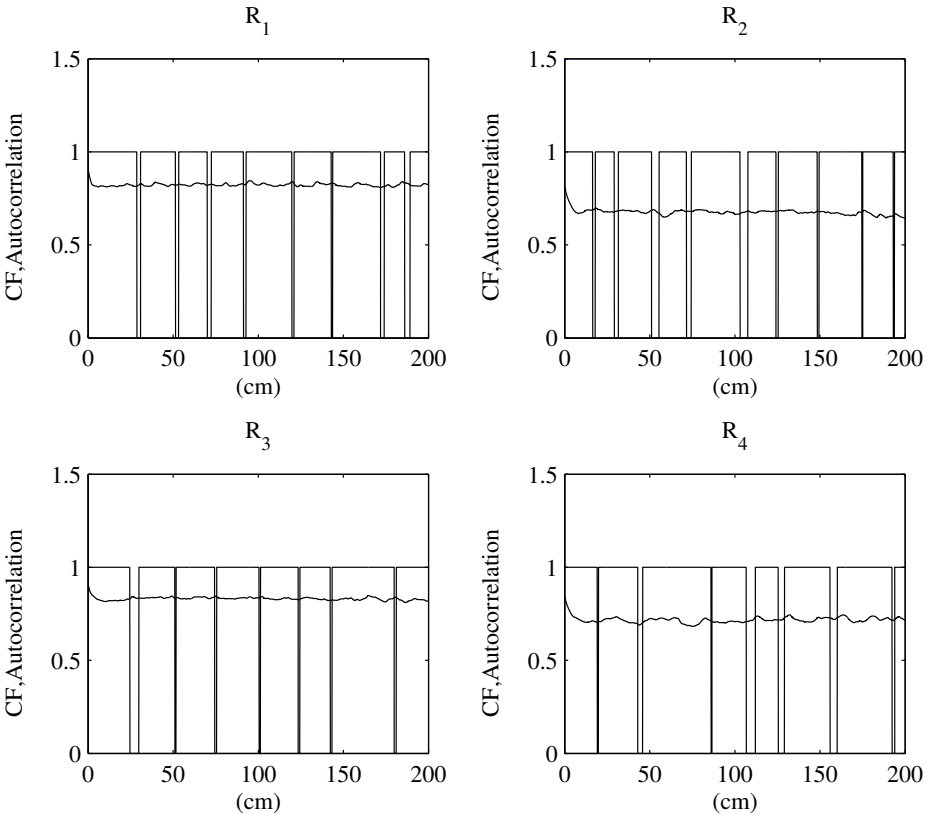
Figure 6. Characteristic functions (CF) and autocorrelation functions relative to the first four rows R_j ($j = 1, \ldots, 4$) of the masonry wall in Figure 2.

where

$$E^2_{Rj} = \lim_{\Delta x \to \infty} \frac{1}{\Delta x} \int_0^{\Delta x} [R_j(x)]^2 \, dx \qquad (9)$$

is the mean square value of $R_j(x)$.

Given Equation (7), E^2_{Rj} is the ratio of stone to stone plus mortar for the row R_j; moreover $AC_{Rj,Rj}(\xi) = E^2_{Rj}$ for $\xi = 0$.

The autocorrelation functions for periodic patterns are illustrated in Figure 5.

For the masonry wall in Figure 2, the characteristic functions $R_j(x)$ and the autocorrelation functions $AC_{Rj,Rj}(\xi)$ relative to the first four rows ($j = 1, 2, 3, 4$) are reported in Figure 6.

Figure 7 shows the autocorrelation functions $AC_{Rj,Rj}(\xi)$ for $j = 1, \ldots, 11$.

Higher autocorrelation functions are observed in the courses with larger stones (the largest correlation corresponds to the case indicated by (A) in Figure 7 relative to row $R_7(x)$); a weaker correlation is observed in courses with smaller stones (these cases are indicated by (C) in Figure 7).

When we consider the characteristic function $C_j(y)$, we obtain the shifted-area
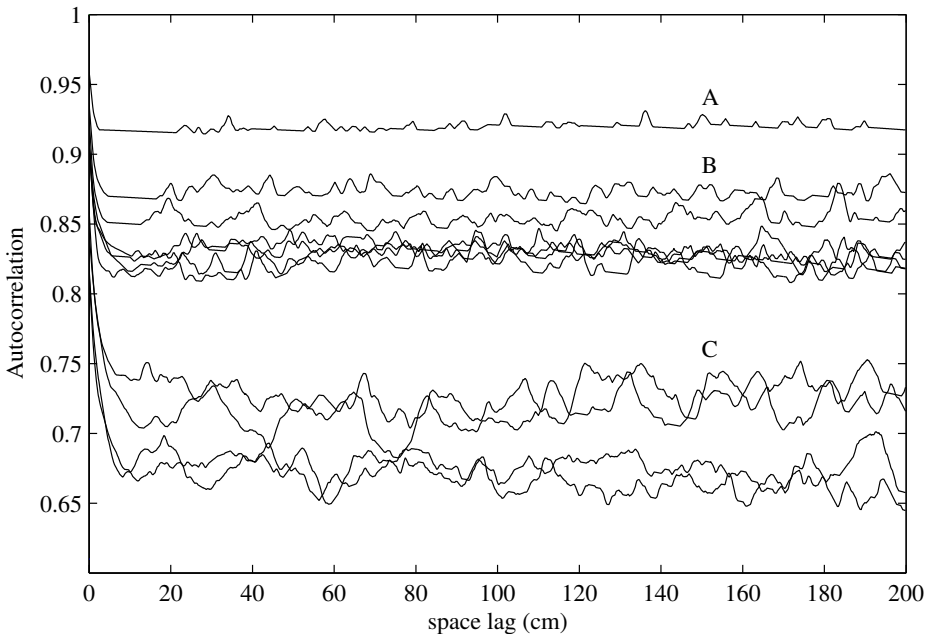
Figure 7. Autocorrelation functions of the first eleven rows R_j ($j = 1, \ldots, 11$) of the masonry wall in Figure 2.

function:

$$A_{Cj,Cj}(\eta) = \lim_{\Delta y \to \infty} \frac{1}{\Delta y} \int_0^{\Delta y} [C_j(y+\eta) - C_j(y)]^2 \, dy, \tag{10}$$

$$AC_{Cj,Cj}(\eta) = \lim_{\Delta y \to \infty} \frac{1}{\Delta y} \int_0^{\Delta y} C_j(y+\eta) C_j(y) \, dy = \frac{2E_{Cj}^2 - A_{Cj,Cj}(\eta)}{2}, \tag{11}$$

$$E_{Cj}^2 = \lim_{\Delta y \to \infty} \frac{1}{\Delta y} \int_0^{\Delta y} [C_j(y)]^2 \, dy. \tag{12}$$

The characteristic functions $C_j(x)$ and the autocorrelation functions $AC_{Cj,Cj}(\eta)$ relative to the first four columns ($j = 1, 2, 3, 4$) of the masonry wall are reported in Figure 8.

Figure 9 shows the autocorrelation functions $AC_{Cj,Cj}(\eta)$ for $j = 1, \ldots, 11$.
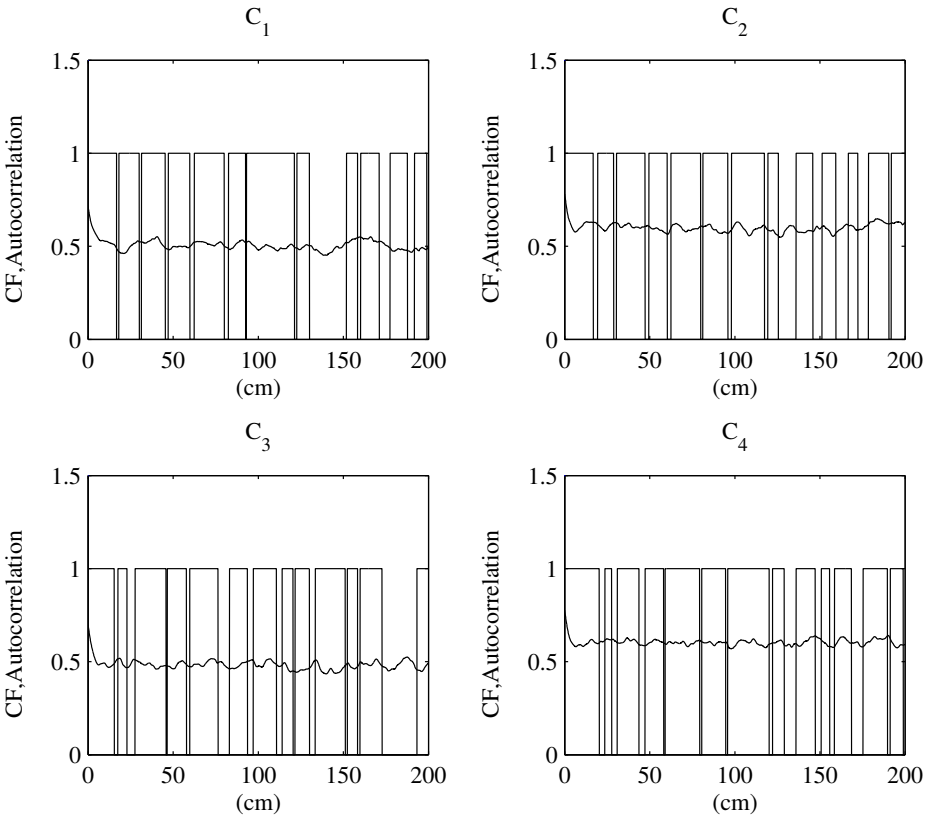
Figure 8. Characteristic functions (CF) and the autocorrelation functions relative to the first four columns C_j ($j = 1, \ldots, 4$) of the masonry wall in Figure 2.
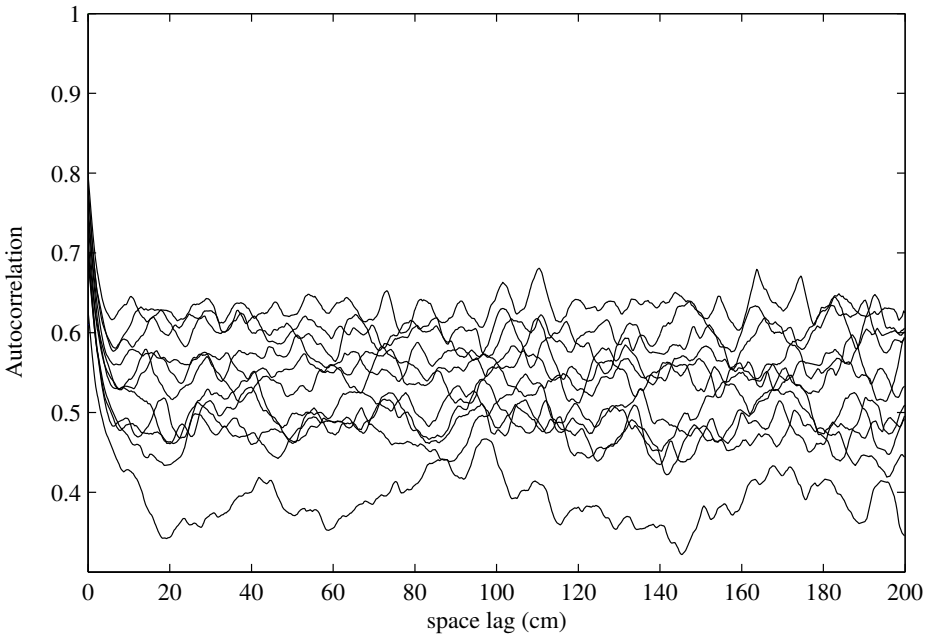
Figure 9. Autocorrelation functions of the first eleven columns
C_j, $j = 1, \ldots, 11$ of the masonry wall in Figure 2.

These functions exhibit analogous behaviors and show a weaker correlation than that of the Figure 7 courses. This is as we expected; in fact, good building practice dictates the use of stones with similar dimensions (width and height) within a single row. Moreover, the head-joint thicknesses are limited, and, conversely, within columns stones differ in height and bed-mortar joints are thicker. These conditions imply that the correlation is stronger within the courses than within the columns. The autocorrelation functions relative to the rows $AC_{Rj,Rj}(\xi)$ and to the columns $AC_{Cj,Cj}(\eta)$ are taken to be the statistical descriptors of the masonry's texture.

The autocorrelation functions $AC_{Rj,Rj}(\xi)$ and $AC_{Cj,Cj}(\eta)$ were used to check the correctness of the weakly homogeneous random field hypothesis. In fact, using different origins to compute the space lags ξ and η showed that the differences among these functions were very limited, both for the rows and for the columns.

Further information about texture can be evinced by:

– the cross-correlation function $CC_{Rj,Rk}(\xi)$ between rows:

$$CC_{Rj,Rk}(\xi) = \lim_{\Delta x \to \infty} \frac{1}{\Delta x} \int_o^{\Delta x} R_j(x + \xi) R_k(x)\, dx$$
$$= \frac{E_{Rj}^2 + E_{Rk}^2 - A_{Rj,Rk}(\xi)}{2}, \tag{13}$$

where

$$A_{Rj,Rk}(\xi) = \lim_{\Delta x \to \infty} \frac{1}{\Delta x} \int_0^{\Delta x} [R_j(x+\xi) - R_k(x)]^2 \, dx; \tag{14}$$

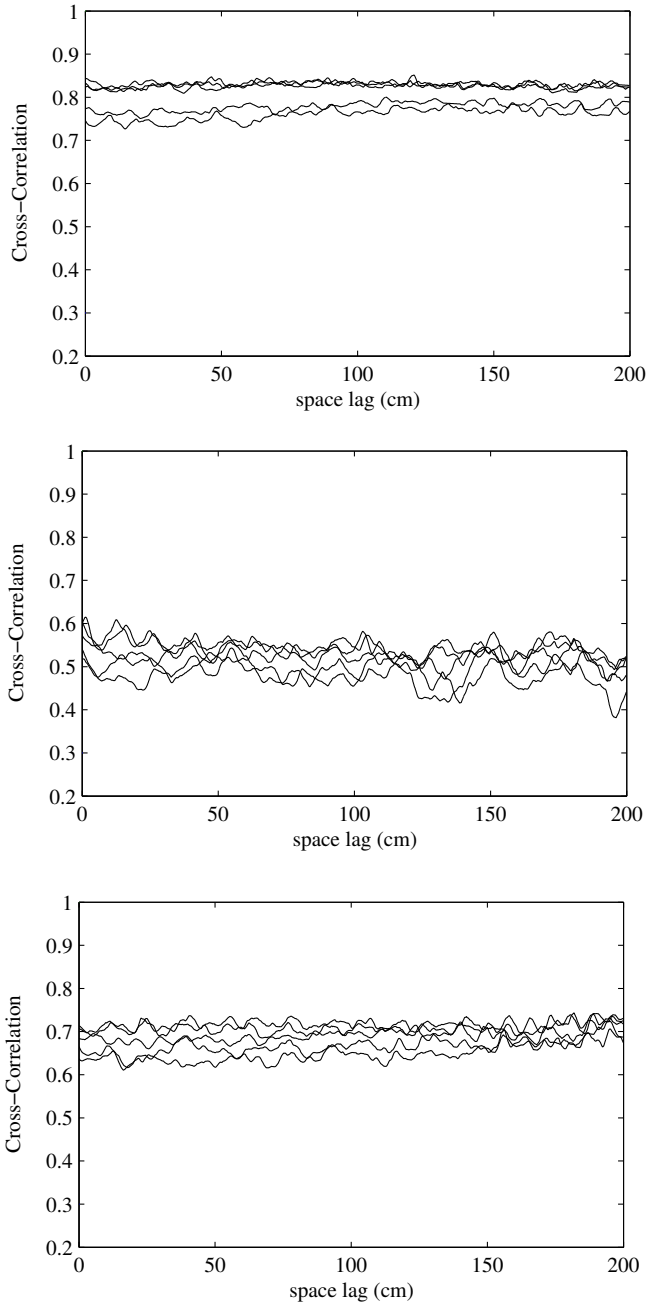– the cross-correlation function $CC_{Cj,Ck}(\xi)$ between columns:

$$\begin{aligned} CC_{Cj,Ck}(\eta) &= \lim_{\Delta y \to \infty} \frac{1}{\Delta y} \int_0^{\Delta y} C_j(y+\eta) C_k(y) dy \\ &= \frac{E_{Cj}^2 + E_{Ck}^2 - A_{Cj,Ck}(\eta)}{2}, \end{aligned} \tag{15}$$

where

$$A_{Cj,Ck}(\eta) = \lim_{\Delta y \to \infty} \frac{1}{\Delta y} \int_0^{\Delta y} [C_j(y+\eta) - C_k(y)]^2 dy; \tag{16}$$

– the cross-correlation function $CC_{Rj,Ck}(\xi)$ between rows and columns:

$$\begin{aligned} CC_{Rj,Ck}(\xi) &= \lim_{\Delta x \to \infty} \frac{1}{\Delta x} \int_0^{\Delta x} R_j(x+\xi) C_k(y) \, dx \\ &= \frac{E_{Rj}^2 + E_{Ck}^2 - A_{Rj,Ck}(\xi)}{2}, \end{aligned} \tag{17}$$

where

$$A_{Rj,Ck}(\xi) = \lim_{\Delta x \to \infty} \frac{1}{\Delta x} \int_0^{\Delta x} [R_j(x+\xi) - C_k(y)]^2 \, dx. \tag{18}$$

The column pairs and row-column pairs show patterns that are less strongly cross-correlated than those of the row pairs (Figure 10).

4. Convergence criteria

In order to determine the elastic moduli of the homogeneous continuum equivalent to the masonry it is necessary to estimate the representative volume element. Following [Cluni and Gusella 2004], this estimate can be performed by using the finite-size test-window method. Here this method is improved, however, by coupling the classical mechanical convergence criterion with a probabilistic convergence criterion which considers the statistical descriptors introduced above.

Consider a "*material window*" with a representative size L placed at any given point in the wall. This window plays the role of the *L-Volume Element* (VE$_L$).

Let $T_{ij,L}^e$ ($\boldsymbol{T}_L^e$ in matrix notation) indicate the elastic stiffness components obtained under essential boundary conditions applied to an VE$_L$, and let $S_{ij,L}^n$ ($\boldsymbol{S}_L^n$) indicate the elastic compliance components obtained under natural conditions. The $T_{ij,L}^n$ natural elastic stiffness components are obtained by inversion: $\boldsymbol{T}_L^n = [\boldsymbol{S}_L^n]^{-1}$.

Figure 10. Cross-correlation functions between rows and columns for the wall in Figure 2. Top: rows R_1 and R_j ($j = 2, \ldots, 6$); middle: columns C_1 and C_j ($j = 2, \ldots, 6$); bottom: row R_1 and columns C_j ($j = 1, \ldots, 5$).

We consider the statistical descriptors relative to the inclusions and texture of the VE_L:

- the samples of stone sizes $\{B_{st}^L\}$ and $\{H_{st}^L\}$ with mean values, standard deviations, and probability density functions: E_{Bst}^L, E_{Hst}^L, Σ_{Bst}^L, Σ_{Hst}^L, P_{Bst}^L, P_{Hst}^L, respectively;

- the samples of mortar joint sizes $\{B_m^L\}$ and $\{H_m^L\}$, with mean values, standard deviations, and probability density function E_{Bm}^L, E_{Hm}^L, Σ_{Bm}^L, Σ_{Hm}^L, P_{Bm}^L, P_{Hm}^L, respectively;

- the autocorrelation functions relative to the rows $AC_{Rj,Rj}^L(\xi)$ and the columns $AC_{Cj,Cj}^L(\eta)$.

Consider increasing the size of the material window

$$L_k \ k = 1, \ldots, i, j, \ldots N \ (j > i \rightarrow L_j > L_i)$$

thus obtaining the sequence VE_{L_k}.

In order to check that the larger window $VE_{\hat{L}}$, with $\hat{L} = L_N$, is an adequate estimate of the representative volume element, the following *mechanical convergence criterion* can be utilized: Given $\Delta c \in R^+$, the sequence VE_{L_k} converges to the representative volume element when the differences between the natural and essential elastic stiffness of the larger material window $VE_{\hat{L}}$ are limited:

$$\max_{ij} \left| \frac{T_{ij,\hat{L}}^e - T_{ij,\hat{L}}^n}{T_{ij}^1} \right| \leq \Delta c, \tag{19}$$

where

$$T_{ij}^1 = \left| T_{ij,L_1}^e - T_{ij,L_1}^n \right|. \tag{20}$$

The exclusively mechanical approach can, nonetheless, introduce an incorrect estimate; in fact Δc is an arbitrary value and the convergence is not uniform (see the following application on the masonry wall in Figure 1).

In order to overcome this limitation the following *probabilistic convergence criterion* can be used: The sequence VE_{L_k} converges to the representative volume element when the statistical descriptors relative to the elements and the texture of the larger material window $VE_{\hat{L}}$ comply with the statistical descriptors of the random field model relative to the masonry wall as a whole. In other words, the material window $VE_{\hat{L}}$ must be *statistically similar* to the random medium model of the masonry as obtained by analyzing the entire wall.

With regard to the statistical descriptors introduced above, the similarity in stone and mortar joint characteristics is checked by minimizing the differences among

mean values, standard deviations, and probability density functions:

$$\left| \Sigma_{\Pi j} - \Sigma_{\Pi j}^{\hat{L}} \right| \leq \Delta c_{\Sigma_{\Pi j}}, \tag{21}$$

$$\int \left| P_{\Pi j} - P_{\Pi j}^{\hat{L}} \right| d\Pi j \leq \Delta c_{P_{\Pi j}}, \tag{22}$$

where $\Pi = B, H$ and $j = st, m$.

Similarity in texture is checked by minimizing the differences among correlation functions. Let $\bar{A}C_{RR}(\xi)$ be the mean autocorrelation function obtained by averaging, for any space lag ξ, the autocorrelation function of the rows of the masonry wall; let $\bar{A}C_{C,C}(\eta)$ be the mean autocorrelation function relative to the columns. Let $\bar{A}C_{R,R}^{\hat{L}}(\xi)$ and $\bar{A}C_{C,C}^{\hat{L}}(\eta)$ be the same quantities relative to the rows and columns of the material window. The similarity is then checked by:

$$\int \left| \bar{A}C_{Rj,Rj}(\xi) - \bar{A}C_{Rj,Rj}^{\hat{L}}(\xi) \right| d\xi \leq \Delta c_{Rj}, \tag{23}$$

$$\int \left| \bar{A}C_{Cj,Cj}(\eta) - \bar{A}C_{Cj,Cj}^{\hat{L}}(\eta) \right| d\eta \leq \Delta c_{Cj}. \tag{24}$$

If these mechanical and probabilistic convergence criteria are met, then the material window $VE_{\hat{L}}$ is an adequate estimate of the *statistically equivalent representative volume element*. In this case the elastic stiffness components $T_{ij}^{\text{hom}} = T_{ij}^{\text{RVE}}$ of the equivalent homogeneous continuum can be estimated by $T_{ij,\hat{L}}^{*}$:

$$T_{ij}^{\text{hom}} = T_{ij}^{\text{RVE}} = T_{ij,\hat{L}}^{*} = \frac{T_{ij,\hat{L}}^{e} + T_{ij,\hat{L}}^{n}}{2}. \tag{25}$$

We mention that the linear properties of random composites can be accurately estimated using volumes subjected to periodic boundary conditions. The periodic boundary conditions give an estimation of the effective elastic moduli which are intermediate between those deriving from displacement and those from traction boundary conditions, as demonstrated numerically by Terada et al. [2000] and theoretically by Sab and B. [2005]. These conditions could be used to improve the approach proposed in the present paper, however, as will be highlighted in the numerical application, the difference between essential and natural evaluations of the elastic moduli decrease very quickly as L increases.

5. Application

The previous method has been applied in estimating the elastic stiffness components of the homogeneous continuum equivalent to the masonry wall in Figure 1.
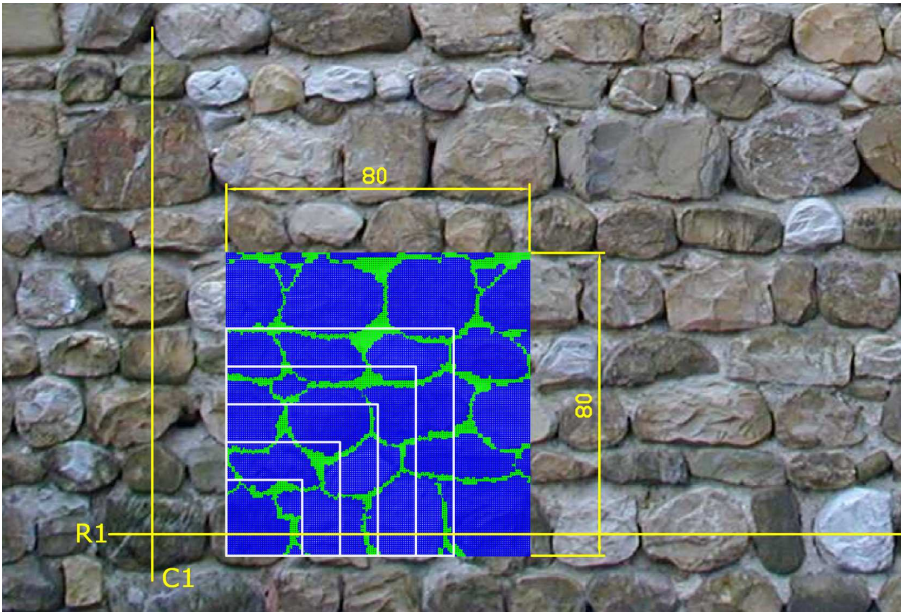
Figure 11. Square material window with $L = 80$ cm and finite element model.

A sequence of square material windows of size L was located in the bottom-left portion. Given the ergodicity hypothesis, the position of the test-windows is arbitrary. The 80 cm window is shown, as an example, in Figure 11.

These windows, which contain both of the composite phases, were modeled by the finite element method using membrane elements measuring 5×5 mm (these finite elements were defined by superimposing a 5×5 mm grid onto the photograph). The mechanical properties of stone and mortar were assumed to be deterministic and are reported in Table 1.

Phase	Material	Young's module E	Poisson's coefficient ν	T_{11} (MPa)	T_{12} (MPa)	T_{22} (MPa)	T_{33} (MPa)
1	Stone	12500	0.20	13021	2604	13021	10417
2	Mortar	1200	0.30	1319	396	1319	923

Table 1. Mechanical properties of the masonry phases: stone and mortar.

5.1. *Mechanical convergence.* Two types of boundary conditions were applied to the material windows:

(a) *Essential*, in terms of displacements u_i (Dirichlet, displacement-controlled),

$$u_i = \varepsilon_{ij}^0 x_j, \tag{26}$$

where ε_{ij}^0 are constant strains and x_j are point coordinates, or

(b) *Natural*, in terms of tractions t_i (Neumann, or stress-controlled),

$$t_i = \sigma_{ij}^0 n_j, \tag{27}$$

where σ_{ij}^0 are constant stresses and n_j are the components of the unit vector outward from the boundary. The average values of the strain in condition (a) are $\bar{\varepsilon}_{ij} = \varepsilon_{ij}^0$, while the average values of the stress in condition (b) are $\bar{\sigma}_{ij} = \sigma_{ij}^0$, where

$$\bar{\varepsilon}_{ij} = \frac{1}{V} \int_V \varepsilon_{ij} \, dV \quad \bar{\sigma}_{ij} = \frac{1}{V} \int_V \sigma_{ij} \, dV. \tag{28}$$

According to Hill [1963], when the volume considered is the representative volume element, then the relation between average stress and strain is the same for both types of boundary conditions (a) and (b).

The masonry is composed of two phases, so that:

$$\bar{\sigma}_{ij} = c_1 \bar{\sigma}_{ij}^{(1)} + c_2 \bar{\sigma}_{ij}^{(1)} = c_1 T_{ijkl}^{(1)} \bar{\varepsilon}_{kl}^{(1)} + c_2 T_{ijkl}^{(1)} \bar{\varepsilon}_{kl}^{(1)}, \tag{29}$$

$$\bar{\varepsilon}_{ij} = c_1 \bar{\varepsilon}_{ij}^{(1)} + c_2 \bar{\varepsilon}_{ij}^{(1)} = c_1 S_{ijkl}^{(1)} \bar{\sigma}_{kl}^{(1)} + c_2 S_{ijkl}^{(1)} \bar{\sigma}_{kl}^{(1)}, \tag{30}$$

where c_1 and c_2 are the fractional concentrations by volume ($c_1 + c_2 = 1$), $T_{ijkl}^{(1)}$ and $T_{ijkl}^{(1)}$ are the elastic stiffness constants, and $S_{ijkl}^{(1)}$ and $S_{ijkl}^{(1)}$ are the elastic compliances of the two phases.

Combining the previous equation we obtain:

$$T_{ijkl}^e \varepsilon_{kl}^0 = T_{ijkl}^{(1)}(\varepsilon_{kl}^0 - c_2 \bar{\varepsilon}_{kl}^{(1)}) + c_2 T_{ijkl}^{(1)} \bar{\varepsilon}_{kl}^{(1)} = T_{ijkl}^{(1)} \varepsilon_{kl}^0 + c_2 (T_{ijkl}^{(1)} - T_{ijkl}^{(1)}) \bar{\varepsilon}_{kl}^{(1)}, \tag{31}$$

$$S_{ijkl}^n \sigma_{kl}^0 = S_{ijkl}^{(1)}(\sigma_{kl}^0 - c_2 \bar{\sigma}_{kl}^{(1)}) + c_2 S_{ijkl}^{(1)} \bar{\sigma}_{kl}^{(1)} = S_{ijkl}^{(1)} \sigma_{kl}^0 + c_2 (S_{ijkl}^{(1)} - S_{ijkl}^{(1)}) \bar{\sigma}_{kl}^{(1)}, \tag{32}$$

where T_{ijkl}^e are the stiffness components under essential conditions and S_{ijkl}^n are the compliance components under natural conditions.

Applying displacement boundary conditions such that $\boldsymbol{\varepsilon}^0 = \boldsymbol{I}^{mn}$ ($\boldsymbol{I}^{mn}$ is the symmetrical matrix with all its components set to 0, except the component mn, which is set to 1) it becomes possible to determine the columns of the stiffness matrix $\boldsymbol{T}^e$.

Applying traction boundary conditions such that $\boldsymbol{\sigma}^0 = \boldsymbol{I}^{mn}$, it becomes possible to determine the columns of the compliance matrix $\boldsymbol{S}^n$. The stiffness matrix relative to natural conditions is obtained from $\boldsymbol{T}^n = [\boldsymbol{S}^n]^{-1}$.

We adopted an orthotropic equivalent continuum:

$$
\begin{pmatrix} \sigma_{11} \\ \sigma_{22} \\ \tau_{12} \end{pmatrix} = \begin{bmatrix} T_{11} & T_{12} & T_{13} \\ T_{21} & T_{22} & T_{23} \\ T_{31} & T_{32} & T_{33} \end{bmatrix} \begin{pmatrix} \varepsilon_{11} \\ \varepsilon_{22} \\ \gamma_{12} \end{pmatrix} = \begin{bmatrix} 2G_{11} + \lambda & \lambda & 0 \\ \lambda & 2G_{22} + \lambda & 0 \\ 0 & 0 & G_{12} \end{bmatrix} \begin{pmatrix} \varepsilon_{11} \\ \varepsilon_{22} \\ \gamma_{12} \end{pmatrix}, \quad (33)
$$

$$
\begin{pmatrix} \varepsilon_{11} \\ \varepsilon_{22} \\ \gamma_{12} \end{pmatrix} = \begin{bmatrix} S_{11} & S_{12} & S_{13} \\ S_{21} & S_{22} & S_{23} \\ S_{31} & S_{32} & S_{33} \end{bmatrix} \begin{pmatrix} \sigma_{11} \\ \sigma_{22} \\ \tau_{12} \end{pmatrix} = \begin{bmatrix} 1/E_{11} & -\nu_{12}/E_{22} & 0 \\ -\nu_{21}/E_{11} & 1/E_{22} & 0 \\ 0 & 0 & 1/G_{12} \end{bmatrix} \begin{pmatrix} \sigma_{11} \\ \sigma_{22} \\ \tau_{12} \end{pmatrix}, \quad (34)
$$

where $1 \equiv x$, $2 \equiv y$, $T_{ij} = T_{ji} \leftrightarrow S_{ij} = S_{ji}$, G_{ij}, λ are Lamé's constants, E_{ij} and ν_{ij} are Young's constants and Poisson's coefficients, respectively:

$$
E_{11} = \frac{(2G_{11} + \lambda)(2G_{22} + \lambda) - \lambda^2}{2G_{22} + \lambda}; \qquad E_{22} = \frac{(2G_{11} + \lambda)(2G_{22} + \lambda) - \lambda^2}{2G_{11} + \lambda};
$$

$$
\nu_{21} = \frac{\lambda}{2G_{11} + \lambda}; \qquad \nu_{12} = \frac{\lambda}{2G_{22} + \lambda}; \qquad \frac{\nu_{12}}{E_{22}} = \frac{\nu_{21}}{E_{11}}. \quad (35)
$$

In order to determine the three columns of the essential stiffness matrix, the following strains and boundary displacements were applied (for details see [Cluni and Gusella 2004]):

$$
\boldsymbol{\varepsilon}_1^0 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \leftrightarrow \begin{pmatrix} u_1 \\ u_2 \end{pmatrix} = \begin{pmatrix} x \\ 0 \end{pmatrix}, \qquad \boldsymbol{\varepsilon}_2^0 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \leftrightarrow \begin{pmatrix} u_1 \\ u_2 \end{pmatrix} = \begin{pmatrix} 0 \\ y \end{pmatrix},
$$

$$
\boldsymbol{\varepsilon}_3^0 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \leftrightarrow \begin{pmatrix} u_1 \\ u_2 \end{pmatrix} = \begin{pmatrix} x \\ y \end{pmatrix}. \quad (36)
$$

In order to determine the three columns of the natural compliance matrix, the following stresses and boundary tractions were applied

$$
\boldsymbol{\sigma}_1^0 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \leftrightarrow \begin{pmatrix} t_1 \\ t_2 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \qquad \boldsymbol{\sigma}_2^0 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \leftrightarrow \begin{pmatrix} t_1 \\ t_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix},
$$

$$
\boldsymbol{\sigma}_3^0 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \leftrightarrow \begin{pmatrix} t_1 \\ t_2 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}. \quad (37)
$$

Since an L-size window is used, the essential stiffness is designated $\boldsymbol{T}_L^e$ and the natural stiffness is designated $\boldsymbol{T}_L^n$.

Introducing

$$T^*_{ij,L} = \frac{T^e_{ij,L} + T^n_{ij,L}}{2} \qquad [\boldsymbol{T}^*_L = (\boldsymbol{T}^e_L + \boldsymbol{T}^n_L)/2], \tag{38}$$

we have [Huet 1990; Sab 1992; Ostoja-Starzewski 1998]:

$$\boldsymbol{T}^R_L \le \boldsymbol{T}^n_L \le \boldsymbol{T}^*_L \le \boldsymbol{T}^e_L \le \boldsymbol{T}^V_L, \tag{39}$$

where $\boldsymbol{T}^R_L$ and $\boldsymbol{T}^V_L$ are the Reuss and Voigt bounds, respectively ($\boldsymbol{A} \le \boldsymbol{B}$ means that $\boldsymbol{v}^T \boldsymbol{A} \boldsymbol{v} \le \boldsymbol{v}^T \boldsymbol{B} \boldsymbol{v}$ for all $\boldsymbol{v} \ne \boldsymbol{0}$).

As the size L of the window increases, the difference between $\boldsymbol{T}^e_L$ and $\boldsymbol{T}^n_L$ decreases. As L goes to infinity, the VE_L converges with the representative volume element, and $\boldsymbol{T}^*_L$ converges with the stiffness matrix of the nonrandom equivalent homogeneous continuum $\boldsymbol{T}^{\text{RVE}} = \boldsymbol{T}^{\text{hom}}$ [Sab 1992]

$$\lim_{L \to \infty} T^*_L = \boldsymbol{T}^{\text{RVE}} = \boldsymbol{T}^{\text{hom}}. \tag{40}$$

In effective applications, the window has a finite size which increases, giving the sequence: $L_k \ k = 1, \ldots, N$ with $i > j \to L_j > L_i$. Applying the mechanical convergence criterion introduced above the material window with $\hat{L} = L_n$ is an adequate estimate of the representative volume element when the equation (3) is verified:

$$\max_{ij} \left| \frac{T^e_{ij,\hat{L}} - T^n_{ij,\hat{L}}}{T^1_{ij}} \right| \le \Delta c, \tag{41}$$

where $T^1_{ij} = |T^e_{ij,L_1} - T^n_{ij,L_1}|$ and $\Delta c \in R^+$ is a fixed admissible error.

In the present application the size L was increased from 20 cm to 130 cm (the 130 cm window is shown in Figure 12).

At first, the numerical results confirmed the hypothesis in Equations (33) and (34); in fact the stiffness components $T_{i3} \ i = 1, 2$ and the compliance components $S_{i3} \ i = 1, 2$ were negligible with respect to the others. Table 2 reports the Young's moduli and Lamé constants relative to essential and natural conditions. The convergence of these characteristics is shown in Figure 13. Poisson's coefficients converge very quickly (for the window with $L = 130$ cm: $\nu_{21} = 0.199$, $\nu_{12} = 0.203$ for essential conditions, and $\nu_{21} = 0.178$, $\nu_{12} = 0.180$ for natural conditions).

5.2. *Probabilistic convergence.*

The previous mechanical convergence does not permit us to state with adequate reliability that the test-window is a good approximation of the representative volume element. Differences between essential and natural evaluations of the elastic moduli decrease very quickly as L increases. The convergence could be accelerated by averaging over several samples of the same

L	Essential Boundary Conditions			Natural Boundary Conditions		
(cm)	E_{11} (MPa)	E_{22} (MPa)	G_{12} (MPa)	E_{11} (MPa)	E_{22} (MPa)	G_{12} (MPa)
20	7807.03	7698.80	3271.50	6408.04	6338.20	2487.00
30	7890.46	7752.42	3162.00	7086.51	6665.78	2710.50
40	7929.02	7852.14	3255.50	7214.07	7290.57	2883.00
50	7894.35	7290.79	3049.50	7401.59	6709.97	2724.00
60	7840.72	7079.91	2917.51	7504.52	6584.28	2625.44
80	7937.80	7055.86	2893.87	7658.00	6601.37	2655.77
100	7813.77	6929.10	2816.60	7543.66	6677.43	2657.16
120	8095.70	7290.24	2932.05	7879.92	7058.18	2784.91
130	8143.23	7282.57	2940.33	7958.80	7046.80	2801.83

Table 2. Mechanical characteristics obtained under essential and natural boundary conditions as L, the side of the square window, increases.
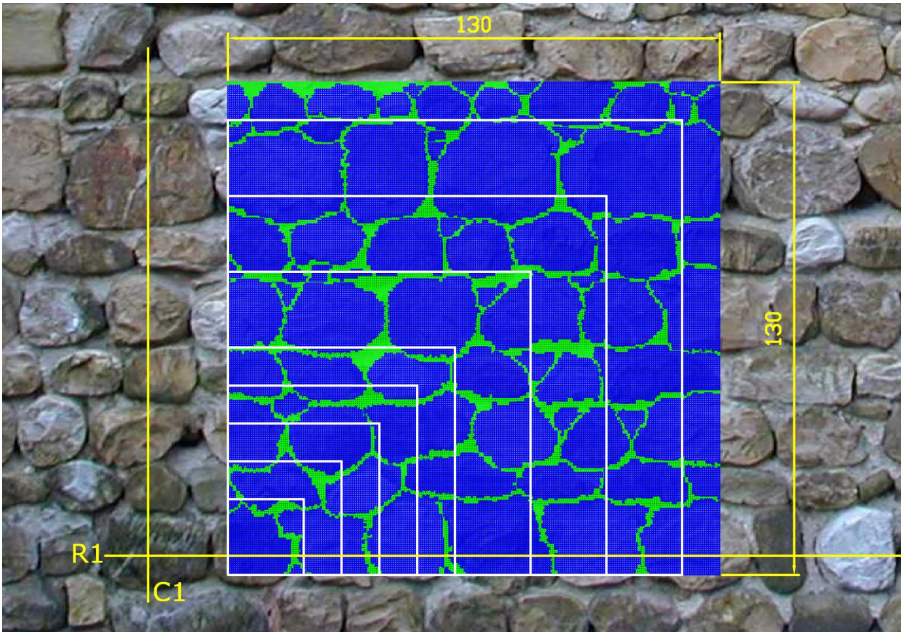
Figure 12. Square material window with $L = 130$ cm and finite element model.

size, as noted in [Cluni and Gusella 2004]; the averaging over samples with different sizes is not appropriate because these samples have different and unknown statistical weights.
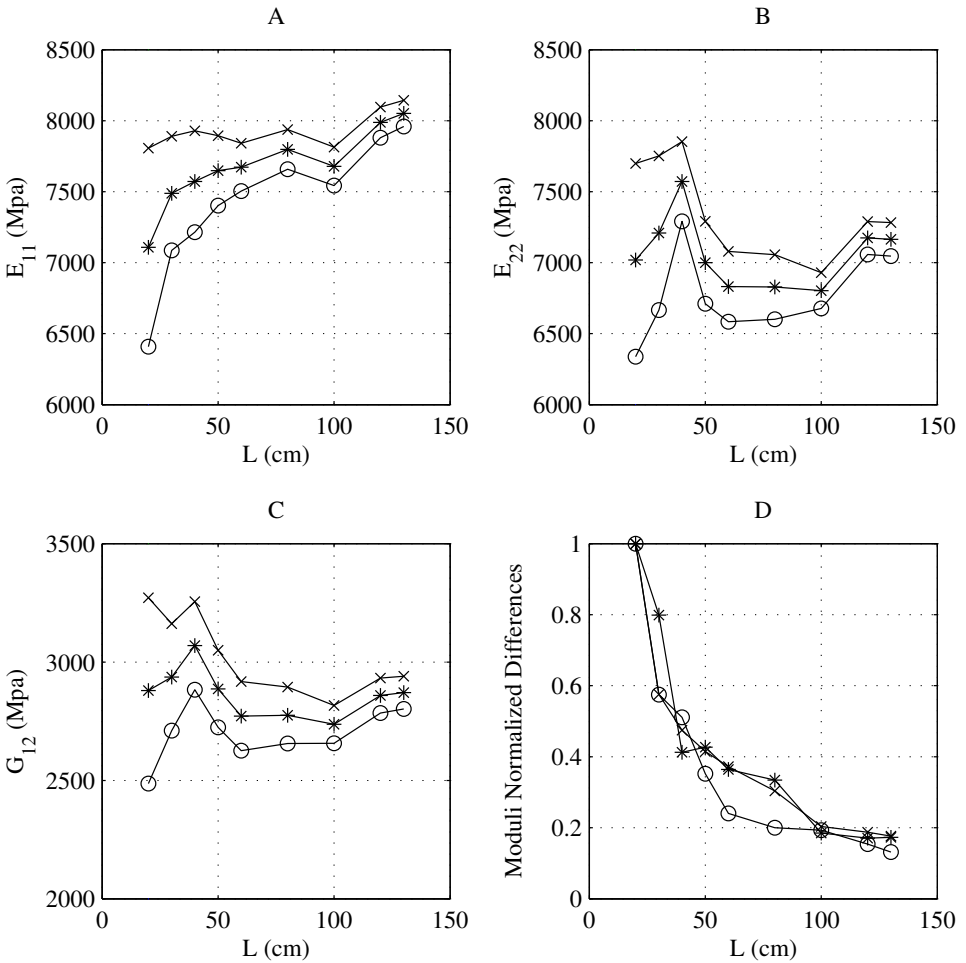
Figure 13. (A), (B) and (C): behavior of the mechanical properties of the test-window as L increases: "$\times$" essential, "$\circ$" natural, "$*$" mean value; (D) normalized differences between essential and natural evaluations for E_{11}, E_{22}, and G_{12}.

Even with a single 80 cm window, the differences between essential and natural moduli are limited. However, for the 130 cm window these differences are quite similar to those noted above, whereas the values of essential and natural elastic moduli and their mean values are significantly different from those of the 80 cm window (Figure 13 (A)–(C)).

This discrepancy can be explained by observing that the 80 cm window does not adequately represent the masonry wall because it does not contain a representative

	Bst		Hst		Bm		Hm	
Entire wall	E_{Bst}	Σ_{Bst}	E_{Hst}	Σ_{Hst}	E_{Bm}	Σ_{Bm}	E_{Hm}	Σ_{Hm}
	18.87	5.86	11.69	4.73	4.20	4.22	1.81	1.29
VE_L	E_{Bst}^L	Σ_{Bst}^L	E_{Hst}^L	Σ_{Hst}^L	E_{Bm}^L	Σ_{Bm}^L	E_{Hm}^L	Σ_{Hm}^L
80×80	21.61	3.45	13.34	5.02	2.79	1.56	2.06	1.62
130×130	19.90	5.57	12.75	4.98	3.70	3.48	1.83	1.25

Table 3. Mean E and standard deviation Σ of the width and the height of the stones, head and bed mortar joints, for the portion of wall shown in Figure 2 and for the material windows with $L = 80\,\text{cm}$ (Figure 11) and $L = 130\,\text{cm}$ (Figure 12).

sample of stones: for example, the course with very large stones (row R_7 in Figure 2) is not taken into account.

It should, therefore, be noted that as the size of the test-window increases, the differences between the essential and natural components decrease, but the convergence to the equivalent homogeneous continuum components is not uniform. In order to overcome this limitation the probabilistic convergence criterion introduced above was applied, taking into account Equations (21), (22), (23) and (25).

Excluding windows with $L < 60\,\text{cm}$, because the number of stones and joints was too small, the means and standard deviations of the sample relative to the stone and mortar joint sizes were compared with those of the entire wall (see Table 3).

Moreover, for $L \geq 80\,\text{cm}$, the probability density functions of these samples were compared with those of the entire wall. Comparisons relative to the width and height of the stones in the windows with $L = 80\,\text{cm}$, the window with $L = 130\,\text{cm}$, and the wall as a whole, are shown in Figure 14. Comparisons relative to the mortar joint sizes are shown in Figure 15.

The window with $L = 130\,\text{cm}$ is adequately representative of the statistical distribution of stones and mortar joints in the masonry (only the sample of the bed joint thickness should be improved). In fact, the differences (Table 3) in statistical moments (mean and standard deviation) and in probability density functions (Figures 14 and 15) between the 130 cm window and the entire wall were found to be sufficiently limited.

Moreover, it is necessary to check that the pattern of the window is representative of the entire wall. This was done by analyzing the autocorrelation functions of the rows and columns. The results relative to the 130 cm window are shown in Figure 16. These functions match those relative to the entire wall (Figure 7 and Figure 9).
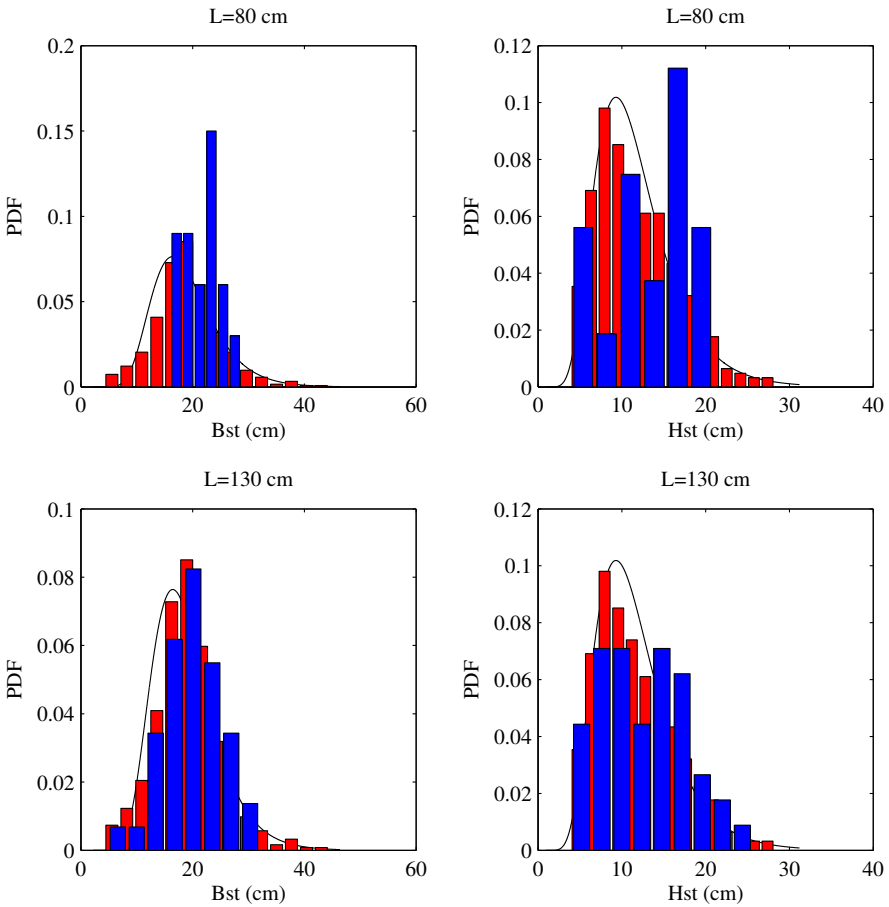
Figure 14. Comparison among probability density functions (PDF) of the width and the height of the stones relative to the windows (blue bars) with $L = 80$ cm and with $L = \hat{L} = 130$ cm, the entire masonry wall (red bars), and the log-normal curve (see Figure 3).

In particular, the difference in mean autocorrelation functions of the rows between the 130 cm window $\bar{A}C_{R,R}^{\hat{L}}(\xi)$ and the entire wall $\bar{A}C_{RR}(\xi)$ proved to be limited; see Figure 16 (A).

A similar result was obtained for the difference in mean autocorrelation functions of the columns: $\bar{A}C_{C,C}^{\hat{L}}(\eta)$ and $\bar{A}C_{C,C}(\eta)$, respectively; see Figure 16 (B).

Further studies would be necessary to gauge the importance of higher-order statistical properties. In any case, the elastic moduli of the equivalent homogeneous continuum are related to the average values of the strain and stress, and,
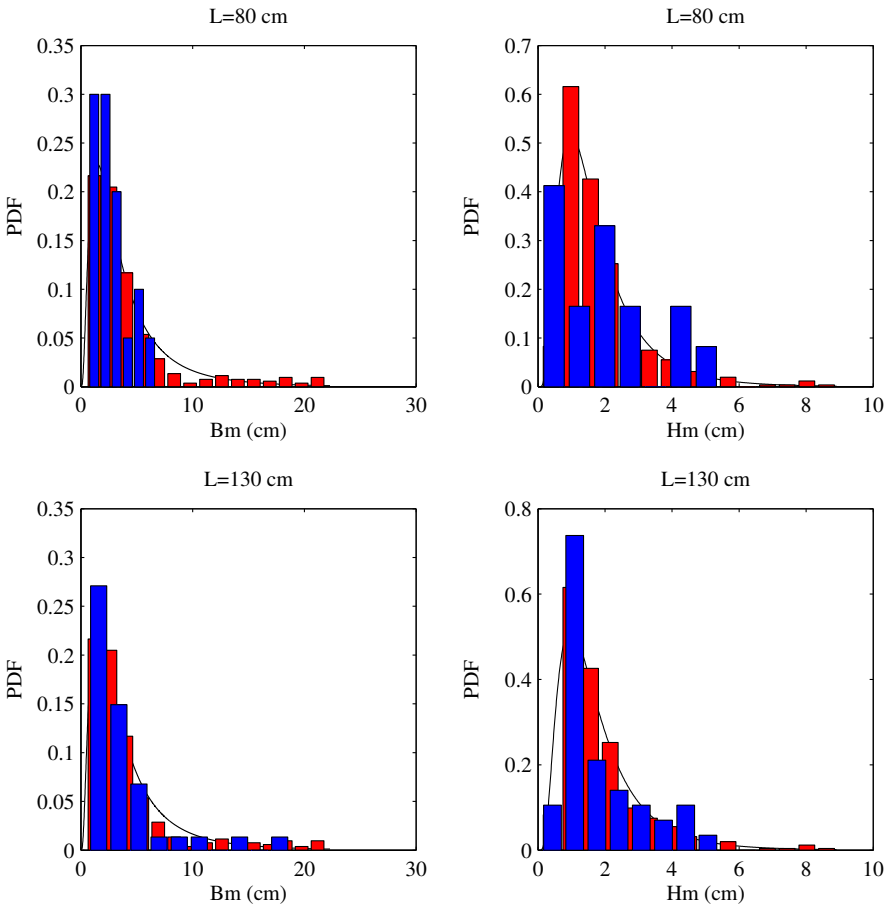
Figure 15. Comparison among probability density functions (PDF) of the characteristics of the mortar joints relative to the windows (blue bars) with $L = 80$ cm and with $L = \hat{L} = 130$ cm, the entire masonry wall (red bars), and the log-normal curve (see Figure 4).

consequently, the checks based on the proposed statistical descriptors would seem to be adequate.

Finally, by taking into account the mechanical and statistical results presented above, it becomes possible to consider the 130 cm window an adequate estimate of the statistically equivalent representative volume element, so that

$$T_{ij}^{\mathrm{hom}} = T_{ij}^{\mathrm{RVE}} = T_{ij,\hat{L}}^{*} = \frac{T_{ij,\hat{L}}^{e} + T_{ij,\hat{L}}^{n}}{2}, \quad \hat{L} = 130 \text{ cm}. \tag{42}$$
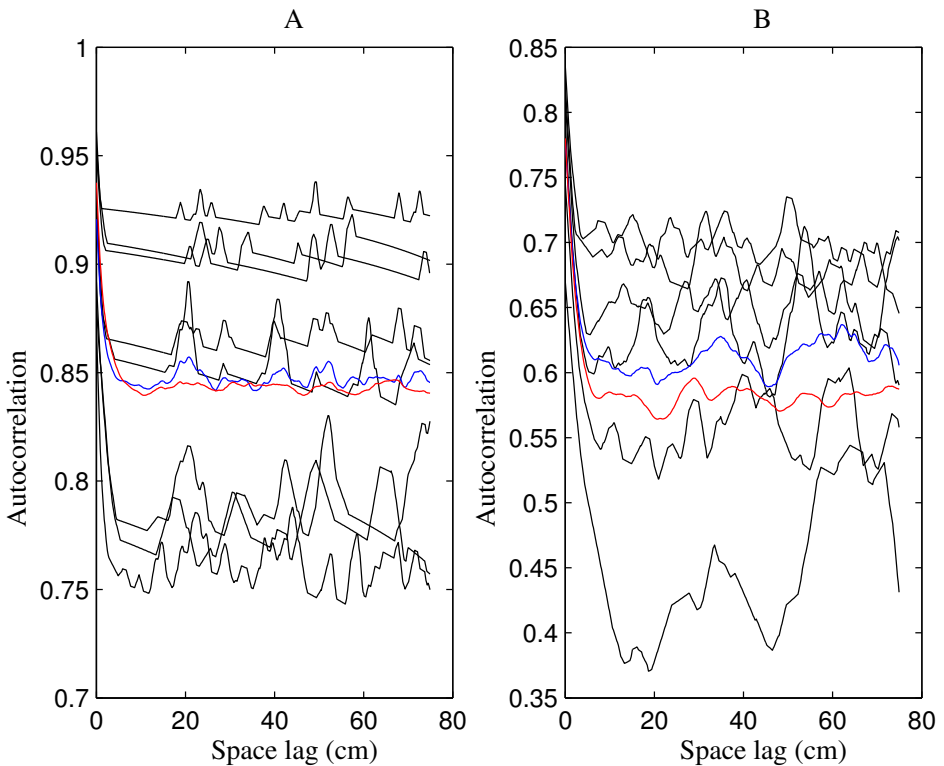
Figure 16. (A) Autocorrelation functions of the rows of the window with $\hat{L} = 130\,\text{cm}$; mean autocorrelation functions: $\bar{A}C_{R,R}^{\hat{L}}(\xi)$ for the window (blue line), $\bar{A}C_{RR}(\xi)$ for the wall (red line); (B) Autocorrelation functions relative to the columns of the window with $\hat{L} = 130\,\text{cm}$; mean autocorrelation functions: $\bar{A}C_{C,C}^{\hat{L}}(\eta)$ for the window (blue line), $\bar{A}C_{C,C}(\eta)$ for wall (red line).

6. Conclusions

This paper deals with the homogenization of masonry with nonperiodic microstructures. The masonry is treated as a random heterogeneous material and statistical descriptors are introduced. On the basis of observations concerning masonry construction, an orthogonal grid of rows and columns is used to determine the geometric characteristics of stones and of head- and bed-mortar joints. This allows one to estimate the statistical moments and the probability density function of these geometrical characteristics.

It is well known that the mechanical behavior of masonry is significantly influenced by its texture. In order to describe this feature, indicator (or characteristic)

functions are introduced to describe the random alternation of stone and mortar in the rows and columns of the orthogonal grid. Moreover, shifted-area functions and correlation functions are taken into account to highlight the random features of the masonry texture.

The modeling of the masonry as a random field was introduced to permit one to improve the finite-size test-window method in estimating the representative volume and the elastic moduli of the equivalent homogeneous medium. Since the classic mechanical criterion, which requires that differences between moduli evaluated under essential and natural conditions be limited, does not assure a uniform convergence of the test-window to the representative volume element, a probabilistic criterion is introduced.

This convergence criterion requires that the material window be statistically similar to the complete masonry wall, where this similarity is assured by the minimizing of differences among the statistical descriptors relative to the window and the entire masonry wall.

When both of these criteria are respected, the window can be considered an adequate estimate of the statistical equivalent representative volume, and the homogeneous continuum moduli can be obtained by averaging those relative to essential and natural boundary conditions.

A numerical application highlights the importance of applying the two criteria jointly and the effectiveness of the method proposed.

References

[Alpa and Monetto 1994] G. Alpa and I. Monetto, "Microstructural model for dry block masonry walls with in-plane loading", *J. Mech. Phys. Solids* **42**:7 (1994), 1159–1175.

[Anthoine 1995] A. Anthoine, "Derivation of the in-plane elastic characteristics of masonry through homogenization theory", *Int. J. Solids Struct.* **32**:2 (1995), 137–163.

[Bensoussan et al. 1978] A. Bensoussan, J.-L. Lions, and G. Papanicolau, *Asymptotic analysis for periodic structures*, North-Holland, Amsterdam, 1978.

[Bochenek and Pyrz 2004] B. Bochenek and R. Pyrz, "Reconstruction of random microstructures – a stochastic optimization problem", *Comput. Mater. Sci.* **31**:1-2 (2004), 93–112.

[Cecchi and Sab 2002] A. Cecchi and K. Sab, "A multi-parameter homogenization study for modeling elastic masonry", *Eur. J. Mech. A: Solids* **21**:2 (2002), 249–268.

[Christensen 1980] R. Christensen, *Mechanics of composite materials*, Wiley, New York, 1980.

[Cluni and Gusella 2004] F. Cluni and V. Gusella, "Homogenization of non-periodic masonry structures", *Int. J. Solids Struct.* **41**:7 (2004), 1911–1923.

[De Buhan and De Felice 1997] P. De Buhan and G. De Felice, "A homogenization approach to the ultimate strength of brick masonry", *J. Mech. Phys. Solids* **45**:7 (1997), 1085–1104.

[Drugan and Willis 1996] W. Drugan and J. Willis, "A micromechanics-based nonlocal constitutive equation and estimates of representative volume element size for elastic composites", *J. Mech. Phys. Solids* **44**:4 (1996), 497–524.

[Hill 1963] R. Hill, "Elastic properties of reinforced solids: some theoretical principles", *J. Mech. Phys. Solids* **11**:5 (1963), 357–372.

[Huet 1990] C. Huet, "Application of variational concepts to size effects in elastic heterogeneous bodies", *J. Mech. Phys. Solids* **38**:6 (1990), 813–841.

[Luciano and Sacco 1997] R. Luciano and E. Sacco, "Homogenization technique and damage model for old masonry material", *Int. J. Solids Struct.* **34**:24 (1997), 3191–3208.

[Maier et al. 1991] G. Maier, A. Nappi, and E. Papa, "On damage and failure of brick masonry", pp. 223–245 in *Experimental and numerical methods in earthquake engineering* (Ispra, 1991), edited by J. Donea and P. M. Jones, Kluwer, Dordrecht, 1991.

[Ostoja-Starzewski 1998] M. Ostoja-Starzewski, "Random field models of heterogeneous materials", *Int. J. Solids Struct.* **35**:19 (1998), 2429–2455.

[Pande et al. 1989] G. Pande, J. X. Liang, and J. Middleton, "Equivalent elastic moduli for brick masonry", *Comput. Geotech.* **8**:3 (1989), 243–265.

[Pietruszczak and Niu 1992] S. Pietruszczak and X. Niu, "A mathematical description of macroscopic behaviour of brick masonry", *Int. J. Solids Struct.* **29**:5 (1992), 531–546.

[Povirk 1995] G. L. Povirk, "Incorporation of microstructural information into models of two-phase materials", *Acta Metall. Mater.* **43**:8 (1995), 3199–3206.

[Sab 1992] K. Sab, "On the homogenization and the simulation of random materials", *Eur. J. Mech. A: Solids* **11**:5 (1992), 585–607.

[Sab and B. 2005] K. Sab and N. B., "Periodization of random media and representative volume element size for linear composites", *C. R. Mecanique* **333**:2 (2005), 187–195.

[Sanchez-Palencia 1980] E. Sanchez-Palencia, *Non homogeneous media and vibration theory*, Lecture Notes in Physics **127**, Springer, Berlin, 1980.

[Suquet 1987] P. M. Suquet, "Elements of homogenization for inelastic solid mechanics", in *Homogenization techniques for composite media* (Udine, 1985), edited by E. Sanchez-Palencia and Z. A., Lecture Notes in Physics **272**, Springer, Berlin, 1987.

[Terada et al. 2000] K. Terada, M. Hori, T. Kyoya, and N. Kikuchi, "Simulation of the multi-scale convergence in computational homogenization approaches", *Int. J. Solids Struct.* **37**:16 (2000), 2285–2311.

[Torquato 2001] S. Torquato, *Random heterogeneous materials*, Springer, New York, 2001.

[Torquato and Stell 1982] S. Torquato and G. Stell, "Microstructure of two-phase random media, I: The n-point probability functions", *J. Chem. Phys.* **77**:4 (1982), 2071–2077.

[Šejnoha and Zeman 2002] M. Šejnoha and J. Zeman, "Overall viscoelastic response of random fibrous composites with statistically quasi uniform distribution of reinforcements", *Comput. Methods Appl. Mech. Eng.* **191**:44 (2002), 5027–5044.

[Šejnoha et al. 2004] M. Šejnoha, J. Zeman, and J. Novak, "Homogenization of random masonry structures, comparison of numerical methods", pp. 13–16 in *17th ASCE Engineering Mechanics Division Conference* (Newark (DE), 2004), edited by J. T. Kirby et al., University of Delaware, Newark, 2004.

[Zeman and Šejnoha 2001] J. Zeman and M. Šejnoha, "Numerical evaluation of effective elastic properties of graphite fiber tow impregnated by polymer matrix", *J. Mech. Phys. Solids* **49**:1 (2001), 69–90.

VITTORIO GUSELLA: guse@unipg.it
*Department of Civil and Environmental Engineering, University of Perugia, Via G. Duranti,
Perugia 06125, Italy*

FEDERICO CLUNI: f.cluni@tiscali.it
Via Cima 36, fraz. Cerqueto, Marsciano (PG) 06052, Italy

THE STRESS-MINIMIZING HOLE IN AN ELASTIC PLATE UNDER REMOTE SHEAR

SHMUEL VIGDERGAUZ

Conformal mappings provide an elegant formulation for planar elastostatic problems. Here, the mapping function coefficients are used in a new manner as design variables in the genetic-algorithm (GA) approach to find a piecewise smooth optimal shape of a single traction-free hole in an elastic plate that minimizes the local stresses under remote shear. This scheme is sufficiently fast and accurate to numerically show that the sought-for shape generates tangential stress of constant absolute value, equal to 30% less than the stress concentration factor (SCF) for the commonly used circular hole. The shape has four symmetrically located corners, and the stress changes sign while remaining finite as it rounds each corner. This is the same shape as the energy-minimizing contour identified in 1986 by the author and Cherkaev for the same load. Other nontrivial examples are given to demonstrate the potential of the approach. Methodologically, this article continues the optimization study first conducted by the author and Cherkaev (*J. Appl. Math. Mech.* **50**:3 (1986), 401–404) and subsequently by Cherkaev et al. (*Internat. J. Solids Structures* (**35**):33, 4391–4410).

1. Introduction

Designing elastic structures to diminish the stresses around construction holes in flat plates remains an actual problem in spite of intensive studies carried out in the area over the last decades. Various stress-reducing technologies, such as auxiliary unloading holes, reinforcement rings and others are known so far, each posing its own elastostatic problem. In most applications, the hole area matters much more than its shape, which thus permits a certain freedom in design. Prompted by this, our concern here is with optimization of the hole shapes to minimize the stress concentration factor (SCF), denoted by $\mathcal{K}$, and defined as the maximum modulus of the tangential stress along the holes, at unit remote load. The lesser the factor, the stronger the hole-weakened construction will be. Technologically, only piecewise smooth holes with a finite number of corner points may be used as stress-minimizers. This is assumed in what follows.

The SCF, though local in nature, is obtainable only from full-scale elastic stress solutions, which are rarely known in a closed form. A nontrivial example is a uniform stress distribution

$$\sigma_{\tau\tau} = \text{Const} \tag{1}$$

occurring at specifically shaped interfaces [Cherepanov 1974]. Here τ stands for the contour unit tangent. These equistress shapes do exist in an infinite plane in any number and mutual arrangement, provided the remote deviatoric load is relatively small, as stated accurately in the next section. In particular, pure bulk gives $\mathcal{K} = 2$ independently of the geometry of the hole set. The equistress condition not only prevents the stress concentration at the hole but also provides the global minimum of $\mathcal{K}$ over all shapes at given bulk-type loading [Vigdergauz 1976].

In the opposite case of pure shear, the $\mathcal{K}$-optimal shapes are yet unknown, even for the simplest configuration of a single traction-free hole. To make progress in this stubborn problem, we replaced the local criterion $\mathcal{K}$ in [Vigdergauz and Cherkayev 1986] by the less severe global criterion of minimizing the hole-induced energy perturbation δW, taken at unit load and related to the hole area. In the effective medium theory, a minimum in δW corresponds to maximum shear rigidity of a dilute planar composite when the holes are far apart and have little influence on each other (see [Torquato 2002], for instance). Derived variationally, this optimality condition then implies constancy of the *absolute* value of the tangential stresses along the sought-for contour

$$|\sigma_{\tau\tau}| = \text{Const.} \tag{2}$$

In contrast to the equistress condition (1), the less restrictive identity (2) may be compatible with remote shear load, provided the stresses change sign across a finite number of angular points. Using this assumption, the resultant near-square hole shape is found numerically [Vigdergauz and Cherkayev 1986] with $\delta W = 3.714 \ldots$.

Conversely, (2) is an immediate result of (1); hence equistress $\mathcal{K}$-optimal shapes, when they exist, also minimize the energy perturbation δW under a fixed load.

Though similar, identities (1) and (2) work differently. The second one is used as a prerequisite in directly finding the energy-minimizing shapes at any far load (see details in [Cherkaev et al. 1998]), while $\mathcal{K}$-optimality of the equistress condition (1) is revealed only *a posteriori* [Vigdergauz 1976]. The proof is based on the maximum module principle, which is not applicable to the shear-type stress field associated with (2). This brings up the following inverse problem:

Among all possible continuous curves, find the shape of a single hole in an elastic plate so as to minimize the factor $\mathcal{K}$ under pure shear at infinity.

The challenge is very interesting mathematically, and its practical importance is in providing the theoretical bound of material behavior. The latter is significant to the designer, who can determine how far the actual structure is from the theoretical optimum. As excellent sources on the direct problem of finding $\mathcal{K}$ for various shapes, we refer to the classical monograph [Savin 1961] and the recently reprinted [Pilkey 1997].

In the absence of the $\mathcal{K}$-optimal precondition, semianalytical and numerical global optimization are the methods of choice in solving the problem. The purpose of this paper is twofold: to present a new numerical optimization scheme, and to produce new optimal solutions not available in the literature.

Computationally, any optimization process involves two main ingredients: the solution of a given direct boundary value problem, which has to be repeated many times, and a minimization scheme. As the first ingredient, we choose the Kolosov–Muskhelishvili potentials $\varphi(z)$ and $\psi(z)$ [Muskhelishvili 1963], together with the conformal mapping of the sought-for contour onto a circle. Thanks to the power of complex variable techniques, this combination provides an effective numerical solver of the direct problem. Especially relevant here is the scheme by Kalandiya [1975], in which the contour stresses are solved from an infinite system of linear algebraic system with easy-to-compute coefficients. In practice, the system size and the Laurent series expansion of the mapping function are both truncated at different finite orders. We have substantially improved the algorithmic performance of the scheme in the following two aspects. First, we show that a finite mapping expansion generates *exactly* a finite-size system, thus allowing to avoid the additional truncation error. Second, the analytical manipulations over $\sigma_{\tau\tau}$ are performed to the maximum extent resulting in a simple rational expression. The latter provide evaluation of the local-type function $\mathcal{K}$ with sufficient accuracy to be incorporated into the genetic algorithm (GA) optimization process rapidly gaining use in elasticity. For design variables we choose coefficients of the mapping function instead of the contour nodal points, used in [Vigdergauz 2001b; 2001a; 2002]. This drastically reduces the number of design variables and allows the analytic calculation of all the integrals that occur. On this basis, the energy-minimizing shapes (2) are numerically shown to remain $\mathcal{K}$-optimal under pure shear as in the equistress case (1) under bulk load. We also calculate the energy-maximizing holes for a small number of mapping coefficients, to illustrate the "worst" possible situation.

In Section 2 we recall basic facts of complex variable theory applied to plane elasticity. Section 3 states the optimization problem and details the mixed GA/conformal mapping solving technique. Section 4 details the novel scheme of evaluating tangential stresses along the hole shape. Numerical comparison with the less accurate approach of Cherkaev et al. [1998] is performed in Section 5. Some analytical consequences are deduced in Section 6. They serve as a benchmark

for GA testing and calibration (Section 7). The numerical results are given and discussed in Section 8. Section 9 contains some final remarks.

2. The Kolosov–Muskhelishvili approach in plane elasticity

Consider the setup in Figure 1. Let an infinite elastic plane E be weakened by a hole with a piecewise smooth boundary L enclosing the origin of xy-plane. The curve L divides the plane in the hole region S_1 of finite area f_1 and the outside region $S_2 = E \setminus S_1$, filled with a linearly elastic phase. Let the plate be remotely loaded by uniform nontangential stresses

$$\sigma_{xx}^0 = P, \quad \sigma_{yy}^0 = Q, \quad \sigma_{xy}^0 = 0. \tag{3}$$

Let S_2 be conformally mapped onto the exterior Σ_2 of the unit circle Σ_1 with the boundary l in the auxiliary plane $F = \Sigma_1 \cup \Sigma_2$ of the complex variable ζ. Up to a scaling factor, the mapping function $\omega(\zeta) : S_2 + L \rightarrow \Sigma_2 + l$ is represented as a
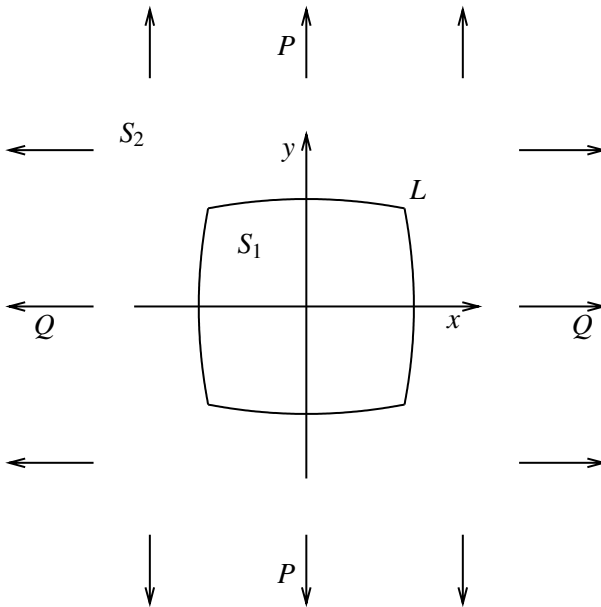
Figure 1. An infinite plate with a traction-free hole under uniform stresses. The cases $P = Q$ and $P = -Q$ correspond to remote bulk and shear, respectively. The piecewise smooth boundary of the hole has a certain rotational symmetry and possibly a finite number of corners.

Laurent series

$$\omega(\zeta) = \zeta + \sum_{k=1}^{\infty} d_k \zeta^{-k}, \tag{4}$$

with

$$f_1 = \pi \left(1 - \sum_{k=1}^{\infty} k |d_k|^2 \right). \tag{5}$$

(See [Ahlfors 1978] for this and subsequent background facts from complex function theory.) Since the map $\omega(\zeta)$ must be one-to-one, its coefficients fall in the intervals

$$-\frac{1}{\sqrt{m}} \leq d_m \leq \frac{1}{\sqrt{m}}, \quad m = 1, 2, \ldots. \tag{6}$$

The Airy stress function remains biharmonic under the map and hence is expressed through the Kolosov–Muskhelishvili (KM) potentials $\Phi_0(\zeta)$, $\Psi_0(\zeta)$, analytic in Σ_2 [Muskhelishvili 1963], with far field asymptotics governed by (3):

$$\Phi_0(\zeta) = B + \Phi(\zeta), \ \Psi_0(\zeta) = \Gamma + \Psi(\zeta); \ \zeta \in \Sigma_2, \quad \Phi(\zeta), \ \Psi(\zeta) = O\left(|\zeta|^{-2}\right), \tag{7a}$$

$$4B = Q + P, \quad 2\Gamma = Q - P, \quad \operatorname{Im} B = \operatorname{Im} \Gamma = 0. \tag{7b}$$

As in (4), they also have convergent series expansions (the summation begins with $n = 2$ to match the asymptotics (7a)):

$$\Phi(\zeta) = \sum_{k=2}^{\infty} a_k \zeta^{-k}, \quad \Psi(\zeta) = \sum_{k=2}^{\infty} b_k \zeta^{-k}; \qquad \zeta \in \Sigma_2 + l, \tag{8}$$

with (see for instance Vigdergauz [2001b])

$$\delta W = 2\pi f_1^{-1} \left(2\Gamma a_2 + B b_2 \right). \tag{9}$$

For simplicity, suppose that the hole boundary is traction-free. Then the stresses $\sigma_{\varrho\varrho}(\xi)$ and $\sigma_{\varrho\theta}(\xi)$ vanish along it: $\xi = \exp i\theta$, $\varrho = 1$ in the plane F, thus forming the boundary condition for the KM potentials

$$-\frac{2}{\xi^2} \overline{\omega'(\xi)} \operatorname{Re} \Phi_0(\xi) + \overline{\omega(\xi)} \Phi_0{}'(\xi) + \omega'(\xi)\Psi_0(\xi) = 0; \quad \xi \in l. \tag{10}$$

The nonzero stress component $\sigma_{\tau\tau}(t(\xi)) \equiv \sigma_{\theta\theta}(\xi)$ along l possesses the form [Muskhelishvili 1963]

$$\sigma_{\theta\theta}(\xi) = 4Re\Phi_0(\xi) = 4B + 4\sum_{k=2}^{\infty} a_k \cos(k\theta); \quad \xi \in l. \tag{11}$$

Of course, the stresses and strains at any point inside Σ_2 are also expressed in $\Phi_0(\zeta)$, $\Psi_0(\zeta)$ [Muskhelishvili 1963]. We omit the formulas here to save room.

Identities (7) and (10) form the boundary value problem in the KM potentials. This problem is uniquely solvable [Muskhelishvili 1963], at least for any piecewise smooth inclusion shape L, which is therefore the only factor defining the problem's solving complexity. Specifically, the equistress principle (1) yields $\sigma_{\varrho\theta}(\xi) = 4B$, and hence (see [Cherepanov 1974])

$$\Phi_0(\zeta) = B, \quad \omega(\zeta) = \zeta - \kappa\zeta^{-1}, \quad \omega'(\zeta)\Psi_0(\zeta) = -B\frac{\kappa\zeta^2 + 1}{\zeta^2 + \kappa}; \quad \kappa \equiv \frac{\Gamma}{2B}.$$

The equistress shape appears to be an ellipse that exists if and only if

$$|\kappa| \equiv |d_1| < 1,$$

or equivalently, thanks to (3) and (7b),

$$\sigma_{xx}^0 \sigma_{yy}^0 \geq 0.$$

3. Problem reformulation, design variables and basic GA scheme

In contrast, under the shear-dominated far field (3) when $\sigma_{xx}^0, \sigma_{yy}^0$ are of opposite signs, the $\mathcal{H}$-optimal hole shape cannot be found in a closed form. Thus, numerical methods are called for. In computational practice, the expansion (4) is necessarily truncated at a finite number N of first terms. With this in view, our optimization problem is reformulated as follows:

At a given finite number N of mapping coefficients and pure shear field $B = 0$, find the $\mathcal{H}$-optimal hole shape on which

$$\mathcal{H} \equiv \max_{t \in L_N} |\sigma_{\tau\tau}(t)| \equiv \max_{\xi \in l} |\sigma_{\theta\theta}(\xi)| \xrightarrow[\{L_N\}]{} \min, \tag{12}$$

where $\{L_N\}$ denotes the set of all curves mapped onto the unit circle by $\omega(\zeta)$ with any admissible finite set $\{d_m, m = 1, \ldots, N\}$, and $d_m = 0, m > N$.

In our opinion, the N-parametric nonlinear optimization problem (12) with the linear restrictions (6) is well suited to be solved by the genetic algorithm advanced in [Holland 1975]. This heuristic method performs a nongradient stochastic search for the global optimum by mimicking the Darwinian principal of survival of the fittest through blind mutation and natural selection over successive generations; see [Gen and Cheng 1997] for a state-of-the-art review and many references. GA specifics, as applied to shape optimization in planar elasticity, are illustrated in [Vigdergauz 2001b; 2001a; 2002] where the radii of nodal points equally spaced along the sought-for contour are directly taken as the GA design variables without a conformal premapping. This scheme helps find the δW-optimal hole shapes, but is not sufficiently accurate to handle the local criterion $\mathcal{H}$. For this reason, optimized contours are represented here through a set of mapping coefficients rather than through nodal points. This not only dramatically reduces the number of required

design variables but also allows the development of an efficient fitness evaluation scheme (see the next sections) which is easily included into a standard GA configuration.

On the other hand, in contrast to the nodal-based shape encoding, a mapped contour may have self-intersections even assuming inequalities (6) on the mapping coefficients. Other uniqueness conditions imply that all the roots of the polynomial $P_{N+1}(\zeta) = \zeta^{N+1}\omega'(\zeta)$ lie strictly inside the unit circle (see [Ahlfors 1978]):

$$P_{N+1}(\zeta) = \zeta^{N+1} - \sum_{m=1}^{N} m d_m \zeta^{N-m} = \prod_{m=1}^{N+1} (\zeta - \lambda_m);$$
$$|\lambda_m| < 1, \quad m = 1, \ldots, N+1. \quad (13)$$

Though more restrictive than (6) these bounds are also only necessary but not sufficient to avoid self-intersections. Mathematically, this is because for $N > 1$, such inequalities provide a one-to-one mapping only locally rather than globally, as exemplified in Figure 2, where the mapping terms are strictly inside the intervals (6).
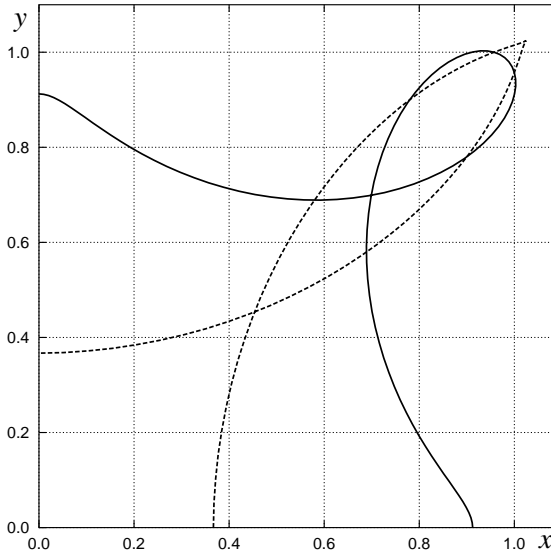
Figure 2. Two-term conformal mapping of the unit circle onto a self-crossing line with square symmetry: $d_3 \approx 0.24451093$, $d_7 \approx 0.15696709$, the largest root modulus is approximately 1.47730100 is outside the circle (the solid line), and $d_3 \approx -0.54182861$, $d_7 \approx -0.09122962$, with all the roots inside the circle (the dotted line). The loop areas enter identity (5) with a negative sign.

To exclude self-intersections, we use the geometric fact that a closed curve

$$\omega(\xi) = \rho(\xi) \exp i\vartheta(\xi), \quad \xi = \exp i\theta \in l, \tag{14}$$

with p-fold rotational symmetry is intersection-free if and only if $\vartheta(\xi)$ is an increasing function of θ in the interval $0 < \theta < \pi/p$. Though seemingly cumbersome, this condition is quickly checked over a discrete set of points along the irreducible part of mapped curves with penalizing their fitness to the extent by which (14) is violated at the first point so detected.

(We note in passing that the nonmonotony of finite-term mappings produces a closed loop of a negative area. This formally results in zero-area curves with self-intersections rather than in physically reasonable zero-area slits. The trivial exception is the case $p = 1$, when the function $\omega(\zeta) = \zeta + \zeta^{-1}$ maps the unit circle to a rectilinear slit

$$L : -2 \leq t \leq 2$$

at the x-axis of the physical plane E. In Section 8 this note is used to explain the numerically found behavior of the energy-maximizing hole shapes.)

Another difficulty of the proposed scheme is that a relatively small number N of mapping coefficients smoothes the shape corners and hence may yield too conservative an optimum. However, it is physically clear that the stresses should be bounded at the "true" corners of the optimal shape, which makes only an infinitely small contribution to the minimized criterion value. Earlier work [Vigdergauz and Cherkayev 1986] and our current results show that this is the case.

4. Fast stress-evaluation scheme

We now refer back to the direct boundary problem (7), (10), the solution of which gives the value of $\mathcal{H}$ for an arbitrarily shaped hole. Our concern here is to maximally extend the analytical transformations before resorting to numerical calculations. To this end, we rework the boundary condition (10) with (7a) as

$$-\frac{2B}{\xi^2}\overline{\omega'(\xi)} + \Gamma\omega'(\xi) - \frac{2}{\xi^2}\overline{\omega'(\xi)}\operatorname{Re}\Phi(\xi) + \overline{\omega(\xi)}\Phi'(\xi) = -\omega'(\xi)\Psi(\xi); \quad \xi \in l. \tag{15}$$

The left-hand side of (15) is the boundary value of an Σ_2-holomorphic function tending to zero at infinity. In turn, this means that its series expansion involves no nonnegative powers in ζ. Substituting (4) and (8) in (15) and zeroing the resulting coefficients of ζ^n, $n \geq 0$, gives an infinite linear algebraic system in a_k, $k \geq 2$:

$$a_{m+2} - \sum_{k=1}^{m}(m-k+1)\bar{d}_{m-k+1}a_k - (m+1)\sum_{k=1}^{\infty}\bar{d}_{m+k+1}\bar{a}_k = A_m; \quad m = 0, 1, \ldots, \tag{16a}$$

$$A_0 = 2B - \Gamma, \quad A_1 = 0, \quad A_m = -2B(m+1)\bar{d}_{m+1}, \quad m \geq 2. \tag{16b}$$

The first sum is omitted in (16a) when $m = 0, 1$.

Remarkably, the second potential $\Psi(\zeta)$ remains outside the system, thus allowing us to separate out the coefficients a_k, which are only needed to compute the boundary stresses $\sigma_{\theta\theta}(\xi)$ via (11). In numerical practice, the mapping expansion (4) and the system size are truncated to finite numbers.

This scheme was proposed by Kalandiya [1975] three decades ago and since then, to the author's best knowledge, it has not yet been studied analytically. Our aim now is to prove that the double truncation is unnecessary, because by taking only the N first items in the mapping expansion (4) the actual system size already shrinks to the same finite value of N.

To make the algebra simpler, assume that the unit far field is only shear ($B = 0$, $\Gamma = 1$) and that the hole shape L is symmetric about the x-axis. Then the coefficients $\{a_k, b_k, d_k\}$ are real and hence

$$\delta W = \pi a_2 f_1^{-1}. \tag{17}$$

Independently of these assumptions, we further note that with $d_n = 0$, $n > N$ the second sum disappears in all equations (16a) from $m = N - 1$ on. Beginning with $m = N$ they form an infinite linear system of finite differences

$$a_{m+2} - \sum_{k=1}^{N} k d_k a_{m-k+1} = 0; \qquad m = N, N+1, \ldots, \tag{18}$$

with constant coefficients $1, 0, -d_1, -2d_2, \ldots, -Nd_N$ which define the finite Laurent expansion of $\omega'(\zeta)$. For this reason, the characteristic roots of (18) coincide with the set of roots $\{\lambda_k\}$ of (13); see [Levy and Lessman 1958]. This is the key point for further analysis.

Suppose first that all the roots of $P_{N+1}(z)$ are different. Then the general solution to the homogeneous system (18) takes the form

$$a_m = D_1 \lambda_1^m + D_2 \lambda_2^m + \ldots + D_{N+1} \lambda_{N+1}^m, \quad m = 1, 2, \ldots, \tag{19}$$

where the arbitrary constants $D_1, \ldots, D_{N+1}$ are to be found by plugging (19) into (16). Substitution of (19) into (8) results in infinite geometrical progressions in ζ, which converge by virtue of inequalities (13). Summing them we get, in view of (13),

$$\Phi(\zeta) = \sum_{m=1}^{N+1} \frac{D_m \lambda_m^2}{\zeta(\zeta - \lambda_m)}. \tag{20}$$

In order to avoid the polynomial roots calculation, we exclude them by performing summation in (20) over m:

$$\Phi(\zeta) = \frac{R_N(\zeta)}{\zeta \omega'(\zeta)}. \tag{21}$$

Here $R_N(\zeta)$ is a new polynomial of degree N in ζ,

$$R_N(\zeta) = r_N \zeta^N + r_{N-1} z^{N-1} + \ldots + r_0. \tag{22}$$

It is easy to see that the coefficients r_m, $m = 0, \ldots, N$ are specified by

$$r_m = (-1)^m \sum_{j=1}^{N+1} D_j \lambda_j p_m^j, \tag{23}$$

where p_m^j stands for all the possible products of m different roots λ_k excluding λ_j and $p_0^j \equiv 1$.

Writing out p_m^j through the coefficients (13) of the polynomial $P_{N+1}(\zeta)$ and making use of (19) we obtain finally

$$r_0 = a_1 = 0, \quad r_1 = a_2,$$
$$r_m = a_{m+1} + \sum_{j=1}^{m-1} (-1)^{m-j} (m - j + 1) d_{m-j+1} a_j; \quad m \geq 2. \tag{24}$$

Therefore, the net expression (21) for the potential $\Phi(\zeta)$ does not explicitly contain the roots λ_m. Lastly, the second potential $\Psi(\zeta)$ is algebraically found from (15).

For the multiple roots λ_m, analogous manipulations lead to the same formulas, (21) and (24). Note that the s-repeated root λ_m enters expression (19) for a_m as

$$D_k^1 \lambda_k + k D_k^2 \lambda_k^2 + \ldots + D_k^{s-1} \lambda_k^s; \tag{25}$$

see [Levy and Lessman 1958]. Substitution of (25) into (8) results in convergent sums of the type

$$S_m = \sum_{i=1}^{m} i^m x^i, \quad |x| < 1,$$

which are found recurrently. Indeed, it is evident that $S_0 = x(1 - x)^{-1}$, while for $m \geq 1$,

$$S_m = \frac{x}{1-x} \left(1 + \sum_{j=1}^{m-1} \binom{m}{j} S_j \right),$$

as follows from the chain of identities

$$S_m = \sum_{i=1}^{m} i^m x^i = x + \sum_{i=1}^{m} (i + 1)^m x^{i+1} = x + x \sum_{j=1}^{m} \binom{m}{j} S_j.$$

Further simplification is made by assuming a possible stress field symmetry which permits the unknowns to be partially eliminated. Say, for a square-symmetric hole ($d_k = 0$ when $k \neq 4j - 3$) and pure shear, only a_{4k-2}, $k = 1, 2, \ldots$, differ from zero.

5. Comparison with the truncated expansion solution

For clarity, we repeat the basic steps of the proposed evaluation scheme.

First, the map function (4) is assumed to contain only a finite number N of terms.

Next, the linear algebraic system of the first N equations from (16) is solved to find the N lowest coefficients of $\Phi(\zeta)$. We remark again that the higher coefficients do not enter the system, which hence is *exact* with no truncation needed. The first term a_2 so found gives the energy increment δW via (9). The more general energy-related Pólya–Szegő matrices are derived in [Movchan and Serkov 1997] exactly in this way.

The final and novel component is finding the function $\Phi(\zeta)$ or, equivalently, the tangential stress distribution $\sigma_{\theta\theta}(\xi)$ along the hole shape by the exact summation of the infinite tail in expansion (8). This is not done in [Cherkaev et al. 1998], where the truncated series

$$\Phi(\zeta) = \sum_{k=2}^{N} a_k \zeta^{-k} \tag{26}$$

is used instead. This brings up the question of assessing the resultant truncation error in the local stresses in dependence on N. To this end we borrow the δW-optimal mapping terms for $n = 3, 7, 11, 15$ found in [Cherkaev et al. 1998] for square symmetry and use (22)–(24) and (26) to compute the exact and truncated $\mathcal{H}$. At a given $\{d_k\}$ (Table 1), the discrepancy between the values (Table 2) is entirely due to the system truncation. The relative error $\Delta\mathcal{H}$ is seen to decrease rather slowly with increasing N. In addition, the truncation leads to spurious oscillations as exemplified in Figure 3 for $N = 23$. This validates the proposed $\mathcal{H}$-evaluation scheme against (26).

N	d_3	d_7	d_{11}	d_{15}	d_{19}	$\delta W_{\min}$
3	-0.13807					3.72792
7	-0.14251	0.01575				3.71725
11	-0.14372	0.01652	-0.00513			3.71532
15	-0.14420	0.01683	-0.00539	0.00239		3.71473
19	-0.14445	0.01699	-0.00521	0.00251	-0.00134	3.71449

Table 1. A single square-symmetric hole under remote shear: the optimal mapping coefficients and the global criterion $\delta W_{\min}$ for different values of N, taken from [Cherkaev et al. 1998].

	Exact Summation		Truncation		
N	value	angle	value	angle	$\Delta\mathcal{K}(\%)$
3	3.29603	31.5	3.51472	0.0	6.63
7	3.26002	37.3	3.32372	22.9	1.95
11	3.25179	39.6	3.29671	30.3	1.38
15	3.24801	40.9	3.28945	34.0	1.27
19	3.24623	41.6	3.28690	36.3	1.25

Table 2. Values of $\mathcal{K}$ and angular locations along the δW-optimal hole shape, under exact summation (Eqs. (22)–(24)) and under truncation (Eq. (26)). See Table 1 for the mapping coefficients.
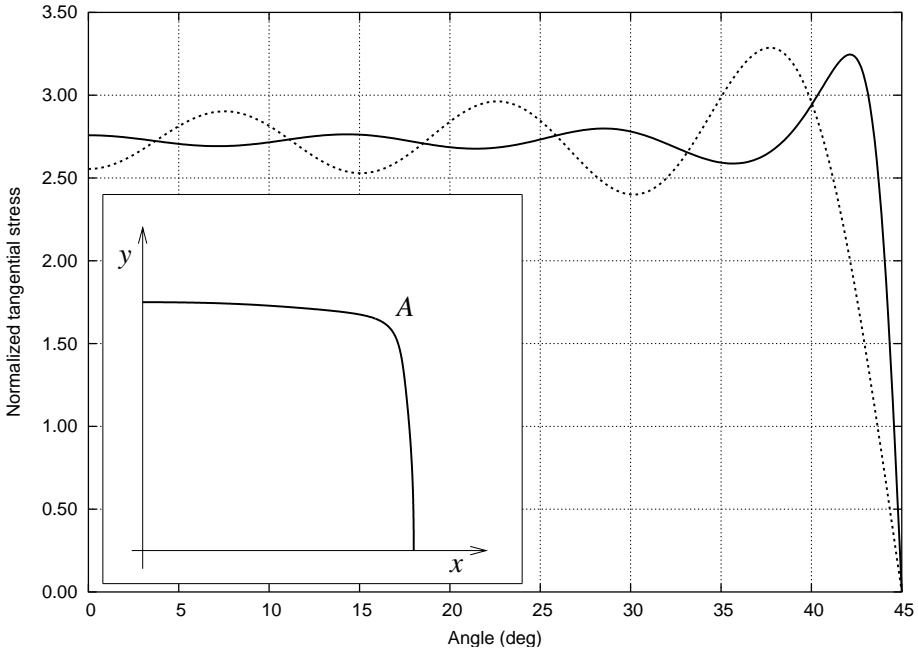
Figure 3. The tangential stress distribution along the δW-optimal hole shape (A) at $N = 23$: Exact summation (the solid line) versus truncation (the dotted line)

6. Analytical consequences of the main formula

In principle, the first N equations (16) with ease-to-compose coefficients can be solved analytically even for rather large N but we consider here only the simplest cases of one- and two-term mappings under pure shear ($B = 0$, $\Gamma = 1$).

One-term mapping. Let a hole with $(p+1)$-fold rotational symmetry ($p \geq 1$) be mapped by the function $\omega(\zeta) = \zeta + d_p \zeta^{-p}$. At $N = p$, the resolving $N \times N$ system (16) reduces to one or two equations, with $R_N(\zeta)$ becoming a monomial or binomial, respectively. Indeed, with (18), (24) and (11), a little algebra applied to (16) yields the following formulas:

Two- and three-fold symmetry ($p = 1, 2$):

$$a_2 = 1, \quad \delta W_2 = \frac{4}{(1 - pd_p^2)} : \quad \min_{d_p} \delta W_p = 4, \quad d_p^{(\min)} = 0, \tag{27a}$$

$$\sigma_{\theta\theta}(\xi) = 4 \frac{\cos 2\theta - pd_p \cos(p-1)\theta}{1 - 2pd_p \cos(p+1)\theta + p^2 d_p^2}. \tag{27b}$$

Square symmetry ($p = 3$):

$$a_2 = \frac{1}{1 - d_3}, \quad \delta W_3 = \frac{4}{(1 - d_3)(1 - 3d_3^2)} : \quad \min_{d_3} \delta W_3 = \frac{9}{\sqrt{2} + 1}, \quad d_3^{(\min)} = \frac{1 - \sqrt{2}}{3},$$

$$\sigma_{\theta\theta}(\xi) = \frac{4(1 - 3d_3) \cos 2\theta}{(1 - d_3)(1 - 6d_3 \cos 4\theta + 9d_3^2)}.$$

Higher symmetry ($p \geq 4$):

$$a_2 = \frac{1}{1 - (p-2)d_p^2}, \quad a_{p-1} = \frac{d_p(p-2)}{1 - (p-2)d_p^2}, \quad \delta W_p = \frac{4}{(1 - (p-2)d_p^2)(1 - pd_p^2)},$$

$$\min_{d_p} \delta W_p = 4, \quad d_p^{(\min)} = 0,$$

$$\sigma_{\theta\theta}(\xi) = 4 \frac{a_2(\cos 2\theta - pd_p \cos(p-1)\theta) + a_{p-1}(\cos(p+2)\theta - pd_p \cos 2\theta)}{1 - 2pd_p \cos(p+1)\theta + p^2 d_p^2}.$$

Excepting the square-symmetric case $p = 3$, the energy minimizing hole under remote shear appears to be a circle with $\mathcal{K}(d_p^{(\min)}) = 4$. Most likely, this is true not only for the one-term approximation but in general too. For triangular ($p = 2$) and hexagonal ($p = 5$) symmetry this fact was conjectured by Torquato et al. [1998] and is verified in Section 8, where the physical reasons behind the specifics of the square-symmetric optimal hole are also discussed.

Next, the value $p = 3$ yields

$$\mathcal{K}(d_3^{(\min)}) = \max_\theta |\sigma_{\theta\theta}(\xi, d_3^{(\min)})|$$

$$= \sigma_{\theta\theta}(\theta_0, d_3^{(\min)}) = \frac{4\alpha(1 - 3d_3^{(\min)})}{(1 - d_3^{(\min)})((1 + 3d_3^{(\min)})^2 - 12\alpha^2 d_3^{(\min)})}; \tag{28}$$

$$\alpha = \sqrt{\tfrac{1}{2}(\sqrt{2} - 1)}, \quad \theta_0 = \tfrac{1}{2} \arccos \alpha.$$

Analytical optimization of $\mathcal{H}(d_p)$ over the mapping term d_p, though routine, is too lengthy. Numerical GA optimization is performed instead (see Section 8).

Further we note that for any p, the energy maximum invariably occurs on the map univalency bound (13)

$$\max_{d_p} d_p^{(\max)} = \frac{1}{p}; \quad \omega_p(\zeta) = \zeta + \frac{1}{p\zeta^{-p}}; \quad \delta W_p = \begin{cases} 8, & p = 2, \\ 9, & p = 3, \\ \dfrac{4p^3}{(p^2-p+2)(p-1)}, & p \geq 4. \end{cases} \quad (29)$$

Therefore, the "worst" (shear energy maximizing) single hole as mapped by one-term function (29) is the $(p+1)$-cusped hypocycloid shown in Section 8. It has $(p+1)$ entrant angles where the tangential stress $\sigma_{\theta\theta}(\xi)$ goes to infinity.

The multi-term worst shapes reveal more complex behavior, which is analyzed numerically and displayed graphically in Section 8.

Two-term mapping. Here, we focus only on the square-symmetric case with $p = 3$, when

$$\omega(\zeta) = \zeta + d_3\zeta^{-3} + d_7\zeta^{-7}.$$

Solving the 2×2 system of the first nontrivial equations (16) yields

$$a_2 = \frac{1}{1 - d_3 - 3d_3d_7 - 5d_7^2}, \quad a_6 = \frac{3d_3 + 5d_7}{1 - d_3 - 3d_3d_7 - 5d_7^2},$$

$$\delta W_3 = \frac{4}{(1 - d_3 - 3d_3d_7 - 5d_7^2)(1 - 3d_3^2 - 7d_7^2)}.$$

The routinely obtained δW_3-minimum conditions take the form

$$3d_3(3d_3 + 10d_7) = 7d_7(1 + 3d_7), \quad (30a)$$

$$6d_3(5d_7^2 + 3d_3d_7 + d_3 - 1) = (1 + 3d_3)(1 - 3d_3^2 - 7d_7^2). \quad (30b)$$

Together with more explicit analytics like (28) and (29), they serve as a benchmark to configure the GA scheme.

7. Testing and calibration of the optimization scheme

The numerical accuracy of the proposed algorithm has been verified by reproducing the δW-related results. They fully coincide with those obtained differently in [Cherkaev et al. 1998] (Table 1). The corresponding evolution of the optimal square-like shape with N is not shown here; see the same reference.

We mention that the design variables d_m, $m = 1, \ldots, N$, are encoded using a discrete n-bit procedure when each coefficient d_m is approximated in view of (6) only by 2^n separate values in the continuous search space $[-1/\sqrt{m}; 1/\sqrt{m}]$.

These values are decoded from a randomly generated integer $P \in [-2^{n-1}; 2^{n-1}]$ as $d_k P = P/(2^{n-1}\sqrt{m})$. The genes for different coefficients are concatenated into a binary string of length Nn, or *chromosome*, which encodes the shape to be evaluated. The constant-size chromosome group so formed is then randomly subject to bitwise crossover and mutations that lead the initial population to the global optimum.

In this context it is of interest to evaluate the GA accuracy in dependence on the number of bits n by comparing with the analytical relations (27a), (29), (30). The results are collected in Table 3. This validates the approach and allows us to calibrate the heuristic GA parameters involved, such as population size and mutation rate. Typical GA settings chosen for the further $\mathcal{H}$-optimizations are given in Table 4.

Relation	Two bytes (16 bits)	Four bytes (32 bits)
(27a)	$< 2.1 \times 10^{-11}$	$< 4.4 \times 10^{-18}$
(29), $p = 3$	$< 7.4 \times 10^{-8}$	$< 2.0 \times 10^{-14}$
(29), $p = 5$	$< 3.8 \times 10^{-9}$	$< 4.8 \times 10^{-15}$
(30a)	$< 4.1 \times 10^{-6}$	$< 5.7 \times 10^{-9}$
(30b)	$< 5.8 \times 10^{-5}$	$< 1.6 \times 10^{-8}$

Table 3. Absolute error produced by GA optimization in cases of known analytical identities: two-byte versus four-bytes encoding.

Gene	Integer $\in [-2^{31}, 2^{31}]$	Individual	Interface shape
Population size	800	Number of genes	6
Initial population	Random	Selection format	Tournament
Elitism	Four best individuals	Termination	1200 iterations
Crossover	1-point	with rate	0.90
Creep mutation	Randomly change a bit	with rate	0.35
Jump mutation	Add random integer ± 64	with rate	0.35

Table 4. Typical GA setup used in our $\mathcal{H}$-optimizations.

8. Numerical results

We now present numerical results of the proposed method for a more complicated $\mathcal{H}$-optimization. Figures 4 and 5 exhibit the N-related evolution of the optimal hole shape and associated stress distributions, respectively. Table 5 shows the optimal parameters in dependence on increasing N. Comparison with the corresponding columns in Tables 1 and 2 indicates that δW- and $\mathcal{H}$-optimizations give rather similar values of the energy increment, while the maximum stresses differ significantly
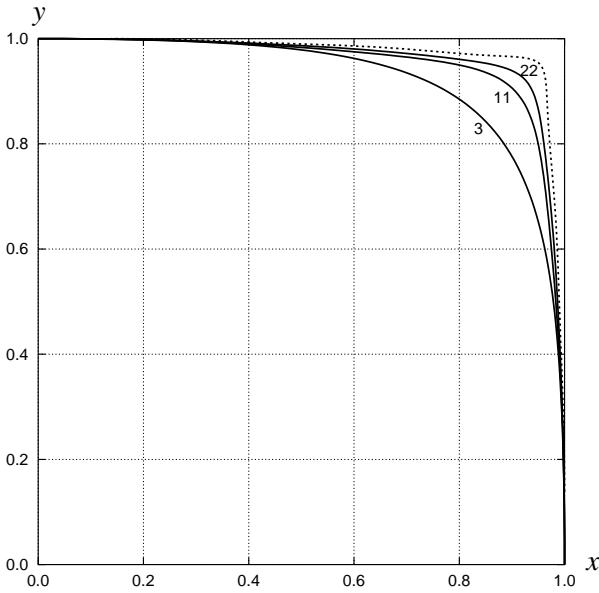
Figure 4. A quarter of the $\mathcal{K}$-optimal hole: evolution with increasing N ($N = 3, 11, 22$). The δW-optimal shape for $N = 23$ is added for comparison (dotted line).
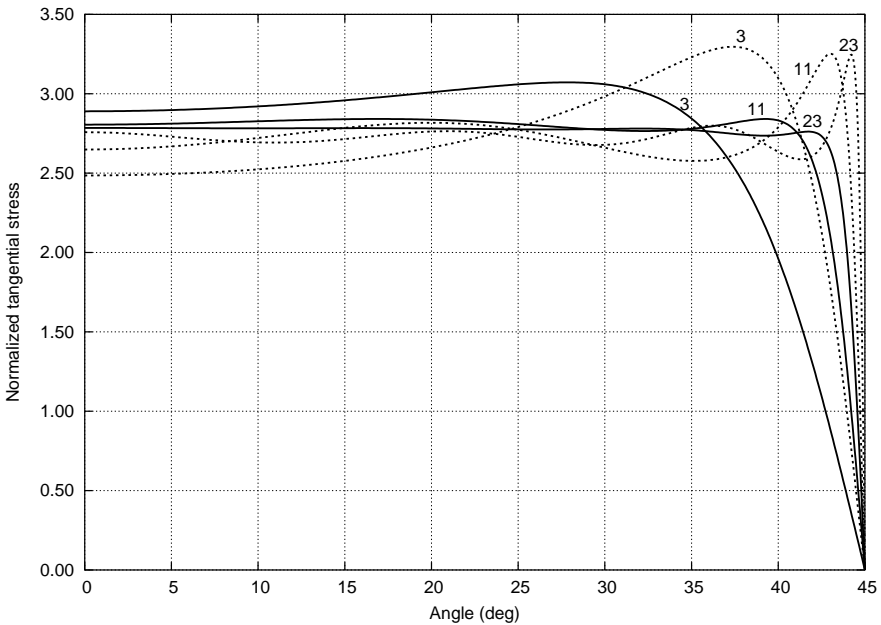
Figure 5. Tangential stress distribution along $\mathcal{K}$- and δW-optimal holes (solid and dotted lines, respectively) for $N = 3, 11, 23$.

N	d_3	d_7	d_{11}	d_{15}	d_{19}	d_{23}	$\mathcal{K}_{min}$	δW
3	−0.09000						3.07165	3.76112
7	−0.11162	0.00751					2.90563	3.73211
11	−0.12182	0.01044	−0.00200				2.84110	3.72304
15	−0.12732	0.01210	−0.00293	0.00076			2.80824	3.71959
19	−0.13049	0.01293	−0.00340	0.00112	−0.00032		2.78843	3.71773
23	−0.13059	0.01292	−0.00338	0.00116	−0.00041	0.00010	2.77936	3.71770

Table 5. A single square-symmetric hole under remote shear: the conformal mapping coefficients and the local criterion $\mathcal{K}_{min}$ resulted from the GA optimization process for different values of N. The corresponding global criterion δW is also shown to compare with its optimal values in Table 1.

(the relative discrepancies being respectively 0.086% and 16.8% at $N = 23$). However, close inspection of the stress distribution along optimal shapes of both types in Figure 5 shows that the discrepancy is concentrated near the forming angular point at $\theta = \pi/4$. In either case the stresses tend to obey the same identity (2). The different nature of the criteria defines the different convergence behavior of the optimal solutions: δW forms the angular point more rapidly, whereas the local $\mathcal{K}$-criterion allows no high-frequency peaks. This is clearly seen in Figure 4. With some caution, we conclude that both optimal solutions are the same, though an analytical proof would be very desirable. This is beyond our scope at the moment.

For triangular and hexagonal symmetry, similar computations lead to an interesting conclusion:

Under remote shear, the δW- and $\mathcal{K}$-optimal hole shape is a circle with $\delta W_{min} = 4$ and $\mathcal{K}_{min} = 4$.

In the absence of angular points, this has been verified with to extremely high accuracy.

The stress distribution along a circle is readily expressed as

$$\Phi_0(\zeta) = \Phi(\zeta) = \zeta^{-2}: \quad \sigma_{\theta\theta}(\xi) = 4\cos 2\theta, \quad \xi = \exp i\theta \in l,$$

(see [Muskhelishvili 1963]) with the sign-changing points lying on the bisectors of the quadrants

$$\theta = \frac{\pi(2j-1)}{4}, \quad j = 1, \ldots, 4,$$

as in the square-symmetric case.

Geometrically, it is clear that other points location compatible with the remote shear antisymmetry may exist only for $(8p - 4)$-symmetric shapes, $p = 2, 3, \ldots,$

with an odd number of the points in a quadrant. However, GA optimization spe-
cially performed for $p = 2$ (dodecagonal symmetry) and $p = 3$ (icosagonal sym-
metry) bring us back to a circle.

Therefore, the prescribed square symmetry gives the only nontrivial example of
the optimal shape where the angular points substantially reduce the values of both
considered criteria as compared to a circle. The bottom row in Table 5 shows that
$\Delta(\mathcal{K}) = 1 - 2.77936/4.0 = 30.5\%$ and $\Delta(\delta W) = 1 - 3.71449/4.0 = 7.1\%$.

Finally, the energy-maximization results are presented in Table 6 and Figure 6.
At the first glance on the shapes it seems that the energy increment δW increases
rapidly witn N due to the distinctive corner points which make finite contribution
to the coefficient a_2 in (17). However, the most effective strategy really performed
here by the GA is to minimize the hole area entering in (17) as a denominator.
Indeed, as noted in the end of Section 3, finite-term mapping may not give a zero-
area curve for $p \geq 2$. Therefore, as the only possible compromise between the
finite number N of mapping terms and prohibited self-intersections the algorithm
identifies the limiting case of the optimal shapes containing entrant angles that are
of pure geometrical nature with no optimization resort. It is supported by the fact
that the coefficient a_2 in Table 6 does not diminish with increasing N.

When N tends to infinity, the optimal mapping should then approximate a cross-
like cut with $(p+1)$ equal arms as exemplified in Figure 6 for $p = 3$. This is the
analitically known case [Ahlfors 1978] with

$$\omega(\zeta) = \zeta\left(1 + \zeta^{-n}\right)^{2/n} = \zeta + \sum_{k=1}^{\infty} \frac{q(q-1)\dots(q-k+1)\zeta^{-(p+1)k+1}}{k!}; \qquad q = \frac{1}{(p+1)}. \quad (31)$$

N	d_3	d_7	d_{11}	d_{15}	d_{19}	a_2	f_1/π	δW
1	0.33333					1.50000	0.66667	9.00000
2	0.42857	−0.14286				1.53139	0.30615	20.00833
3	0.43301	−0.14193	0.0631			1.56067	0.25450	24.52922
4	0.45091	−0.11683	0.0738	−0.0470		1.60610	0.20225	31.76542
5	0.46864	−0.11875	0.05391	−0.04401	0.02510	1.61902	0.16943	38.22265
∞	0.5	−0.125	0.0625	−0.03906	0.02734		0.0	∞

Table 6. A single square-symmetric hole under remote shear: the
maximal mapping coefficients, the global criterion δW_{max} and its
components a_2, f_1 found by the GA approach. The first row
emerges analytically from (29) at $p = 3$. The last row contains
the limiting data of the cross-like slit (31).
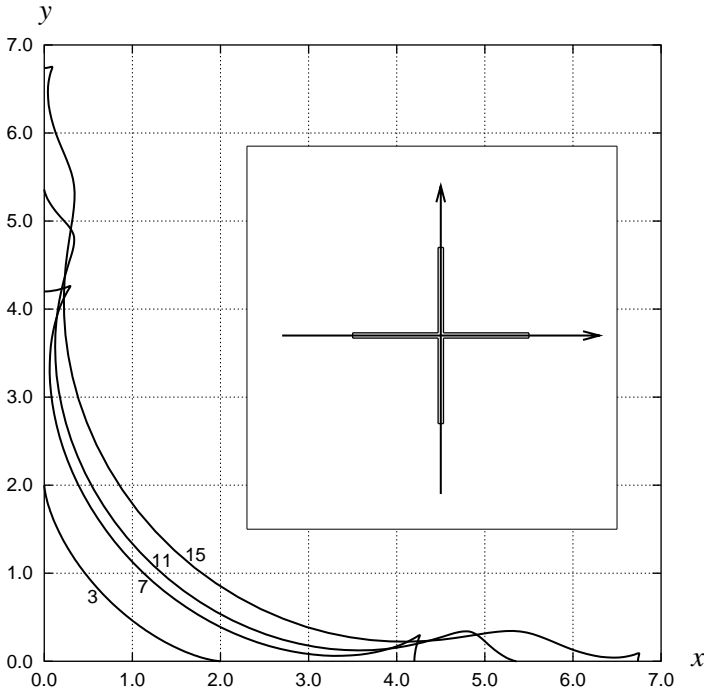
Figure 6. The energy-*maximizing* hole under square symmetry and remote shear: evolution to the limiting cross-like slit with increasing N. For $N = 3$ we have a four-cusped hypocycloid (29).

This reasoning is justified by the additional calculations in which the hole area was optimized instead of the energy. The resultant mapping terms practically coincide with those in Table 6. In turn, the latter converge, though slowly, to the cross-like limit (31).

9. Concluding remarks

The main goal of this work was to try a new GA scheme of shape optimization in plane elasticity when the conformal mapping is used instead of the direct nodal representation. In combination with the advanced fitness evaluation the proposed approach has enabled us to obtain new and significant results. This encoding may be also effective in other optimization problems governed by the conformal-invariant Laplace equation such as those in electrostatics. On the other hand, the current results are confined to the particular model with a single inhomogeneity. It is still unknown how this GA configuration will work in in the more realistic situation of multiple inclusions whose interaction hampers both the analytical and numerical manipulations. We hope to pursue this problem in further publications.

References

[Ahlfors 1978] L. V. Ahlfors, *Complex analysis: an introduction to the theory of analytic functions of one complex variable*, Third ed., McGraw-Hill, New York, 1978. MR 80c:30001

[Cherepanov 1974] G. P. Cherepanov, "Inverse problems of the plane theory of elasticity", *Prikl. Mat. Mekh.* **38**:6 (1974), 963–979. In Russian; translated in *J. Appl. Math. Mech.* **38**:6 (1974), 915–931. MR 52 #7254 Zbl 0315.73106

[Cherkaev et al. 1998] A. V. Cherkaev, Y. Grabovsky, A. B. Movchan, and S. K. Serkov, "The cavity of the optimal shape under the shear stresses", *Int. J. Solids Struct.* **35**:33 (1998), 4391–4410. MR 99e:73040

[Gen and Cheng 1997] M. Gen and R. Cheng, *Genetic algorithms and engineering design*, Wiley, New York, 1997.

[Holland 1975] J. H. Holland, *Adaptation in natural and artificial systems*, University of Michigan Press, Ann Arbor, MI, 1975. MR 55 #14256

[Kalandiya 1975] A. I. Kalandiya, *Mathematical methods of two-dimensional elasticity*, Mir Publishers, Moscow, 1975. MR 53 #4676

[Levy and Lessman 1958] H. Levy and F. Lessman, *Finite difference equations*, Pitman, London, 1958. Reprinted McMillan, New York, 1961, and Dover, New York, 1992. MR 21 #5827

[Movchan and Serkov 1997] A. B. Movchan and S. K. Serkov, "The Pólya–Szegő matrices in asymptotic models of dilute composites", *Eur. J. Appl. Math.* **8**:6 (1997), 595–621. MR 98m:73004

[Muskhelishvili 1963] N. I. Muskhelishvili, *Some basic problems of the mathematical theory of elasticity: Fundamental equations, plane theory of elasticity, torsion and bending*, 2nd English ed., Noordhoff, Groningen, 1963. Reprinted 1975; translated from the 4th Russian edition. MR 31 #920

[Pilkey 1997] W. D. Pilkey, *Peterson's stress concentration factors*, 2nd ed., Wiley, New York, 1997.

[Savin 1961] G. N. Savin, *Stress concentration around holes*, Pergamon, New York, 1961. MR 23 #B27

[Torquato 2002] S. Torquato, *Random heterogeneous materials*, Interdisciplinary Applied Mathematics **16**, Springer, New York, 2002. MR 2002k:82082

[Torquato et al. 1998] S. Torquato, L. V. Gibiansky, M. Silva, and L. Gibson, "Effective mechanical and transport properties of cellular solids", *Int. J. Mech. Sci.* **40**:1 (1998), 71–82.

[Vigdergauz 1976] S. B. Vigdergauz, "Integral equation of the inverse problem of the plane theory of elasticity", *Prikl. Mat. Mekh.* **40**:3 (1976), 566–569. In Russian; translated in *J. Appl. Math. Mech.* **40**:3 (1976), 518-522.

[Vigdergauz 2001a] S. B. Vigdergauz, "The effective properties of a perforated elastic plate: Numerical optimization by genetic algorithm", *Int. J. Solids Struct.* **38**:48-49 (2001), 8593–8616.

[Vigdergauz 2001b] S. B. Vigdergauz, "Genetic algorithm perspective to identify energy optimizing inclusions in an elastic plate", *Int. J. Solids Struct.* **38**:38-39 (2001), 6851–6867.

[Vigdergauz 2002] S. B. Vigdergauz, "Genetic algorithm optimization of the effective Young moduli in a perforated plate", *Struct. Multidiscip. O.* **24**:2 (2002), 106–117.

[Vigdergauz and Cherkayev 1986] S. B. Vigdergauz and A. V. Cherkayev, "A hole in a plate, optimal for its biaxial extension-compression", *Prikl. Mat. Mekh.* **50**:3 (1986), 524–528. In Russian; translated in *J. Appl. Math. Mech.* **50**:3 (1986), 401–404. MR 88g:73073 Zbl 0621.73022

SHMUEL VIGDERGAUZ: smuel@iec.co.il
R&D Division, The Israel Electric Corporation, Ltd., P.O.Box 10, Haifa 31000, Israel

CONTENTS

Volume 1, no. 1 and no. 2

SUBMISSION GUIDELINES

ORIGINALITY

Authors may submit manuscripts in PDF format on-line. Submission of a manuscript acknowledges that the manuscript is *original and has neither previously, nor simultaneously, in whole or in part, been submitted elsewhere.* Information regarding the preparation of manuscripts is provided below. Correspondence by email is requested for convenience and speed. For further information, write to:

Marie-Louise Steele
Division of Mechanics and Computation
Durand Building, Room 262
Stanford University
Stanford CA 94305

LANGUAGE

Manuscripts must be in English. A brief abstract of about 150 words or less must be included. The abstract should be self-contained and not make any reference to the bibliography. Also required are keywords and subject classification for the article, and, for each author, postal address, affiliation (if appropriate), and email address if available. A home-page URL is optional.

FORMAT

Authors are encouraged to use LaTeX and the standard article class, but submissions in other varieties of TeX, and, exceptionally in other formats, are acceptable. Electronic submissions are strongly encouraged in PDF format only; after the refereeing process we will ask you to submit all source material.

REFERENCES

Bibliographical references should be listed alphabetically at the end of the paper and include the title of the article. All references in the bibliography should be cited in the text. The use of BibTeX is preferred but not required. Tags will be converted to the house format (see a current issue for examples), however, in the manuscript, the citation should be by first author's last name and year of publication, e.g. "as shown by Kramer, et al. (1994)". Links will be provided to all literature with known web locations and authors are encoraged to provide their own links on top of the ones provided by the editorial process.

FIGURES

Figures prepared electronically should be submitted in Encapsulated PostScript (EPS) or in a form that can be converted to EPS, such as GnuPlot, Maple, or Mathematica. Many drawing tools such as Adobe Illustrator and Aldus FreeHand can produce EPS output. Figures containing bitmaps should be generated at the highest possible resolution. If there is doubt whether a particular figure is in an acceptable format, the authors should check with production by sending an email to:

production@mathscipub.org

Each figure should be captioned and numbered so that it can float. Small figures occupying no more than three lines of vertical space can be kept in the text ("the curve looks like this:"). It is acceptable to submit a manuscript with all figures at the end, if their placement is specified in the text by means of comments such as "Place Figure 1 here". The same considerations apply to tables.

WHITE SPACE

Forced line breaks or page breaks should not be inserted in the document. There is no point in your trying to optimize line and page breaks in the original manuscript. The manuscript will be reformatted to use the journal's preferred fonts and layout.

PROOFS

Page proofs will be made available to authors (or to the designated corresponding author) at a web site in PDF format. Failure to acknowledge the receipt of proofs or to return corrections within the requested deadline may cause publication to be postponed.