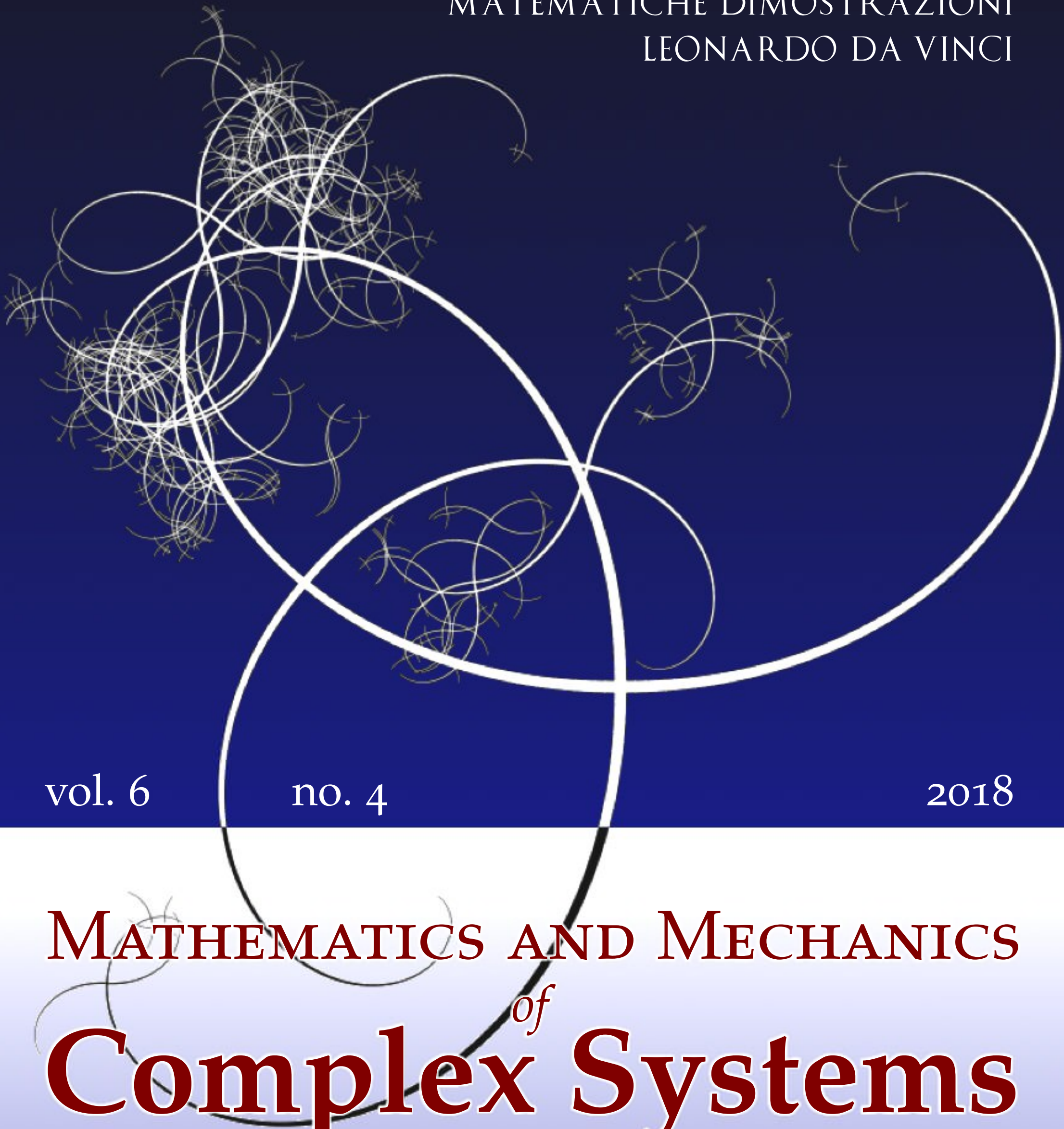
NISSUNA UMANA INVESTIGAZIONE SI PUO DIMANDARE
VERA SCIENZIA S'ESSA NON PASSA PER LE
MATEMATICHE DIMOSTRAZIONI
LEONARDO DA VINCI



vol. 6 no. 4 2018

# Mathematics and Mechanics *of* Complex Systems

# MATHEMATICS AND MECHANICS OF COMPLEX SYSTEMS

msp.org/memocs

MEMOCS (ISSN 2325-3444 electronic, 2326-7186 printed) is a journal of the International Research Center for the Mathematics and Mechanics of Complex Systems at the Università dell’Aquila, Italy.

Cover image: “Tangle” by © John Horigan; produced using the *Context Free* program (contextfreeart.org).

PUBLISHED BY
mathematical sciences publishers
nonprofit scientific publishing
http://msp.org/

MATHEMATICS AND MECHANICS OF COMPLEX SYSTEMS
Vol. 6, No. 4, 2018
dx.doi.org/10.2140/memocs.2018.6.267

# A MODEL FOR INTERFACES AND ITS MESOSCOPIC LIMIT

MICHELE ALEANDRI AND VENANZIO DI GIULIO

We study a system of $N$ layers with a Kac horizontal interaction of parameter $\gamma > 0$ and a Kac vertical interaction of parameter $\gamma^{1/2}$. We shall prove that the limit free energy functional is the rate function of the large deviations of the Gibbs measure (of a canonical constrained magnetization). The limit free energy functional is achieved as a $\Gamma$-limit for $\gamma \to 0$ for magnetizations with fixed average. Among all such magnetizations there exists a quasiconstant magnetization that minimizes the energy.

## 1. Introduction

Equilibrium and dynamics of interfaces is a very well studied issue both in physics and mathematics. In several instances to simplify the problem it is supposed that the interface is a graph, an assumption which is not at all unrealistic if the interface is studied locally. In the SOS models of statistical mechanics the interface is a graph over a lattice $\mathbb{Z}^d$; namely for each site $i \in \mathbb{Z}^d$ we draw a vertical line and the position of the interface on the line (its height) is represented by a real-valued spin $S_i$. One then introduces a Hamiltonian which describes the interactions among the spins so that the equilibrium properties of the interface are derived from the Gibbs properties of the Hamiltonian. The difficulty in this approach arises from the fact that the Hamiltonian is massless, which corresponds to the fact that vertical translations of the interface do not cost energy. The theory of DLR states is then quite more involved than in the classical Ising model; a breakthrough was achieved in [Funaki and Spohn 1997], followed by many other papers.

In this paper we take a step back towards microscopic scalar; namely we suppose that on each horizontal line there is an Ising system so that instead of a real-valued spin $S_i$ we have a configuration $\sigma(x, i)$, $x \in \mathbb{Z}$, of $\pm 1$-valued spins. We actually consider a finite system with $i = 1, \dots, N$ and $x \in [0, L] \cap \mathbb{Z}$, $L = \gamma^{-1}\ell$, $\gamma > 0$, $\ell > 0$ ($L$ an integer). To simulate a phase transition the spins on each horizontal line interact via a Kac potential $J_\gamma(x, y)$ (the same on each line), whose strength is 1 and whose range is $\gamma^{-1}$ (see Section 2 for a precise definition). The spins

**Communicated by Raffaele Esposito**.
*MSC2010:* 82B24.
*Keywords:* Kac potential, mesoscopic limit, $\Gamma$-convergence.

between nearest neighbor horizontal lines (say $(x, i)$ and $(y, i+1)$) interact via the Kac potential $\lambda J_{\gamma^{1/2}}(x, y)$, $\lambda > 0$; that is, the vertical interaction is much more local than the horizontal one.

We study the mesoscopic limit $\gamma \to 0$. The mesoscopic state of the system is a collection $m \equiv \{m(r, i) : r \in [0, \ell], i = 1, \dots, N\}$ of measurable functions with values in $[-1, 1]$. Its statistical properties are then described by a free energy functional $F(m)$. According to the Gibbs theory such a functional is the limit as $\gamma \to 0$ of $-1/\beta$ times the log of a constrained partition function where the spin configurations are required to be "close" to the mesoscopic state $m$ (this involves a coarse grain procedure which is specified in Section 2). This is not as in the classical Lebowitz–Penrose [1966; Penrose and Lebowitz 1971] procedure because there are two scales, $\gamma^{-1}$ for the horizontal interaction and $\gamma^{-1/2}$ for the vertical one. Thus, there could be oscillations on the scale $\gamma^{-1/2}$ which do not appear in $m$ because the latter is defined by averages over $\approx \gamma^{-1}$ but which could affect the free energy of $m$. These oscillations actually do not occur if $\lambda$ is small; indeed by Theorem 4.1 the optimal profile is quasiconstant on the scale $\gamma^{-\alpha}$ with $\alpha \in (0, 1)$. However, if $\lambda$ is large enough we can provide an example where such a phenomena occurs.

The paper is organized as follows. In Section 2 we introduce the microscopic and mesoscopic models and enunciate the main results. In Section 3 we introduce the coarse graining procedure used to prove the Lebowitz–Penrose limit. In Section 4 we prove a key result, that is, Theorem 4.1, in which we provide a technique to minimize the free energy. This theorem is needed to prove the main results in Section 2. In Section 5 we prove the Lebowitz–Penrose limit for our model. In Section 6 we prove the Γ-limit result. The proofs of Theorem 2.4 and Proposition 3.1 are deferred to Appendix A. In Appendix B, finally, we illustrate the case in which Theorem 2.3 fails for the parameter $\lambda$ large enough.

Similar model have been studied in [Cassandro et al. 2016; Fontes et al. 2014; 2015]. A numerical investigation of the mesoscopic limit for lattice gas model was also recently tackled in [Colangeli et al. 2016; 2017].

This work is the first step of a research program pointed towards the characterization of the surface tension associated to free energy in the thermodynamic limit.

## 2. Model and main results

We consider an Ising spin system in a rectangle $T_{L,N} = \{(x, i) \in \mathbb{Z}^2 : x \in [0, L-1], i \in [1, N]\}$, $L = \gamma^{-1}\ell$, with $\gamma^{-1} \in \{2^n : n \in \mathbb{N}\}$ and $\ell \in \{2^k : k \in \mathbb{Z}\}$. We will eventually take the limit $\gamma \to 0$ keeping $\ell$ and $N$ fixed. We denote by $\sigma$ a spin configuration $\sigma = \{\sigma(x, i) \in \{-1, 1\} : (x, i) \in T_{L,N}\} \in \{-1, 1\}^{T_{L,N}}$, and since we will consider periodic boundary conditions we extend periodically $\sigma$ to a configuration on $\mathbb{Z}^2$ (denoted by the same symbol) by setting $\sigma(x,i)=\sigma(y,j)$ if $(x,i)\sim(y,j)$ where

$$(x,i)\sim(y,j) \quad \text{if } y=x+kL \text{ and } j=i+k'N,\ k,k'\in\mathbb{Z}. \tag{2-1}$$

The interaction among spins is given by a highly anisotropic Kac potential which will be defined in terms of a function $J(r)$, $r\in\mathbb{R}$: we suppose that $J(r)$ is a nonnegative $C^2$ function with $\int J(r)\,dr=1$ supported by $|r|\le 1$. We then define for any $x$, $y$ in $\mathbb{R}$

$$J_{\gamma^{1/2}}(x,y):=\gamma^{1/2}J(\gamma^{1/2}|x-y|), \qquad J_\gamma(x,y):=\gamma J(\gamma|x-y|). \tag{2-2}$$

The Hamiltonian of the system (with periodic boundary conditions) is then defined as

$$\begin{aligned}H_{\gamma,\lambda}(\sigma)=\sum_{i=1}^{N}\Bigg[-\frac{1}{2}\sum_{x,y\in[0,L-1]\cap\mathbb{Z}}&\big\{\mathbf{1}_{\{x\neq y\}}J_\gamma(x,y)\sigma(x,i)\sigma(y,i)\\ &\quad-\lambda J_{\gamma^{1/2}}(x,y)\sigma(x,i)(\sigma(y,i-1)+\sigma(y,i+1))\big\}\\ -\sum_{\substack{x\in[0,L-1]\cap\mathbb{Z}\\ y\notin[0,L-1]\cap\mathbb{Z}}}&\big\{J_\gamma(x,y)\sigma(x,i)\sigma(y,i)\\ &\quad-\lambda J_{\gamma^{1/2}}(x,y)(\sigma(y,i-1)+\sigma(y,i+1))\big\}\Bigg].\end{aligned} \tag{2-3}$$

Thus, the range of the vertical interaction is much shorter than the range of the horizontal one.

We denote by $\mu_{\beta,\gamma,\lambda}$ the Gibbs measure at inverse temperature $\beta$:

$$\mu_{\beta,\gamma,\lambda}(\sigma)=\frac{e^{-\beta H_{\gamma,\lambda}(\sigma)}}{Z_{\beta,\gamma,\lambda}}$$

with

$$Z_{\beta,\gamma,\lambda}=\sum_{\sigma}e^{-\beta H_{\gamma,\lambda}(\sigma)},$$

being interested in the mesoscopic limit $\gamma\to 0$. The aim is to compute the limiting free energy and the probability of mesoscopic states.

A mesoscopic state is a measurable function $m$ on $T_{\ell,N}=[0,\ell]\times\{1,\dots,N\}$ with values in $[-1,1]$. We extend $m$ periodically by setting $m(r,i)=m(r',j)$ if $(r,i)\sim(r',j)$, which means $r'=r+k\ell$ and $j=i+k'N$, $k,k'\in\mathbb{Z}$. The correspondence between spin configurations $\sigma$ and mesoscopic states $m$ is via coarse graining, namely by comparing averages. The "microscopic length" used for averaging is $\gamma^{-\alpha}$, $\alpha\in(0,1)$, and to avoid taking integer parts we suppose $\alpha$ a rational number. We tacitly suppose that $\gamma$ is small enough so that $\gamma^{-1}$ and therefore also $L$ are integer multiples of $\gamma^{-\alpha}$.

**Definition 2.1** (partition and empirical averages). Let $\alpha$ and $\gamma$ as above. We define for any $k \in \mathbb{Z}$

$$C_{k,i}^{(\alpha)} := \{(x, i) \in \mathbb{R} \times \mathbb{Z} : k\gamma^{-\alpha} \le x < (k+1)\gamma^{-\alpha}\}.$$

The collection $\mathscr{C}^{(\alpha)}$ of all $C_{k,i}^{(\alpha)}$ defines a partition of $\mathbb{R} \times \mathbb{Z}$. Moreover, $\mathscr{C}^{(\alpha)} \cap \mathbb{Z}^2$ paves exactly $T_{L,N}$: namely any $C_{k,i}^{(\alpha)} \cap \mathbb{Z}^2$ is either contained in $T_{L,N}$ or in its complement.

Given a spin configuration $\sigma$ we then define

$$\sigma^{(\alpha)}(x, i) := \gamma^{\alpha} \sum_{y \in C_{k,i}^{(\alpha)} \cap \mathbb{Z}^2} \sigma(y, i), \quad \text{where } k \text{ is such that } (x, i) \in C_{k,i}^{(\alpha)} \tag{2-4}$$

and $\sigma^{(\alpha)}$ is a function with values in $M^{(\alpha)}$ where

$$M^{(\alpha)} := \left\{-1, -1 + \frac{2}{\gamma^{-\alpha}}, \dots, 1 - \frac{2}{\gamma^{-\alpha}}, 1\right\}. \tag{2-5}$$

Analogously, given a mesoscopic state $m \in L^{\infty}(T_{\ell,N}; [-1, 1])$ we set

$$m^{(\alpha)}(x, i) := \gamma^{\alpha} \int_{k\gamma^{-\alpha}}^{(k+1)\gamma^{-\alpha}} m(\gamma r, i)\, dr \tag{2-6}$$

where $k$ is such that $(x, i) \in C_{k,i}^{(\alpha)}$ and $m^{(\alpha)}$ is a function with values in $[-1, 1]$.

We next specify in which sense a spin configuration $\sigma$ "recognizes" a mesoscopic state $m$ and use this notion to define the free energy and the probability associated to a mesoscopic state.

**Definition 2.2.** $\sigma$ "recognizes" $m$, and we write $\sigma \approx^{\alpha} m$ if

$$|\sigma^{(\alpha)}(x, i) - m^{(\alpha)}(x, i)| \le 2\gamma^{\alpha} \quad \text{for all } (x, i) \in T_{L,N} \tag{2-7}$$

(recall that, by flipping a spin, $\sigma^{(\alpha)}(x, i)$ changes by $2\gamma^{\alpha}$). We then define the finite volume free energy of the mesoscopic state $m$ as

$$F_{\beta,\gamma,\lambda}^{(\alpha)}(m) := -\frac{1}{\beta\gamma^{-1}} \log Z_{\beta,\gamma,\lambda}^{(\alpha)}(m), \tag{2-8}$$

where

$$Z_{\beta,\gamma,\lambda}^{(\alpha)}(m) := Z_{\beta,\gamma,\lambda}(\{\sigma \approx^{\alpha} m\}) = \sum_{\sigma : \sigma \approx^{\alpha} m} e^{-\beta H_{\gamma,\lambda}(\sigma)}.$$

Analogously we define the Gibbs probability of the mesoscopic state $m$ as

$$\mu_{\beta,\lambda,\gamma}[\sigma \approx^{\alpha} m] = \frac{Z_{\beta,\gamma,\lambda}^{(\alpha)}(m)}{Z_{\beta,\gamma,\lambda}}.$$

The main result in this paper is this:

**Theorem 2.3.** *For any* $\alpha \in (0, 1)$, *any* $\lambda \in (0, 1/(8\beta))$, *and any mesoscopic state* $m \in L^\infty(T_{\ell,N}, [-1, 1])$,

$$\lim_{\gamma\to 0} F^{(\alpha)}_{\beta,\gamma,\lambda}(m) = F_{\beta,\lambda}(m) \tag{2-9}$$

*where*

$$\begin{aligned} F_{\beta,\lambda}(m) = -\frac{1}{2}\sum_{i=1}^{N}\int_0^\ell\int_0^\ell J(r,r')m(r,i)m(r',i)\,dr\,dr' \\ -\frac{\lambda}{2}\sum_{i=1}^{N}\int_0^\ell m(r,i)(m(r,i+1)+m(r,i-1))\,dr \\ -\frac{1}{\beta}\sum_{i=1}^{N}\int_0^\ell I(m(r,i))\,dr \end{aligned} \tag{2-10}$$

*and*

$$I(m) = -\frac{1+m}{2}\log\frac{1+m}{2} - \frac{1-m}{2}\log\frac{1-m}{2}. \tag{2-11}$$

The following two theorems are essentially a corollary of Theorem 2.3. The first one is about free energy.

**Theorem 2.4.** *Let* $0 < \lambda < 1/(8\beta)$ *and* $\alpha \in (0, 1)$. *Then*

$$-\lim_{\gamma\to 0}\frac{1}{\beta\gamma^{-1}}\log Z_{\beta,\gamma,\lambda} = \inf_{m\in L^\infty(T_{\ell,N};[-1,1])} F_{\beta,\lambda}(m). \tag{2-12}$$

*Moreover*, *if* $\beta(1+2\lambda) > 1$, *then* (*recalling* (2-11) *for notation*)

$$\inf_m F_{\beta,\lambda}(m) = N\ell\left(-\frac{b}{2}m_{b\beta}^2 - \frac{I(m_{b\beta})}{\beta}\right), \quad b = 1+2\lambda, \tag{2-13}$$

*where* $m_{b\beta}$ *is the positive solution of the equation*

$$m_{b\beta} = \tanh\{\beta\lambda m_{b\beta}\}. \tag{2-14}$$

*If instead* $\beta(1+2\lambda) \le 1$, *then*

$$\inf_m F_{\beta,\lambda}(m) = \frac{N\ell}{\beta}\log(\tfrac{1}{2}). \tag{2-15}$$

The next theorem is about large deviations; on the general issue see for instance [Ellis 2006].

**Theorem 2.5.** *Let* $0 < \lambda < 1/(8\beta)$, $\alpha \in (0, 1)$, *and* $m \in L^\infty(T_{\ell,N}; [-1, 1])$ *be a mesoscopic state*; *then*

$$\lim_{\gamma\to 0}\gamma\log\mu_{\beta,\lambda,\gamma}[\sigma \approx^{(\alpha)} m] = -(F_{\beta,\lambda}(m) - \inf_{m'} F_{\beta,\lambda}(m')). \tag{2-16}$$

The theorems are proved in the next sections; here we make some remarks on Theorem 2.3. We note in particular that the limit free energy of a mesoscopic state is independent of the coarse graining parameter $\alpha$, a fact to some extent unexpected.

The point is that the partition function $Z^{(\alpha)}_{\beta,\gamma,\lambda}(m)$ is clearly an increasing function of $\alpha$ because the constraint $\sigma \approx^{\alpha} m$ is weakened when increasing $\alpha$. In particular the result contained in Theorem 2.3 shows that this effect is negligible in the limit $\gamma \to 0$. The basic idea in the proofs goes back to Lebowitz and Penrose, and it is based on a coarse graining with grain lengths which must be large with respect to the lattice spacing but small with respect to the range of the interaction. Following Lebowitz and Penrose we use a coarse graining with grain length $\gamma^{-\alpha'}$ with $\alpha' < \frac{1}{2}$ and $\gamma^{-\alpha'} \le \gamma^{-\alpha}$. We then obtain an estimate for the logarithm of the partition function characterized to the leading orders (as $\gamma \to 0$) by a nonrescaled functional

$$\begin{aligned}\bar F_{\beta,\gamma,\lambda}(\bar m)\\ &= -\frac{1}{2}\sum_{i=1}^{N}\int_0^{\gamma^{-1}\ell}\int_0^{\gamma^{-1}\ell} J_\gamma(r,r')\bar m(r,i)\bar m(r',i)\,dr\,dr'\\ &\quad -\frac{\lambda}{2}\sum_{i=1}^{N}\int_0^{\gamma^{-1}\ell}\int_0^{\gamma^{-1}\ell} J_{\gamma^{1/2}}(r,r')\bar m(r,i)(\bar m(r',i-1)+\bar m(r',i+1))\,dr\,dr'\\ &\quad -\frac{1}{\beta}\sum_{i=1}^{N}\int_0^{\gamma^{-1}\ell} I(\bar m(r,i))\,dr, \qquad (2\text{-}17)\end{aligned}$$

where $\bar m$ is constant on the scale $\gamma^{-\alpha'}$ used in the coarse graining.

To simplify the argument, let us assume that the mesoscopic profile $m(r,i)=0$ for all $r$ and $i$. If the constraint $\sigma \approx^{(\alpha)} m$ with $\alpha < \frac{1}{2}$, then by letting $\alpha' = \alpha$ (so that $\bar m \equiv 0$) the functional $\bar F$ becomes $F$ after rescaling.

If instead $\alpha > \frac{1}{2}$, we cannot take $\alpha' = \alpha$ and there may be vertical energy gains via suitable oscillations of the magnetization within the constraint $\sigma \approx^{(\alpha)} m$. This is not just a theoretical possibility as it may indeed occur when $\lambda$ is large. Let $\beta\lambda > 1$; then

$$\bar F_{\beta,\gamma,\lambda}(m \equiv 0) = -N\ell\frac{I(0)}{\beta}.$$

Fix $\bar m = +m_{\beta\lambda}$ (the positive solution of (2-14)) in the left half of each interval of length $\gamma^{-\alpha}$ and equal to $-m_{\beta\lambda}$ in the right half. Note that $\bar m$ satisfies the constraint $\{\sigma \approx^{\alpha} m\}$. In Appendix B we prove that the rescaled free energy of $\bar m$ in the limit of $\gamma \to 0$ is equal to

$$N\ell\biggl(-\lambda m^2_{\beta\lambda} - \frac{I(m_{\beta\lambda})}{\beta}\biggr),$$

which is smaller than $\bar F_{\beta,\gamma,\lambda}(0)$.

Instead when $\lambda$ is small as in Theorem 2.3, then the optimal $\bar{m}$ is constant on the scale $\gamma^{-\alpha}$ (when $\gamma \to 0$). The proof of Theorem 2.3 is then reduced to prove that the functional in (2-17) $\Gamma$-converges [Braides 2002] to the functional in (2-10).

## 3. Coarse graining procedure

In this section we prove some estimates for the logarithm of the partition function $\log Z^{(\alpha)}_{\beta,\lambda,\gamma}(m)$ in terms of $\bar{F}_{\beta,\lambda,\gamma}$ defined in (2-17). These estimates will be used in the Lebowitz–Penrose limit discussed in the Section 5. A different coarse graining procedure from the classical Lebowitz–Penrose result will be used. This is needed due to the presence of two different scales of interaction along the horizontal and vertical directions.

The partition function $Z_{\beta,\lambda,\gamma}(\,\cdot\,)$ is defined on the space of the configurations while $\bar{F}_{\beta,\lambda,\gamma}(\,\cdot\,)$ is defined on the space of measurable functions. Recalling Definition 2.2, we consider $\mathcal{M}^{(\alpha)}$ the space of all functions which are constant on $\{C^{(\alpha)}_{i,k}\}_{i,k\in\mathbb{Z}}$ with values in $M^{(\alpha)}$. For each empirical average $m^{(\alpha)}(\,\cdot\,)$ there exists a function $\bar{m} \in \mathcal{M}^{(\alpha)}$ such that $|m^{(\alpha)}(x,i) - \bar{m}(x,i)| \le 2\gamma^{-\alpha}$ for all $(x,i) \in T_{\ell,N}$. Furthermore, given a function $\bar{m} \in \mathcal{M}^{(\alpha)}$ we define the set

$$\{\sigma^{(\alpha)} := \bar{m}\} = \{\sigma \in \{-1,1\}^{T_{L,N}} : \sigma^{(\alpha)}(x,i) = \bar{m}(x,i) \text{ for all } (x,i) \in T_{L,N}\}.$$

The next results are the basic steps in establishing the Lebowitz–Penrose limit.

**Proposition 3.1.** *For any $\alpha \in (0,\frac{1}{2})$, there is a constant $c > 0$ such that for any $\bar{m} \in \mathcal{M}^{(\alpha)}$*

$$\log Z_{\beta,\gamma,\lambda}(\{\sigma^{(\alpha)} = \bar{m}\}) \le -\beta \bar{F}_{\beta,\gamma,\lambda}(\bar{m}) + \beta c \epsilon(\gamma,\lambda)|T_{L,N}|, \tag{3-1}$$

$$\log Z_{\beta,\gamma,\lambda}(\{\sigma^{(\alpha)} = \bar{m}\}) \ge -\beta \bar{F}_{\beta,\gamma,\lambda}(\bar{m}) - \beta c \epsilon(\gamma,\lambda)|T_{L,N}|, \tag{3-2}$$

*where $\bar{F}_{\beta,\gamma,\lambda}$ is defined in (2-17) and*

$$\epsilon(\gamma,\lambda) := \lambda\gamma^{1/2-\alpha} + \gamma^{\alpha}\log\gamma^{-\alpha}. \tag{3-3}$$

The proof, which follows the standard techniques, we postpone to Appendix A. For such choice of $\bar{m}$ the set $\{\sigma \approx^{\alpha} m\} = \{\sigma^{(\alpha)} = \bar{m}\}$, and for any $\mathcal{A} \subseteq \mathcal{M}^{(\alpha)}$ we define

$$Z^{(\alpha)}_{\beta,\gamma,\lambda}(\mathcal{A}) = \sum_{m\in\mathcal{A}} Z_{\beta,\gamma,\lambda}(\{\sigma^{(\alpha)} = m\}).$$

**Proposition 3.2.** *For any $\alpha \in (0,\frac{1}{2})$, there is a constant $c > 0$ such that for any $\mathcal{A} \subseteq \mathcal{M}^{(\alpha)}$*

$$\log Z^{(\alpha)}_{\beta,\gamma,\lambda}(\mathcal{A}) \le -\beta \inf_{\bar{m}\in\mathcal{A}} \bar{F}_{\beta,\gamma,\lambda}(\bar{m}) + \beta c \epsilon(\gamma,\lambda)|T_{L,N}|, \tag{3-4}$$

$$\log Z^{(\alpha)}_{\beta,\gamma,\lambda}(\mathcal{A}) \ge -\beta \inf_{\bar{m}\in\mathcal{A}} \bar{F}_{\beta,\gamma,\lambda}(\bar{m}) - \beta c \epsilon(\gamma,\lambda)|T_{L,N}|. \tag{3-5}$$

*Proof.* The proof is the same as that of Theorem 4.2.2.2 in [Presutti 2009]. □

Now we consider the case $\alpha > \frac{1}{2}$. We cannot directly apply Proposition 3.1 since the length of the vertical interaction is less than the length of the coarse graining. The idea is to write the fixed average of $m$ on the scale of $\gamma^{-\alpha}$ as an average after the coarse graining of scale $\gamma^{-\alpha'}$, $\alpha' \in (0, \frac{1}{2})$.

For $\bar{m}_\alpha \in \mathcal{M}^{(\alpha)}$ we define the set

$$\mathcal{A}_{\bar{m}_\alpha} = \left\{ \bar{m}_{\alpha'} \in \mathcal{M}^{(\alpha')} : \frac{1}{\gamma^{-\alpha}} \int_{C^{(\alpha)}_{k,i}} \bar{m}_{\alpha'}(r', i)\, dr' = \bar{m}_\alpha(r, i) \text{ for all } (r, i) \in T_{\ell,N} \right\}. \tag{3-6}$$

Using the above definition we prove a same result as in Proposition 3.1:

**Proposition 3.3.** *For any* $\alpha \in (\frac{1}{2}, 1)$, *there is a constant* $c > 0$ *such that for any* $\bar{m}_\alpha \in \mathcal{M}^{(\alpha)}$

$$\log Z_{\beta,\gamma,\lambda}(\{\sigma^{(\alpha)} = \bar{m}_\alpha\}) \le -\beta \inf_{\bar{m}_{\alpha'} \in \mathcal{A}_{\bar{m}_\alpha}} \bar{F}_{\beta,\gamma,\lambda}(\bar{m}_{\alpha'}) + \beta c \epsilon(\gamma, \lambda) |T_{L,N}|, \tag{3-7}$$

$$\log Z_{\beta,\gamma,\lambda}(\{\sigma^{(\alpha)} = \bar{m}_\alpha\}) \ge -\beta \inf_{\bar{m}_{\alpha'} \in \mathcal{A}_{\bar{m}_\alpha}} \bar{F}_{\beta,\gamma,\lambda}(\bar{m}_{\alpha'}) - \beta c \epsilon(\gamma, \lambda) |T_{L,N}|, \tag{3-8}$$

*where* $\bar{F}_{\beta,\gamma,\lambda}$ *is defined in* (2-17) *and* $\epsilon(\gamma, \lambda)$ *in* (3-3).

*Proof.* The proof follows by Propositions 3.1 and 3.2 □

## 4. Minimizer of the free energy functional

In this section we prove a technical result needed to prove Theorem 2.3. This key theorem tells us that a minimizer of the free energy functional under the constraint $⨍_\Lambda m^\Lambda = s$ is a "quasiconstant" function in a subset of $\Lambda \subset T_{\ell,N}$ (see (4-1)). So there are not oscillations that can affect the minimum of the free energy.

Fix $k \in \eta\mathbb{Z} \cap [0, \ell]$, with $\eta = \gamma \lceil \gamma^{-\alpha} \rceil$. We define the set $\Lambda_k = [k\eta, (k+1)\eta] \times \{1, \dots, N\}$, its complement $\Lambda_k^c = T_{\ell,N} \setminus \Lambda_k$, and the set $\Lambda_{k,i}$ as the restriction of $\Lambda_k$ to the $i$-th column.

Let $m^{\Lambda_k} \in L^\infty(\Lambda_k, [-1, 1])$; we define the *free energy functional* restricted to $\Lambda_k$

$$\begin{aligned} F^{\Lambda_k}_{\beta,\gamma,\lambda}(m^{\Lambda_k}) = -\frac{1}{2} \sum_{i=1}^N \int_{\Lambda_{k,i}} m^{\Lambda_k}(r, i) \bigg[ & \int_{\Lambda_{k,i}} J(r, r') m^{\Lambda_k}(r', i)\, dr' \\ & + \lambda \int_{\Lambda_{k,i}} J_{\gamma^{-1/2}}(r, r') (m^{\Lambda_k}(r', i-1) + m^{\Lambda_k}(r', i+1))\, dr' \bigg] dr \\ & - \frac{1}{\beta} \sum_{i=1}^N \int_{\Lambda_{k,i}} I(m^{\Lambda_k}(r, i))\, dr. \end{aligned}$$

Let $m^{\Lambda_k^c} \in L^\infty(\Lambda_k^c, [-1, 1])$; we define the *conditioned free energy functional*

$$
\begin{aligned}
&F^{\Lambda_k}_{\beta,\gamma,\lambda}(m^{\Lambda_k}\,|\,m^{\Lambda_k^c})\\
&\quad = F^{\Lambda_k}_{\beta,\gamma,\lambda}(m^{\Lambda_k}) - \sum_{i=1}^{N}\int_{\Lambda_{k,i}} m^{\Lambda_k}(r,i)\int_{\Lambda^c_{k,i}} J(r,r')m^{\Lambda_k^c}(r',i)\,dr'\,dr\\
&\qquad\qquad - \lambda\sum_{i=1}^{N}\int_{\Lambda_{k,i}} m^{\Lambda_k}(r,i)\biggl[\int_{\Lambda^c_{k,i-1}} J_{\gamma^{-1/2}}(r,r')m^{\Lambda_k^c}(r',i-1)\,dr'\\
&\qquad\qquad\qquad\qquad + \int_{\Lambda^c_{k,i+1}} J_{\gamma^{-1/2}}(r,r')m^{\Lambda_k^c}(r',i+1)\,dr'\biggr]\,dr,
\end{aligned}
$$

where the set $\Lambda^c_{k,i}$ is the set $\Lambda^c_k$ restricted to the $i$-th column.

The following theorem is the most relevant contribution of this work.

**Theorem 4.1.** *Take* $\gamma > 0$, *and define* $\eta := \gamma[\gamma^{-\alpha}] = \gamma^{1/2}(1+\gamma^{-\varepsilon}\gamma^{-\delta}\zeta)$ *where* $\varepsilon \in (0,\frac{1}{2})$, $\delta \in (0, \frac{1}{2}-\varepsilon]$, *and* $\zeta > 0$ *is small enough.*

*If* $\beta(\eta\|J\|_\infty + 2\lambda) \le \frac{1}{4}$, *then for all* $k \in \eta\mathbb{Z}\cap[0,\ell]$, $s\in[-1,1]^N$, *and* $m^{\Lambda_k^c} \in L^\infty(\Lambda^c_k,[-1,1])$ *there exists a unique* $\phi^{\Lambda_k} \in L^\infty(\Lambda_k,[-1,1])$ *such that* $⨍_{\Lambda_{k,i}}\phi^{\Lambda_k} = s_i$ *for all* $i$, *and*

$$F^{\Lambda_k}_{\beta,\gamma,\lambda}(m^{\Lambda_k}\,|\,m^{\Lambda_k^c}) \ge F^{\Lambda_k}_{\beta,\gamma,\lambda}(\phi^{\Lambda_k}\,|\,m^{\Lambda_k^c}),$$

*for any* $m^{\Lambda_k}\in L^\infty(\Lambda_k,[-1,1])$ *such that* $⨍_{\Lambda_{k,i}} m_{\Lambda_k} = s_i$ *for all* $i$.

*Moreover, there exists a constant* $C>0$ *such that, for any* $r\in\bar\Lambda_{k,i}$,

$$|\phi^{\Lambda_k}(r,i) - s_i| \le C\|\nabla_r\phi\|_{\infty,\bar\Lambda_{k,i}}\,\eta \tag{4-1}$$

*where* $\bar\Lambda_{k,i} = [k\eta + \gamma^{1/2}(1+\gamma^{-\varepsilon}), (k+1)\eta - \gamma^{1/2}(1+\gamma^{-\varepsilon})]$.

*Proof.* If $s_i = \pm 1$ for all $i$, we have $m^{\Lambda_k} = \pm 1$ almost everywhere and the theorem follows easily. Now we take $|s_i|<1$ for all $i$ and we use Lagrange multipliers. In the following we omit the dependence on $k$ and we keep only the dependence on the column $i$; then we take $\Lambda = \Lambda_k$ and $\Lambda_i = \Lambda_{k,i}$.

For all $h\in\mathbb{R}^N$ we define

$$F^{\Lambda,h}_{\beta,\gamma,\lambda}(m^\Lambda, m^{\Lambda^c}) = F^{\Lambda}_{\beta,\gamma,\lambda}(m^\Lambda\,|\,m^{\Lambda^c}) - \sum_{i=1}^{N} h_i\int_{\Lambda_i} m^\Lambda(r,i)\,dr.$$

For all $r\in[k\eta,(k+1)\eta)$ we define the vectors

$$\underline{m}^\Lambda(r) = (m_i^\Lambda(r))_{i=1}^N = (m^\Lambda(r,1),\dots,m^\Lambda(r,N))$$

and

$$(J*\underline{m}^\Lambda)(r) = ((J*m_i^\Lambda)(r))_{i=1}^N = \biggl(\int_{\Lambda_i} J(r,r')m^\Lambda(r,i)\,dr\biggr)_{i=1}^N.$$

In this notation the free energy becomes

$$
\begin{aligned}
&F^{\Lambda,h}_{\beta,\gamma,\lambda}(m^\Lambda \mid m^{\Lambda^c})\\
&= -\frac{1}{2}\sum_{i=1}^N \int_{\Lambda_i} m_i^\Lambda(r)((J * m_i^\Lambda)(r) + \lambda(J_{\gamma^{-1/2}} * (m_{i-1}^\Lambda + m_{i+1}^\Lambda))(r))\,dr\\
&\quad - \sum_{i=1}^N \int_{\Lambda_i} m_i^\Lambda(r)[(J * m_i^{\Lambda^c})(r) - \lambda((J_{\gamma^{-1/2}} * m_{i+1}^{\Lambda^c})(r) + (J_{\gamma^{-1/2}} * m_{i-1}^{\Lambda^c})(r))]\,dr\\
&\quad - \sum_{i=1}^N h_i \int_{\Lambda_i} m_i^\Lambda(r)\,dr - \frac{1}{\beta}\sum_{i=1}^N \int_{\Lambda_i} I(m_i^\Lambda(r))\,dr.
\end{aligned}
$$

Let $A_h(\underline{m}^\Lambda) = (A_i(\underline{m}^\Lambda))_{i=1}^N$, where

$$
\begin{aligned}
A_i(\underline{m}^\Lambda) &= \tanh(\beta[J * (m_i^\Lambda + m_i^{\Lambda^c}) + \lambda J_{\gamma^{-1/2}} * (m_{i-1}^\Lambda + m_{i-1}^{\Lambda^c} + m_{i+1}^\Lambda + m_{i+1}^{\Lambda^c}) + h_i])\\
&= A_i(m_i, m_{i+1}, m_{i-1}).
\end{aligned}
$$

From general results[1] the infimum of $F^{\Lambda,h}_{\beta,\gamma,\lambda}(\,\cdot \mid m^{\Lambda^c})$ is a minimum attained on functions such that $A_h(\underline{\psi}^\Lambda) = \underline{\psi}^\Lambda$. Thus, the set

$$
\mathcal{G}_{h,m^{\Lambda^c}} = \{\underline{\psi}^\Lambda \in L^\infty(\Lambda, [-1, 1]^N) : \underline{\psi}^\Lambda = A_h(\underline{\psi}^\Lambda)\}
$$

is nonempty. We want to show that $\mathcal{G}$ is actually a singleton.

**Step 1.** *$A_h$ is a contraction.*

*Proof.* We define the norm $\|A_h(\underline{m}^\Lambda)\|_{\infty,N} = \max_{\{i=1,\dots,N\}} \|A_i(\underline{m}^\Lambda)\|_\infty$. Given $\underline{m}^\Lambda$, $\underline{m}'^\Lambda$ we have, by the triangle inequality, the Lagrange theorem, and properties of $J$,

$$
\|A_i(\underline{m}^\Lambda) - A_i(\underline{m}'^\Lambda)\|_\infty \le \beta(\eta\|J\|_\infty + 2\lambda)\|\underline{m}^\Lambda - \underline{m}'^\Lambda\|_{\infty,N}.
$$

We observe that in this framework we can identify the set $\Lambda_{i+1}$ with the set $\Lambda_i$, and with an abuse of notation we call it $\Lambda$. Then $A$ is a contraction and there exists a unique fixed point $\underline{\phi}^{\Lambda,h}$ such that

$$
\underline{\phi}^{\Lambda,h} = \lim_{n\to\infty} A_h(\underline{u}_n) \quad \text{with } \underline{u}_n = A_h(\underline{u}_{n-1}) \text{ and } \underline{u}_0 = s\mathbf{1}_\Lambda.
$$

The convergence is in the sup norm, and then it is uniform in $h$. □

**Step 2.** *$\underline{\phi}^{\Lambda,h}$ is differentiable in $h$.*

*Proof.* We prove by induction on $n$ that $\underline{u}_n$ is differentiable in $h$ with derivative

$$
\frac{\partial}{\partial h_j} u_i^{n,\Lambda} = p^{i,n}\left[J * \frac{\partial}{\partial h_j} u_i^{n-1,\Lambda} + \lambda J_{\gamma^{-1/2}} * \left(\frac{\partial}{\partial h_j} u_{i-1}^{n-1,\Lambda} + \frac{\partial}{\partial h_j} u_{i+1}^{n-1,\Lambda}\right) + \delta_{i,j}\right] \tag{4-2}
$$

[1] See [Presutti 2009, Theorem 6.2.6.2].

where

$$p^{i,n} = \beta \cosh^{-2}\big[\beta\big(J * (u_i^{\Lambda,n-1} + m_i^{\Lambda^c}) \\ + \lambda J_{\gamma^{-1/2}} * (u_{i-1}^{\Lambda,n-1} + u_{i+1}^{\Lambda,n-1} + m_{i-1}^{\Lambda^c} + m_{i+1}^{\Lambda^c}) + h_i\big)\big].$$

Indeed $D\underline{u}_0 = 0$ and if $\underline{u}_{n-1}$ is differentiable, $D\underline{u}_n$ exists and it is given by (4-2). Suppose $\|\frac{\partial}{\partial h_j} u_i^{n-1,\Lambda}\|_\infty \le 2\beta$; then

$$\left\| \frac{\partial}{\partial h_j} u_i^{n,\Lambda} \right\|_\infty \le \beta(\|J\|_\infty \eta 2\beta + 4\lambda\beta + 1)$$

by hypothesis $2\beta(\|J\|_\infty \eta + 2\lambda) + 1 \le 2$.

Then $\underline{\phi}^{\Lambda,h}$ is differentiable on $h$ and

$$\nabla_h \underline{\phi}^{\Lambda,h} = \lim_{n\to\infty} \nabla_h \underline{u}_n. \qquad \square$$

**Step 3.** *For all $\lambda$ small enough, there exists exactly one function $h(\lambda)$ such that*

- *$\underline{\phi}^{\Lambda,h}$ is the minimum of $F^{\Lambda,h}_{\beta,\gamma,\lambda}$ and*
- *$H(\lambda, h(\lambda)) = \fint_\Lambda \underline{\phi}^{\Lambda,h}\, dr - s = 0$.*

*Proof.* If $\lambda = 0$, every column is independent of the other columns; then for each column we can find $h_i^0$ such that $\fint_\Lambda \phi^{\lambda,h_i^0}\, dr = s_i$.[2] This implies that $H(0, h^0) = 0$. In order to apply the implicit function theorem, we prove the invertibility of $\frac{\partial H(\lambda,h(\lambda))}{\partial h}$. We start by explicitly writing the derivative

$$\begin{aligned}\frac{\partial H}{\partial h} &= \fint_\Lambda \frac{\partial}{\partial h} A_h(\underline{\phi}^{\Lambda,h})\, dr \\ &= \fint_\Lambda \frac{\partial}{\partial h}\big(\tanh\big\{\beta\big[J * (\phi_i^\Lambda + m_i^{\Lambda^c}) \\ &\qquad + \lambda J_{\gamma^{-1/2}} * (\phi_{i-1}^\Lambda + m_{i-1}^{\Lambda^c} + \phi_{i+1}^\Lambda + m_{i+1}^{\Lambda^c}) + h_i\big]\big\}\big)_{i=1}^N\, dr.\end{aligned}$$

We define the square matrices $P, K \in M_N(\mathbb{R})$:

$$P_{i,j} = \begin{cases} 0 & \text{if } i \ne j, \\ p_i & \text{if } i = j, \end{cases} \qquad K_{i,j} = p_i[J * b_{i,j} + \lambda J_{\gamma^{-1/2}} * (b_{i+1,j} + b_{i-1,j})]$$

where $b_{i,j} = \frac{\partial \phi_i^{\Lambda,h}}{\partial h_j}$ and

$$p_i = \beta \cosh^{-2}[\beta(J * (\phi_i^{\Lambda,h} + m_i^{\Lambda^c}) + \lambda J_{\gamma^{-1/2}} * (\phi_{i-1}^{\Lambda,h} + \phi_{i+1}^{\Lambda,h} + m_{i-1}^{\Lambda^c} + m_{i+1}^{\Lambda^c}) + h_i)].$$

We write the derivative with respect to $h$ of $H$ in terms of $P$ and $K$:

$$\frac{\partial}{\partial h} H = \fint_\Lambda (K + P)\, dr = \bar{K} + \bar{P} = \bar{P}(\bar{P}^{-1}\bar{K} + I).$$

[2] This follows from [Presutti 2009, Theorem 6.4.1.1].

We observe that

$$\begin{aligned}|(\bar{P}^{-1}\bar{K})_{i,j}| &\le \frac{1}{\fint_\Lambda p_i\,dr}\fint_\Lambda |K_{i,j}|\,dr\\ &\le \frac{1}{\fint_\Lambda p_i\,dr}\fint_\Lambda p_i\|(P^{-1}K)_{i,j}\|_\infty\,dr \le \|(P^{-1}K)_{i,j}\|_\infty.\end{aligned}$$

To prove the existence of the matrix $(\bar{P}^{-1}\bar{K}+I)^{-1}$ we show that $\sum_{n=0}^{\infty}(-P^{-1}K)^n < \infty$, proving that $\sup_i \sum_{j=0}^{N}(P^{-1}K)_{i,j} \le c < 1$.

We give an estimate for $\|b_{i,j}\|_\infty$. We recall that

$$b_{i,j} = p_i(J * b_{i,j} + \lambda J_{\gamma^{-1/2}} * (b_{i-1,j} + b_{i+1,j}) + \delta_{i,j}).$$

Then

$$\|b_{i,j}\|_\infty \le \frac{\|p_i\|_\infty}{1-\|p_i\|_\infty\eta\|J\|_\infty}(\lambda(\|b_{i-1,j}\|_\infty + \|b_{i+1,j}\|_\infty) + \delta_{i,j}). \tag{4-3}$$

We define for all $i \in \{1, \dots, N\}$ and $a \in \mathbb{N}$

$$\begin{aligned} q_i &= \frac{\|p_i\|_\infty}{1-\|p_i\|_\infty\eta\|J\|_\infty},\\ \Omega_a^i &= \{\sigma \in \mathbb{Z}^a : \sigma(0) = i,\ \sigma(k) \equiv \sigma(k-1) \pm 1 \bmod N\}.\end{aligned}$$

We observe that

$$q_i \le \frac{\beta}{1-\beta\eta\|J\|_\infty}, \qquad |\Omega_a^i| = 2^a.$$

Iterating the inequality (4-3) $a$ times we obtain

$$\begin{aligned}\|b_{i,j}\|_\infty &\le \sum_{\sigma\in\Omega_a^i} q_{\sigma(0)}\cdots q_{\sigma(a)}\|b_{\sigma(a),j}\|_\infty\lambda^a + \sum_{n=0}^{a-1}\sum_{\sigma\in\Omega_n^i} q_{\sigma(0)}\cdots q_{\sigma(n)}\lambda^n\delta_{\sigma(n),j}\\ &\le \left(\frac{2\beta\lambda}{1-\eta\|J\|_\infty\beta}\right)^a\|b_{\sigma(a),j}\|_\infty + q_{\sigma(0)}\sum_{n=0}^{a-1}\left(\frac{2\beta\lambda}{1-\eta\|J\|_\infty\beta}\right)^n\delta_{\sigma(n),j}.\end{aligned}$$

If the number of iterations is big enough, then the Dirac delta is 1. We define $n := n(i-j)$ where

$$n(i-j) = \begin{cases} |i-j| & \text{if } |i-j| \le \lceil N/2 \rceil,\\ N-|i-j| & \text{otherwise.}\end{cases} \tag{4-4}$$

Then $\delta_{\sigma(n),j} = 1$ if the number of iterations is at least $n(i-j)$. Taking the limit $a \to \infty$,

$$\|b_{i,j}\|_\infty \le q_{\sigma(0)}\sum_{n=[i-j]}^{\infty}\left(\frac{2\beta\lambda}{1-\eta\|J\|_\infty\beta}\right)^n\delta_{\sigma(n),j} \le q_i\frac{\theta^{n(i-j)}}{1-\theta}, \tag{4-5}$$

with $\theta = 2\beta\lambda/(1-\eta\|J\|_\infty\beta) < 1$.

Let $r \in \Lambda$, and consider

$$(P^{-1}K)_{i,j}(r) = (J * b_{i,j})(r) + \lambda(J_{\gamma^{-1/2}} * (b_{i+1,j} + b_{i-1,j}))(r).$$

Using (4-5) we have

$$\sum_{j=1}^{N} \|(P^{-1}K)_{i,j}\|_\infty \le \frac{\alpha}{1-\theta} \sum_{j=1}^{N} \frac{\theta^{(\min_{r=i-1,i,i+1} n(r-j))}}{1-\theta} \le \frac{2\alpha}{(1-\theta)^2}$$

where $\alpha = (\eta\|J\|_\infty + 2\lambda)\beta/(1-\eta\|J\|_\infty\beta)$.

Now keeping in mind that $\beta(\eta\|J\|_\infty + 2\lambda) \le \frac{1}{4}$ we obtain

$$\frac{2\alpha}{(1-\theta)^2} \le \tfrac{8}{9}(1-\eta\|J\|_\infty\beta).$$

For $\eta$ small enough the matrix $(\bar{P}^{-1}\bar{K} + I)$ can be inverted and we find the function $h(\lambda) = h$ such that $H(\lambda, h(\lambda)) = 0$. For $\underline{m}^\Lambda$ such that $⨍_\Lambda \underline{m}^\Lambda = s$ the conditioned free energy

$$\begin{aligned} F^{\Lambda,h}_{\beta,\gamma,\lambda}(m^\Lambda \,|\, m^{\Lambda^c}) &= F^{\Lambda,h}_{\beta,\gamma,\lambda}(m^\Lambda \,|\, m^{\Lambda^c}) + h\sum_{i=1}^{N} |\Lambda_i| s_i \\ &\ge F^{\Lambda,h}_{\beta,\gamma,\lambda}(\underline{\phi}^{\Lambda,h} \,|\, m^{\Lambda^c}) + h\sum_{i=1}^{N} |\Lambda_i| s_i \\ &= F^{\Lambda,h}_{\beta,\gamma,\lambda}(\underline{\phi}^{\Lambda,h} \,|\, m^{\Lambda^c}). \end{aligned}$$

□

We now prove the last part of the theorem. Let $\bar{\Lambda}_i = (k\eta + \gamma^{1/2}(1+\gamma^{-\varepsilon}), (k+1)\eta - \gamma^{1/2}(1+\gamma^{-\varepsilon}))$, and define

$$\bar{s}_i = ⨍_{\bar{\Lambda}_i} \phi_i^{\Lambda,h}(r)\,dr.$$

We observe that, for such a constant $c > 0$,

$$\begin{aligned} |s_i - \bar{s}_i| &\le \left(\frac{|\bar{\Lambda}_i|}{|\Lambda|} - 1\right) ⨍_{\bar{\Lambda}_i} |\phi_i^{\Lambda,h}(r)|\,dr + \frac{|\Lambda \setminus \bar{\Lambda}_i|}{|\Lambda|} ⨍_{\Lambda\setminus\bar{\Lambda}_i} |\phi_i^{\Lambda,h}(r)|\,dr \\ &\le c\gamma^\delta \le c\eta \end{aligned} \tag{4-6}$$

because of the choice of $\eta$. Fix $r' \in \bar{\Lambda}_i$:

$$|\bar{s}_i - \phi_i^{\Lambda,h}(r')| \le C \left\| \frac{\partial}{\partial r}\phi_i^{\Lambda,h} \right\|_\infty \eta,$$

where $\left\| \frac{\partial}{\partial r}\phi_i^{\Lambda,h} \right\|_\infty = \sup_{r\in\bar{\Lambda}_i} \left| \frac{\partial}{\partial r}\phi_i^{\Lambda,h}(r) \right|$.

It remains to prove that $\|\frac{\partial}{\partial r}\phi_i^{\Lambda,h}\|_\infty < \infty$. We shall use the recursive formula

$$\begin{aligned}\frac{\partial}{\partial r}\phi_i^{\Lambda,h}(r)\\ = p_i\left[\frac{\partial}{\partial r}J * (\phi_i^{\Lambda,h} + m_i^{\Lambda^c}) + \lambda\frac{\partial}{\partial r}J_{\gamma^{-1/2}} * (\phi_{i-1}^{\Lambda,h} + \phi_{i+1}^{\Lambda,h} + m_{i-1}^{\Lambda^c} + m_{i+1}^{\Lambda^c})\right].\end{aligned} \tag{4-7}$$

If we iterate (4-7) $a$ times we obtain

$$\begin{aligned}\frac{\partial}{\partial r}\phi_i^{\Lambda,h}(r) = \sum_{\sigma\in\Omega_a^i} p_{\sigma(0)}\cdots p_{\sigma(a)}\lambda^a J_{\gamma^{-1/2}} *^{(a-1)}\cdots * \frac{\partial J}{\partial r} * (\phi_{\sigma(a)}^{\Lambda,h} + m_{\sigma(a)}^{\Lambda^c})\\ + \sum_{n=0}^{a-1}\sum_{\sigma\in\Omega_n^i} p_{\sigma(0)}\cdots p_{\sigma(n)}\lambda^{n+1} J_{\gamma^{-1/2}} *^{(n)}\cdots * \frac{\partial}{\partial r}\phi_{\sigma(n)}^{\Lambda,h}.\end{aligned}$$

Observing that, at each iteration $n$, if $n < \gamma^{-\varepsilon}$, then $(J_{\gamma^{-1/2}} *^{(n)}\cdots * m_{\sigma(n)}^{\Lambda^c})(r) = 0$ by the choice of the set $\bar{\Lambda}$. Taking the norm,

$$\left\|\frac{\partial}{\partial r}\phi_i^{\Lambda,h}\right\|_\infty \le \eta\|\nabla J\|_\infty\beta\sum_{n=0}^{a}(2\beta\lambda)^n + (2\beta\lambda)^a\|\nabla J\|_\infty 2\gamma^{-1/2}$$

where we took the derivative of $J_{\gamma^{-1/2}}$ in the last term indexed by $a$. If $a = \lceil\gamma^{-\varepsilon}\rceil$, then

$$\lim_{\gamma\to 0}(2\beta\lambda)^a\|\nabla J\|_\infty 2\gamma^{-1/2} = 0$$

and

$$\left\|\frac{\partial}{\partial r}\phi_i^{\Lambda,h}\right\|_\infty \le c'\eta\|\nabla J\|_\infty\beta < \infty,$$

for some constant $c' > 0$. Equation (4-6) gives

$$|\phi_i^{\Lambda}(r) - s_i| \le C\eta \quad \text{for all } r \in \bar{\Lambda}_i,$$

and the theorem is proved. □

## 5. The Lebowitz–Penrose limit

*Proof of Theorem 2.3.* The proof is divided into two parts: $\alpha < \frac{1}{2}$ and $\alpha > \frac{1}{2}$. We will use the results of the previous sections in both cases. While for $\alpha < \frac{1}{2}$ we can use them straightforwardly in the Lebowitz–Penrose procedure, for $\alpha > \frac{1}{2}$ the technical Theorem 4.1 is needed in order to control the fluctuations.

**Case 1** ($\alpha \in (0, \frac{1}{2})$). *Let $m \in L^\infty(T_{\ell,N}; [-1, 1])$; we prove that*

$$\lim_{\gamma\to 0} F_{\beta,\gamma,\lambda}^{(\alpha)}(m) = F_{\beta,\lambda}(m). \tag{5-1}$$

*Proof.* Given a mesoscopic state $m$ we choose a function $\bar{m}_\alpha \in \mathcal{M}^{(\alpha)}$ that "recognizes" $m$ (see Definition 2.2):

$$\lim_{\gamma\to 0} F^{(\alpha)}_{\beta,\gamma,\lambda}(m) = \lim_{\gamma\to 0} -\frac{1}{\beta\gamma^{-1}} \log Z_{\beta,\gamma,\lambda}(\{\sigma^{(\alpha)} = \bar{m}_\alpha\}).$$

We apply Proposition 3.1 and the change of coordinates $m_\alpha(r, i) := \bar{m}_\alpha(\gamma^{-1} r, i)$. We shall show that $|F_{\beta,\gamma,\lambda}(m^{(\alpha)}) - F_{\beta,\lambda}(m)| \to 0$ as $\gamma \to 0$ where

$$\begin{aligned} F_{\beta,\gamma,\lambda}(m) = -\frac{1}{2}\sum_{i=1}^{N}\int_0^\ell m(r,i)\biggl(\int_0^\ell J(r,r')m(r',i)\,dr' \\ + \lambda\int_0^\ell J_{\gamma^{-1/2}}(r,r')(m(r',i-1)+m(r',i+1))\,dr'\biggr)\,dr \\ -\frac{1}{\beta}\sum_{i=1}^{N}\int_0^\ell I(m(r,i))\,dr. \end{aligned} \tag{5-2}$$

By the Lebesgue differentiation theorem [Rudin 1987] we know that $m^{(\alpha)} \xrightarrow{L^1} m$ (see (2-6)); thus, by the triangle inequality the limit can be divided into three parts. The *first term* is

$$\begin{aligned} \left|\int_0^\ell\int_0^\ell J_{\gamma^{-1/2}}(r,r')m^{(\alpha)}(r,i)m^{(\alpha)}(r',i-1)\,dr\,dr' - \int_0^\ell m(r,i)m(r,i-1)\,dr\right| \\ \le \left|\int_0^\ell m^{(\alpha)}(r,i)\biggl[\int_0^\ell J_{\gamma^{-1/2}}(r,r')m^{(\alpha)}(r',i-1)\,dr' - m(r,i-1)\biggr]\,dr\right| \\ + \left|\int_0^\ell m(r,i-1)[m^{(\alpha)}(r,i)-m(r,i)]\,dr\right|. \end{aligned}$$

We observe that

$$J_{\gamma^{-1/2}} * m(r) = \int_0^\ell J_{\gamma^{-1/2}}(r,r')m(r')\,dr'$$

converges to the Dirac delta as $\gamma \to 0$; then by the dominated convergence theorem

$$\left|\int_0^\ell\int_0^\ell J_{\gamma^{-1/2}}(r,r')m^{(\alpha)}(r,i)m^{(\alpha)}(r',i-1)\,dr\,dr' - \int_0^\ell m(r,i)m(r,i-1)\,dr\right| \to 0.$$

The other two terms converge to 0 by the dominated convergence theorem. And (5-1) is proved. □

**Case 2** ($\alpha \in (\frac{1}{2}, 1)$). *By Proposition 3.3 and using the same notations introduced in the case* $\alpha \in (0, \frac{1}{2})$, *we have*

$$\lim_{\gamma\to 0} F^{(\alpha)}_{\beta,\gamma,\lambda}(m) = \lim_{\gamma\to 0} \gamma \inf_{\bar{m}_{\alpha'} \in \mathcal{A}_{\bar{m}_\alpha}} \bar{F}_{\beta,\gamma,\lambda}(\bar{m}_{\alpha'}) = \lim_{\gamma\to 0} \inf_{\substack{m_{\alpha'}(\cdot) = \bar{m}_{\alpha'}(\gamma^{-1}\cdot) \\ \bar{m}_{\alpha'} \in \mathcal{A}_{\bar{m}_\alpha}}} F_{\beta,\gamma,\lambda}(m_{\alpha'}).$$

In order to pass the limit through the infimum, we need to prove a result of Γ-convergence. Let us start defining a notion of convergence:

**Definition 5.1** ($\star$-convergence). Set $\eta := \gamma[\gamma^{-\alpha}]$; then for all sequences $\{m_\gamma\}$ and $m$ in $L^\infty(T_{\ell,N}, [-1,1])$ we say that $m_\gamma \xrightarrow{\star} m$ if

$$\lim_{\gamma\to 0} \sum_{i=1}^{N} \sum_{k=1}^{\lceil \ell/\eta \rceil} \left| \fint_{D_{k,i}^{(\alpha)}} m_\gamma(r', i)\, dr' - \fint_{D_{k,i}^{(\alpha)}} m(r,i)\, dr \right| = 0 \tag{5-3}$$

where $D_{k,i}^{(\alpha)} = \{(x,i) \in \mathbb{R}\times\mathbb{Z} : k\eta \le x \le (k+1)\eta\}$.

We can write (5-3) as $\lim_\gamma d(m_\gamma, m) = 0$, since $d$ is actually a distance.

**Remark.** Following the same notation of the case $\alpha < \frac{1}{2}$, we observe that the sequence $\{m_\alpha\}_\gamma \xrightarrow{\star} m$ as $\gamma \to 0$.

The following Γ-convergence result will be proved in the next section.

**Proposition 5.2.** *Let $F_{\beta,\gamma,\lambda}$ be as in (5-2) and $F_{\beta,\lambda}$ as in (2-10). Then*

$$F_{\beta,\lambda} = \Gamma \lim_{\gamma\to 0} F_{\beta,\gamma,\lambda}$$

*according to the $\star$-convergence.*

We move to the last part of proof of Theorem 2.3.

We start by proving the *lower bound*. For each $\delta > 0$ we can take $\gamma$ small enough such that there exists a function $m_\alpha(\cdot) = \bar{m}_\alpha(\gamma^{-1}\cdot)$ such that $d(m_\alpha, m) < \delta$; then

$$\inf_{\substack{m_{\alpha'}(\cdot)=\bar{m}_{\alpha'}(\gamma^{-1}\cdot) \\ \bar{m}_{\alpha'} \in \mathcal{A}_{\bar{m}_\alpha}}} F_{\beta,\gamma,\lambda}(m_{\alpha'}) \ge \inf_{\substack{m' \in L^\infty(T_{\ell,N},[-1,+1]) \\ d(m',m)<\delta}} F_{\beta,\gamma,\lambda}(m').$$

Taking the infimum limit with respect to $\gamma$ and the supremum with respect to $\delta$,

$$\begin{aligned} \liminf_{\gamma\to 0} F^{(\alpha)}_{\beta,\gamma,\lambda}(m) &\ge \sup_{\delta>0} \liminf_{\gamma\to 0} \inf_{\substack{m' \in L^\infty(T_{\ell,N},[-1,+1]) \\ d(m',m)<\delta}} F_{\beta,\gamma,\lambda}(m') \\ &\ge F_{\beta,\lambda}(m). \end{aligned}$$

The last inequality follows by definition of the Γ-limit[3] and Proposition 5.2.

Now we consider the *upper bound*. Let $\tilde{\bar{m}}_{\alpha'}$ be the closest element in $\mathcal{A}_{\bar{m}_\alpha}$ to $\bar{m}_\alpha$, namely

$$|\bar{m}_\alpha(r,i) - \tilde{\bar{m}}_{\alpha'}(r,i)| \le \frac{2}{\gamma^{-\alpha'}} \quad \text{for all } (r,i) \in T_{L,N}. \tag{5-4}$$

[3] See [Braides 2002].

Such a magnetization exists by the definition of the sets $M^{(\alpha')}$ and $M^{(\alpha)}$. We define $\tilde{m}_{\alpha'}(r,i)=\tilde{\bar{m}}_{\alpha'}(\gamma^{-1}r,i)$, for all $(r,i)$; then

$$\begin{aligned}\limsup_{\gamma\to 0} F^{(\alpha)}_{\beta,\gamma,\lambda}(m) &= \limsup_{\gamma\to 0} \inf_{\substack{m_{\alpha'}(\,\cdot\,)=\bar{m}_{\alpha'}(\gamma^{-1}\,\cdot\,)\\ \bar{m}_{\alpha'}\in\mathscr{A}_{\bar{m}_\alpha}}} F_{\beta,\gamma,\lambda}(m_{\alpha'})\\ &\le \limsup_{\gamma\to 0} F_{\beta,\gamma,\lambda}(\tilde{m}_{\alpha'})\\ &\le F_{\beta,\lambda}(m).\end{aligned}$$

The last inequality follows from Proposition 5.2 and (5-4). □

## 6. Γ-limit

In this section we shall prove the existence of the Γ-limit of $F_{\beta,\gamma,\lambda}$. Definition 5.1 of $\star$-convergence involves the average of $m$ on sets of length $\gamma[\gamma^{-\alpha}]=\eta$. This implies a constraint on the minimizer of the free energy functional, and at this level we use the Theorem 4.1.

*Proof of Proposition 5.2.* We start with the *lower bound*: for all $\{m_\gamma\}$ such that $m_\gamma \overset{\star}{\to} m$,

$$\liminf_{\gamma\to 0} F_{\beta,\gamma,\lambda}(m_\gamma) \ge F_{\beta,\lambda}(m).$$

Fix $\eta$, $\Lambda_k$ as in Theorem 4.1; we set $n=\ell/\eta$ and observe that $T_{\ell,N}=\bigcup_{k=1}^n \Lambda_k$. We define $m_\gamma^{\Lambda_k} := m_\gamma|_{\Lambda_k}$, the restriction of $m_\gamma$ to the set $\Lambda_k$. Fix $\Lambda_1$; then by Theorem 4.1 there exists $\phi_\gamma^{\Lambda_1}$ such that

$$\begin{aligned}F_{\beta,\gamma,\lambda}(m_\gamma) &= F^{\Lambda_1}_{\beta,\gamma,\lambda}(m_\gamma^{\Lambda_1}\,|\,m_\gamma^{\Lambda_1^c}) + F^{\Lambda_1^c}_{\beta,\gamma,\lambda}(m_\gamma^{\Lambda_1^c})\\ &\ge F^{\Lambda_1}_{\beta,\gamma,\lambda}(\phi_\gamma^{\Lambda_1}\,|\,m_\gamma^{\Lambda_1^c}) + F^{\Lambda_1^c}_{\beta,\gamma,\lambda}(m_\gamma^{\Lambda_1^c})\\ &= F^{\Lambda_1}_{\beta,\gamma,\lambda}(m_\gamma^1)\end{aligned}$$

where $m_\gamma^1 = m_\gamma \mathbb{1}_{\Lambda_1^c} + \phi_\gamma^{\Lambda_1}\mathbb{1}_{\Lambda_1}$ and $\mathbb{1}_A$ is the indicator function of the set $A$.

We iterate this procedure for each $k$, and we define $m_\gamma^{1,\dots,n}$; then

$$F_\gamma(m_\gamma) \ge F_\gamma(m_\gamma^{1,\dots,n}).$$

**Lemma 6.1.** *Let $m_\gamma^{1,\dots,n}$ be as above. We have*

$$\lim_{\gamma\to 0}\sum_{i=1}^N \int_0^\ell |m_\gamma^{1,\dots,n}(r,i)-m(r,i)|\,dr = 0.$$

*Proof.* For any $i \in \{1, \dots, N\}$ we split the integral following the partition given by $\Lambda_{k,i}$, and we consider

$$\sum_{k=1}^{n} \int_{\Lambda_{k,i}} \left| m_\gamma^{1,\dots,n}(r,i) \pm \fint_{\Lambda_{k,i}} m_\gamma^{1,\dots,n}(r',i)\,dr' - m(r,i) \right| dr.$$

Applying the triangle inequality, the fist term converges by definition of $m_\gamma^{1,\dots,n}$ and the Lebesgue differentiation theorem. The second term can be estimated as

$$\begin{aligned}\sum_{k=1}^{n} \int_{\Lambda_{k,i}} &\left| m_\gamma^{1,\dots,n}(r,i) - \fint_{\Lambda_{k,i}} m_\gamma^{1,\dots,n}(r',i)\,dr' \right| dr \\ &\le \sum_{k=1}^{n} \int_{\bar\Lambda_{k,i}} |m_\gamma^{1,\dots,n}(r,i) - s|\,dr + \int_{\bar E_k} \left| m_\gamma^{1,\dots,n}(r,i) - \fint_{\Lambda_{k,i}} m_\gamma^{1,\dots,n}(r',i)\,dr' \right| dr \\ &\le \sum_{k=1}^{n} |\Lambda_{k,i}|\eta c + |\bar E_k| c'\end{aligned}$$

where

$$\bar\Lambda_{k,i} = (k\eta + \gamma^{1/2}(1+\gamma^{-\varepsilon}), (k+1)\eta - \gamma^{1/2}(1+\gamma^{-\varepsilon}))$$

and $\bar E_k = \Lambda_{k,i} \setminus \bar\Lambda_{k,i}$. To finish the proof we just observe that the size of $\bar E_k$ is of the order of $\gamma^{1/2}(1+\gamma^{-\varepsilon})$. □

To prove the lower bound we separately consider the convergence of the three terms of $F_{\beta,\gamma,\lambda}$.

The *first term* is

$$\sum_{i=1}^{N} \frac{1}{2} \left| \int_0^\ell \int_0^\ell J(r,r')(m_\gamma^{1,\dots,n}(r,i) m_\gamma^{1,\dots,n}(r',i) - m(r,i)m(r',i))\,dr\,dr' \right|.$$

Using the triangle inequality with $m_\gamma^{1,\dots,n}(r,i)m(r,i)$, the convergence follows from Lemma 6.1.

The *second term* is

$$\begin{aligned}\sum_{i=1}^{N} \frac{\lambda}{2} \bigg| \int_0^\ell \int_0^\ell m_\gamma^{1,\dots,n}(r,i) J_{\gamma^{-1/2}}(r,r')[m_\gamma^{1,\dots,n}(r',i+1) + m_\gamma^{1,\dots,n}(r',i-1)]\,dr\,dr' \\ - \int_0^\ell m(r,i)[m(r,i+1) + m(r,i-1)]\,dr \bigg|.\end{aligned}$$

We only discuss the term $m(\cdot,i)m(\cdot,i+1)$ because for the other term the proof is analogous. We sum and subtract for each $i$ the term $m_\gamma^{1,\dots,n}(r, i+1)$; then

$$\left|\int_0^\ell m_\gamma^{1,\dots,n}(r,i)\left[\int_0^\ell J_{\gamma^{-1/2}}(r,r')m_\gamma^{1,\dots,n}(r',i+1)\,dr' \pm m_\gamma^{1,\dots,n}(r,i+1)\right]dr \right.$$
$$\left. - \int_0^\ell m(r,i)m(r,i+1)\,dr\right|.$$

We split the first integral in the sum of integrals $\sum_{k=1}^n \int_{\bar{\bar{\Lambda}}_{k,i}} + \int_{\bar{\bar{E}}_k}$ where

$$\bar{\bar{\Lambda}}_{k,i} = (k\eta + 2\gamma^{1/2}(1+\gamma^{-\varepsilon}),\ (k+1)\eta - 2\gamma^{1/2}(1+\gamma^{-\varepsilon}))$$

and $\bar{\bar{E}}_k = \Lambda_{k,i} \setminus \bar{\bar{\Lambda}}_{k,i}$. If $r \in \bar{\bar{\Lambda}}_{k,i}$, we have that $\int_{\bar{\Lambda}_{k,i}} J_{\gamma^{-1/2}}(r,r')\,dr' = 1$; then

$$\sum_{k=1}^n \left|\int_{\bar{\bar{\Lambda}}_{k,i}} m_\gamma^{1,\dots,n}(r,i) \int_{\bar{\Lambda}_{k,i}} J_{\gamma^{-1/2}}(r,r')(m_\gamma^{1,\dots,n}(r',i+1) - m_\gamma^{1,\dots,n}(r,i+1))\,dr'\,dr\right|$$
$$\le \sum_{k=1}^n \int_{\bar{\bar{\Lambda}}_{k,i}} \int_{\bar{\Lambda}_{k,i}} J_{\gamma^{-1/2}}(r,r')|m_\gamma^{1,\dots,n}(r',i+1) - m_\gamma^{1,\dots,n}(r,i+1)|\,dr'\,dr$$
$$\le \sum_{k=1}^n |\bar{\bar{\Lambda}}_{k,i}|\eta c$$

because $m_\gamma^{1,\dots,n}$ is almost constant in $\bar{\Lambda}_{k,i}$. While integrating over $\bar{\bar{E}}_k$,

$$\sum_{k=1}^n \left|\int_{\bar{\bar{E}}_k} m_\gamma^{1,\dots,n}(r,i) \int_{\Lambda_{k,i}} J_{\gamma^{-1/2}}(r,r')(m_\gamma^{1,\dots,n}(r',i+1) - m_\gamma^{1,\dots,n}(r,i+1))\,dr'\,dr\right|$$
$$\le \sum_{k=1}^n \int_{\bar{\bar{E}}_k} \int_{\overline{\overline{E^\star}}_k} J_{\gamma^{-1/2}}(r,r')|m_\gamma^{1,\dots,n}(r',i+1) - m_\gamma^{1,\dots,n}(r,i+1)|\,dr'\,dr$$
$$\le c\sum_{k=1}^n |\bar{\bar{E}}_k||\overline{\overline{E^\star}}_k|\gamma^{-1/2}\|J\|_\infty$$

with $c > 0$ a constant. The size of $|\overline{\overline{E^\star}}_k|$ is of the same order as $|\bar{\bar{E}}_k| + 2\gamma^{1/2}$. In the end the term

$$\sum_{i=1}^N \frac{\lambda}{2}\left|\int_0^\ell (m_\gamma^{1,\dots,n}(r,i)m_\gamma^{1,\dots,n}(r,i+1) - m(r,i)m(r,i+1))\,dr\right|$$

can be estimated using Lemma 6.1. All the other terms that we did not consider can be estimated in the same way.

For the *third term*, we consider $\liminf_\gamma F_{\beta,\gamma,\lambda}(m_\gamma^{1,\dots,n})$. Then

$$\begin{aligned}\liminf_{\gamma\to 0}\sum_{i=1}^{N}-\frac{1}{\beta}\int_0^{\ell} I(m_\gamma^{1,\dots,n})(r,i)\,dr\\ &=\liminf_{\gamma\to 0}\sum_{i=1}^{N}-\frac{1}{\beta}\sum_{k=1}^{n}\int_{\Lambda_{k,i}} I(m_\gamma^{1,\dots,n}(r,i))\,dr\\ &\geq\liminf_{\gamma\to 0}\sum_{i=1}^{N}-\frac{1}{\beta}\sum_{k=1}^{n}|\Lambda_{k,i}|I\biggl(⨍_{\Lambda_{k,i}} m_\gamma^{1,\dots,n}(r,i)\,dr\biggr)\\ &=\liminf_{\gamma\to 0}\sum_{i=1}^{N}-\frac{1}{\beta}\int_0^{\ell} I\biggl(⨍_{\Lambda_{k(r),i}} m_\gamma^{1,\dots,n}(r',i)\,dr'\biggr)\,dr\end{aligned}$$

by Jensen's inequality. We write the sum over $k$ as an integral over $r$ observing that the function $I$ is constant for all $r$ in the set $\Lambda_{k,i}$. Moreover, there exists a subsequence $\{m_{\gamma_j}^{1,\dots,n}\}$ that achieves the infimum limit:

$$\begin{aligned}\liminf_{\gamma\to 0}\sum_{i=1}^{N}-\frac{1}{\beta}\int_0^{\ell} I\biggl(⨍_{\Lambda_{k(r),i}} m_\gamma^{1,\dots,n}(r',i)\,dr'\biggr)\,dr\\ &=\lim_{\gamma_j\to 0}\sum_{i=1}^{N}-\frac{1}{\beta}\int_0^{\ell} I\biggl(⨍_{\Lambda_{k_j(r),i}} m_{\gamma_j}^{1,\dots,n}(r',i)\,dr'\biggr)\,dr.\end{aligned}$$

Let $\tilde m_\gamma(r,i)=⨍_{\Lambda_{k_j(r),i}} m_{\gamma_j}^{1,\dots,n}(r',i)\,dr'$; then by Lemma 6.1 $\tilde m_\gamma \xrightarrow{L^1} m$ and there exists a subsubsequence, which we denote again $\{m_{\gamma_j}^{1,\dots,n}\}$, that converges to $m$ almost everywhere. Then by the dominated convergence theorem

$$\lim_{\gamma\to 0}-\frac{1}{\beta}\int_0^{\ell} I\biggl(⨍_{\Lambda_{k_j(r),i}} m_{\gamma_j}^{1,\dots,n}(r',i)\,dr'\biggr)\,dr=-\frac{1}{\beta}\int_0^{\ell} I(m(r,i))\,dr$$

and

$$\begin{aligned}\liminf_{\gamma\to 0}F_{\beta,\gamma,\lambda}(m_\gamma)&\geq\liminf_{\gamma\to 0}F_{\beta,\gamma,\lambda}(m_\gamma^{1,\dots,n})\\ &=\lim_{\gamma_j\to 0}F_{\beta,\gamma,\lambda}(m_{\gamma_j}^{1,\dots,n})\\ &=F_{\beta,\lambda}(m).\end{aligned}$$

Now we prove the *upper bound*. There exists a sequence $\{m_\gamma\}$ such that $m_\gamma \xrightarrow{\star} m$ and

$$\lim_{\gamma\to 0}F_{\beta,\gamma,\lambda}(m_\gamma)=F_{\beta,\lambda}(m).$$

We take $m_\gamma = s = \fint_{\Lambda_{k,i}} m(r,i)\,dr$ on $\Lambda_{k,i} \subset [0,\ell]$ for all $k$. Then from the dominated convergence theorem

$$\lim_{\gamma\to 0} |F_{\beta,\gamma,\lambda}(m_\gamma) - F_{\beta,\lambda}(m)| = 0.$$

And $F_{\beta,\lambda}(m)$ is the $\Gamma$-limit of $F_{\beta,\gamma,\lambda}(m_\gamma)$. □

## Appendix A: Proofs of Proposition 3.1 and Theorem 2.4

*Proof of Proposition 3.1.* The proof follows the guidelines of Section 4.2.2 of [Presutti 2009] taking care of the two different scales of interaction $\gamma^{-1}$ and $\gamma^{-1/2}$.

We define

$$U_{\gamma,\lambda}(\bar m) = \bar F_{\beta,\gamma,\lambda}(\bar m) + \sum_{i=1}^{N} \frac{1}{\beta} \int_0^{\gamma^{-1}\ell} I(\bar m(r,i))\,dr$$

where $\bar F_{\beta,\gamma,\lambda}$ is defined as in (2-17). We want to estimate $|H_{\gamma,\lambda}(\sigma) - U_{\gamma,\lambda}(\sigma^{(\alpha)})|$, and we start taking $(x,i), (y,j) \in T_{L,N}$ and defining

$$\hat J_\gamma((x,i),(y,j)) = J_\gamma(x,y)\mathbb{1}_{i=j} + J_{\gamma^{1/2}}(x,y)\mathbb{1}_{i\neq j}.$$

Recall that for each point $(x,i)$ there is an integer $k$ such that $(x,i) \in C^{(\alpha)}_{k,i}$. Let

$$\hat J^{(\alpha)}_\gamma((x,i),(y,j)) = \int_{C^{(\alpha)}_{k,i}\times C^{(\alpha)}_{h,j}} \hat J_\gamma((r,i),(r',j))\,dr\,dr'.$$

We want to give a bound of $|\hat J_\gamma((x,i),(y,j)) - \hat J^{(\alpha)}_\gamma((x,i),(y,j))|$. We consider only the worst case, namely the vertical interaction, $i \neq j$. In this case

$$\begin{aligned} |J_{\gamma^{1/2}}(x,y) - \hat J^{(\alpha)}_\gamma((x,i),(y,j))| &\le \fint_{C^{(\alpha)}_{k,i}\times C^{(\alpha)}_{h,j}} |J_{\gamma^{1/2}}(x,y) - J_{\gamma^{1/2}}(r,r')|\,dr\,dr' \\ &\le c\gamma^{1-\alpha}\mathbb{1}_{|x-y|\le 2\gamma^{-1/2}}. \end{aligned}$$

Let $C$ and $C'$ be two elements in the partition $\mathscr{C}^{(\alpha)}$, and consider two points $(r,i) \in C$ and $(r',i') \in C'$. As in the previous estimate, we consider the worst case. If $i \neq j$, by the estimate above

$$\begin{aligned} \Bigg| \sum_{(x,i)\in C} \sum_{(y,j)\in C'} \mathbb{1}_{|(x,i)\neq(y,j)|} \hat J_\gamma((x,i),(y,j))\sigma(x,i)\sigma(y,j) \\ - \sum_{(x,i)\in C} \sum_{(y,j)\in C'} \mathbb{1}_{|(x,i)\neq(y,j)|} \hat J^{(\alpha)}_\gamma((x,i),(y,j))\sigma(x,i)\sigma(y,j) \Bigg| \\ \le c'|C|^2\gamma^{1-\alpha}\mathbb{1}_{|r-r'|\le 3\gamma^{-1/2}}. \end{aligned} \tag{A-1}$$

Then

$$|H_{\gamma,\lambda}(\sigma) - U_{\gamma,\lambda}(\sigma^{(\alpha)})| \le c'|T_{L,N}|\lambda\gamma^{1/2-\alpha}. \tag{A-2}$$

We prove (3-1) writing the definition of the partition function

$$\log(Z_{\beta,\gamma,\lambda}(\{\sigma^{(\alpha)} = \overline{m}\})) \le \beta\lambda\gamma^{1/2-\alpha}|T_{L,N}|c - \beta U_{\gamma,\lambda}(\overline{m}) + \log(\operatorname{card}\{\sigma^{(\alpha)} = \overline{m}\}).$$

We can observe that

$$\begin{aligned}\log(\operatorname{card}\{\sigma^{(\alpha)} = \overline{m}\}) &= \log\biggl(\prod_{C_{i,k}} \operatorname{card}\biggl\{\sigma \in \{-1, 1\}^{C_{k,i}} : \sum_{(x,i)\in C_{k,i}} \sigma(x, i) = \overline{m}(r, i)\gamma^{-\alpha} \text{ for all } (r, i)\biggr\}\biggr)\\ &= \log\biggl(\prod_{C_{k,i}} e^{|C_{k,i}|I_{C_{k,i}}(\overline{m}(r,i))}\biggr)\end{aligned}$$

and[4]

$$|I_{C_{k,i}}(\overline{m}(r, i)) - I(\overline{m}(r, i))| \le c\gamma^{\alpha} \log \gamma^{-\alpha}. \tag{A-3}$$

At the end collecting the previous inequalities we have

$$\log(Z_{\beta,\gamma,\lambda}(\{\sigma^{(\alpha)} = \overline{m}\})) \le -\beta \overline{F}_{\beta,\gamma,\lambda}(\overline{m}) + \beta c|T_{L,N}|(\lambda\gamma^{1/2-\alpha} + \gamma^{\alpha} \log(\gamma^{-\alpha})).$$

The inequality (3-2) is proved in a similar way, so the proposition is proved. □

*Proof of Theorem 2.4.* We start by introducing the following proposition.

**Proposition A.1** (Lebowitz–Penrose limit). *Let* $Z_{\beta,\gamma,\lambda} := Z_{\beta,\gamma,\lambda}(\mathcal{M}^{(\alpha')})$ *with* $\alpha' \in (0, \frac{1}{2})$; *then*

$$\lim_{\gamma\to 0} \frac{1}{\beta|T_{L,N}|} \log Z_{\beta,\gamma,\lambda} = p_{\beta,\lambda}$$

*where* $p_{\beta,\lambda} = \sup_{m\in[-1,+1]}\{-\phi_{\beta,\lambda}(m)\}$ *and*

$$\phi_{\beta,\lambda}(m) = -\frac{1+2\lambda}{2}m^2 - \frac{1}{\beta}I(m). \tag{A-4}$$

*Proof.* For the proof see Theorem 4.2.1.1 in [Presutti 2009]. □

We consider

$$\begin{aligned}\gamma \log \mu_{\beta,\gamma,\lambda}[\sigma \approx^{\alpha} m] &= \gamma \log\biggl(\frac{Z^{(\alpha)}_{\beta,\gamma,\lambda}(m)}{\sum_{m'\in M^{(\alpha)}} Z^{(\alpha)}_{\beta,\gamma,\lambda}(m')}\biggr)\\ &= \gamma \log(Z^{(\alpha)}_{\beta,\gamma,\lambda}(m)) - \gamma \log\biggl(\sum_{m'\in\mathcal{M}^{(\alpha)}} Z^{(\alpha)}_{\beta,\gamma,\lambda}(m')\biggr).\end{aligned}$$

[4] The definition of $I_{C_{k,i}}$ and the inequality (A-3) can be found in Appendix A of [Presutti 2009]

By Theorem 2.3 for $\alpha \in (0, 1)$

$$\lim_{\gamma\to 0} \gamma \log(Z^{(\alpha)}_{\beta,\gamma,\lambda}(m)) = -F_{\beta,\lambda}(m).$$

If $\alpha < \frac{1}{2}$ by Propositions 3.1 and A.1, we have that

$$\lim_{\gamma\to 0} -\frac{\gamma}{\beta} \log Z_{\beta,\gamma,\lambda} = \inf_{m'\in\mathcal{M}^{(\alpha)}} F_{\beta,\lambda}(m').$$

For $\alpha > \frac{1}{2}$, instead,

$$\begin{aligned}
-\gamma \log \sum_{m'\in\mathcal{M}^{(\alpha)}} Z^{(\alpha)}_{\beta,\gamma,\lambda}(m') &= -\gamma \log \sum_{m'\in\mathcal{M}^{(\alpha)}} \sum_{\sigma:\sigma\approx^{\alpha} m'} e^{-\beta H_{\gamma,\lambda}(\sigma)} \\
&= -\gamma \log \sum_{m'\in\mathcal{M}^{(\alpha)}} \sum_{m_{\alpha'}\in\mathcal{A}_{m'_\alpha}} \sum_{\sigma:\sigma\approx^{\alpha'} m_{\alpha'}} e^{-\beta H_{\gamma,\lambda}(\sigma)} \\
&= -\gamma \log(Z_{\beta,\gamma,\lambda})
\end{aligned}$$

observing that

$$\inf_{m'} F_{\beta,\lambda}(m') = \sup_{h\in[-1,+1]} \{-\phi_{\beta,\lambda}(h)\} \cdot \ell N = p_{\beta,\lambda}\ell N.$$

Then

$$\lim_{\gamma\to 0} \gamma \log \mu_{\beta,\lambda,\gamma}[\sigma \approx^{(\alpha)} m] = -(F_{\beta,\lambda}(m) - \inf_{m'} F_{\beta,\lambda}(m')). \qquad \square$$

## Appendix B: A counterexample

In this appendix we shall show that Theorem 2.3 cannot be extended to the case $\beta\lambda > 1$; indeed for the mesoscopic state $m \equiv 0$

$$\liminf_{\gamma\to 0} F^{(\alpha)}_{\beta,\gamma,\lambda}(0) < F_{\beta,\lambda}(0).$$

If $\alpha > \frac{1}{2}$ we can take a sequence $m_\alpha$ where $m_\alpha$ is equal, in the first half of each interval $D^{(\alpha)}_r$, to $m_{\beta\lambda}$ and in the second half to $-m_{\beta\lambda}$; we obtain $m^{(\alpha)}_\alpha \equiv 0$. Recalling Definition 5.1, we have that $m_\alpha \overset{\star}{\to} m \equiv 0$. By the definition of $F^{(\alpha)}_{\beta,\gamma,\lambda}$ and Proposition 3.2,

$$F^{(\alpha)}_{\beta,\gamma,\lambda}(0) \le \frac{1}{\gamma^{-1}} \overline{F}_{\beta,\gamma,\lambda}(\overline{m}_\alpha) + \epsilon(\gamma, \lambda).$$

Now we observe that

$$\begin{aligned}\overline{F}_{\beta,\gamma,\lambda}(\overline{m}_\alpha)=\frac{1}{4}\sum_{i=1}^{N}\int_0^{\gamma^{-1}\ell}\int_0^{\gamma^{-1}\ell} J_\gamma(r,r')[\overline{m}_\alpha(r,i)-\overline{m}_\alpha(r',i)]^2\,dr'\,dr\\ +\frac{\lambda}{4}\sum_{i=1}^{N}\int_0^{\gamma^{-1}\ell}\int_0^{\gamma^{-1}\ell} J_{\gamma^{1/2}}(r,r')\big([\overline{m}_\alpha(r,i)-\overline{m}_\alpha(r',i-1)]^2\\ +[\overline{m}_\alpha(r,i)-\overline{m}_\alpha(r',i+1)]^2\big)\,dr'\,dr+\sum_{i=1}^{N}\int_0^{\gamma^{-1}\ell}\phi_{\beta,\lambda}(\overline{m}_\alpha(r,i))\,dr\end{aligned}$$

where $\phi_{\beta,\lambda}$ as in (A-4). Since $\overline{m}_\alpha$ is the same on each line, we have

$$\begin{aligned}\overline{F}_{\beta,\gamma,\lambda}(\overline{m}_\alpha)=\frac{N}{4}\int_0^{\gamma^{-1}\ell}\int_{r-\gamma^{-1}}^{r+\gamma^{-1}} J_\gamma(r,r')[\overline{m}_\alpha(r,1)-\overline{m}_\alpha(r',1)]^2\,dr'\,dr\\ +\frac{\lambda N}{2}\int_0^{\gamma^{-1}\ell}\int_{r-\gamma^{-1/2}}^{r+\gamma^{-1/2}} J_{\gamma^{1/2}}(r,r')[\overline{m}_\alpha(r,1)-\overline{m}_\alpha(r',1)]^2\,dr'\,dr\\ +N\int_0^{\gamma^{-1}\ell}\phi_{\beta,\lambda}(\overline{m}_\alpha(r,1))\,dr.\end{aligned}$$

By the symmetry of $J$ and the definition of $\overline{m}_\alpha$ we have

$$\frac{1}{\gamma^{-1}}\overline{F}_{\beta,\gamma,\lambda}(\overline{m}_\alpha)\le\frac{N}{2}\ell m_{\beta\lambda}^2+N8\lambda\gamma^{\alpha-1/2}m_{\beta\lambda}^2-N\ell\frac{1+2\lambda}{2}\overline{m}_{\beta\lambda}^2-\frac{N\ell}{\beta}I(\overline{m}_{\beta\lambda}).$$

Then

$$\begin{aligned}\liminf_{\gamma\to 0}\frac{1}{\gamma^{-1}}\overline{F}_{\beta,\gamma,\lambda}(0)\le N\ell\bigg(-\lambda\overline{m}_{\beta\lambda}^2-\frac{I(\overline{m}_{\beta\lambda})}{\beta}\bigg)\\ <-N\ell\frac{I(0)}{\beta}=F_{\beta,\lambda}(0).\end{aligned}$$

## Acknowledgements

The authors wish to express their thanks to Professor Errico Presutti and Professor Anna De Masi for suggesting the problem and for many stimulating conversations.

## References

[Braides 2002] A. Braides, *Γ-convergence for beginners*, Oxford Lecture Series in Mathematics and its Applications **22**, Oxford University, 2002.

[Cassandro et al. 2016] M. Cassandro, M. Colangeli, and E. Presutti, "Highly anisotropic scaling limits", *J. Stat. Phys.* **162**:4 (2016), 997–1030.

[Colangeli et al. 2016] M. Colangeli, A. De Masi, and E. Presutti, "Latent heat and the Fourier law", *Phys. Lett.* **380**:20 (2016), 1710–1713.

[Colangeli et al. 2017] M. Colangeli, A. De Masi, and E. Presutti, "Particle models with self sustained current", *J. Stat. Phys.* **167**:5 (2017), 1081–1111.

[Ellis 2006] R. S. Ellis, *Entropy, large deviations, and statistical mechanics*, Springer, 2006.

[Fontes et al. 2014] L. R. Fontes, D. H. U. Marchetti, I. Merola, E. Presutti, and M. E. Vares, "Phase transitions in layered systems", *J. Stat. Phys.* **157**:3 (2014), 407–421.

[Fontes et al. 2015] L. R. Fontes, D. H. U. Marchetti, I. Merola, E. Presutti, and M. E. Vares, "Layered systems at the mean field critical temperature", *J. Stat. Phys.* **161**:1 (2015), 91–122.

[Funaki and Spohn 1997] T. Funaki and H. Spohn, "Motion by mean curvature from the Ginzburg–Landau $\nabla\phi$ interface model", *Comm. Math. Phys.* **185**:1 (1997), 1–36.

[Lebowitz and Penrose 1966] J. L. Lebowitz and O. Penrose, "Rigorous treatment of the van der Waals–Maxwell theory of the liquid-vapor transition", *J. Mathematical Phys.* **7** (1966), 98–113.

[Penrose and Lebowitz 1971] O. Penrose and J. L. Lebowitz, "Rigorous treatment of metastable states in the van der Waals–Maxwell theory", *J. Statist. Phys.* **3** (1971), 211–236.

[Presutti 2009] E. Presutti, *Scaling limits in statistical mechanics and microstructures in continuum mechanics*, Springer, 2009.

[Rudin 1987] W. Rudin, *Real and complex analysis*, 3rd ed., McGraw-Hill, 1987.

Received 11 Jul 2017. Revised 21 Jun 2018. Accepted 24 Jul 2018.

MICHELE ALEANDRI: michele.aleandri@gssi.it
*Gran Sasso Science Institute, L'Aquila, Italy*

VENANZIO DI GIULIO: venanzio.digiulio@graduate.univaq.it
*Università degli Studi dell'Aquila, L'Aquila, Italy*

MATHEMATICS AND MECHANICS OF COMPLEX SYSTEMS
Vol. 6, No. 4, 2018
dx.doi.org/10.2140/memocs.2018.6.293

# OPTIMAL ORTHOTROPY AND DENSITY DISTRIBUTION OF TWO-DIMENSIONAL STRUCTURES

NARINDRA RANAIVOMIARANA, FRANÇOIS-XAVIER IRISARRI,
DIMITRI BETTEBGHOR AND BORIS DESMORAT

This paper describes an optimization methodology giving simultaneously the optimal spatial material distribution and the optimal material orthotropy distribution in a two-dimensional space. The spatial material distribution is parametrized by a density variable that defines the presence or absence of material. A general orthotropic material is parametrized by the polar invariants of the elasticity tensor. The criterion is the compliance that measures the global structural stiffness. The numerical procedure iterates successively between local minimizations and finite element calculations. Thanks to the polar method, the local minimizations are solved explicitly providing analytical solutions. An optimization of a beam shows the effectiveness of the method in finding concurrently the optimal shape and the optimal material.

## 1. Introduction

Reducing cost and weight of structures is a permanent challenge for the aeronautics industry. In this scope, topology optimization is used for the mass minimization problem [Allaire and Delgado 2016]. It gives an ideal repartition of material considering, for instance, global stiffness or eigenfrequency of a structure under loads and boundary conditions. The optimal shape or layout of the structure is then obtained. Topology optimization is widely used for isotropic materials [Bendsøe and Sigmund 2003; Sigmund and Maute 2013] such as metallic ones for example, but it does not optimize the material behavior, e.g., the anisotropy. The mass of the structure can also be reduced by optimizing the material that composes it. Composite structure optimization [Ghiasi et al. 2009; 2010; Sørensen and Lund 2013; Peeters et al. 2015] is used to design the material at each point of the structure. For instance, the optimal layup of laminates is sought by changing the orientations of plies, the thickness, or the stacking sequence with heuristic [Irisarri et al. 2009] or

**Communicated by Francesco dell'Isola**.
The authors acknowledge STELIA Aerospace for funding this project.
*MSC2010:* 74P10.
*Keywords:* topology optimization, SIMP, polar method, distributed orthotropy, material design.

gradient-based methods [Sørensen and Lund 2013]. The composite optimization is generally done with a predefined shape of structure. Thus, topology optimization gives an optimal distribution of material [Rojas-Labanda and Stolpe 2015] without considering its optimal anisotropy and composite structure optimization [Ghiasi et al. 2009; 2010; Sørensen and Lund 2013; Peeters et al. 2015] gives an optimal anisotropy of the material without considering the optimal shape of the structure. Nonetheless, the shape and the material of the structure are closely related. To obtain an ideal structure, it is necessary to optimize the structure by considering the optimal spatial material distribution and the optimal material anisotropy distribution at the same time.

Rion and Bruyneel [2007] treat topology optimization of orthotropic material by considering fiber orientations in the optimization. The determination of the boundaries of the structure combined with that of optimal fiber path is treated in [Peeters et al. 2015], where the stiffness is parametrized by lamination parameters. Allaire and Delgado [2016] optimize laminated composite plates where the shape of each layer is determined concurrently with the stacking sequence. In this paper, we present an optimization methodology giving simultaneously the optimal shape and the optimal orthotropy distribution of the structure. The optimization is made on a general homogenized orthotropic material.

Parametrization of the shape and the anisotropy is necessary. First we choose the density method to parametrize the shape of the structure. The density variable determines at each point of the structure whether there is material or a void. The anisotropy of the material is characterized by its elasticity tensor. As we work on a general orthotropic material, we consider the homogenized elasticity tensor defined in a thermodynamically admissible domain. The elasticity tensor can be described by nine Cartesian coefficients. Since the material orthotropy varies through the structure, one should define a general frame to express the elasticity tensor. However, the use of Cartesian representation is cumbersome when changing frame. The polar method, introduced by Verchery [1982], uses invariants by change of frame to describe the elasticity tensor. Thanks to its simplicity, change of frame is done by changing angles. We choose the polar invariants of the elasticity tensor as a design variable.

The criterion in structural optimization may be for instance eigenfrequency, buckling, or compliance. In this work, the stiffness of the structure which is measured by the external work (compliance) is maximized. The optimization problem, which is based on variational methods similar to those that are used in continuum mechanics [Boutin et al. 2017; Andreaus et al. 2016], is equivalent to minimizing the compliance. Convex approximation methods such as MMA and GCMMA ((globally convergent) method of moving asymptotes) [Svanberg 1987; Zillober 1993] and descent algorithm methods such as SQP (sequential quadratic programming) [Arora and Belegundu 1984; Schittkowski 1985] and IPOPT (interior point optimizer) [Wächter and Biegler 2006] need the evaluation of the objective function as well as its gradient. The optimality criteria method computes the optimal values of design variables by expressing optimality conditions. Therefore, the optimality criteria method is less expensive than the methods above in term of numerical cost. This is the reason why a method similar to optimality criteria is used in this work. The algorithm used to solve the numerical problem is the alternate directions algorithm [Allaire and Kohn 1993]. One iterates between local minimizations with respect to the design variables and global minimizations corresponding to finite element calculations. Numerical results show the effectiveness of the method.

## 2. Problem formulation: simultaneous optimization of the material density and anisotropy

***Parametrization of the distributed material density and anisotropy.*** The shape of the structure is parametrized by a density field variable $\rho(x)$. This density variable defines at each point $x$ of the structure whether there is a material ($\rho(x) = 1$) or a void ($\rho(x) = 0$). Here $\rho(x)$ takes any value in $[\rho_{\min}, 1]$ while, in order to avoid singularity of the elasticity tensors, the lowest admissible value $\rho_{\min}$ is generally set to $10^{-3}$. Allowing $\rho(x)$ to be valued in the interval $[\rho_{\min}, 1]$ involves intermediate densities appearing in the optimum topologies. These intermediate densities involve gray areas that are difficult to interpret because they correspond to a mixture of void and material. To suppress gray areas, the density $\rho(x)$ is forced to tend to either $\rho_{\min}$ or 1. The so-called SIMP method (solid isotropic material with penalization) [Bendsøe 1989] is used. This method uses an exponent $p \geq 2$ in order to interpolate the density $\rho(x)$. Optimized stiffness tensor $\underline{\underline{C}}(x)$ and compliance tensor $\underline{\underline{S}}(x)$ are expressed as functions of the considered material stiffness tensor $\underline{\underline{C_0}}(x)$ and compliance tensor $\underline{\underline{S_0}}(x)$:

$$\underline{\underline{C}}(x) = \rho(x)^p \underline{\underline{C_0}} \quad \Longleftrightarrow \quad \underline{\underline{S}}(x) = \frac{1}{\rho(x)^p} \underline{\underline{S_0}}. \tag{2-1}$$

The elasticity tensor defines the stiffness properties of the anisotropic material. In the present work, spatial variations of the material anisotropy are allowed. A parametrization that allows one to express the elasticity tensor in a general frame in a simple way is necessary. Change of frame is cumbersome using the Cartesian representation. The polar method permits one to write the elasticity tensor with its intrinsic properties using tensor invariants. By doing so, changing frame becomes simple as one needs only to rotate an angle with respect to the frame. Thus, we choose to express the stiffness tensor with its polar invariants for an orthotropic material under assumption of plane stress. As the out-of-plane terms of the stress tensor vanish, the relation between the stress tensor and the strain tensor in the considered plane can be expressed only with the in-plane terms by introducing the reduced stiffness tensor $\underline{Q}$. Equations 2-2 show the relation between the polar components ($T_0$, $T_1$, $R_0$, $R_1$, $\Phi_0$, and $\Phi_1$) and the Cartesian ones of the reduced stiffness tensor $\underline{Q}$ [Julien 2010; Vincenti and Desmorat 2011]:

$$\begin{aligned}
Q_{1111} &= \phantom{-}T_0 + 2T_1 + R_0 \cos 4\Phi_0 + 4R_1 \cos 2\Phi_1,\\
Q_{1122} &= -T_0 + 2T_1 - R_0 \cos 4\Phi_0,\\
Q_{1112} &= \phantom{-T_0 + 2T_1 +{}} R_0 \sin 4\Phi_0 + 2R_1 \sin 2\Phi_1,\\
Q_{2222} &= \phantom{-}T_0 + 2T_1 + R_0 \cos 4\Phi_0 - 4R_1 \cos 2\Phi_1,\\
Q_{2212} &= \phantom{-T_0 + 2T_1} - R_0 \sin 4\Phi_0 + 2R_1 \sin 2\Phi_1,\\
Q_{1212} &= \phantom{-}T_0 \phantom{+ 2T_1} - R_0 \cos \Phi_0.
\end{aligned} \tag{2-2}$$

Each Cartesian component of the reduced stiffness tensor is expressed with isotropic terms $T_0$, $T_1$ that do not depend on the orientation of the material and anisotropic terms $R_0 e^{4i\Phi_0}$, $R_1 e^{2i\Phi_1}$ that depend on the orientations $\Phi_0$, $\Phi_1$ of the material. The change of frame is done by changing these angles. The polar invariants are the moduli $T_0$, $T_1$, $R_0$, $R_1$ and the angle $\Phi_0 - \Phi_1$. The isotropic parts do not influence the anisotropy of the material; thus, $T_0$, $T_1$ are supposed to remain constant (in composite laminated plates made of identical unidirectional layers (with the same material and same thickness in each layer), the homogenized isotropic part $T_0$, $T_1$ of the laminate is equal to the isotropic part $T_0^{\mathrm{EL}}$, $T_1^{\mathrm{EL}}$ of the elementary layer [Jibawy et al. 2011]). The material optimization is performed with respect to the anisotropic parts $R_0$, $R_1$, $\Phi_1$.

Figure 1 shows the decomposition of the reduced stress tensor's first Cartesian component $Q_{1111}$ for a composite made of long and straight carbon fibers in an epoxy matrix ($E_{\mathrm{L}} = 112\,\mathrm{GPa}$, $E_{\mathrm{T}} = 8.2\,\mathrm{GPa}$, $G_{\mathrm{LT}} = 4.5\,\mathrm{GPa}$, and $\nu_{\mathrm{LT}} = 0.3\,\mathrm{GPa}$). The stiffness is expressed as the sum of terms that do not depend on the material orientation, $T_0$ and $T_1$, and terms that depend on the material orientation, $R_0$ and $R_1$. The $R_0$ and $R_1$ terms can take negative values (dashed lines) due to the cosine function (see (2-2)) and are $\frac{\pi}{4}$- and $\frac{\pi}{2}$-periodic, respectively. The material orientation is equal to 0°. The apparent stiffness $Q_{1111}$ is maximized at 0° as the $R_0$ and $R_1$ terms are both positive in this direction. It is minimized at 45° because the $R_0$ and $R_1$ are respectively negative and null. When $R_1$ vanishes, there are only $\frac{\pi}{4}$-periodic terms: the material is square symmetric.

***Optimization constraint: maximum volume and thermodynamical admissibility of the material.*** The optimization constraints are written in terms of the total volume amount of the structure and of the anisotropic part of the polar invariants by

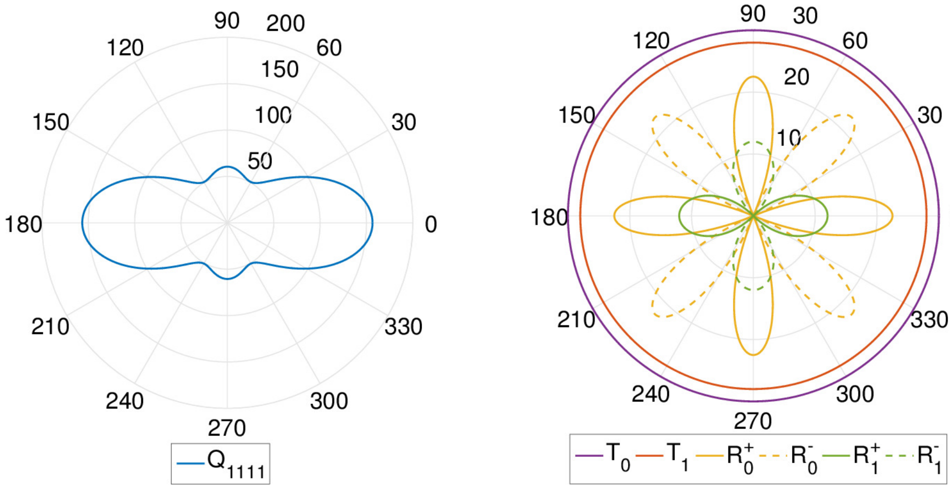


**Figure 1.** Left: representation of the first Cartesian component $Q_{1111}$ of the reduced stress tensor $\underline{\underline{Q}}$, in any orientation. Right: its decomposition into a sum of polar invariant terms $T_0$, $T_1$, $R_0$, $R_1$.

expressing their bounds. During the optimization, a target volume $V_0$ is defined for the structure. The volume $V$ is equal to the material density $\rho(x)$ integrated in the domain $\Omega$. At each step of the optimization, the volume must satisfy the optimization constraint

$$V = \int_{\Omega} \rho(x)\, dx = V_0. \tag{2-3}$$

The material to be designed is imposed to be orthotropic. For an orthotropic material,

$$\Phi_0 - \Phi_1 = K\frac{\Pi}{4} \quad \text{with } K = 0, 1. \tag{2-4}$$

The orthotropic material used in this paper is taken to be as general as possible: the optimized orthotropic material is thermodynamically admissible, which means that the stiffness tensor is positive definite. The analytical bounds of the polar invariants are [Vannucci 2005]

$$\begin{cases} T_0 > 0, \\ T_1 > 0, \\ T_0 > R_0, \\ T_0 T_1 > R_1^2, \\ T_1(T_0^2 - R_0^2) > 2R_1^2(T_0 - R_0 \cos 4(\Phi_0 - \Phi_1)). \end{cases} \tag{2-5}$$

***Double minimization of the complementary energy.*** In topology optimization, criteria such as buckling, frequency, or compliance may be considered; see for instance [Deaton and Grandhi 2014]. In this paper, we aim at maximizing the global structural stiffness measured by the compliance which is the external work. The criterion is written as

$$\text{Criterion} = \int_{\Omega} \boldsymbol{f} \cdot \boldsymbol{u} \, dV + \int_{\Gamma_1} \boldsymbol{F} \cdot \boldsymbol{u} \, dS. \tag{2-6}$$

The domain $\Omega$ is split into two boundaries: $\Gamma_0$ where a zero displacement is imposed and $\Gamma_1$ where a surface load $\boldsymbol{F}$ is applied. Then $\boldsymbol{f}$ is the volume load and $\boldsymbol{u}$ the displacement vector. The more the structure is rigid, the lower is the external work. Thus, maximizing the global structural stiffness means minimizing the compliance. Moreover, the compliance is equal to double the complementary energy. The optimization is made with respect to the density and the anisotropic part of the stiffness tensor polar invariants:

$$\min_{\{\rho, R_0, R_1, \Phi_1\}} \int_{\Omega} \boldsymbol{f} \cdot \boldsymbol{u} \, dV + \int_{\Gamma_1} \boldsymbol{F} \cdot \boldsymbol{u} \, dS = \min_{\{\rho, R_0, R_1, \Phi_1\}} \int_{\Omega} \underline{\sigma} : \underline{\underline{C}}^{-1} : \underline{\sigma} \, dV. \tag{2-7}$$

The complementary energy theorem claims that the complementary energy can be written as the minimization of a positive quantity with respect to the statically admissible (SA) stress field $\underline{\tau}$:

$$\int_{\Omega} \underline{\sigma} : \underline{\underline{C}}^{-1} : \underline{\sigma} \, dV = \min_{\underline{\tau} \text{ SA}} \int_{\Omega} \underline{\tau} : \underline{\underline{C}}^{-1} : \underline{\tau} \, dV. \tag{2-8}$$

The stress field $\underline{\tau}$ satisfies the elasticity problem (P), with assumption of small strains and small displacements:

$$\begin{cases} \operatorname{div} \underline{\tau} + \boldsymbol{f} = 0 & \text{in } \Omega, \\ \underline{\tau} \cdot \boldsymbol{n} = \boldsymbol{F} & \text{on } \Gamma_1, \\ \underline{\tau} = \underline{\underline{C}} : \underline{\epsilon}(\boldsymbol{u}) & \text{in } \Omega, \\ \boldsymbol{u} = 0 & \text{on } \Gamma_0, \end{cases} \tag{P}$$

where $\underline{\epsilon}(\boldsymbol{u}) = \frac{1}{2}(\underline{\nabla \boldsymbol{u}} + \underline{\nabla \boldsymbol{u}^T})$ is the strain tensor. By replacing the expression of the complementary energy in (2-7), the optimization problem is written as a double minimization with respect to the design variables $\{\rho, R_0, R_1, \Phi_1\}$ and to the stress field $\underline{\tau}$. The density variable is subject to a maximal volume constraint, and the polar invariants of the stiffness tensor are constrained by thermodynamic bounds:

$$\min_{\{\rho, R_0, R_1, \Phi_1\}} \min_{\underline{\tau} \text{ SA}} \int_{\Omega} \underline{\tau} : \underline{\underline{C}}^{-1} : \underline{\tau} dV \tag{2-9}$$

with

$$\begin{cases} \int_{\Omega} \rho(x) \, dx = V_0, \\ T_0 > R_0, \\ T_0 T_1 > R_1^2, \\ T_1(T_0^2 - R_0^2) > 2R_1^2(T_0 - R_0 \cos 4(\Phi_0 - \Phi_1)), \\ \Phi_0 - \Phi_1 = K(\Pi/4), \quad K = 0, 1. \end{cases} \tag{C}$$

## 3. Complementary energy minimization using the alternate direction algorithm

***Local minimizations of the complementary energy.*** Since the design variables $\{\rho, R_0, R_1, \Phi_1\}$ are subject only to algebraic constraints, the minimization with respect to them can be put inside the integral:

$$\min_{\underset{\sim}{\tau}\ \mathrm{SA}} \int_\Omega \min_{\{\rho, R_0, R_1, \Phi_1\}} \underset{\sim}{\tau} : \underset{\approx}{C}^{-1} : \underset{\sim}{\tau}\, dV \quad \text{with (C)}. \tag{3-1}$$

The minimization of the complementary energy with respect to the design variables is solved locally in each point of the domain, for a fixed stress state. Since the density variable $\rho(x)$ and the anisotropy variables $\{R_0, R_1, \Phi_1\}$ are independent, the minimization is split into two steps. First the complementary energy is minimized with respect to the anisotropy variables, taking into account the algebraic constraints related to thermodynamic bounds. Second, the minimization with respect to the density variable is performed.

The complementary energy can be written as a simple function of the polar invariants of the stiffness tensor and the stress tensor. Calculating its derivative is then straightforward. Hence, the minimization of the complementary energy with respect to the anisotropy variables is done analytically. The optimal values of $\{R_0, R_1, \Phi_1\}$ depending on the stress field are given in [Julien 2010] and are shown in Table 1, introducing the ratio $R/|T|$ where $R$ and $T$ are the spherical and deviatoric parts of the stress tensor, respectively. The optimal orthotropic material orientation is in the same direction as the principal direction of the stress tensor with maximal absolute value. The optimal values of polar invariants $R_0$ and $R_1$ depend on the ratio $R/|T|$.

The volume constraint is taken into account in the minimization step with respect to the density variable $\rho(x)$ through the introduction of a Lagrangian multiplier $k$:

$$\min_{\rho} \frac{1}{\rho(x)^p} \underset{\sim}{\tau} : \underset{\approx}{C}^{-1}(R_0^{\mathrm{opt}}, R_1^{\mathrm{opt}}, \Phi_1^{\mathrm{opt}}) : \underset{\sim}{\tau} + k\rho(x). \tag{3-2}$$

<table>
<tr><td>$X = R/|T|$</td><td colspan="3">$0$     $\sqrt{T_0/(2T_1)}$     $\sqrt{T_0/T_1}$     $+\infty$</td></tr>
<tr><td>$\Phi_1^{\mathrm{opt}}$</td><td colspan="3">$\mathrm{Dir}\{\max(|\sigma_{\mathrm{I}}|, |\sigma_{\mathrm{II}}|)\}$</td></tr>
<tr><td>$R_0^{\mathrm{opt}}$</td><td>$0 \le R_0^{\mathrm{opt}} < T_0$</td><td>$2T_1X^2 - T_0 < R_0^{\mathrm{opt}} < T_0$</td><td>$T_0^-$</td></tr>
<tr><td>$R_1^{\mathrm{opt}}$</td><td colspan="2">$T_1X$</td><td>$T_0^-/X$</td></tr>
</table>

**Table 1.** Optimal values [Julien 2010, Table 3.8] of the polar invariants $\{R_0, R_1, \Phi_1\}$ depending on the stress field, in the case $\Phi_0 - \Phi_1 = K\frac{\pi}{4} = 0$ ($K = 0$).

The minimum of the local energy is attained by setting to zero the variation of (3-2) with respect to the density field, which yields

$$\rho(x)=\left(\frac{p\underline{\tau}:\underline{\underline{C}}^{-1}(R_0^{\text{opt}},R_1^{\text{opt}},\Phi_1^{\text{opt}}):\underline{\tau}}{k}\right)^{1/(p+1)}. \tag{3-3}$$

The value of $k$ is calculated so that the volume constraint is satisfied.

***Optimization algorithm.*** The double minimization is solved with a fixed point method by considering the optimality conditions. At each iteration of the optimization, the minimization with respect to the design variables $\{\rho, R_0, R_1, \Phi_1\}$ is first performed; then the minimization with respect to the stress field $\underline{\tau}$ is operated. The minimization with respect to the stress field $\underline{\tau}$ corresponds to a finite element analysis thanks to the complementary energy theorem. The minimizations are treated alternatively and separately. This method is an extension of the alternate direction algorithm introduced in [Allaire and Kohn 1993]. Thanks to the polar method, the local complementary energy is written with simple expressions. Hence, the local minimizations are solved analytically.

The advantage of the alternate direction algorithm is its simplicity and low numerical cost as the method iterates between local minimizations solved analytically and finite element calculations of stresses. The work in [Desmorat 2013] shows also the convergence of the algorithm for a compliance minimization problem. The cost of the algorithm is directly related to the finite element calculation cost. Finally, the algorithm can take into account a large number of variables.

## 4. Numerical results

Numerical results are presented in this section to prove the efficiency of the method. The optimization is made for a two-dimensional orthotropic linear elastic material. A support beam from a civil aircraft produced by Messerschmidt-Bölkow-Blohm, called the MBB beam, is optimized here. The beam carries the floor in the fuselage of an Airbus airliner. Maximizing its rigidity has become a classical problem in topology optimization (see [Zhou and Rozvany 1991] for example). The design domain is a rectangle clamped with respect to the $x$-axis at the left side and with respect to the $y$-axis at the bottom of the right side (orange-colored dot in Figure 2). A load is applied on the top of the left side. The domain size is 40 mm × 20 mm discretized with a rectangular 80 × 40 mesh. The volume constraint is fixed at 50% of the total volume. The initial density is set to 1 in every element of the mesh. The initial material is an isotropic material where the values of $T_0$ and $T_1$ correspond to the isotropic part of a monolayered composite made of long and straight carbon fibers in an epoxy matrix: $T_0 = 26.88$ GPa and $T_1 = 24.74$ GPa.

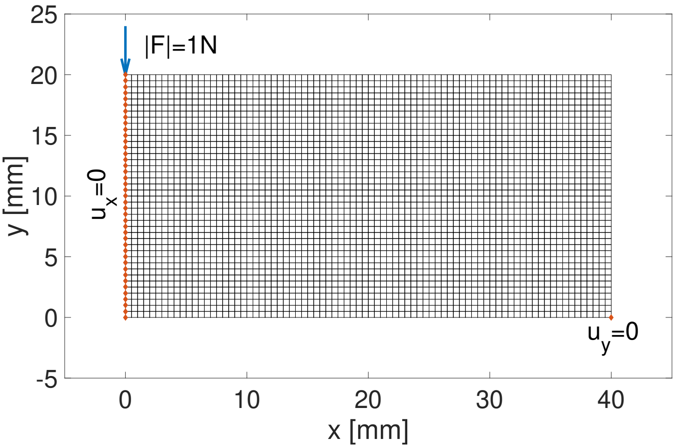


**Figure 2.** Boundary conditions for the MBB beam problem.

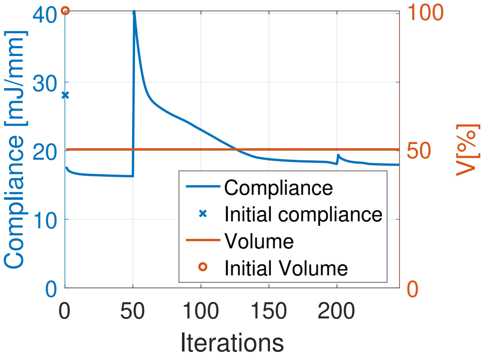


**Figure 3.** Compliance and volume with respect to optimization iterations.

***Convergence.*** The compliance and the volume are displayed as functions of the iterations in Figure 3. The strategy of penalizing the density is made in three steps during which the exponent $p$ in $\rho(x)^p$ is increased gradually. First, the convex problem corresponding to $p = 1$ is treated. The convexity of the problem when taking $p = 1$ is proved theoretically in [Allaire et al. 1997]. This means that the solution at the end of the iterations where $p = 1$ is a global minimum, making the solution independent of the initialization. Second, starting from this global minimum point, the solution is forced to be a $0/1$ layout by increasing $p$ to 3. Finally, $p$ is taken to be equal to 5 to suppress definitely intermediate density.

Except for the first iteration, the volume does not change through the iterations as it is constrained here to be equal to 50% of the total feasible volume. The compliance decreases at each of three steps ($p=1$, $p=3$, and $p=5$). At each step, convergence is reached when the variation of compliance between two consecutive

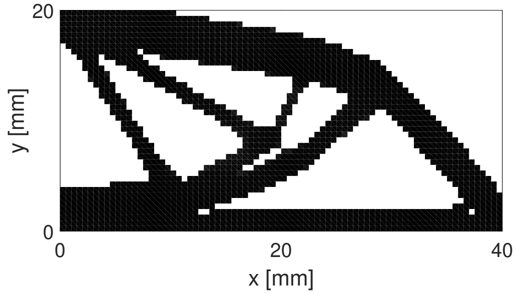


**Figure 4.** Optimal topology of the MBB beam with 50% volume amount.

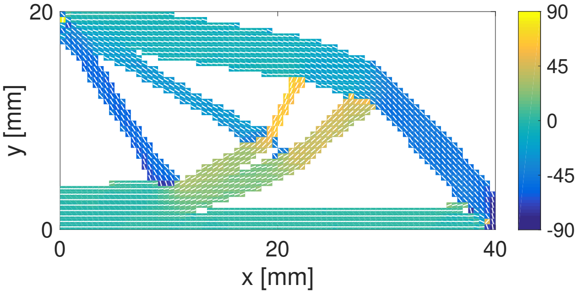


**Figure 5.** Optimal distribution of orthotropy direction.

iterations is less than 0.1% and the variation of the local densities is less than 0.01%. The compliance increases when the value of $p$ is increased because the structure becomes suddenly less stiff when the interpolation of intermediate density values is changed. We can observe that, at the end of the optimization, the compliance has converged.

***Optimal distribution of density.*** Figure 4 shows the optimal shape of the structure where black represents the presence of material and white its absence. The material is pictured when the density value is above 0.9. To avoid numerical problems such as checkerboard, a filter is used: the density of an element depends on the density of its neighbors so that there is no sharp discontinuity of the density in the structure. The neighbor elements that influence the considered element are defined by a radius filter. The filtering method used in this work is similar to the method of filtering sensitivities [Bendsøe and Sigmund 2003]. A structured mesh is used in this work. The filter radius permits us to suppress the checkerboard problem. For a given value of the radius, it has been observed numerically that the mesh dependency of the optimal topology seems to vanish when the element size is small enough compared to the filter radius. However, the optimal shape depends on values of the radius filter, which can be interpreted as a minimal bar width.

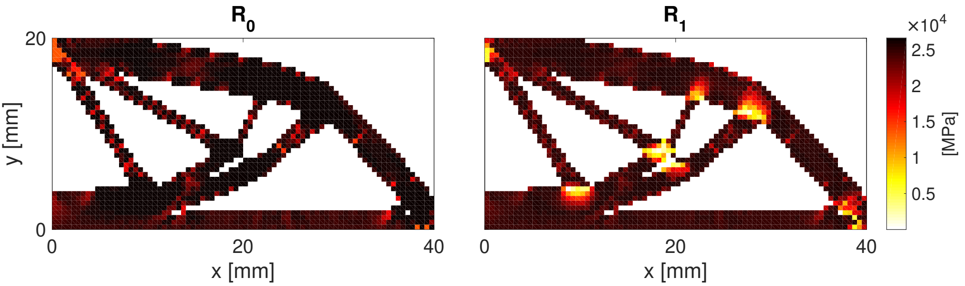


**Figure 6.** Optimal distribution of the stiffness tensor anisotropic polar invariants $R_0$ (left) and $R_1$ (right).

***Optimal distribution of orthotropy.*** In the optimal shape, the orthotropy is distributed: the material orthotropy changes continuously inside the structure. The optimal orthotropy direction $\Phi_1^{\text{opt}}$ is presented in Figure 5. It is aligned with the stress principal direction. The direction changes continuously throughout the structure as the stress field is continuous, except on the areas that are solicited in shearing, where a bar intersect another one. In these areas, the optimized material is square symmetric (i.e., $R_1 = 0$). The apparent stiffness, having the same value in $\Phi_1$ modulo $\pi/4$, is continuous in space in the optimal design. We illustrate the distribution of the moduli $R_0$ and $R_1$ in Figure 6. The $R_0$ values are set to be quasiconstant whereas the $R_1$ values vary from 0 to 25 GPa. We can see that in the areas where $R_1$ are minimum, the shearing is maximum. The optimal materials in these areas where $R_1 = 0$ are square symmetric materials, stiffened in two orthogonal directions. When $R_1$ is maximum, the optimal material is stiffened in one direction because it is solicited mostly in traction or compression.

## 5. Conclusion

The proposed methodology presented in this paper concurrently gives the spatial material distribution and the material anisotropy distribution by minimizing the compliance. The optimization strategy is based on an optimality criteria method in which one iterates successively and separately between local minimizations and finite element calculations. In order to avoid mesh size dependency, it could be of interest to develop such an optimization procedure with the use of some generalized continuum theories. Parametrizing the shape of the structure with a density variable and the anisotropy of the material with polar invariants allows for solving the local minimizations analytically. Thus, the computational cost of the algorithm corresponds to the finite element calculations. The method is straightforward to implement and gives coherent results from a mechanical viewpoint. Indeed, the optimal material where the structure is loaded in shear is square symmetric, because it has to be stiffened in two orthogonal directions. Areas loaded in traction or compression have an optimal material stiffened in one direction only. The presented optimization methodology is very promising when considering real composite material distribution, as the only change to be performed will be on the admissible set of polar parameters that should take into account the feasibility of the considered composite material.

## References

[Allaire and Delgado 2016] G. Allaire and G. Delgado, "Stacking sequence and shape optimization of laminated composite plates via a level-set method", *J. Mech. Phys. Solids* **97** (2016), 168–196.

[Allaire and Kohn 1993] G. Allaire and R. V. Kohn, "Optimal design for minimum weight and compliance in plane stress using extremal microstructures", *European J. Mech. A Solids* **12**:6 (1993), 839–878.

[Allaire et al. 1997] G. Allaire, E. Bonnetier, G. Francfort, and F. Jouve, "Shape optimization by the homogenization method", *Numer. Math.* **76**:1 (1997), 27–68.

[Andreaus et al. 2016] U. Andreaus, F. dell'Isola, I. Giorgio, L. Placidi, T. Lekszycki, and N. L. Rizzi, "Numerical simulations of classical problems in two-dimensional (non) linear second gradient elasticity", *Internat. J. Engrg. Sci.* **108** (2016), 34–50.

[Arora and Belegundu 1984] J. S. Arora and A. D. Belegundu, "Structural optimization by mathematical programming methods", *AIAA J.* **22**:6 (1984), 854–856.

[Bendsøe 1989] M. P. Bendsøe, "Optimal shape design as a material distribution problem", *Struct. Opt.* **1**:4 (1989), 193–202.

[Bendsøe and Sigmund 2003] M. P. Bendsøe and O. Sigmund, *Topology optimization: theory, methods and applications*, Springer, 2003.

[Boutin et al. 2017] C. Boutin, F. dell'Isola, I. Giorgio, and L. Placidi, "Linear pantographic sheets: asymptotic micro-macro models identification", *Math. Mech. Complex Syst.* **5**:2 (2017), 127–162.

[Deaton and Grandhi 2014] J. D. Deaton and R. V. Grandhi, "A survey of structural and multidisciplinary continuum topology optimization: post 2000", *Struct. Multidiscip. Optim.* **49**:1 (2014), 1–38.

[Desmorat 2013] B. Desmorat, "Structural rigidity optimization with an initial design dependent stress field: application to thermo-elastic stress loads", *Eur. J. Mech. A Solids* **37** (2013), 150–159.

[Ghiasi et al. 2009] H. Ghiasi, D. Pasini, and L. Lessard, "Optimum stacking sequence design of composite materials, I: Constant stiffness design", *Compos. Struct.* **90**:1 (2009), 1–11.

[Ghiasi et al. 2010] H. Ghiasi, K. Fayazbakhsh, D. Pasini, and L. Lessard, "Optimum stacking sequence design of composite materials, II: Variable stiffness design", *Compos. Struct.* **93**:1 (2010), 1–13.

[Irisarri et al. 2009] F.-X. Irisarri, D. H. Bassir, N. Carrere, and J.-F. Maire, "Multiobjective stacking sequence optimization for laminated composite structures", *Compos. Sci. Tech.* **69**:7–8 (2009), 983–990.

[Jibawy et al. 2011] A. Jibawy, C. Julien, B. Desmorat, A. Vincenti, and F. Léné, "Hierarchical structural optimization of laminated plates using polar representation", *Int. J. Solids Struct.* **48**:18 (2011), 2576–2584.

[Julien 2010] C. Julien, *Conception optimale de l'anisotropie dans les structures stratifiées à rigidité variable par la méthode polaire-génétique*, Ph.D. thesis, UPMC, 2010.

[Peeters et al. 2015] D. Peeters, D. van Baalen, and M. Abdallah, "Combining topology and lamination parameter optimisation", *Struct. Multidiscip. Optim.* **52**:1 (2015), 105–120.

[Rion and Bruyneel 2007] V. Rion and M. Bruyneel, "Topology optimization of membranes made of orthotropic material", pp. 107–120 in *Collection of papers from Professor Nguyen Dang Hung's former students*, edited by G. DeSaxcé and N. Moës, Vietnam National University, 2007.

[Rojas-Labanda and Stolpe 2015] S. Rojas-Labanda and M. Stolpe, "Benchmarking optimization solvers for structural topology optimization", *Struct. Multidiscip. Optim.* **52**:3 (2015), 527–547.

[Schittkowski 1985] K. Schittkowski, "Software for mathematical programming", pp. 383–451 in *Computational mathematical programming* (Bad Windsheim, Germany, 1984), edited by K. Schittkowski, NATO Adv. Sci. Inst. Ser. F Comput. Systems Sci. **15**, Springer, 1985.

[Sigmund and Maute 2013] O. Sigmund and K. Maute, "Topology optimization approaches", *Struct. Multidiscip. Optim.* **48**:6 (2013), 1031–1055.

[Sørensen and Lund 2013] S. N. Sørensen and E. Lund, "Topology and thickness optimization of laminated composites including manufacturing constraints", *Struct. Multidiscip. Optim.* **48**:2 (2013), 249–265.

[Svanberg 1987] K. Svanberg, "The method of moving asymptotes—a new method for structural optimization", *Internat. J. Numer. Methods Engrg.* **24**:2 (1987), 359–373.

[Vannucci 2005] P. Vannucci, "Plane anisotropy by the polar method", *Meccanica* **40**:4-6 (2005), 437–454.

[Verchery 1982] G. Verchery, "Les invariants des tenseurs d'ordre 4 du type de l'élasticité", pp. 93–104 in *Comportment méchanique des solides anisotropes* (Villard-de-Lans, France, 1979), edited by J.-P. Boehler, Colloques Int. Centre Nat. Recherche Sci. **295**, Springer, 1982.

[Vincenti and Desmorat 2011] A. Vincenti and B. Desmorat, "Optimal orthotropy for minimum elastic energy by the polar method", *J. Elasticity* **102**:1 (2011), 55–78.

[Wächter and Biegler 2006] A. Wächter and L. T. Biegler, "On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming", *Math. Program.* **106**:1, Ser. A (2006), 25–57.

[Zhou and Rozvany 1991] M. Zhou and G. I. N. Rozvany, "The COC algorithm, II: Topological, geometrical and generalized shape optimization", *Comp. Meth. Appl. Mech. Eng.* **89**:1–3 (1991), 309–336.

[Zillober 1993] C. Zillober, "A globally convergent version of the method of moving asymptotes", *Struct. Opt.* **6**:3 (1993), 166–174.

Received 12 Oct 2017. Revised 7 May 2018. Accepted 14 Jun 2018.

NARINDRA RANAIVOMIARANA: narindra.ranaivomiarana@onera.fr
*Centre de Châtillon, Office National d'Etudes et de Recherches Aérospatiales, Chatillon, France*

FRANÇOIS-XAVIER IRISARRI: francois-xavier.irisarri@onera.fr
*Centre de Châtillon, Office National d'Etudes et de Recherches Aérospatiales, Chatillon, France*

DIMITRI BETTEBGHOR: dimitri.bettebghor@onera.fr
*Centre de Châtillon, Office National d'Etudes et de Recherches Aérospatiales, Chatillon, France*

BORIS DESMORAT: boris.desmorat@sorbonne-universite.fr
*Institut Jean Le Rond d'Alembert, Sorbonne Université, Paris, France*

MATHEMATICS AND MECHANICS OF COMPLEX SYSTEMS
Vol. 6, No. 4, 2018
dx.doi.org/10.2140/memocs.2018.6.307

# A MULTIPHYSICS STIMULUS FOR CONTINUUM MECHANICS BONE REMODELING

DANIEL GEORGE, RACHELE ALLENA AND YVES RÉMOND

Bone remodelling is a complex phenomenon during which old and damage bone is removed and replaced with new one to ensure the physiological functions of the skeletal system. It involves many biological, mechanical, chemical processes at different scales. The objective of the present work is to predict the kinetics of bone density evolution by taking into account both the mechanical and the biological frameworks. In order to do so, we propose a new computational model in which the global stimulus triggering bone remodelling is the result of the contribution of a mechanical (i.e. external loads and consequent strain energy), a cellular (i.e. osteoblasts and osteoclasts activities) and a molecular (i.e. oxygen and glucose supply) stimulus. The evolution of the bone density depends on the overall behaviour of the global stimulus. More specifically, when the global stimulus is positive, bone synthesis occurs, whereas when the global stimulus is negative, resorption takes place. Although the theoretical model has been applied on a very simple two-dimensional geometry, the final results provide new insights on the influence of each stimulus on the bone remodelling process. In particular, we confirm that mechanics plays a critical role and affects the kinetics of bone reconstruction, but it highly depends on the biological events and the distribution of bone density.

## 1. Introduction

Bone is a continually renewed material [Frost 1987]. Trying to model its evolution has been going on for a long time since the early works of Wolff [Cowin 1986]. Every year, 5% of trabecular bone and 20% of cortical bone is renewed under applied external mechanical loads and the prediction of bone remodeling, or bone density evolution, using numerical models requires the use of appropriate theories accounting for the specific mechanophysiological phenomena occurring within the bone microstructure. Many studies have followed since; see for example [Beaupré et al. 1990; Turner 1998; Pivonka et al. 2008; Pivonka and Komarova 2010]. Recently, a number of models have tried to combine multiphysics and multiscales

**Communicated by Francesco dell'Isola**.
*MSC2010:* 65M99.
*Keywords:* bone remodeling, multiphysical stimulus, oxygen, glucose, cell motility.

theoretical numerical studies [Lekszycki 2002; Madeo et al. 2011; 2012; Lekszycki and dell'Isola 2012; Andreaus et al. 2014b; Giorgio et al. 2016; 2017; Scala et al. 2016; George et al. 2017b] to represent bone density evolution. Still, many difficulties remain in the precise understanding of the mechanotransduction processes [Lemaire et al. 2011; Sansalone et al. 2015] driving this evolution, without even accounting that in most cases, bone reconstruction also depends on initial healing stages of vascular growth together with nutrient supply [Bednarczyk and Lekszycki 2016; Lu and Lekszycki 2016].

Bone remodeling, being the result of numerous mechanobiological mechanisms, is often presented through a so-called mechanobiological stimulus, based on strain energy density, describing a variation from a state of equilibrium [Lekszycki 2002; Lekszycki and dell'Isola 2012; Scala et al. 2016]. However, for good prediction of bone remodeling, it is necessary, not only to account for the mechanical aspects, but also to account for other external sources such as biological, electrical, neurological,… involved in the process, that can be triggered by genetic or epigenetic factors, and allowing to simultaneously control their impact on the overall response of the system as well as their interactions. For these signals, the development of a thermodynamically consistent model [Martin et al. 2017] is required together with adequate homogenization procedures [Rémond et al. 2016]. The biology also needs to be adequately quantified (for example, the kinetics of bone resorption being 4 times more important than the kinetics of bone reconstruction; see [Burr and Allen 2013, pp. 85–86]) through specific multiscale theoretical models [Lemaire et al. 2006; 2010; 2015].

For example, in orthodontic bone remodeling, the applied mechanical forces on the teeth (ranging from 0.5 N to 2.5 N [Wagner et al. 2017]) lead to the alteration of the cell differentiation and activation due to oxygen percentage variation by the periodontal ligament being partially deformed. Hence, the variations in vascularization blood flow in the periodontal ligament and thus in the supply chain of nutrients and oxygen could be used to predict cell recruitment, proliferation and migration leading to the bone remodelling process.

In this work, a continuous theoretical numerical model is presented and used to predict bone kinetics reconstruction as a function of coupled mechanical and biological sources, of the corresponding constitutive laws, of their mutual interactions as well as of the kinetics of each process. The external sources used here to calculate the mechanobiological stimulus are: (i) the mechanical energy accounting for the mechanical loads sustained by the bone cells and triggering bone density evolution, (ii) the concentration of nutriments (oxygen and glucose) expressed as a function of the developed hydrostatic pressure, and (iii) the cells activity triggered by specific levels of oxygen and glucose concentration due to the applied mechanical load. The cells recruiting and migration is described via two diffusion equations [Allena and Maini 2014; Schmitt et al. 2015; Frame et al. 2017] and the bone density variation in time is calculated by the rates of bone synthesis and resorption respectively, depending on the positiveness of the defined coupled mechanobiological stimulus [George et al. 2017a].

## 2. Model development

**2.1.** ***Theory.*** Without specific external loading conditions, the bone is in a state of mechanobiological equilibrium (under gravity) in the so-called “lazy zone” where little remodeling occurs. When external mechanical load is applied, the system is perturbed and goes out of the “lazy zone”. The modified load conditions are at the origin of the creation of a coupled mechanobiological signal that will activate bone remodelling. We define this signal [George et al. 2017a] by introducing a Lagrangian configuration $B_L \subset \mathbb{R}^3$ [Madeo et al. 2011; 2012; Scala et al. 2016], and a suitably regular kinematical field $\chi(\boldsymbol{X}, t)$ that associates to any material point $\boldsymbol{X} \in B_L$ its current position $\boldsymbol{x}$ at time $t$. The image of the function $\chi$ gives at any time $t$, the current shape of the body also called Eulerian configuration. We also introduce the displacement $u(\boldsymbol{X}, t) = \chi(\boldsymbol{X}, t) - \boldsymbol{X}$, the transformation gradient $\boldsymbol{F} = \nabla\chi(\boldsymbol{X}, t)$, and the Green–Lagrange deformation tensor $\boldsymbol{E} = (\boldsymbol{F}^T \cdot \boldsymbol{F} - I)/2$. In the present work, only the linearized part $\varepsilon$ of $\boldsymbol{E}$ is considered.

Then the global stimulus variation $\Delta S$ is expressed on the Lagrangian configuration $B_L$ in the form

$$\Delta S(\boldsymbol{X}, t) = \prod_{i=1}^{n} \alpha_i S_i(\boldsymbol{X}, t), \tag{1}$$

where $t$ is the time, $n$ is the total number of external sources $S_i$ (i.e. mechanical, biological (cellular, nutrients, ...), electrical, ...) involved in the process and $\alpha_i$ are their weighting coefficients, triggered by genetic or epigenetic factors, allowing to simultaneously control their impact on the overall response of the system as well as their interactions.

In this work, we consider the following external sources: $S_{\text{mech}}$, which includes the applied mechanical load through the mechanical energy developed within the system to trigger the biological actions; $S_{\text{mol}}$, which coincides with glucose and oxygen supply necessary for cell survival and work contribution; $S_{\text{cell}}$, which corresponds to the osteoblasts and osteoclasts recruiting and migration.

(i) The mechanical stimulus $S_{\text{mech}}$ is expressed through the “standard” definition of the mechanical strain energy and accounts for the applied forces and loads sustained by bone cells. It is defined with

$$\alpha_{\mathrm{mech}}\, S_{\mathrm{mech}}(\boldsymbol{X}, t) = \alpha_{\mathrm{mech}} \int_{\Omega} U(\boldsymbol{X_0}, t)\, d(\boldsymbol{X_0}, t) \exp(-D_{\mathrm{mech}} \|\chi(\boldsymbol{X}) - \chi(\boldsymbol{X}_0)\|)\, d\boldsymbol{X}_0, \tag{2}$$

with $\Omega$ the domain of interest, $\alpha_{\mathrm{mech}}$ a weighting coefficient, $D_{\mathrm{mech}}$ the inverse of a characteristic distance accounting for the independent effect of the source, $U$ the strain energy density dependent on the Green–Lagrange deformation tensor $\boldsymbol{\varepsilon}$, and $d$ being a function of the bone mass density expressed as $d(\boldsymbol{X_0}, t) = \eta(\rho_{\mathrm{b}})/\rho_{\mathrm{b,max}}$ with $\eta \in [0, 1]$, where $\rho_{\mathrm{b}}$ is the bone density and $\rho_{\mathrm{b,max}}$ its maximum allowed value, being the density of compact bone (corresponding to minimum porosity).

(ii) The molecular stimulus $S_{\mathrm{mol}}$ is defined with

$$\alpha_{\mathrm{mol}}\, S_{\mathrm{mol}}(\boldsymbol{X}, t) = \alpha_{\mathrm{mol}} \int_{\Omega} (\alpha_{\mathrm{O}_2} c_{\mathrm{O}_2} + \alpha_{\mathrm{CHO}}\, c_{\mathrm{CHO}}) \exp(-D_{\mathrm{mol}} \|\chi(\boldsymbol{X}) - \chi(\boldsymbol{X}_0)\|)\, d\boldsymbol{X}_0, \tag{3}$$

with $D_{\mathrm{mol}}$ the inverse of a characteristic distance, $\alpha_{\mathrm{O}_2}$ and $\alpha_{\mathrm{CHO}}$ the weighting coefficients for $c_{\mathrm{O}_2}$ and $c_{\mathrm{CHO}}$, the concentrations of oxygen and glucose, satisfying two partial differential equations (PDEs) as a function of the hydrostatic pressure as follows:

$$\boldsymbol{D}_{\mathrm{O}_2} \frac{\partial c_{\mathrm{O}_2}}{\partial t} = 0, \tag{4}$$

and

$$\boldsymbol{D}_{\mathrm{CHO}} \frac{\partial c_{\mathrm{CHO}}}{\partial t} = 0, \tag{5}$$

where

$$\boldsymbol{D}_{\mathrm{O}_2} = \boldsymbol{D}_{\mathrm{CHO}} = \mathrm{Tr}(\varepsilon) + \phi(\varepsilon_I \boldsymbol{\theta}_I \otimes \boldsymbol{\theta}_I + \varepsilon_{II} \boldsymbol{\theta}_{II} \otimes \boldsymbol{\theta}_{II}), \tag{6}$$

with Tr the trace of a tensor, $\phi$ a scalar, $\varepsilon_I$ and $\varepsilon_{II}$ and $\boldsymbol{\theta}_I$ and $\boldsymbol{\theta}_{II}$ the principal strains and directions and $\otimes$ the tensor product. In (4) and (5), it is assumed that no external sources are present, only diffusion of the concentration through the geometry is present via a heterogeneous initial distribution.

(iii) The cellular stimulus $S_{\mathrm{cell}}$ defined by the osteoblasts and osteoclasts activity and triggered by specific levels of oxygen and glucose concentration together with the intensity of the mechanical force applied is given by

$$\alpha_{\mathrm{cell}} S_{\mathrm{cell}}(\boldsymbol{X}, t) = \alpha_{\mathrm{cell}} \int_{\Omega} (\alpha_{\mathrm{ob}}\, c_{\mathrm{ob}} - \alpha_{\mathrm{oc}}\, c_{\mathrm{oc}}) \exp(-D_{\mathrm{cell}} \|\chi(\boldsymbol{X}) - \chi(\boldsymbol{X}_0)\|)\, d\boldsymbol{X}_0, \tag{7}$$

where $D_{\mathrm{cell}}$ is the inverse of a characteristic distance, $\alpha_{\mathrm{ob}}$ and $\alpha_{\mathrm{oc}}$ are the weighting coefficients for the concentrations $c_{\mathrm{ob}}$ and $c_{\mathrm{oc}}$ of the osteoblasts and osteoclasts respectively evolving with respect to time via two diffusion-reaction equations [Allena and Maini 2014; Schmitt et al. 2015] as

$$\frac{\partial c_{\mathrm{ob}}}{\partial t} = (1-\rho_{\mathrm{b}})(\operatorname{div} \boldsymbol{D}_{\mathrm{ob}} \nabla c_{\mathrm{ob}} + \beta_{\mathrm{oc}} \operatorname{Tr}(\varepsilon) c_{\mathrm{oc}}), \tag{8}$$

$$\frac{\partial c_{\mathrm{oc}}}{\partial t} = (1-\rho_{\mathrm{b}})[\operatorname{div} \boldsymbol{D}_{\mathrm{oc}} \nabla c_{\mathrm{oc}} + (k_{\mathrm{oc}} - \beta_{\mathrm{oc}} \operatorname{Tr}(\varepsilon)) c_{\mathrm{oc}}], \tag{9}$$

where div and $\nabla$ are the divergence and gradient operators, the diffusion tensors $\boldsymbol{D}_{\mathrm{ob}}$ and $\boldsymbol{D}_{\mathrm{oc}}$ are defined as in (6), $k_{\mathrm{ob}}$ and $k_{\mathrm{oc}}$ are the osteoblasts and osteoclasts proliferation rates, respectively with $k_{\mathrm{ob}}$ equal to $\beta_{\mathrm{oc}}$, the osteoclasts differentiation rate. In (8) and (9), as osteoblast proliferation showed to be dependent on the applied mechanical strain [Ignatius et al. 2005; Ehrlich and Lanyon 2002], we assume on a first approximation that it is directly dependent on the volume variation the structure through the trace of epsilon. Complementarily, as the resorption of osteoclasts immediately triggers the proliferation of osteoblasts, a similar kinetic was defined for osteoclasts.

For the above PDEs (equations (4), (5), (8), (9)), a zero flux boundary condition is applied on the external free surfaces as it is supposed that there is no exchange with the outer system.

The variation of bone density $\rho_{\mathrm{b}}$ is described by a first order ordinary differential equation with respect to time given by

$$\frac{\partial \rho_{\mathrm{b}}}{\partial t} = \mathscr{A}_{\mathrm{b}}(\rho_{\mathrm{b}})[s_{\mathrm{b}}(\Delta S_{+}) + r_{\mathrm{b}}(\Delta S_{-})], \tag{10}$$

where $r_{\mathrm{b}}$ and $s_{\mathrm{b}}$ are the rates for bone resorption and synthesis respectively, depending on the positive ($\Delta S_{+}$) and the negative ($\Delta S_{-}$) value of the global stimulus $\Delta S$. $\mathscr{A}_{\mathrm{b}}$ is a function of the bone porosity controlling the intensity of the bone remodeling process that needs to be defined experimentally.

Here, we consider the bone as an isotropic linear elastic material whose Young modulus $E_{\mathrm{b}}$ is given by $E_{\mathrm{b}} = E_{\mathrm{b0}} \rho_{\mathrm{b}}^{3}$ [Currey 1988; Rho et al. 1995] where $E_{\mathrm{b0}}$ is the initial Young modulus of the bone. The global static equilibrium of the system is expressed with the usual equation $\operatorname{div} \boldsymbol{\sigma} + \boldsymbol{f_v} = \mathbf{0}$, with $\boldsymbol{\sigma}$ and $\boldsymbol{f_v}$ the Cauchy stress and the body forces, respectively. Finally, most of the model parameters defined in the current framework should be experimentally quantified. Some theoretical works have been carried out [Placidi et al. 2015; Misra and Poorsolhjouy 2015] trying to identify these parameters, but appropriately designed experiments should be developed in order to provide confident numerical predictions.

The proposed theoretical model was implemented using the Multiphysics Finite Element (FE) code COMSOL Multiphysics® to predict bone kinetics reconstruction when applied to different mechanobiological stimuli.

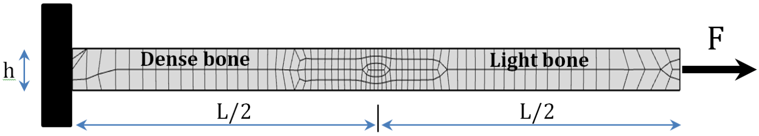


**Figure 1.** Definition of the model geometry, boundary conditions and associated FE mesh.

**2.2.** ***Application.*** Following [Andreaus et al. 2014b; Giorgio et al. 2016], where the bone reconstruction kinetics was studied on a simple two-dimensional (2D) geometry, to compare the obtained results and assess the coupling between the defined variables of the model, the analytical framework described in Section 2.1 is applied similarly on a 2D cantilever beam. The beam of length $L$ and height $h$ is submitted to a tension $F$ at the right side and clamped on the left side (see Figure 1).

As for the initial conditions of the problem, the left half side of the beam is filled with dense bone ($\rho_\text{b} = 0.6$), whereas the right half side of the beam, where the external load $F$ is applied, is constituted by light bone ($\rho_\text{b} = 0.1$) and is assumed to represent a bone substitute or graft. Thus, when mechanical load is applied, migration of cells and nutriments occurs from left to right with bone density increasing in both regions. The input data of the model are listed in Table 1.

These parameters were defined without a priori knowledge of the biological quantifications of the *in vivo* conditions and could therefore require to be tuned for a better approximation of real life conditions. Also, the global stimulus $\Delta S$ was artificially amplified by a multiplication factor to reduce the computation time and accelerate the bone density kinetics evolution (which is of the order of 3 months for real bone) while ensuring consistent results.

## 3. Results and discussion

The concentrations evolutions for osteoclasts, osteoblasts, oxygen and glucose are presented in Figure 2. From the start of the analysis, the concentrations evolve with non-linear distributions and show a clear diffusion from the left to the right of the beam leading to an increase of them on the right side of the beam.

The oxygen and glucose concentration are diffusing quicker than osteoblasts and osteoclasts as their final distribution through the length of the beam is constant at the end of the analysis ($c_{\text{O}_2} = 0.1$ at $c_\text{CHO} = 0.05$), which is not the case for osteoblasts ($0.08 < c_\text{ob} < 0.1$). The osteoclasts completely disappear over time since they differentiate into osteoblasts (initial concentration of 0.05 versus final concentration of $6 \times 10^{-7}$).

The calculated individual and global stimuli, together with bone density evolution over time are presented in Figure 3.

| Symbol | Description | Value |
|---|---|---|
| $L$ | Total length of the beam | 50 mm |
| $h$ | Width of the beam | 2 mm |
| $E_{b0}$ | Initial Young modulus of the bone | 20 GPa |
| $v_b$ | Poisson ratio of the bone | 0.3 |
| $D_{mech}$ | Characteristic distance for the mechanical stimulus | 3 mm |
| $D_{mol}$ | Characteristic distance for the molecular stimulus | 3 mm |
| $D_{cell}$ | Characteristic distance for the cellular stimulus | 3 mm |
| $\alpha_{mech}$ | Weighting coefficient for the mechanical stimulus | 1 |
| $\alpha_{O_2}$ | Weighting coefficient for the oxygen molecular stimulus | 5 |
| $\alpha_{CHO}$ | Weighting coefficient for the glucose molecular stimulus | 5 |
| $\alpha_{ob}$ | Weighting coefficient for the osteoblast cellular stimulus | 5 |
| $\alpha_{oc}$ | Weighting coefficient for the osteoclast cellular stimulus | 5 |
| $\phi$ | Diffusion tensor scalar | 10 |
| $k_{ocl}$ | Osteoclasts proliferation rate | 3 |
| $\beta_{ob}$ | Osteoclasts differentiation rate | 15 |
| $s_b$ | Bone synthesis rate | 1 |
| $r_b$ | Bone resorption rate | 4 |
| $c_{ob}$ | Initial concentration of osteoblasts on the left of the beam | 10% vol |
| $c_{oc}$ | Initial concentration of osteoclasts on the left of the beam | 5% vol |
| $c_{O_2}$ | Initial concentration of oxygen on the left of the beam | 20% vol |
| $c_{CHO}$ | Initial concentration of glucose on the left of the beam | 10% vol |

**Table 1.** Main parameters of the model.

The mechanical stimulus shows a peak of about $2.8 \cdot 10^{-3}$ J/m$^3$ at the beginning of the analysis which increases up to $5.7 \cdot 10^{-3}$ J/m$^3$, propagates towards the right end side due to the external load imposed as the bone reconstruction occurs, and finally decreases to about $5.7 \cdot 10^{-5}$ J/m$^3$ at the end of the analysis as the bone density reaches its maximum value through the whole beam. Such distribution and kinetics are directly dependent on the kinetics of bone remodelling as the bone density increases on the right side of the beam from left to right following this peak where the maximum of the strain energy density is located. In parallel, the cellular stimulus displays a parabolic profile over the left-hand side of the beam, where cells are initially located, with a maximal value of $2.8 \cdot 10^{-6}$ at the beginning of the analysis since a higher bone density is defined on this domain, while it is equal to 0 on the right side as no cells are present. During the analysis, cells migrate from left to right with a decrease of osteoclasts concentration and an increase of osteoblasts one. Cell stimulus increases on the left side up to a value of $6.96 \cdot 10^{-6}$, decreases again at the end of the analysis on the left at $4.5 \cdot 10^{-3}$, and increases continuously

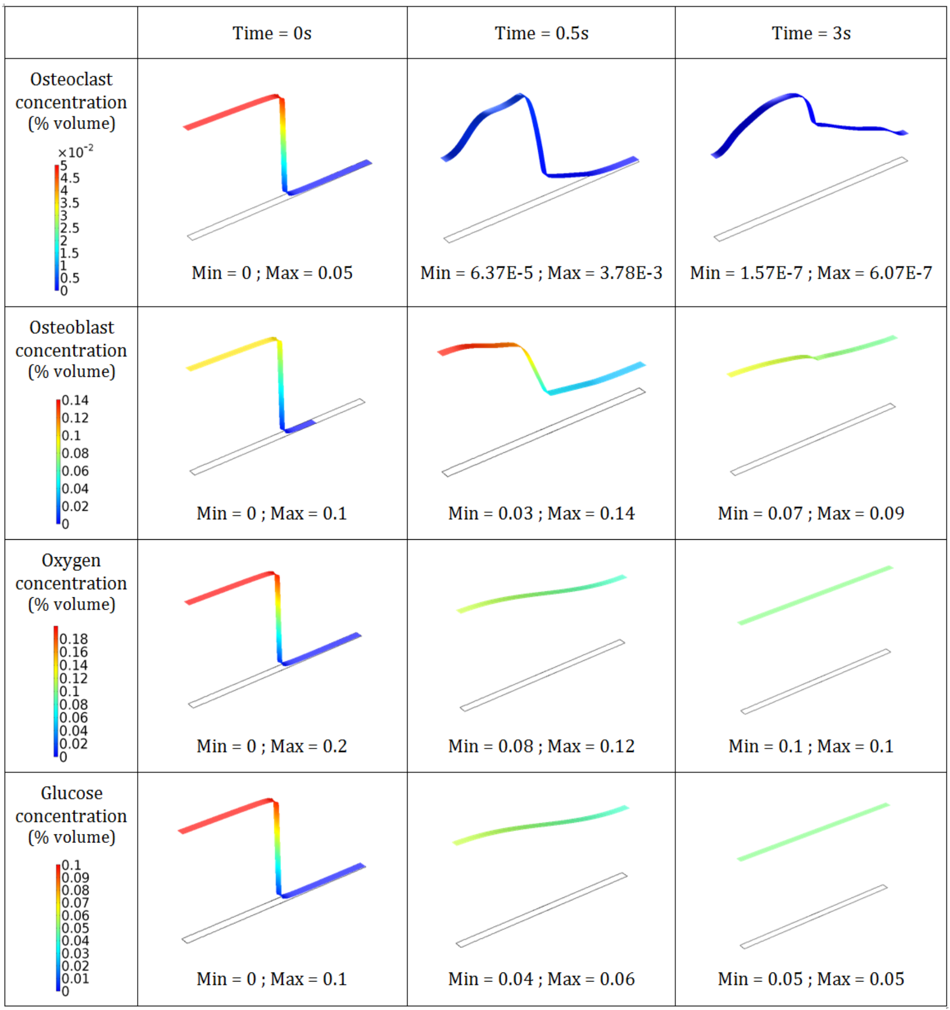


**Figure 2.** Evolution of the osteoclasts, osteoblasts, oxygen and glucose concentration over time.

on the right up to $3.5 \cdot 10^{-3}$ with a non-linear distribution. Correspondingly, the molecular stimulus follows a similar trend as the cellular one but with different kinetics. The initial maximal value is $5.6 \cdot 10^{-6}$, and it decreases over time on the left side down to $3.14 \cdot 10^{-6}$. Identically, it becomes more uniform over the whole beam at the end of analysis, with minimal values at the two extremities ($1 \cdot 10^{-6}$).

For the total stimulus, being the result of the multiplication effects of each stimulus, we still observe the peak value of the mechanical stimulus as it is much larger than the biological ones. However, it is also non-zero everywhere else due to the molecular and cellular contributions. The maximal mechanical stimulus seems

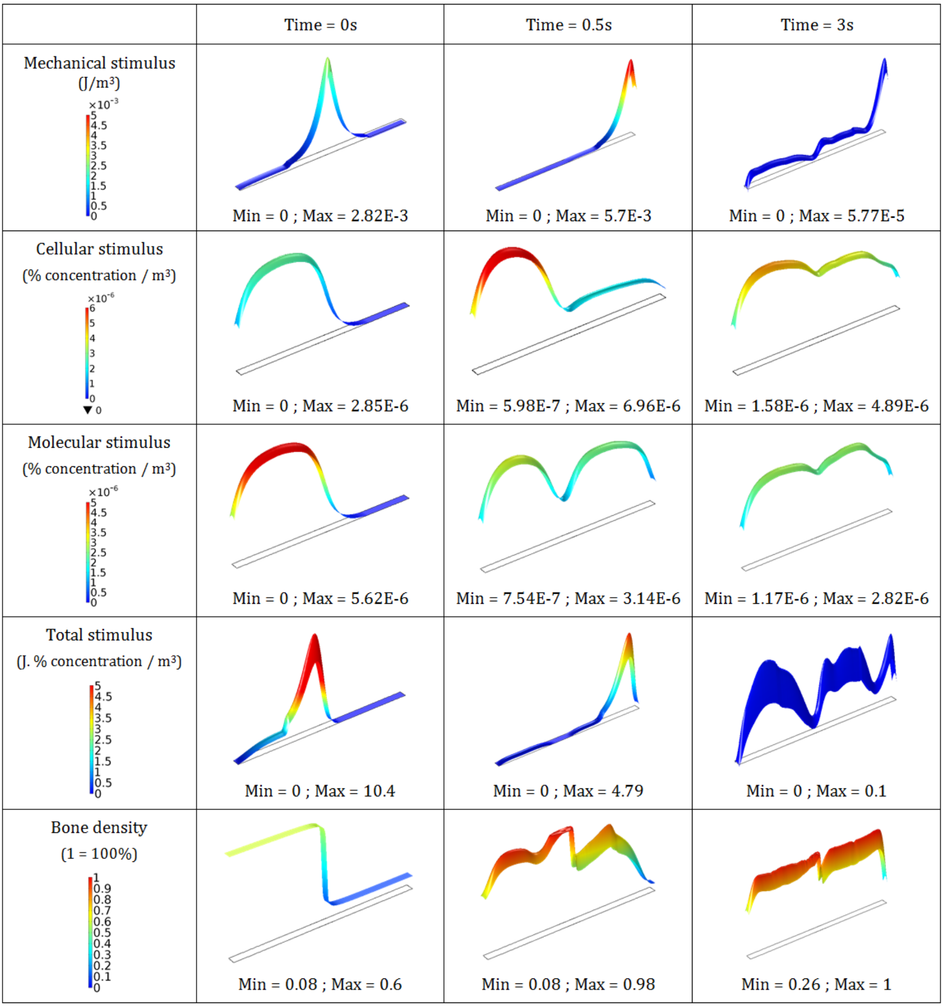


**Figure 3.** Time evolution of each stimulus (mechanical, cellular and molecular), the total stimulus and the bone density.

to be the main driving factor on the effect of the kinetics reconstruction on the right side of the beam. As the initial bone densities are set to 0.6 on the left side and 0.1 on the right side, at the beginning of the analysis we observe a bone density evolution on both sides being triggered by the biological contribution mainly on the left side (due to weak mechanical stimulus, but higher bone density and biological stimulus), and by the mechanical stimulus mainly on the right side (due to its high value and weak bone density with no biology contribution). Once bone density has reached a certain level (mostly reconstructed bone everywhere corresponding to an approximate value of 0.7), the influence of the mechanical stimulus decreases since the structure undergoes a smaller strain and therefore a smaller mechanical energy is developed. Then, the biological effects become thereafter much more important and play a key role in the evolution of bone density. This impact seems to occur over longer periods of time (relative to the mechanical time kinetics) and is clearly visible on the global stimulus at the end of the analysis and on the two bone density distributions during the analysis. The mechanical stimulus moves from the mid-length to the right hand side of the beam without inhibiting the increase of the bone density on the left side where it tends towards zero. Also, the bone density is recovered on the left side due to the biological impact of the stimulus and continues to increase to reach an almost maximum density even after the mechanical effect has dropped at the end of the analysis. The above proposed model being continuous, it is assumed that cell distribution is also continuously distributed through the entire geometry, even with heterogeneous distribution. Accounting for the spatial range of cell influence requires integrating the microstructure distribution [Andreaus et al. 2014a]. This contribution needs to be integrated in future works. Finally, although the mechanical stimulus seems to play a critical role in the bone reconstruction kinetics, it also shows to be highly dependent on the biological contributions and certainly coupled with the bone density impact. In fact, high bone density leads to small strains and therefore to small mechanical stimulus for a given applied mechanical load. This has a direct impact on the cellular response within the structure as higher density (lower porosity) leads to lower cell density (and distribution) and lower density (higher porosity) leads to higher cell density (and distribution) with trabecular bone structure. These effects should also be integrated since the trabecular bone kinetics requires a more specifically adapted thermodynamically consistent model as described for example in [Ganghoffer 2012; 2016; Goda et al. 2016; Louna et al. 2017], and be homogenized in order to obtain a better macroscopic prediction. Nonetheless, the above presented mechanobiological couplings would also need to be integrated within these local frameworks in order to identify precisely the influence of the biology in the bone reconstruction kinetics.

## 4. Conclusion

In the present paper, a new coupled multiphysics model is proposed to compute the mechanobiological stimulus for continuum mechanics bone reconstruction, by taking into account specific mechanical (i.e. external loads) and biological (i.e. cellular migration and differentiation and nutriments supply) phenomena. The final results highlight the respective contributions of each process on the kinetics of bone density evolution. Each effect shows to have an important impact although the model parameters still require adequate quantification for better representation of specific medical applications.

## References

[Allena and Maini 2014] R. Allena and P. K. Maini, "Reaction-diffusion finite element model of lateral line primordium migration to explore cell leadership", *Bull. Math. Biol.* **76**:12 (2014), 3028–3050.

[Andreaus et al. 2014a] U. Andreaus, M. Colloca, and D. Iacoviello, "Optimal bone density distributions: numerical analysis of the osteocyte spatial influence in bone remodeling", *Comput. Methods Programs Biomed.* **113**:1 (2014), 80–91.

[Andreaus et al. 2014b] U. Andreaus, I. Giorgio, and T. Lekszycki, "A 2-D continuum model of a mixture of bone tissue and bio-resorbable material for simulating mass density redistribution under load slowly variable in time", *J. Appl. Math. Mech.* **94**:12 (2014), 978–1000.

[Beaupré et al. 1990] G. S. Beaupré, T. E. Orr, and D. R. Carter, "An approach for time-dependent bone modeling and remodeling-application: A preliminary remodeling simulation", *J. Orthop. Res.* **8**:5 (1990), 662–670.

[Bednarczyk and Lekszycki 2016] E. Bednarczyk and T. Lekszycki, "A novel mathematical model for growth of capillaries and nutrient supply with application to prediction of osteophyte onset", *Z. Angew. Math. Phys.* **67** (2016), 94.

[Burr and Allen 2013] D. B. Burr and M. R. Allen (editors), *Basic and applied bone biology*, Academic Press, 2013. Available at https://tinyurl.com/y8j9bkmz.

[Cowin 1986] S. C. Cowin, "Wolff's law of trabecular architecture at remodeling equilibrium", *J. Biomech. Eng.* (*ASME*) **108**:1 (1986), 83–88.

[Currey 1988] J. D. Currey, "The effect of porosity and mineral content on the Young's modulus of elasticity of compact bone", *J. Biomech.* **21**:2 (1988), 131–139.

[Ehrlich and Lanyon 2002] P. J. Ehrlich and L. E. Lanyon, "Mechanical strain and bone cell function: a review", *Osteoporos. Int.* **13**:9 (2002), 688–700.

[Frame et al. 2017] J. Frame, P.-Y. Rohan, L. Corté, and R. Allena, "A mechano-biological model of mulit-tissue evolution in bone", *Contin. Mech. Therm.* (2017), 1–31.

[Frost 1987] H. M. Frost, "Bone 'mass' and the 'mechanostat': a proposal", *Anat. Rec.* **219**:1 (1987), 1–9.

[Ganghoffer 2012] J.-F. Ganghoffer, "A contribution to the mechanics and thermodynamics of surface growth: application to bone external remodeling", *Int. J. Eng. Sci.* **50**:1 (2012), 166–191.

[Ganghoffer et al. 2016] J.-F. Ganghoffer, R. Rahouadj, J. Boisse, and S. Forest, "Phase field approaches of bone remodelling based on TIP", *J. Non Equilib. Thermodyn.* **41**:1 (2016), 49–75.

[George et al. 2017a] D. George, R. Allena, and Y. Rémond, "Mechanobiological stimuli for bone remodeling: mechanical energy, cell nutriments and mobility", *Comput. Methods Biomech. Biomed. Engin.* **20**:sup1 (2017), 91–92.

[George et al. 2017b] D. George, C. Spingarn, C. Dissaux, M. Nierenberger, R. A. Rahman, and Y. Rémond, "Examples of multiscale and multiphysics numerical modeling of biological tissues", *Biomed. Mater. Eng.* **28**:s1 (2017), S15–S27.

[Giorgio et al. 2016] I. Giorgio, U. Andreaus, D. Scerrato, and F. dell'Isola, "A visco-poroelastic model of functional adaptation in bones reconstructed with bio-resorbable materials", *Biomech. Model. Mechanobiol.* **15**:5 (2016), 1325–1343.

[Giorgio et al. 2017] I. Giorgio, U. Andreaus, F. dell'Isola, and T. Lekszycki, "Viscous second gradient porous materials for bones reconstructed with bio-resorbable grafts", *Extrem. Mechan. Letters* **13** (2017), 141–147.

[Goda et al. 2016] I. Goda, J.-F. Ganghoffer, and G. Maurice, "Combined bone internal and external remodeling based on Eshelby stress", *Int. J. Solids Struct.* **94-95** (2016), 138–157.

[Ignatius et al. 2005] A. Ignatius, H. Blessing, A. Liedert, C. Schmidt, C. Neidlinger-Wilke, D. Kaspar, B. Friemert, and L. Clase, "Tissue engineering of bone: effects of mechanical strain on osteoblastic cells in type I collagen matrices", *Biomater.* **26**:3 (2005), 311–318.

[Lekszycki 2002] T. Lekszycki, "Modeling of bone adaptation based on an optimal response hypothesis", *Meccanica* (*Milano*) **37**:4-5 (2002), 343–354.

[Lekszycki and dell'Isola 2012] T. Lekszycki and F. dell'Isola, "A mixture model with evolving mass densities for describing synthesis and resorption phenomena in bones reconstructed with bioresorbable materials", *J. Appl. Math. Mech.* **92**:6 (2012), 426–444.

[Lemaire et al. 2006] T. Lemaire, S. Naïli, and A. Rémond, "Multiscale analysis of the coupled effects governing the movement of interstitial fluid in cortical bone", *Biomech. Model. Mechanobiol.* **5**:1 (2006), 39–52.

[Lemaire et al. 2010] T. Lemaire, S. Naili, and V. Sansalone, "Multiphysical modelling of fluid transport through osteo-articular media", *An. Acad. Bras. Ciências* **82**:1 (2010), 127–144.

[Lemaire et al. 2011] T. Lemaire, E. Capiez-Lernout, J. Kaiser, N. S., and V. Sansalone, "What is the importance of multiphysical phenomena in bone remodelling signals expression? A multiscale perspective", *J. Mech. Behav. Biomed. Mater.* **4**:6 (2011), 909–920.

[Lemaire et al. 2015] T. Lemaire, J. Kaiser, S. Naili, and V. Sansalone, "Three-scale multiphysics modeling of transport phenomena within cortical bone", *Math. Probl. Eng.* (2015), 398970.

[Louna et al. 2017] Z. Louna, I. Goda, J.-F. Ganghoffer, and S. Benhadid, "Formulation of an effective growth response of trabecular bone based on micromechanical analyses at the trabecular level", *Arch. Appl. Mech.* **87**:3 (2017), 457–477.

[Lu and Lekszycki 2016] Y. Lu and T. Lekszycki, "A novel coupled system of non-local integro-differential equations modelling Young's modulus evolution, nutrients' supply and consumption during bone fracture healing", *Z. Angew. Math. Phys.* **67** (2016), 111.

[Madeo et al. 2011] A. Madeo, T. Lekszycki, and F. dell'Isola, "A continuum model for the bio-mechanical interactions between living tissue and bio-resorbable graft after bone reconstructive surgery", *Comptes Rendus Mécanique* **339**:10 (2011), 625–640.

[Madeo et al. 2012] A. Madeo, D. George, T. Lekszycki, M. Nierenberger, and Y. Rémond, "A second gradient continuum model accounting for some effects of micro-structure on reconstructed bone remodelling", *Comptes Rendus Mécanique* **340**:8 (2012), 575–589.

[Martin et al. 2017] M. Martin, T. Lemaire, G. Haïat, P. Pivonka, and V. Sansalone, "A thermodynamically consistent model of bone rotary remodeling: a 2D study", *Comput. Methods Biomech. Biomed. Engin.* **20**:sup1 (2017), 127–128.

[Misra and Poorsolhjouy 2015] A. Misra and P. Poorsolhjouy, "Identification of higher-order elastic constants for grain assemblies based upon granular micromechanics", *Math. Mech. Comp. Syst.* **3**:3 (2015), 285–308.

[Pivonka and Komarova 2010] P. Pivonka and S. V. Komarova, "Mathematical modeling in bone biology: From intracellular signaling to tissue mechanics", *Bone* **47**:2 (2010), 181–189.

[Pivonka et al. 2008] P. Pivonka, J. Zimak, D. W. Smith, B. S. Gardiner, C. R. Dunstan, N. A. Sims, T. J. Martin, and G. R. Mundy, "Model structure and control of bone remodeling: A theoretical study", *Bone* **43**:2 (2008), 249–263.

[Placidi et al. 2015] L. Placidi, U. Andreaus, A. Della Corte, and T. Lekszycki, "Gedanken experiments for the determination of two-dimensional linear second gradient elasticity coefficients", *Z. Angew. Math. Phys.* **66**:6 (2015), 3699–3725.

[Rémond et al. 2016] Y. Rémond, S. Ahzi, M. Baniassadi, and M. Garmestani, *Applied RVE reconstruction and homogenization of heterogeneous materials*, Wiley-ISTE, 2016.

[Rho et al. 1995] J. Y. Rho, M. C. Hobatho, and R. B. Ashman, "Relations of mechanical properties to density and CT numbers in human bone", *Med. Eng. Phys.* **17**:5 (1995), 347–355.

[Sansalone et al. 2015] V. Sansalone, D. Gagliardi, C. Descelier, G. Haïat, and S. Naili, "On the uncertainty propagation in multiscale modeling of cortical bone elasticity", *Comput. Methods Biomech. Biomed. Engin.* **18** (2015), 2054–2055.

[Scala et al. 2016] I. Scala, C. Spingarn, A. Rémond, Y. Madeo, and D. George, "Mechanically-driven bone remodeling simulation: Application to LIPUS treated rat calvarial defects", *Math. Mech. Solids* **22**:10 (2016), 1976–1988.

[Schmitt et al. 2015] M. Schmitt, R. Allena, T. Schouman, S. Frasca, J. M. Collombet, X. Holy, and P. Rouch, "Diffusion model to describe osteogenesis within a porous titanium scaffold", *Comput. Methods Biomech. Biomed. Engin.* **19**:2 (2015), 171–179.

[Turner 1998] C. H. Turner, "Three rules for bone adaptation to mechanical stimuli", *Bone* **23**:5 (1998), 399–407.

[Wagner et al. 2017] D. Wagner, Y. Bolender, Y. Rémond, and D. George, "Mechanical equilibrium of forces and moments applied on orthodontic brackets of a dental arch: correlation with literature data on two and three adjacent teeth", *Biomed. Mater. Eng.* **28**:s1 (2017), S169–S177.

Received 16 Dec 2017. Revised 26 Mar 2018. Accepted 13 May 2018.

DANIEL GEORGE: george@unistra.fr
*ICube Laboratory, University of Strasbourg/CNRS, Strasbourg, France*

RACHELE ALLENA: rachele.allena@ensam.eu
*Institute de Biomécanique Humaine George Charpak/LBM, Arts et Métiers, ParisTech, Paris, France*

YVES RÉMOND: remond@unistra.fr
*ICube Laboratory, University of Strasbourg/CNRS, Strasbourg, France*

MATHEMATICS AND MECHANICS OF COMPLEX SYSTEMS
Vol. 6, No. 4, 2018
dx.doi.org/10.2140/memocs.2018.6.321

# ON LINEAR NON-LOCAL THERMO-VISCOELASTIC WAVES IN FLUIDS

JOE D. GODDARD

The following is an elaboration on the linear non-local model of viscoelastic fluids proposed in a previous work (*Int. J. Eng. Sci.* **48** (2010), 1279–1288). As a recapitulation of that work, the basic theory is presented in terms of the temporal frequency and spatial wave number in the Laplace–Fourier domain. Taylor-series expansions in these variables provides a weakly non-local theory in spatio-temporal gradients that is more comprehensive than the "bi-velocity" model of Brenner. The linearized Chapman–Enskog kinetic theory is shown to provide a confirmation of the more general theory, from which one can reconstruct a fully non-local integral model.

Following the work of Davis and Brenner (*J. Acoust. Soc. Am.* **132** (2012), 2963–2969), the general theory is employed to derive dispersion relations for acoustic, thermal and shear-wave propagation in compressible viscoelastic fluids. At Burnett order the Chapman–Enskog theory gives a cubic polynomial in wave number squared which reduces in the dissipative quasi-static limit to a quadratic like that given by the classical Navier–Stokes–Fourier model and the bi-velocity modification of that model.

With minor modification, the present analysis applies to viscoelastic shear and dilatational wave propagation in solids with higher-gradient and Cosserat effects, where it may, for example, find application to the field of rotational seismology.

## 1. Introduction

Following [Goddard 2010], hereinafter referred to as [G10], we consider a linear, fully non-local model for the thermo-mechanics of fluids. As was the case with [G10], the present paper is motivated in part by the ideas of the late H. Brenner, who wrote extensively at the end of his career [2009; 2012] on the possible breakdown of the classical Navier–Stokes–Fourier model of momentum and heat flux arising

**Communicated by Francesco dell'Isola**.
*MSC2010:* 35Q72, 44A30, 74JXX.
*PACS2010:* primary 02.30.Nw, 05.20.Dd; secondary 02.30.Uu, 43.20.+g, 47.10.-g, 47.10.ab.
*Keywords:* non-locality, thermo-viscoelastic waves, bivelocity fluid mechanics, Maxwell–Cattaneo relaxation, Chapman–Enskog kinetic theory.

from strong inhomogenieties due to large temperature or density gradients. This ostensibly motivates his revised constitutive theory, “bivelocity fluid mechanics”, based on the notion that barycentric velocity, associated with material inertia and kinetic energy, is not appropriate for the description of internal stress in a fluid or solid. In its place, he proposes a “volume” or “work” velocity, together with various constitutive models for the “diffuse volume flux” representing the difference between the two velocities. This stratagem serves *inter alia* to introduce higher spatial gradients of temperature and velocity into the constitutive theory.

An alternative perspective is offered in [G10], where it is argued that the above revision is necessitated by the breakdown of the thermo-mechanically *simple material* of Coleman [1961; 1964] and that Brenner’s constitutive theory is a restricted version of a more general non-local theory. Such a theory, anticipated in numerous previous works (see [Eringen 2002] and references therein), was sketched out in [G10], which leaves unanswered certain questions regarding the magnitude of material-specific length and time scales involved in the breakdown of the classical model and the inadequacy of the bi-velocity model as a strictly linear theory.

The purpose of the present work is to elucidate further the above questions, by considering specific models that are fully non-local in both space and time, i.e., models which involve long-range interactions in space combined with long-range history effects in time. In particular, we show that the linear model which emerges at “Burnett order” in the classical Chapman–Enskog kinetic theory is a special case of the general model. As appreciated by others [Müller and Ruggeri 1998], this kinetic theory involves relaxation effects of the type described earlier by Maxwell’s viscoelasticity [1867] and later by Cattaneo’s retarded thermal conductivity [1948]. As we shall also show, Brenner’s theory represents a restriction to the dissipative response arising on time scales longer that the Maxwell–Cattaneo relaxation times. Also, it is shown that a fully non-local model can be reconstructed from the linearized Chapman–Enskog theory.

Acoustic wave propagation represents a plausible testing ground for non-local thermo-mechanical effects, as already recognized in [Davis and Brenner 2012]; the present paper provides an extension of that work. It also presents an extension of [G10] that identifies the hyperstresses conjugate to higher velocity gradients. However, the applications to wave propagation are restricted to the linear momentum balance, with no account taken of higher-order inertial terms. Finally, we establish a connection to various non-local models of wave propagation in complex solids.

## 2. Fourier–Laplace representation: Recapitulation of previous work

Following the analysis of [G10], we recall that Fourier representations embody the notion of wave-number dependent transport coefficients, capturing the dispersive effects associated with higher gradients. When extended to the time domain by means of the Laplace transform, one obtains a similar description of frequency effects in materials with memory[1]. Hence, the transform

$$\hat{\boldsymbol{\psi}}(\boldsymbol{k},s)=\hat{\boldsymbol{\psi}}_t(\boldsymbol{k},s)=\frac{1}{\sqrt{8\pi^3}}\int_{\mathfrak{R}}\int_0^{\infty} e^{-\imath\boldsymbol{k}\cdot\boldsymbol{x}-st'}\boldsymbol{\psi}(\boldsymbol{x},t-t')\,\mathrm{d}V(\boldsymbol{x})\,\mathrm{d}t' \tag{1}$$

provides a *localized* description in Fourier space $(\boldsymbol{k},s)$ of a spatio-temporally *de-localized* field in physcial space, $\boldsymbol{\psi}_t(\boldsymbol{x}',t')=\boldsymbol{\psi}(\boldsymbol{x}',t-t')$, $\quad \boldsymbol{x}'\in\mathfrak{R}$, $\quad t'\geq 0$, and *vice versa*. Accordingly, a causal, non-local and linear constitutive equation between two sets of tensor fields

$$\boldsymbol{\Phi}(\boldsymbol{x},t)=\{\boldsymbol{\varphi}^{(1)},\dots,\boldsymbol{\varphi}^{(m)}\}(\boldsymbol{x},t) \quad \text{and} \quad \boldsymbol{\Psi}(\boldsymbol{x},t)=\{\boldsymbol{\psi}^{(1)},\dots,\boldsymbol{\psi}^{(m)}\}(\boldsymbol{x},t-t'), \tag{2}$$

for $t'\geq 0$, of the type pursued by Eringen [1992; 2002], can be represented by the linear form:

$$\widehat{\boldsymbol{\Phi}}(\boldsymbol{k},s)=\widehat{\boldsymbol{\mathfrak{L}}}(\boldsymbol{k},s)\widehat{\boldsymbol{\Psi}}(\boldsymbol{k},s) \tag{3}$$

where $\hat{\boldsymbol{\mathfrak{L}}}$ represents a matrix of tensor moduli. This relation is tantamount to the spectral theory of commutative linear operators with $\{\imath\boldsymbol{k},s\}\to\{\nabla,\partial_t\}$, and the time-honored Fourier–Laplace transforms provide a concrete algebraic representation.

With $\widehat{\boldsymbol{\Phi}}(\boldsymbol{x},t)=\{\hat{\boldsymbol{\sigma}},\hat{\boldsymbol{q}}\}$ representing stress $\hat{\boldsymbol{\sigma}}$ and heat flux $\hat{\boldsymbol{q}}$ in (2), one obtains a linear non-local theory of thermo-viscoelasticity. We recall that Eringen [2002, Section 7] proposes a simpler non-local theory for viscous incompressible fluids with uncoupled heat flux, a theory that was overlooked in [G10].

If we adopt a scaling in which $\boldsymbol{k}$ and $s$ are replaced by non-dimensional forms $\lambda_0\boldsymbol{k}$ and $\tau_0 s$, with $\lambda_0$ and $\tau_0$ denoting, respectively, appropriate material length and time scales, then $k=|\boldsymbol{k}|$ and $s$ represent, respectively, a *Knudsen* and a *Deborah* number. Hence, one obtains a *weakly non-local* spatio-temporal models from the Taylor-series expansion of $\hat{\boldsymbol{\mathfrak{L}}}(\boldsymbol{k},s)$ about the spatially uniform steady state $k=0, s=0$. The expansion in $\boldsymbol{k}$ is, to terms $O(k^2)$, tantamount to the Burnett expansion of kinetic theory [Müller and Ruggeri 1998], whereas the expansion in $s$ represents the "retarded motions" of [Coleman and Noll 1961; Coleman 1964]. In particular, the *simple fluid* emerges at $O(k)$ in $\boldsymbol{k}$. Dissipative response, defining the Navier–Stokes–Fourier regime, arises for $s\to 0$ at $O(1)$ for $\hat{\boldsymbol{q}}$ and at $O(s)$ for $\hat{\boldsymbol{\sigma}}$, provided we take $\hat{\boldsymbol{v}}$ to be $O(s)$, i.e.,

$$\boldsymbol{v}=\partial_t\boldsymbol{u}, \quad \text{with} \quad \therefore \quad \hat{\boldsymbol{v}}=s\hat{\boldsymbol{u}}, \tag{4}$$

where $\boldsymbol{u}$ denotes material displacement from the positions at $t=0$.

[1] Since our "Fourier–Laplace" transform involves what is essentially a Fourier transform on $t=[0,\infty)$ with complex wave vector $s$, we could as well employ the terminology "Fourier transform" and "Fourier space".

Following [G10]we consider a spatially nonlocal linear viscoelastic fluid in which the stress $\boldsymbol{\sigma}$ relative to a spatially uniform equilibrium pressure $p_0$ (replacing the deviatoric stress $\boldsymbol{\tau}$ of [G10], which deals with incompressible fluids) and the heat flux $\boldsymbol{q}$, are represented by $\hat{\boldsymbol{\sigma}}$ $[= \hat{\sigma}_{ij}]$ and $\hat{\boldsymbol{q}}$ $[= \hat{q}_i]$ as functions of velocity $\hat{\boldsymbol{v}}$ $[= \hat{v}_i]$ and departure $\hat{\theta} = \hat{T} - T_0/s$ from the uniform absolute temperature $T_0$ of the equilibrium base state. With Cartesian tensor components displayed for clarity, these can be written as in [G10]:

$$\begin{aligned} \hat{\boldsymbol{q}}\,[= \hat{q}_i] &= \hat{\boldsymbol{L}}^{(11)}\hat{\theta} + \hat{\boldsymbol{L}}^{(12)}\hat{\boldsymbol{v}} \quad \big[\overset{\text{def}}{=} \hat{L}^{(11)}\hat{\theta} + \hat{L}^{(12)}_{ij}\hat{v}_j\big], \\ \hat{\boldsymbol{\sigma}}\,[\hat{\sigma}_{ij}] &= \hat{\boldsymbol{L}}^{(21)}\hat{\theta} + \hat{\boldsymbol{L}}^{(22)}\hat{\boldsymbol{v}} \quad \big[\overset{\text{def}}{=} \hat{L}^{(21)}_{ij}\hat{\theta} + \hat{L}^{(22)}_{ijl}\hat{v}_l\big], \end{aligned} \tag{5}$$

where the tensor coefficients $\hat{L}$ depend on the complex frequency $s$ and wave vector $\boldsymbol{k}$. Here as in the following, we indicate components of tensors on a given Cartesian system by means of square brackets [ ], and the Cartesian summation convention is employed. We further employ colons to denote contraction of the trailing components of a prefactor with all the components of the postfactor, with the conventional dot for the scalar product of vectors.

For isotropic materials, the various tensors in (5) must be isotropic functions of the wave vector and, for the case of symmetric stress assumed here, can be written down explicitly as in [G10]:

$$\begin{aligned} \hat{L}^{(11)}_i &= \hat{A}k_i, \\ \hat{L}^{(12)}_{ij} &= \hat{B}\delta_{ij} + \hat{C}k_ik_j, \\ \hat{L}^{(21)}_{ij} &= \hat{D}\delta_{ij} + \hat{E}k_ik_j, \\ \hat{L}^{(22)}_{ijl} &= \hat{F}\delta_{ij}k_l \; + \hat{G}(\delta_{il}k_j + \delta_{jl}k_i) + \hat{H}k_ik_jk_l, \end{aligned} \tag{6}$$

where the scalar coefficients $\hat{A}$, $\hat{B}, \dots, \hat{H}$ are functions of $s$ and $k^2$, where $k^2 = k_ik_i$ defines a generally complex quantity, since we shall admit complex wave vectors $\boldsymbol{k}$. Also, we have added carats to the coefficients defined in [G10], in order to distinguish them from their physical-space images considered below.

In direct tensor notation, the preceding relations become

$$\begin{aligned} \hat{\boldsymbol{q}} &= \hat{A}\boldsymbol{k}\hat{\theta} + (\hat{B}\mathbf{1} + \hat{C}\boldsymbol{k}\otimes\boldsymbol{k})\hat{\boldsymbol{v}}, \\ \hat{\boldsymbol{\sigma}} &= (\hat{D}\mathbf{1} + \hat{E}\boldsymbol{k}\otimes\boldsymbol{k})\hat{\theta} + \hat{F}(\boldsymbol{k}\cdot\hat{\boldsymbol{v}})\mathbf{1} + \hat{G}(\boldsymbol{k}\otimes\hat{\boldsymbol{v}} + \hat{\boldsymbol{v}}\otimes\boldsymbol{k}) + \hat{H}(\boldsymbol{k}\cdot\hat{\boldsymbol{v}})\boldsymbol{k}\otimes\boldsymbol{k}. \end{aligned} \tag{7}$$

Now, the requirement of real $\boldsymbol{q}$ and $\boldsymbol{\sigma}$ implies that the coefficient, say, $\hat{K}_n$ of the general term in (7):

$$\hat{K}_n(k, s)\boldsymbol{k}^n, \quad \text{where} \quad \boldsymbol{k}^n = \boldsymbol{k}^{n-1}\otimes\boldsymbol{k}, \quad n = 1, 2, \dots, \quad \boldsymbol{k}^0 = 1, \tag{8}$$

must satisfy $\hat{K}_n^*(k^2, s) = (-1)^n \hat{K}_n(k^{*2}, s)$, where asterisks denote complex conjugates here and below. Hence, the coefficients of even (odd) order terms in $\boldsymbol{k}$ must be essentially real (imaginary). (By *essentially real* we mean a function $\mathbb{R} \to \mathbb{R}$, i.e., a function that is real-valued when its arguments are real, whereas essentially imaginary is a function $\mathbb{R} \to \imath\mathbb{R}$, i.e., $\imath$ times an essentially real function.)

In line with the above remarks, and following [G10], we obtain from (6) a weakly nonlocal theory in space by means of the wave-number expansions of $\hat{K} = \hat{A}, \hat{B}, \ldots, \hat{H}$ of the form

$$\hat{K} = \hat{K}_0(s) + \hat{K}_1(s)k^2 + \hat{K}_2(s)k^4 + \cdots, \tag{9}$$

where $\hat{K}_m(s)$ are independent of $\boldsymbol{k}$. As pointed out in [G10], Gallilean invariance of heat flux requires that $\hat{B}_0 = 0$ which, by a general form of Onsager symmetry, implies that $\hat{D}_0$=0. We shall show presently that the latter result arises from a properly restricted form of that symmetry.

Note that the stress defined by $(7)_2$ represents a non-local quantity whose expansion in $\boldsymbol{k}$ defines a hierarchy of hyperstresses. In particularly, by an extension of the dissipative forms discussed by [Goddard and Lee 2017] we have

$$\hat{\boldsymbol{\sigma}} = \sum_{m \geq 1} \hat{\boldsymbol{\sigma}}^{(m)} : (-\imath \boldsymbol{k})^{m-1}, \quad i.e., \quad \hat{\sigma}_{ij} = \sum_{m \geq 1} \hat{\sigma}^{(m)}_{ij, j_1, \ldots, j_{m-1}} (-\imath \boldsymbol{k})^{m-1}_{j_1, \ldots, j_{m-1}},$$

where $\hat{\boldsymbol{\sigma}}^{(1)}$ is Cauchy stress and the $\hat{\boldsymbol{\sigma}}^{(m)}, \quad m > 1$, is the hyperstress conjugate to $(\imath \boldsymbol{k})^m \hat{\boldsymbol{v}}$.

Now, if both $\hat{B}_0$ and $\hat{D}_0$ vanish, then (5) and (7) reduce to a standard form in which $\{\nabla\theta, \mathrm{Sym}(\nabla \boldsymbol{v})\}$ represent nine forces conjugate to nine fluxes $\{\boldsymbol{q}, \boldsymbol{\sigma}\}$. In that case, the local dissipation rate is given by:

$$\boldsymbol{\sigma} : \nabla \boldsymbol{v} - \frac{\boldsymbol{q}}{T_0} \cdot \nabla\theta \geq 0, \tag{10}$$

in the dissipative regime, where $\boldsymbol{\sigma}, \boldsymbol{q}$ are strictly dissipative. Thus, by the *Parseval–Plancherel theorem*, the global dissipation becomes[2]

$$\begin{aligned} &\int_{\mathfrak{R}} \left( \boldsymbol{\sigma} : \nabla \boldsymbol{v} - \frac{\boldsymbol{q}}{T_0} \cdot \nabla\theta \right) \mathrm{d}V(\boldsymbol{x}) \\ &\qquad = -\imath \int_{\hat{\mathfrak{R}}} \left( \hat{\boldsymbol{\sigma}} : \boldsymbol{k}^* \hat{\boldsymbol{v}}^* - \hat{\boldsymbol{q}} \cdot \boldsymbol{k}^* \frac{\hat{\theta}^*}{T_0} \right) \mathrm{d}\hat{V}(\boldsymbol{k}) \\ &\qquad = -\imath \int_{\hat{\mathfrak{R}}} \left( \hat{L}^{(21)}_{ij} k_i^* \hat{v}_j^* \frac{\hat{\theta}}{T_0} + \hat{L}^{(22)}_{ijl} k_i^* \hat{v}_j^* \hat{v}_l - \hat{L}^{(11)}_i k_i^* \frac{|\hat{\theta}|^2}{T_0^2} - \hat{L}^{(12)}_{ij} k_i^* \hat{v}_j \frac{\hat{\theta}^*}{T_0} \right) \mathrm{d}\hat{V}(\boldsymbol{k}) \\ &\qquad \geq 0, \end{aligned} \tag{11}$$

[2] after extension to complex $\boldsymbol{k}$ and correction of a typographical error of [G10]

where $\mathcal{R}$ is the spatial region occupied by the fluid and $\hat{\mathcal{R}}$ is its Fourier image (i.e., the transform of its indicator function). Based on an unwarranted restriction to real-valued transforms $\hat{\theta}$, $\hat{\boldsymbol{v}}$, it is erroneously concluded in [G10] that the general Onsager symmetry $L_{ij}^{(21)} = L_{ij}^{(12)}$ eliminates dissipative coupling between temperature and velocity.

As a more restricted form of Onsager symmetry, note that (7) gives

$$\begin{aligned} &\hat{p} = \hat{I}\hat{\theta} + \hat{J}\boldsymbol{k}\cdot\hat{\boldsymbol{v}}, \quad \text{where } \hat{I} = -(\hat{D} + \hat{E}k^2/3),\ \hat{J} = -(\hat{F} + 2\hat{G}/3 + \hat{H}k^2/3), \\ &\boldsymbol{k}\cdot\hat{\boldsymbol{q}} = \hat{A}k^2\hat{\theta} + (\hat{B} + \hat{C}k^2)\boldsymbol{k}\cdot\hat{\boldsymbol{v}}, \end{aligned} \tag{12}$$

with $J \to J_0 = -\imath\beta_0$ for $k \to 0$, where $\beta_0$ denotes the standard bulk (or "volume") viscosity. Thus, in the dissipative regime, the quantities

$$\theta\nabla\cdot\boldsymbol{q}/T_0 - p\nabla\cdot\boldsymbol{v} \geq 0, \quad \text{or} \quad \imath(\hat{\theta}^*\boldsymbol{k}\cdot\hat{\boldsymbol{q}}/T_0 - \hat{p}^*\boldsymbol{k}\cdot\hat{\boldsymbol{v}}) \geq 0$$

represent the dissipation rate. The significance of the term involving pressure work is obvious, while the other term is essentially the potential Carnot work dissipated locally by irreversible heat flow, since $\theta/T_0 = -(1 - T/T_0)$. Hence, the Onsager symmetry of the linear relations (12) requires that

$$(\hat{B} + \hat{C}k^2)/T_0 = -\hat{I} = (\hat{D} + \hat{E}k^2/3), \quad \text{and} \quad \therefore \quad \hat{D}_0 = \hat{B}_0 = 0 \tag{13}$$

*in the dissipative regime*, a necessary restriction on the more general form proposed in [G10].

As they stand, the relations (7) represent linear non-local thermo-viscoelasticity, with the $\boldsymbol{x}$-$t$ images of the coefficients $\hat{K} = \hat{A}, \hat{B}, \hat{C}, \ldots$ providing the kernels of integral operators acting on $\mathfrak{f} = \{\theta, \boldsymbol{v}\}$. According to (8) these assume the form

$$\hat{K}_n\boldsymbol{k}^n(\cdot)\hat{\mathfrak{f}} \to (-\imath)^n \int_{t'=0}^{\infty}\int_{\mathcal{R}'} K_n(\boldsymbol{x}', t')\nabla^n(\cdot)\mathfrak{f}(\boldsymbol{x} - \boldsymbol{x}', t - t')\,\mathrm{d}V(\boldsymbol{x}')\,\mathrm{d}t', \tag{14}$$

where $(\cdot)$ represents an optional dot product or contraction. Moreover, since the coefficients $\hat{K}$ are functions of $\boldsymbol{k}$ that depend only on $k^2$, they admit simplified inverse spatial transforms, as discussed in the Appendix. We now consider the special cases of the general theory represented by the kinetic theory of gases and by Brenner's bivelocity model.

## 3. Linearized kinetic theory of gases

As a slight variant on the kinetic-theory results given by Chapman and Cowling [1960, p. 410], Müller and Ruggeri [1998, p. 74] give the following implicit forms for heat flux and shear stress in a monatomic gas:

$$\begin{aligned}
\boldsymbol{q} &= -\tau_q\big\{-\tfrac{5}{2}Rp\nabla\theta+\dot{\boldsymbol{q}}+\boldsymbol{q}\cdot\nabla\boldsymbol{v}-R\theta\nabla\cdot\boldsymbol{\sigma}-\tfrac{7}{5}\boldsymbol{q}\nabla\cdot\boldsymbol{v}-\tfrac{4}{5}\boldsymbol{q}\cdot\nabla\boldsymbol{v}+\tfrac{7}{2}R\boldsymbol{\sigma}\cdot\nabla\theta+\tfrac{\boldsymbol{\sigma}}{\rho}\cdot\nabla p\big\},\\
\boldsymbol{\sigma} &= \tau_\sigma\big\{p[\nabla\boldsymbol{v}+(\nabla\boldsymbol{v})^T-\tfrac{2}{3}\nabla\cdot\boldsymbol{v}\mathbf{1}]-\dot{\boldsymbol{\sigma}}-2[\boldsymbol{\sigma}\nabla\boldsymbol{v}+(\boldsymbol{\sigma}\nabla\boldsymbol{v})^T]\\
&\qquad +\tfrac{2}{5}[\nabla\boldsymbol{q}+(\nabla\boldsymbol{q})^T-\tfrac{2}{3}\nabla\cdot\boldsymbol{q}\mathbf{1}]-\boldsymbol{\sigma}\nabla\cdot\boldsymbol{v}\big\}, \qquad (15)
\end{aligned}$$

where $\tau_\sigma$ and $\tau_q$ are the respective relaxation times for stress and heat flux, and $R$ is the species-specific gas constant in the ideal-gas law $R=\rho T/p$. By means of the the leading linear terms in (15), we identify the Newtonian shear viscosity and the Fourier conductivity, respectively, as

$$\begin{aligned}
\mu &= p\tau_\sigma,\\
\kappa &= 5Rp\tau_q/2 = 5(\tau_q/\tau_\sigma)R\mu/2.
\end{aligned} \qquad (16)$$

Taking $\tau_q = 3\tau_\sigma/2$ one recovers a standard approximation $\kappa \doteq 5c_V\mu/2$ for smooth spherically-symmetric molecules [Chapman and Cowling 1960, p. 273] with specific heat $c_V = 3R/2$.

It is a straighhtforward matter to linearize the equations (15) about a uniform state of density $\rho_0$, temperature $T_0$ and pressure $p_0 = p_{\text{eq}}(\rho_0, T_0)$, since terms involving products of quantities that vanish in the uniform state do not contribute to the linearized equations. The function $p_{\text{eq}}$ introduced here represents the equilibrium equation of state, which is of course given by the above ideal-gas law for dilute gases, but we allow here a more general equation of state.

Letting

$$\begin{aligned}
\mu &= \rho_0\nu = \frac{\mu_0}{(1+\tau_{\sigma 0}s)}, & \tau &= \frac{4\tau_{\sigma 0}}{5(1+\tau_{\sigma 0}s)},\\
\kappa &= \rho_0 c_{p0}\alpha = \frac{\kappa_0}{(1+\tau_{q0}s)}, & f_0 &= \frac{2T_0}{5p_0},
\end{aligned} \qquad (17)$$

one finds that the linearized equations take on this compact form in Fourier space:

$$\begin{aligned}
&& \hat{\boldsymbol{\sigma}} &= 2\mu\hat{\boldsymbol{\epsilon}} + \imath\tau[(\boldsymbol{k}\otimes\hat{\boldsymbol{q}}+\hat{\boldsymbol{q}}\otimes\boldsymbol{k})/2-(\hat{\boldsymbol{q}}\cdot\boldsymbol{k})\mathbf{1}/3]\\
&\text{and} & \hat{\boldsymbol{q}} &= -\imath\kappa\boldsymbol{k}\hat{\theta}+\imath\kappa f_0\boldsymbol{k}\cdot\hat{\boldsymbol{\sigma}}, \qquad (18)\\
&\text{where} & \hat{\boldsymbol{\epsilon}} &= \imath[(\boldsymbol{k}\otimes\hat{\boldsymbol{v}}+\hat{\boldsymbol{v}}\otimes\boldsymbol{k})/2-(\boldsymbol{k}\cdot\hat{\boldsymbol{v}})\mathbf{1}/3].
\end{aligned}$$

After a bit of algebra, one can solve equations (18) for $\hat{\boldsymbol{q}}$, $\hat{\boldsymbol{\sigma}}$ in terms of $\hat{\theta}$, $\hat{\boldsymbol{v}}$, to give

$$\begin{aligned}
\hat{\boldsymbol{q}} &= -\kappa\{\imath\boldsymbol{k}\hat{\theta}f_1+\mu f_0 f_2[k^2\hat{\boldsymbol{v}}+f_1\boldsymbol{k}\otimes\boldsymbol{k}\cdot\hat{\boldsymbol{v}}/2]\},\\
\hat{\boldsymbol{\sigma}} &= 2\imath\mu f_2[(\boldsymbol{k}\otimes\hat{\boldsymbol{v}}+\hat{\boldsymbol{v}}\otimes\boldsymbol{k})/2-(\boldsymbol{k}\cdot\hat{\boldsymbol{v}})\mathbf{1}/3]+\kappa\tau f_1(\boldsymbol{k}\otimes\boldsymbol{k}-k^2\mathbf{1}/3)[\hat{\theta}-\imath\mu f_0 f_2(\boldsymbol{k}\cdot\hat{\boldsymbol{v}})/3],\\
&\text{where} \quad f_1 = [1+2(\lambda k)^2/3]^{-1}, \quad f_2 = [1+(\lambda k)^2/2]^{-1}, \quad \lambda = \sqrt{\kappa\tau f_0}. \qquad (19)
\end{aligned}$$

The terms in $\hat{\theta}$ appearing in the expression for $\hat{\boldsymbol{\sigma}}$ represent a non-local form of Maxwell's celebrated thermal stress [Maxwell 1879, Eqs. (53)–(54)] in a rarefied gas. According to Maxwell's kinetic theory, a good estimate of the magnitude of this stress relative to the Newtonian viscous stress is $\nu_0|\nabla^2\theta|/T_0\sqrt{\mathrm{tr}(\boldsymbol{\epsilon}^2)}$, where $\nu_0 = \mu_0/\rho_0$ is the kinematic viscosity and $\sqrt{\mathrm{tr}(\boldsymbol{\epsilon}^2)}$ the effective shear rate. Hence, according to the kinetic theory, the thermal stress will generally to be important only in the slow shearing of a rarified gas, as pointed out by Maxwell and noted in [G10].

Comparing with the general forms (7), one finds that the coefficients $\hat{A}, \hat{B}, \dots, \hat{H}$ are given by

$$\begin{aligned} &\hat{A} = -\imath\kappa f_1, && \hat{B} = -\kappa\mu k^2 f_0 f_2, && \hat{C} = -\kappa\mu f_0 f_1 f_2/3, \\ &\hat{E} = \kappa\tau f_1, && \hat{G} = \imath\mu f_2, && \hat{H} = -\imath\kappa\tau\mu f_0 f_1 f_2/3, \end{aligned} \tag{20}$$

and it is easy to obtain expansions in $k^2$ of the type (9).

According to the kinetic theory of dilute monatomic gases, the irreversible contribution to pressure vanishes [Chapman and Cowling 1960; Müller and Ruggeri 1998], implying that the coefficients $\hat{I}$, $\hat{K}$ in (12) are zero and hence that

$$\hat{D} = -\hat{E}k^2/3 \quad \text{and} \quad \hat{F} = -2\hat{G}/3 - \hat{H}k^2/3, \tag{21}$$

determining the remaining coefficients $\hat{D}$, $\hat{F}$. However, one should not expect these relations to hold for more general fluids, such as liquids and polyatomic gases, whose bulk viscosity $\beta_0 = \imath\hat{J}_0 = -\imath(\hat{F}_0 + 2/3\hat{G}_0)$ is generally non-zero.

The terms in (17) of the form $(1 + \tau s)$ represent exponential relaxation in the time domain. As such, they describe Maxwell's viscoelasticity and Cataneo's heat conduction, which admit both mechanical shear waves and heat waves, reflecting a breakdown of purely diffusive, dissipative response on time scales $\tau$. We recall that Ignaczak and Ostoja-Starzewski [2009] give a comprehensive treatment of the local theory of finite thermoelastic wave speeds, represented by terms $O(k)$ in (19). By contrast, and as anticipated above, we expect dissipative response to arise in the small Deborah number limit $De = \tau_0 s \ll 1$.

It is shown in the Appendix that one can analytically determine the inverse transforms of the coefficients in (20) by means of the formula (42), thereby providing the kernels in the integral operator (14). This provides a fully non-local model which should be much superior to weakly non-local models involving a sequence of higher spatial gradients, since integral operators, in contrast to differential operators, are generally bounded. This is especially significant in the neighborhood of singularities, as illustrated by the well-known work of Eringen [2002, Section 6.14] on crack-tip stresses in linear elasticity.

## 4. Bivelocity model

Here, we analyze here a recent version of Brenner's bi-velocity model [2009], in order to compare it with the linear theory proposed above. Given that Brenner's modeling rests heavily on appeals to linear irreversible thermodynamics ("LIT"), it is appropriate to employ a fully linearized version of the type employed in the present paper and, also, to restrict the analysis to dissipative response.

Since Brenner employs somewhat special variables and notation, we have included Table 1 below to clarify the relation of his variables to those employed in the present work.

We adopt that form of Brenner's model which he deems appropriate to creeping (inertialess) flow[4], as represented by Eqs. (2.7), (2.12) and (2.13) of [Brenner 2009]. In the present notation, these become

$$\begin{aligned}
\boldsymbol{q} &= -\kappa_0\nabla\theta + L_{12}\nabla p_{eq} - p_{eq}\,\boldsymbol{j}_w \doteq -\kappa_0\nabla\theta + L_{12}[(\partial_\theta p)_0\nabla\theta + (\partial_\rho p)_0\nabla\rho] - p_0\,\boldsymbol{j}_w\\
\boldsymbol{j}_w &= -L_{21}T^{-1}\nabla\theta + L_{22}\nabla p_{eq} \doteq -L_{21}T_0^{-1}\nabla\theta + L_{22}[(\partial_\theta p)_0\nabla\theta + (\partial_\rho p)_0\nabla\rho]\\
\boldsymbol{\sigma} &= 2\mu_0\overline{\nabla\boldsymbol{v}_w} = 2\mu_0[\overline{\nabla\boldsymbol{v}} + \overline{\nabla\boldsymbol{j}}_w], \qquad p = -\beta_0\nabla\cdot\boldsymbol{v}_w = -\beta_0\nabla\cdot(\boldsymbol{v}+\boldsymbol{j}_w),
\end{aligned} \tag{22}$$

where overbars represent symmetric deviators and $\doteq$ denotes the approximation arising from linearization about the uniform base state employed elsewhere in the present article. In Brenner's model, the coefficients $L_{ij}$ are assumed to describe a dissipative linear system, with corresponding Onsager symmetry $L_{21} = L_{12}$.

[4] Otherwise, his constitutive equations appear to contain inertial terms that are hard to reconcile with the principle of material frame indifference.

| Quantity | [Brenner 2009] | Present |
|---|---|---|
| absolute temperature | $T$ | $T_0+\theta$ |
| barycentric velocity | $\boldsymbol{v}_m$ | $\boldsymbol{v}$ |
| "work"[3] or "volume" velocity | $\boldsymbol{v}_w$ | $\boldsymbol{v}_w$ |
| diffuse "volume" flux | $\boldsymbol{j}_w = \boldsymbol{v}_w - \boldsymbol{v}_m$ | $\boldsymbol{j}_w = \boldsymbol{v}_w - \boldsymbol{v}$ |
| pressure tensor | $\boldsymbol{P}$ | $p_{eq}\mathbf{1} - \boldsymbol{\sigma}$ |
| pressure | $\overline{p} = \mathrm{tr}(\boldsymbol{P})/3$ | $p = p_{eq} - \mathrm{tr}(\boldsymbol{\sigma})/3$ |
| "thermodynamic" pressure | $p$ | $p_{eq}$ |
| shear stress | $\boldsymbol{T}$ | $\boldsymbol{\tau} = \boldsymbol{\sigma} + p\boldsymbol{I}$ |
| heat flux | $\boldsymbol{j}_u$ | $\boldsymbol{q}$ |
| "entropic" heat flux | $\boldsymbol{q} = \boldsymbol{j}_u + p\,\boldsymbol{j}_w$ | $\boldsymbol{q} + p_{eq}\,\boldsymbol{j}_w$ |
| thermal conductivity for $\boldsymbol{q}$ | $k$ | $\kappa_0$ |
| shear and bulk viscosity | $\eta$, $\zeta$ | $\mu_0$, $\beta_0$ |

**Table 1.** Variables and notation.

To compare with the present constitutive theory, it suffices to eliminate $\boldsymbol{j}_w$ from (22), and it is algebraically expedient to express these relations as Fourier–Laplace transforms. Account taken of the linearized mass balance (cf. Eqs. (27) below), one thereby obtains relations of the form (7) and (12), with

$$\begin{aligned}
&\hat{A} = \imath\,\{[L_{12} - p_0 L_{22}](\partial_\theta p)_0 - \kappa_0 + L_{12} p_0/T_0\}\,, \; \hat{B} = 0,\\
&\hat{C} = [L_{12} - p_0 L_{22}]\rho_0(\partial_\rho p)_0/s, \quad \hat{E} = -2\mu_0[L_{22}(\partial_\theta p)_0 - L_{12}/T_0],\\
&\hat{G} = \imath\mu_0, \quad \hat{H} = 2\imath\mu_0\rho_0(\partial_\rho p)_0 L_{22}/s,\\
&\hat{I} = \beta_0[L_{22}(\partial_\theta p)_0 - L_{12}/T_0]k^2 = -(\hat{D} + \hat{E}k^2/3),\\
&\hat{J} = -\imath\beta_0[1 + \rho_0(\partial_\rho p)_0 L_{22}k^2/s] = -(\hat{F} + 2\hat{G}/3 + \hat{H}k^2/3),
\end{aligned} \tag{23}$$

from which it follows that

$$\beta_0 = 2\mu_0/3, \quad \hat{D} = 0. \quad \hat{F} = 0 \tag{24}$$

Note that [Brenner 2009] takes

$$L_{12} = T_0\alpha_0\beta_0, \quad \text{where} \quad \alpha_0 = \kappa_0/\rho_0 c_{p0}, \; \beta_0 = -(\partial_\theta\rho)_{p0}/\rho_0 = (\partial_\theta p)_0/\rho_0(\partial_\rho p)_0,$$

involving the thermal diffusivity $\alpha_0$ and isobaric coefficient of thermal expansion $\beta_0$. With certain reservations, he then takes $L_{22} = \alpha_0\beta_0/(\partial_\theta p)_0$, which would imply that $L_{22}(\partial_\theta p)_0 - L_{12}/T_0$ and, hence, $\hat{I}$ and $\hat{E}$ vanish in (23). Thus, the Maxwell thermal stress represented by the term $\hat{E}$ in (7) also vanishes.

Brenner does not invoke the restrictions on the coefficients of viscosity $\mu_0$, $\beta_0$ that are required for consistency with the general model proposed in this work.

## 5. Application to linear thermo-acoustic waves

For the uniform fluid at rest, we adopt mechanical and caloric equations of state connecting equilibrium pressure and specific internal energy to temperature and density:

$$p = p_{\mathrm{eq}}(\theta, \rho) \quad \text{and} \quad \varepsilon = \varepsilon_{\mathrm{eq}}(\theta, \rho), \tag{25}$$

with

$$\partial_\theta\varepsilon_{\mathrm{eq}} = c_v, \quad \partial_\rho\varepsilon_{\mathrm{eq}} = \frac{1}{\rho^2}\left[p - \theta(\partial_\theta p_{\mathrm{eq}})\right], \tag{26}$$

where $c_v$ denotes the isochoric specific heat.

The present treatment of temperature and density as independent variables is inspired by the modern literature on continuum thermodynamics, where various intensive variables are given as derivatives of Helmholtz free energy. It seems to us more natural than the formulation based on pressure and entropy adopted in standard treatises on acoustics [Pierce 1981] but in any case can be easily converted to the latter. Accordingly, we shall refer to the "entropy mode" identified by [Pierce 1981, p. 523], and subsequently by [Davis and Brenner 2012], as the "thermal mode", noting that the modal amplitudes are simply related by a constant of proportionality [Pierce 1981, Eq. (10-3.16)] according to the linear theory which follows.

Thus, with subscripts 0, 1 referring, respectively to a uniform equilibrium state and a small perturbation on that state, such that $\zeta = \zeta_0 + \zeta_1$, for any variable $\zeta$, the linearized balances of momentum, mass, and energy reduce in the absence of body forces or radiant energy transfer to:

$$\begin{aligned} \rho_0 \partial_t \boldsymbol{v}_1 &= -(\partial_\theta p)_0 \nabla\theta_1 - (\partial_\rho p)_0 \nabla\rho_1 + \nabla\cdot\boldsymbol{\sigma}_1, \\ \text{where}\quad (\partial_z p)_0 &= \partial_z p_{\text{eq}}\big|_{T_0,\rho_0}, \quad z = \theta,\, \rho, \quad \partial_t\rho_1 = -\rho_0\nabla\cdot\boldsymbol{v}_1, \\ \text{and}\quad \rho_0 c_{v0}\partial_t\theta_1 &= \nabla\cdot\boldsymbol{q}_1 - T_0(\partial_\theta p)_0\nabla\cdot\boldsymbol{v}_1. \end{aligned} \tag{27}$$

This stated, we shall now drop the subscript 1 on perturbations, as done implicitly in the preceding discussion, where (7) provides constitutive equations for the perturbed heat flux and stress $\boldsymbol{q}$ and $\boldsymbol{\sigma}$ in terms of $\theta$ and $\rho$.

Other than an assumption of a dissipative regime for small $s$, we shall not consider in detail the restrictions on the constitutive model arising from the entropy balance (the Clausius–Duhem inequality) and the related "extended thermodynamics" [Müller and Ruggeri 1998]. However, we note that if heat flux is neglected from (27) the last two members of (27) yield the condition of constant equilibrium entropy $\eta_{\text{eq}}$:

$$\rho_0\partial_t\eta_{\text{eq}} = \rho_0 c_{v0}\partial_t\theta/T_0 - (\partial_\theta p)_0\partial_t\rho/\rho_0 = 0, \tag{28}$$

whereas the actual entropy $\eta$ may generally increase owing to thermo-mechanical dissipation.

Modulo inhomogeneous terms arising from initial values of $\theta$, $\rho$, $\boldsymbol{v}$, the Fourier–Laplace transforms of (27) reduce to the linear homogeneous form:

$$\begin{aligned} \rho_0 s\hat{\boldsymbol{v}} + \imath(\partial_\theta p)_0\hat{\theta}\boldsymbol{k} + \frac{\rho_0}{s}(\partial_\rho p)_0\boldsymbol{k}\boldsymbol{k}\cdot\hat{\boldsymbol{v}} - \imath\hat{\boldsymbol{\sigma}}\boldsymbol{k} &= \boldsymbol{0}, \\ \rho_0 c_{v0}\hat{\theta} + \imath\boldsymbol{k}\cdot\hat{\boldsymbol{q}} + \imath T_0(\partial_\theta p)_0\boldsymbol{k}\cdot\hat{\boldsymbol{v}} &= 0. \end{aligned} \tag{29}$$

Substitution of (7) into (29) yields a set of four linear equations in $\hat{\theta}$, $\hat{\boldsymbol{v}}$. However, these can be reduced to a set of two linear equations in $\hat{\theta}$, $\nabla\cdot\boldsymbol{v}$ by employing the "divergence" form obtained by taking the dot product of $\boldsymbol{k}$ with the first member of (29). The determinantal equation results then in the *dispersion relation* for the resultant compressive modes:

$$\begin{aligned} &[\hat{D} + \hat{E}k^2 - (\partial_\theta p)_0]\,[\hat{B} + \hat{C}k^2 + T_0(\partial_\theta p)_0]k^2 \\ &\qquad - [\rho_0 c_{v0}s + \imath\hat{A}k^2]\,[\rho_0 s - \{2\imath\hat{G} + \imath\hat{F} - \rho_0(\partial_\rho p)_0/s\}k^2 - \imath\hat{H}k^4] = 0. \end{aligned} \tag{30}$$

In the application of this relation to the time-periodic waves with temporal frequency $\omega$ it is understood that $s = -i\omega$ here and below.

In addition to the modes described by (30) there exists a decoupled "vorticity" or shearing mode involving the vorticity $\boldsymbol{w} = \nabla \times \boldsymbol{v}$ [Davis and Brenner 2012; Pierce 1990]. By means of the Fourier representation $\hat{\boldsymbol{w}} = \imath \boldsymbol{k} \times \hat{\boldsymbol{v}}$ and the cross product of $\boldsymbol{k}$ with the first member of (29), one obtains

$$[\rho_0 s - \imath \hat{G}(k^2, s)k^2]\hat{\boldsymbol{w}} = \boldsymbol{0}, \tag{31}$$

which has an immediate interpretation in terms of the inverse $G(\boldsymbol{x}, t)$ of the transform $\hat{G}$. As indicated by the analysis in the Appendix, (31) describes shear waves on time scales $\tau_{\sigma 0}$. By contrast, in the dissipative regime that emerges on longer time scales one obtains strongly damped diffusive modes [Davis and Brenner 2012; Pierce 1990].

In sum, given the Fourier–Laplace inverses $A(\boldsymbol{x}, t)$, $B(\boldsymbol{x}, t), \ldots, H(\boldsymbol{x}, t)$, the relations (30)-(31) provide a fully non-local model of linear signal propagation, including long-range memory effects in time. Clearly, a more restricted form is required for most practical applications. Thus, the retention of terms up to $O(k^2)$ in (7) reduces the (30) to

$$\begin{aligned}[\hat{D}_0 + (\hat{D}_1 + \hat{E}_0)k^2 - (\partial_\theta p)_0]\,[(\hat{B}_1 + \hat{C}_0)k^2 + T_0(\partial_\theta p)_0]k^2 \\ - [\rho_0 c_{v0} s + \imath \hat{A}_0 k^2]\,[\rho_0 s - \{2\imath \hat{G}_0 + \imath \hat{F}_0 - \rho_0(\partial_\rho p)_0/s\}k^2] = 0,\end{aligned} \tag{32}$$

which involves the five distinct coefficients $\hat{A}_0$, $\hat{B}_1+\hat{C}_0$, $\hat{D}_0$, $\hat{D}_1+\hat{E}_0$ and $\hat{F}_0+2\hat{G}_0$, with dependence on $s$ representing relaxation effects in the time domain.

By a slight extension of the kinetic theory of Section 3, four of the coefficients appearing in (32) and the coefficient appearing in the limiting form of (31),

$$[\rho_0 s - \imath \hat{G}_0(s)k^2]\hat{\boldsymbol{w}} = \boldsymbol{0}, \tag{33}$$

are given respectively by

$$\begin{aligned} &\imath \hat{A}_0 = \frac{\kappa_0}{1+\tau_{q0}s}, \qquad \hat{B}_1 + \hat{C}_0 = -\frac{4\kappa_0\mu_0 f_0}{3(1+\tau_{\sigma 0}s)(1+\tau_{q0}s)}, \\ &\hat{D}_1 + \hat{E}_0 = \frac{8\kappa_0\tau_{\sigma 0}}{5(1+\tau_{q0}s)}, \quad \imath \hat{F}_0 + 2\hat{G}_0 = \frac{\beta_0 + 4\mu_0/3}{(1+\tau_{\sigma 0}s)}, \quad \hat{G}_0 = \frac{\imath \mu_0}{1+\tau_{\sigma 0}s}. \end{aligned} \tag{34}$$

Note that the form of $\hat{G}_0$ and (33) imply elastic shear waves in at high frequencies $s \to \infty$, thereby eliminating infinite propagation speeds associated with the dissipative limit $s \to 0$ [Davis and Brenner 2012].

Following [Brenner 2009] and [Davis and Brenner 2012], we have included a bulk viscosity coefficient $\beta_0$, but which now involves elastic relaxation[5]. The coefficient $\beta_0$ vanishes according to the monatomic kinetic theory, as does the remaining unspecified coefficient $\hat{D}_0(s)$. Otherwise, we note from (7) that $\hat{D}_0(s)$ involves a non-equilibrium response of pressure to temperature variation. We further note that a similar relaxation effect in the temperature-energy response would be obtained upon replacing the specific heat $c_{v0}$ by an $s$-dependent term $\hat{c}_{v0}(s)$, analogous to the formalism proposed in [Goddard 1992].

In the dissipative model obtained by neglecting terms $\tau s$ and taking $\hat{D}_0 = 0$ (by the Onsager symmetry discussed above), the dispersion relation (34) reduces to a cubic in both $s$ and $k^2$, whereas the model considered by [Davis and Brenner 2012] is cubic in $s$ but quadratic in $k^2$.

For the classical Navier–Stokes–Fourier model we have

$$\begin{aligned} \hat{A} &= \hat{A}_0 = -\imath\kappa_0, & \hat{B} &= \hat{C} = \hat{D} = \hat{H} = 0, \\ \hat{G} &= \hat{G}_0 = \imath\mu_0, & \hat{F}_0 + 2\hat{G}_0 &= -\imath(\beta_0 + 4\mu_0/3), \end{aligned} \tag{35}$$

and the dispersion relation (32) reduces to

$$\begin{gathered} [s+\alpha_0\gamma k^2]\,[s^2-(\beta_0+4\mu_0/3)sk^2/\rho_0+(\partial_\rho p)_0 k^2]-(\gamma-1)k^2 s=0, \quad \text{where} \\ \alpha_0=\kappa_0/\rho_0 c_{p0} \quad \text{and} \quad \gamma=c_{p0}/c_{v0}=c_S^2/c_T^2=1+T_0(\partial_\theta p)_0^2/\rho_0^2 c_{v0}(\partial_\rho p)_0, \end{gathered} \tag{36}$$

with $c_S$ and $c_T$ denoting, respectively, the isentropic and isothermal speeds of sound, whose ratio is given by the specific heat ratio $\gamma$. It is easy to show that (36) is identical with the form given in [Davis and Brenner 2012, Eq. (11)] if $(\beta_0 + 4\mu_0/3)/\rho_0$ is replaced by the equivalent quantity $(2\nu_0 + \lambda_0)$ in their analysis and $-\rho_0^2(\partial_\rho p)_0$ is identified as the isothermal compressibility.

The more general versions (32) and (34) can be written in the non-dimensional form as

$$\begin{aligned} &(a\tilde{k}^2+b)(c\tilde{k}^2+d)\tilde{k}^2\tilde{s}-(\tilde{s}+e\tilde{k}^2)(\tilde{s}^2+f\tilde{k}^2\tilde{s}+g\tilde{k}^2)=0, \\ &\quad \text{where} \quad \tilde{k}^2=\tau_{\sigma 0}\nu_0 k^2, \quad \tilde{s}=\tau_{\sigma 0}s, \\ &\quad \text{with} \quad a=8\gamma\alpha/5\nu_0, \quad b=-(\partial_\theta p)_0/\rho_0 c_{v0}, \\ &\qquad\qquad c=-8\alpha\nu\rho_0 c_{p0}T_0/15\nu_0^2 p_0, \quad d=T_0(\partial_\theta p)_0\tau_{\sigma 0}/\mu_0, \quad e=\gamma\alpha/\nu_0, \\ &\qquad\qquad f=-(\beta+4\mu/3)/\mu_0, \quad g=(\partial_\rho p)_0\tau_{\sigma 0}/\nu_0=c_T^2\tau_{\sigma 0}/\nu_0. \end{aligned} \tag{37}$$

Note that $\tilde{k}^2$ involves a squared length $\tilde{\lambda}_0^2 = \tau_{\sigma 0}\nu_0$, which is related by a factor $|c|$ to that introduced in the Appendix. Note also that for dilute gases all the coefficients $a, b, \ldots, g$ are of order unity, so that the polynomial in the first equation of (37)

[5] The quantity $(\beta_0 + 4/3\mu_0)/\rho_0$ is equal to the quantity $\lambda + 2\nu$ in equations (14) and (23) of [Davis and Brenner 2012], who employ the unconventional designation of $\rho_0\lambda$ in their equation (2) as bulk viscosity.

is “well-tempered”, that is has derivatives that are all of comparable magnitude for arguments near unity.

Casting (37) in the standard form of a cubic equation in $z = \tilde{k}^2$:

$$\mathbb{A}z^3 + \mathbb{B}z^2 + \mathbb{C}z + \mathbb{D} = 0, \quad \text{with} \tag{38}$$
$$\mathbb{A} = ac\tilde{s}, \quad \mathbb{B} = (ad + bc - ef)\tilde{s} - eg, \quad \mathbb{C} = (bd - g)\tilde{s} - (e + f)\tilde{s}^2, \quad \mathbb{D} = -\tilde{s}^3,$$

the three roots are given by the well-known formula

$$z_k = -\frac{\mathbb{B}}{3\mathbb{A}}\left[1 + 2(1 - 3\mathbb{A}\mathbb{C}/\mathbb{B}^2)^{1/2}\cos\frac{2k\pi + \phi}{3}\right] \quad \text{for} \quad k = 0, 1, 2,$$
$$\text{with} \quad \phi = \cos^{-1}\zeta, \quad \text{where} \quad \zeta = \frac{1 - 9\mathbb{A}\mathbb{C}/2\mathbb{B}^2 + 27\mathbb{A}^2\mathbb{D}/2\mathbb{B}^3}{(1 - 3\mathbb{A}\mathbb{C}/\mathbb{B}^2)^{-3/2}}. \tag{39}$$

The quantities involved in (39) are generally complex, and we can express the complex circular function appearing there in terms of elementary functions as

$$\cos\frac{2k\pi + \phi}{3} = \tfrac{1}{2}\big(e^{(2k\pi+\phi)\imath/3} + e^{-(2k\pi+\phi)\imath/3}\big)$$
$$= \tfrac{1}{2}\big(e^{2k\pi\imath/3}\big[\zeta + \imath\sqrt{1-\zeta^2}\,\big]^{1/3} + e^{-2k\pi\imath/3}\big[\zeta - \imath\sqrt{1-\zeta^2}\,\big]^{1/3}\big), \quad k = 0, 1, 2, \tag{40}$$

or, by means of yet other well-known formulae [Abramowitz and Stegun 1965, equations 15.1.3-19] in terms of hypergeometric functions $F = {}_2F_1$ as

$$\cos\frac{2k\pi + \phi}{3} = -\frac{1}{2}\Big[\cos\frac{\phi}{3} \mp \sqrt{3}\sin\frac{\phi}{3}\Big], \quad \text{for} \quad k = 1, 2, \quad \text{where}$$
$$\cos\frac{\phi}{3} = F\left(-\tfrac{1}{6}, \tfrac{1}{6}; \tfrac{1}{2}; 1 - \zeta^2\right), \quad \sin\frac{\phi}{3} = \tfrac{1}{3}\sqrt{1-\zeta^2}\,F\left(\tfrac{1}{3}, \tfrac{2}{3}; \tfrac{3}{2}; 1 - \zeta^2\right), \tag{41}$$

with appropriate branch cuts for $\sqrt{1-\zeta^2}$ and with $\zeta = \cos\phi$ given by the last equation of (39).

***Comparison to the bi-velocity model.*** We recall that the classical dispersion relation as well as the modification proposed by [Davis and Brenner 2012] involve a quadratic equation for $z$ in lieu of (38). It is clear that such a quadratic arises from (38) for $|\tilde{s}| \ll 1$, which is characteristic of the dissipative regime represented by the previous studies. Indeed, by neglecting terms $O(\tilde{s}^3)$, one obtains a quadratic similar to that given by equation (23) of [Davis and Brenner 2012], with generally different coefficients. This gives a dispersion relation that is quadratic in both $k^2$ and $s$ representing a PDE that is quadratic in $\nabla^2$ and $\partial_t$, for which [Davis and Brenner 2012] offer some special solutions that suggest experiments to arrive at the correct coefficients in the dispersion relations.

## Extension to solids and Cosserat media

With minor modifications the preceding analysis applies to graded (also known as "higher-gradient") isotropic linear thermo-viscoelastic solids. For this purpose, it suffices to allow for static stress by taking account of (4) and including terms that behave like $s^{-1}$ for $s \to 0$ in the coefficients $\hat{D}, \dots, \hat{H}$ in $(7)_2$. It can be noted that the coefficient $\hat{D}$ serves to describe both static and dynamic thermoelasticity and, as with the fluids considered above, the static contribution can be included in an equation of state for equilibrium pressure $p_{\text{eq}}(\theta, \rho)$.

Without pursuing the algebraic details, we note that the strain-gradient theory of [Mindlin 1964], anticipated by the seminal works of Piola [dell'Isola et al. 2015], yields dispersion relations for both dilatational and shear waves that are quadratics in $k^2$ [Mindlin 1964, (9.34)] with a much simpler form than (32) and (37).

As an extension of [Mindlin 1964], one may treat a more general Cosserat thermo-viscoelasticity by addition to the list of variables in (2) and (5) the Cosserat rotation vector $\boldsymbol{\vartheta} = [\vartheta_i]$ and the moment stress $\boldsymbol{\sigma}^{(2)} = [\sigma_{ij}^{(2)}]$, conjugate to $\nabla\boldsymbol{\vartheta}$, and by replacing the stress $\boldsymbol{\sigma}$ by a non-symmetric tensor with antisymmetric part defining a vector conjugate to $\boldsymbol{\vartheta}$. Under the rubric of micropolar elasticity, [Eringen 1984] has already given a comprehensive analysis for the isothermal case that leads to a cubic in $k^2$ as dispersion relation, and [Abreu et al. 2017] provide a similar analysis with a view to the emerging field of rotational seismology.

Finally, we note that the present type of analysis can be extended to anisotropic media like those considered by [Suiker et al. 2001] by appropriate symmetry restrictions and modification of the relations (6). One possibility is to employ the joint isotropic invariants of the wave vector $\boldsymbol{k}$ and a set of structure tensors to capture the anisotropy [Cowin 1985; Man and Goddard 2016].

## Conclusions

The abstract provides a generally adequate summary of the present work. It is worth emphasizing that the Burnett-order linearized Chapman–Enskog kinetic theory is subsumed by the general wave-number expansions proposed in the present work, which gives more general thermo-viscous response than that of Brenner's bi-velocity model, while also allowing for themo-viscoelastic behavior.

As matter for future work, it would be interesting to consider the utility of non-local models in resolving certain fluid-mechanical singularities, such as three-phase contact lines, which bear a certain resemblance to the linear-elastic singularities around crack tips addressed by the non-local elasticity of [Eringen 2002].

## Appendix: Inverse transforms

We recall that the inverse Fourier transform of a function $\hat{f}(k)$, with $k^2 = k_i k_i$, is a function of $r = |\boldsymbol{x}|$ given by the *radial* form [Gradshteyn and Ryzhik 2000]

$$\begin{aligned} f(\boldsymbol{x}) &= \frac{1}{\sqrt{8\pi^3}} \int e^{\imath \boldsymbol{k}\cdot\boldsymbol{x}} \hat{f}(k) \sin\hat{\vartheta} k^2 \,\mathrm{d}k \,\mathrm{d}\hat{\vartheta}\,\mathrm{d}\hat{\varphi} \\ &= -\frac{1}{\sqrt{2\pi}} \int_{k=0}^{\infty} \int_0^{\pi} e^{\imath k r \cos\hat{\vartheta}} \hat{f}(k) \sin\hat{\vartheta}\,\mathrm{d}\hat{\vartheta} k^2 \,\mathrm{d}k \\ &= \sqrt{\frac{2}{\pi}} \int_0^{\infty} \operatorname{sinc}(kr) \hat{f}(k) k^2 \,\mathrm{d}k, \quad \text{where } \operatorname{sinc}(z) = z^{-1}\sin z. \end{aligned} \tag{42}$$

One can use (42) to derive the inverse transforms $K(t,\boldsymbol{x}) = A(t,\boldsymbol{x}), B(t,\boldsymbol{x}), \dots$ of the coefficients (8), noting that the functions $f_1$, $f_2$ in (19) can be written for $i = 1, 2$ as

$$\begin{aligned} &f_i = (1+\lambda_i^2 k^2)^{-1}, \quad \text{where} \quad \lambda_i^2 = b_i \lambda_0^2 (1+\tau_1 s)^{-1}(1+\tau_2 s)^{-1}, \\ &\lambda_0^2 = 4\kappa_0 \tau_1 f_0/5, \quad b_1 = 1/2, \quad b_2 = 2/3, \quad \tau_1 = \tau_{\sigma 0}, \quad \tau_2 = \tau_{q0}, \end{aligned} \tag{43}$$

Now, the coefficients in (8) can all be expressed as affine forms in $f_1$, $f_2$, since

$$k^2 f = k^2(1+\lambda^2 k^2)^{-1} = (1-f)/\lambda^2, \quad \text{and} \quad f_1 f_2 = \frac{b_1}{b_1 - b_2} f_1 + \frac{b_2}{b_2 - b_1} f_2 \tag{44}$$

First, note that substitution of $\hat{f} = (1+\lambda^2 k^2)^{-1}$ into (42) gives

$$f(\boldsymbol{x}) = \sqrt{\frac{2}{\pi r^2}} \int_0^{\infty} \frac{k}{1+\lambda^2 k^2} \sin kr \,\mathrm{d}k = \sqrt{\frac{\pi}{2\lambda^4 r^2}} \exp\Big(-\frac{r}{\lambda}\Big).$$

Second, note that

$$\exp\Big(-\frac{r}{\lambda}\Big) = \exp\big(-\gamma\sqrt{s'^2 - a^2}\big), \qquad \text{where} \quad s' = s + \frac{\tau_1+\tau_2}{2\tau_1\tau_2}, \quad a = \frac{\tau_2 - \tau_1}{2\tau_1\tau_2}, \quad \gamma = \Big(\frac{\tau_1\tau_2}{b}\Big)^{1/2} r.$$

However, the inverse Laplace transform [Abramowitz and Stegun 1965]

$$g(t) = \mathscr{L}^{-1}\big\{\exp\big(-\gamma\sqrt{s^2-a^2}\big)\big\} = \delta(t-\gamma) + \frac{a\gamma}{\sqrt{t^2-\gamma^2}} I_1\big(a\sqrt{t^2-\gamma^2}\big) u(t-\gamma),$$

where $I_1(z)$ is the Bessel function of the second kind, $u(t)$ the Heaviside function and $\delta(t) = u'(t)$ the Dirac delta, gives

$$h(t) = \mathscr{L}^{-1}\big\{\exp\big(-\gamma\sqrt{s^2-a^2}\big)\big\} = \exp\Big(-\frac{\tau_1+\tau_2}{2\tau_1\tau_2} t\Big) g(t).$$

Thus, the coefficients in (20) are seen to involve various powers of $1+\tau_1 s$ and $1+\tau_2 s$ multiplying the above transforms, so that the inverse Laplace transform of the resulting products can in principle be obtained by convolution.

## Acknowledgement

I wish to acknowledge several helpful conversations with Professor A. M. J. Davis during the early stages of this work. Also, I am grateful for the hospitality accorded me during extended stays in the Laboratoire PMMH and the Institut Langevin of the École supérieure de physique et chimie industrielles (ESPCI), Paris, where part of this work was completed.

## References

[Abramowitz and Stegun 1965] M. Abramowitz and I. A. Stegun (editors), *Handbook of mathematical functions*, U.S. National Bureau of Standards, New York, 1965.

[Abreu et al. 2017] R. Abreu, J. Kamm, and A.-S. Reiß, "Micropolar modelling of rotational waves in seismology", *Geophys. J. Int* **210**:2 (2017), 1021–1046.

[Brenner 2009] H. Brenner, "Bi-velocity hydrodynamics: single-component fluids", *Internat. J. Engrg. Sci* **47**:9 (2009), 930–958.

[Brenner 2012] H. Brenner, "Beyond Navier–Stokes", *Internat. J. Engrg. Sci* **54** (2012), 67–98.

[Cattaneo 1948] C. Cattaneo, "Sulla conduzione de calore", *Atti del Semin. Mat. e Fis. Univ. Modena* **3**:3 (1948), 3–22.

[Chapman and Cowling 1960] S. Chapman and T. G. Cowling, *The mathematical theory of non-uniform gases: an account of the kinetic theory of viscosity, thermal conduction, and diffusion in gases*, Cambridge University Press, New York, 1960.

[Coleman 1964] B. D. Coleman, "Thermodynamics of materials with memory", *Arch. Rational Mech. Anal* **17** (1964), 1–46.

[Coleman and Noll 1961] B. D. Coleman and W. Noll, "Foundations of linear viscoelasticity", *Rev. Mod. Phys* **33** (1961), 239–249.

[Cowin 1985] S. C. Cowin, "The relationship between the elasticity tensor and the fabric tensor", *Mech. Materials* **4**:2 (1985), 137–147.

[Davis and Brenner 2012] A. M. J. Davis and H. Brenner, "Thermal and viscous effects on sound waves: revised classical theory", *J. Acoust. Soc. Am.* **132**:5 (2012), 2963–2969.

[dell'Isola et al. 2015] F. dell'Isola, U. Andreaus, and L. Placidi, "At the origins and in the vanguard of peridynamics, non-local and higher-gradient continuum mechanics: an underestimated and still topical contribution of Gabrio Piola", *Math. Mech. Solids* **20**:8 (2015), 887–928.

[Eringen 1984] A. C. Eringen, "Plane waves in nonlocal micropolar elasticity", *Int. J. of Eng. Sci.* **22**:8-10 (1984), 1113–1121.

[Eringen 1992] A. C. Eringen, "Vistas of nonlocal continuum physics", *Internat. J. Engrg. Sci* **30**:10 (1992), 1551–1565.

[Eringen 2002] A. C. Eringen, *Nonlocal continuum field theories*, Springer, 2002.

[Goddard 1992] J. D. Goddard, "History effects in transient diffusion through heterogeneous media", *Ind. Eng. Chem. Res.* **31**:3 (1992), 713–721.

[Goddard 2010] J. D. Goddard, "On material velocities and non-locality in the thermo-mechanics of continua", *Int. J. Eng. Sci.* **48**:11 (2010), 1279–1288. .

[Goddard and Lee 2017] J. D. Goddard and J. Lee, "On the stability of the $\mu(I)$ rheology for granular flow", *J. Fluid Mech* **833** (2017), 302–331.

[Gradshteyn and Ryzhik 2000] I. S. Gradshteyn and I. M. Ryzhik (editors), *Table of integrals, series, and products*, Academic Press, New York, 2000.

[Ignaczak and Ostoja-Starzewski 2009] J. Ignaczak and M. Ostoja-Starzewski, *Thermoelasticity with finite wave speeds*, Oxford University Press, New York, 2009.

[Man and Goddard 2016] C.-S. Man and J. Goddard, "Remarks on isotropic extension of anisotropic constitutive functions via structural tensors", *Math. Mech. Solids* **23**:4 (2016), 554–563.

[Maxwell 1867] J. C. Maxwell, "On the dynamical theory of gases", *Phil. Trans.* **157** (1867), 49–88.

[Maxwell 1879] J. C. Maxwell, "Stresses in rarefied gases arising from inequalities of temperature", *Phil. Trans.* **170** (1879), 231–256.

[Mindlin 1964] R. S. Mindlin, "Micro-structure in linear elasticity", *Arch. Ratl. Mech. Anal.* **16**:1 (1964), 51–78.

[Müller and Ruggeri 1998] I. Müller and T. Ruggeri, *Rational extended thermodynamics*, 2nd ed., Springer Tracts in Natural Philosophy **37**, Springer, 1998.

[Pierce 1981] A. D. Pierce, *Acoustics: an introduction to its physical principles and applications*, McGraw-Hill, New York, 1981.

[Pierce 1990] A. D. Pierce, "Wave equation for sound in fluids with unsteady inhomogeneous flow", *J. Acoust. Soc. Am.* **87**:6 (1990), 2292–2299.

[Suiker et al. 2001] A. Suiker, A. Metrikine, and R. D. Borst, "Comparison of wave propagation characteristics of the Cosserat continuum model and corresponding discrete lattice models", *Int. J. Solids Structs.* **38**:9 (2001), 1563–1583.

Received 11 Mar 2018. Accepted 24 Jul 2018.

JOE D. GODDARD: jgoddard@ucsd.edu
*Department of Mechanical and Aerospace Engineering, University of California, San Diego,*
*9500 Gilman Drive, La Jolla, CA 92093-0411, United States*

MATHEMATICS AND MECHANICS OF COMPLEX SYSTEMS
Vol. 6, No. 4, 2018
dx.doi.org/10.2140/memocs.2018.6.339

# HETEROGENEOUS DIRECTIONS OF ORTHOTROPY IN THREE-DIMENSIONAL STRUCTURES: FINITE ELEMENT DESCRIPTION BASED ON DIFFUSION EQUATIONS

RACHELE ALLENA AND CHRISTOPHE CLUZEL

Heterogeneous materials such as bone or woven composites show mesostructures whose constitutive elements are all oriented locally in the same direction and channel the stress flow throughout the mechanical structure. The interfaces between such constitutive elements and the matrix are regions of potential degradations. Then, when building a numerical model, one has to take into account the local systems of orthotropic coordinates in order to properly describe the damage behavior of such materials. This can be a difficult task if the orthotropic directions constantly change across the complex three-dimensional geometry as is the case for bone structures or woven composites. In the present paper, we propose a finite element technique to estimate the continuum field of orthotropic directions based on the main hypothesis that they are mainly triggered by the external surface of the structure itself and the boundary conditions. We employ two diffusion equations, with specific boundary conditions, to build the radial and the initial longitudinal unit vectors. Then, to ensure the orthonormality of the basis, we compute the longitudinal, the circumferential, and the radial vectors via a series of vector products. To validate the numerical results, a comparison with the average directions of the experimentally observed Haversian canals is used. Our method is applied here to a human femur.

## 1. Introduction

In order to simulate the mechanical behavior of heterogeneous structures such as bone or composites, computed tomography (CT) or $\mu$CT images allow one to build their three-dimensional (3D) real and complex geometries [Rémond et al. 2016] and the associated finite element (FE) meshes. Nonetheless, there exist only a few numerical tools able to describe a continuum field of anisotropic directions varying throughout the 3D structure.

**Communicated by Francesco dell'Isola**.
*MSC2010:* 00AXX.
*Keywords:* orthotropic materials, cortical bone, diffusion equations.

Here, we propose an approach which enables one to estimate via an FE technique the directions of orthotropy (i.e., longitudinal, circumferential, and radial) in 3D structures, which may be compared to beams, assuming that the orthotropy of their mesostructure is mainly triggered by their external surface and the boundary conditions. Such a method is based on [Allena and Aubry 2011], in which a system of Laplacian equations is employed to define the orientation of the cylindrical coordinates across 3D thin membranes. To support the numerical results, a comparison with $\mu$CT obtained with data is performed. In [Cluzel and Allena 2015], we applied our method only to a femoral diaphysis, while here the whole 3D cortical domain of a human femur is considered. Additionally, we validate our numerical approach by comparing it to the experimental data previously obtained in [Cluzel and Allena 2018].

**1.1.** ***Cortical bone anisotropy.*** Cortical bone shows a very significant anisotropy at different scales [Rho et al. 1998], and as described in [Rho 1996; Bernard et al. 2013], at the macroscale the elastic behavior is orthotropic. Additionally, microcracks seem to be involved at each length scale as a function of the loading (i.e., tension, compression, or torsion) and to trigger the damage mechanics of bone, although they are described in a local system of coordinates linked to the main directions of the mesostructure [Vashishth 2007].

In [Herman et al. 2010], two types of mechanical degradation of the cortical bone are described: one is linked to linear microcracks, which are 10 to 100 $\mu$m long, and the other is a set of diffused microcracks, which are 1 to 2 $\mu$m long. In [Seref-Ferlengez et al. 2015] these two networks are still observed, but the authors suggest that only the linear microcracks influence the evolution of the elastic behavior of the cortical bone and may be involved in the remodeling process.

At this level, the osteons play an important role and more particularly the cement line appears to be a weak interface likely to stop or divert the microcracks [O'Brien et al. 2007]. From a quantitative point of view, Wasserman et al. [2008] showed that microcracks are almost parallel to the osteons and this is independent from the age of the specimen. Given such a scenario, to describe the degradation or the failure behavior of the cortical bone, one may employ anisotropic criteria which are associated to specific mechanisms. For instance, in [Doblaré et al. 2004; Cowin and He 2005] anisotropic and macroscopic criteria are presented, some of them based on approaches that have been previously developed for composites [Tsai and Wu 1971]. Similarly, the fracture toughness is anisotropic and linked to the osteons' direction [Ural and Vashishth 2007].

Although it has been shown that to precisely describe the global response of a bone structure to different loadings it is necessary to take into account the orthotropic behavior of the cortical bone, many authors still use an isotropic elastic model or adopt isotropic failure criteria such as von Mises. In [Bessho et al. 2009; Duchemin et al. 2008], the objective is to localize the fractures and to do so the constitutive behavior of the bone is described as isotropic and heterogeneous. Báča et al. [2008] showed that an isotropic, elastic, and heterogeneous model allows proper quantification of the global displacements of a femur. Nevertheless, for an accurate description of the stress in the case of nonphysiological loads (i.e., prosthesis) or in order to obtain a better understanding of the damage mechanisms, the anisotropy of the bone must be taken into account.

From a numerical point of view, the employment of an orthotropic model remains rather difficult since two main challenges arise: the higher number of material parameters to be introduced and the description of the field of orthotropic coordinates throughout 3D complex geometries. Nonetheless, a few attempts can be found in the literature.

In [Peng et al. 2006], for both spongy and cortical bone, a transversely isotropic model is employed and the local systems of coordinates are described with respect to the superior-inferior axis of the structure without taking into account the potential variations in the neck or in the head. In [Taylor et al. 2002] or [Ün and Çalık 2016], the femoral diaphysis is described as an orthotropic material in a cylindrical coordinate system. Additionally, Ün and Çalık [2016] employ a discrete description of the orthotropic field via a finite number of subvolumes in the diaphysis. In [Báča et al. 2008], the macroscopic bone mesh is manually decomposed into small domains in order to take into account the anisotropy directions detected in vitro. Hambli et al. [2012] proposed an orthotropic damage model to describe the mechanical behavior of the proximal spongy domain of the femur in two dimensions. The orthotropic directions are associated to the principal stress directions obtained through a previous simulation involving a compression load on the top of the femur. A further approach can be found in [Doblaré and García 2001; Gómez-Benito et al. 2005] where the orientation of the orthotropic coordinates is continuously updated thanks to a remodeling model [García et al. 2001]. The simulation runs until the orthotropy directions coincide with the principal stress directions for a typical physiological load. Finally, in [Spingarn et al. 2017], anisotropy is also considered via a remodeling model, but at the mesoscale and in trabecular bone.

As mentioned earlier, in this paper we propose an FE method to approximate the orthotropic field of 3D structures such as the human femur. In the following sections we describe the numerical approach used to build the field of orthotropic directions. In Section 2.1, the segmentation technique adopted to obtain the femur 3D geometry (Section 2.1.1) as well as the diffusion equations employed to determine directions of orthotropy numerically (Section 2.1.2) are detailed. The main results are presented and compared to the experimental data obtained in [Cluzel and Allena 2018] in Section 3. In Section 4, we discuss our numerical outcomes with

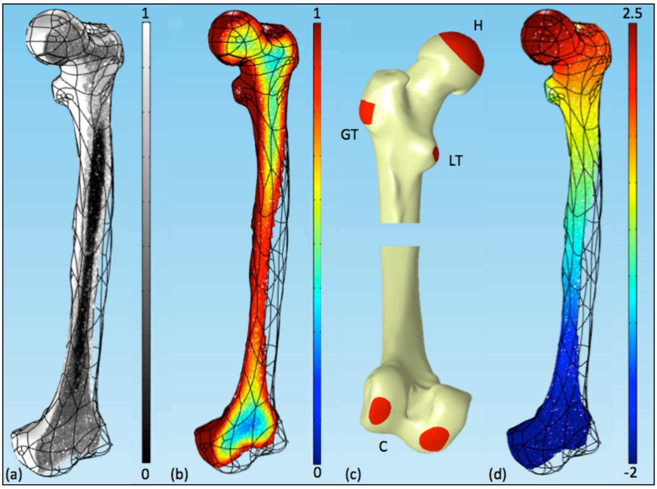


**Figure 1.** Sagittal sections of the femur showing the gray level (a) and the concentrations of $u_3$ (b) and $u_{10}$ (d), and boundary conditions for the diffusion problem providing the unit longitudinal vector $\nabla u_{10}$ (c).

respect to previous works in the literature and some limitations and perspectives of the work are also considered.

## 2. Material and methods

### 2.1. *FE approximation of the orthotropic field.*

**2.1.1.** *3D personalized geometry of the femur.* A left human male femur (91 years old) was collected and frozen at $-20^\circ$ in a plastic bag. Once defrosted, the femur was cleaned by a clinician to remove soft tissues around it and dried with ethanol. Then, it was CT-scanned with a calibration phantom by a GE LightSpeed Pro 16 at pixel spacing of 0.875 mm and slice thickness of 1.25 mm. CT scans provided the normalized gray level (GL) values varying between 0 and 1.

The femur has been semiautomatically segmented in Avizo to find its external surface and to mesh the 3D volume in COMSOL Multiphysics (Figure 1(a)). Additionally, by defining a specific threshold on the GL, here fixed at 0.7, it has been possible to write two characteristic functions ($h_{\text{cort}}$ and $h_{\text{spong}}$) to distinguish between the cortical ($\Omega_{\text{cort}}$) and the spongy ($\Omega_{\text{spong}}$) 3D FE domains:

$$h_{\text{cort}} = \begin{cases} 1 & \text{if GL} \geq 0.7, \\ 0 & \text{otherwise,} \end{cases} \tag{1}$$

$$h_{\text{spong}} = \begin{cases} 1 & \text{if GL} < 0.7, \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

Some artifacts induced by the presence of remaining soft tissues or due to unexpected porosities may be found on the external femoral surface. Nevertheless, in order to ensure a minimal amount of cortical bone, the GL values have been automatically set to 1 across a thickness of 0.5 mm from the external surface of the femur to the inner volume.

**2.1.2.** *Numerical technique to determine the main orthotropy directions.* In this section, the technique used to determine the numerical system of orthotropic coordinates $R_{\mathrm{FE}} = \{\boldsymbol{i}_1, \boldsymbol{i}_2, \boldsymbol{i}_3\}$ is detailed. We adopt and adapt a method previously proposed in [Allena and Aubry 2011]. Such an approach was first used to parametrize very thin 3D objects, such as the cortical bone. Two diffusion equations are employed, and the orientations of the concentration gradients provide the orthotropic directions.

Based on the assumption that the osteons are mainly parallel to the external surface of the 3D structure, the first diffusion equation defines the evolution of the concentration $u_3$ and allows defining the vector $\nabla u_3$ across the thickness of the cortical domain $\Omega_{\mathrm{cort}}$:

$$\begin{cases} c\,\mathrm{div}(\nabla u_3) = -\kappa h_{\mathrm{spong}}, & \\ u_3 = 1 & \text{on } \partial\Omega_{\mathrm{ext}}, \end{cases} \tag{3}$$

where div is the divergence and $\nabla$ is the gradient, $c = 10^{12}$, and the source term $-\kappa h_{\mathrm{spong}}$, where $\kappa = 10^6$, enables the introduction of a flow from the exterior to the interior of the femur. The concentration of $u_3$ across the femur is illustrated in Figure 1(b). The isosurfaces $u_3 = \mathrm{const}$ do not cross the outer boundary due to the maximum principle [Courant 1962], and they are parallel surfaces.

Thus, an approximate normalized vector $\boldsymbol{i}_3$ can be computed as

$$\boldsymbol{i}_3 \simeq \frac{\nabla u_3}{\|\nabla u_3\|} \tag{4}$$

with $\|\cdot\|$ the Euclidean norm of a vector. Assuming that in a 3D structure such as the femur there exists a strong relationship between the external loads and the directions of the osteons in the cortical bone [Wolff 1892], the second diffusion equation describes the evolution of the concentration $u_{10}$ and allows the description of the initial longitudinal direction $\nabla u_{10}$:

$$\begin{cases} \mathrm{div}[(a h_{\mathrm{cort}} + b)\nabla u_{10}] = 0, & \\ u_{10} = -2 & \text{on } \partial\Omega_{\mathrm{C}}, \\ u_{10} = 2.7 & \text{on } \partial\Omega_{\mathrm{H}}, \\ u_{10} = 0 & \text{on } \partial\Omega_{\mathrm{LT}}, \\ u_{10} = 1.8 & \text{on } \partial\Omega_{\mathrm{GT}}, \\ \frac{\partial u_{10}}{\partial \boldsymbol{n}} = 0 & \text{everywhere else}, \end{cases} \tag{5}$$

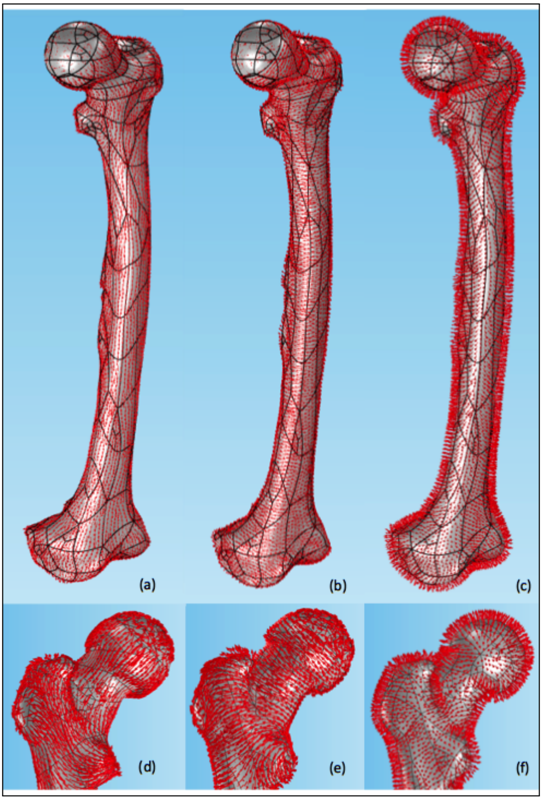


**Figure 2.** Numerical unit vectors $\boldsymbol{i}_1$ (left), $\boldsymbol{i}_2$ (center), and $\boldsymbol{i}_3$ (right). Anterior view of the femur (top), and posterior view of the upper femur (bottom).

where $a$ and $b$ are two constants that weight the influence of the spongy and the cortical bone. For our problem, $a \gg b$ (i.e., $a = 10^{15}$ and $b = 10^{11}$) to trigger a very high diffusion across the cortical domain $\Omega_{\text{cort}}$ and a very low diffusion in the spongy domain $\Omega_{\text{spong}}$. The surfaces $\partial\Omega_{\text{C}}$, $\partial\Omega_{\text{H}}$, $\partial\Omega_{\text{LT}}$, and $\partial\Omega_{\text{GT}}$ are represented in Figure 1(c), with C indicating the condyles, and allow one to mimic the muscular anchoring surfaces at the extremities of the femur as has been done in [Huiskes et al. 1987; Doblaré and García 2001; Hambli et al. 2012]. A more accurate description of the directions of orthotropy, which are related to the distribution of the physiological stresses throughout the femur [Petrtýl et al. 1996], would require additional anchoring regions along the diaphysis [Duda et al. 1998]. It also has to be said that the values of the boundary conditions in (5) are not representative of the average physiological loads, but they have rather been optimized to best fit the $\mu$CT observations. The concentration of $u_{10}$ across the femur is shown in Figure 1(d).

| $S_i$ | region | $h_c$ (mm) | error | anisotropy mode |
|---|---|---|---|---|
| $S_1$ | anterior diaphysis | | 19.3° | orthotropy |
| $S_2$ | anterior diaphysis | $\approx 4.1$ | 11.5° | orthotropy / transverse isotropy |
| $S_4$ | great trochanter | $\approx 0.8$ | 26.5° | orthotropy |
| $S_5$ | great trochanter | $\approx 0.8$ | 21.1° | orthotropy |
| $S_8$ | less trochanter | $\approx 2.5$ | 62.8° | orthotropy |
| $S_{10}$ | neck | $\approx 0.7$ | 19.7° | orthotropy |
| $S_{11}$ | neck | | 12.8° | orthotropy |
| $S_{15}$ | posterior diaphysis | $\approx 5.3$ | 6.6° | transverse isotropy |
| $S_{16}$ | posterior diaphysis | $\approx 5.7$ | 3.7° | orthotropy / transverse isotropy |
| $S_{17}$ | posterior diaphysis | $\approx 6.1$ | 10.2° | orthotropy / transverse isotropy |
| $S_{18}$ | posterior diaphysis | $\approx 6.0$ | 9° | orthotropy / transverse isotropy |

**Table 1.** Estimated error for the available cortical specimens ($h_c$ is the cortical thickness). All specimens are cortical.

In the same spirit as for $\boldsymbol{i}_3$, we can compute the normalized longitudinal vector

$$\boldsymbol{i}_{10} \simeq \frac{\nabla u_{10}}{\|\nabla u_{10}\|}. \tag{6}$$

By a simple cross product, we are able to obtain the circumferential vector

$$\boldsymbol{i}_2 = \frac{\boldsymbol{i}_3 \wedge \boldsymbol{i}_{10}}{\|\boldsymbol{i}_3 \wedge \boldsymbol{i}_{10}\|}. \tag{7}$$

We now have three vectors: the longitudinal ($\boldsymbol{i}_{10}$), the circumferential ($\boldsymbol{i}_2$), and the radial ($\boldsymbol{i}_3$). Nevertheless, to ensure the orthogonality of the basis, we need to recompute the longitudinal vector $\boldsymbol{i}_{10}$ to obtain

$$\boldsymbol{i}_1 = \frac{\boldsymbol{i}_2 \wedge \boldsymbol{i}_3}{\|\boldsymbol{i}_2 \wedge \boldsymbol{i}_3\|}. \tag{8}$$

The diffusion equations are integrated over the 3D personalized geometry of the femur through a FE discretization.

## 3. Results

**3.1.** ***FE computation of the directions of orthotropy.*** The main objective of the FE model is to provide a good approximation of the field of orthotropic coordinates across the femur via the set of diffusion equations presented in Section 2.1.2. In Figure 2, we show the global trend of the three unit vectors $\boldsymbol{i}_1$ (longitudinal), $\boldsymbol{i}_2$ (circumferential), and $\boldsymbol{i}_3$ (radial).

In order to validate the numerical approach, a more precise comparison between the numerical and the experimental orthotropic directions is necessary. In [Cluzel

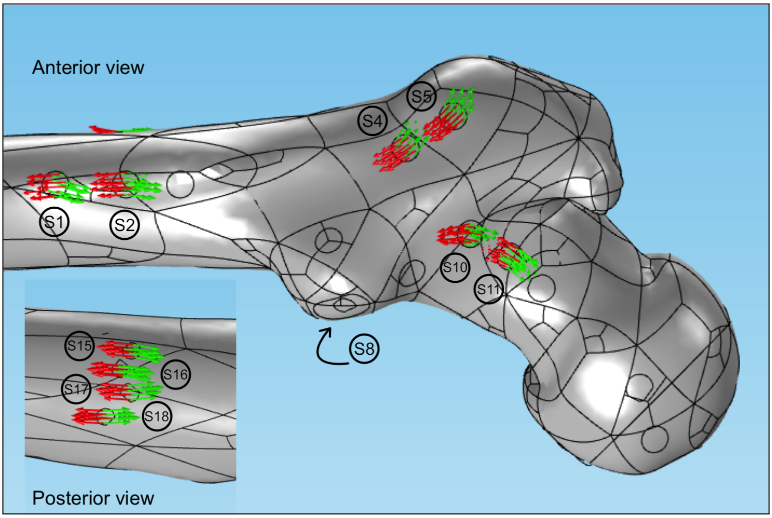


**Figure 3.** Comparison between the numerical longitudinal vector $\boldsymbol{i}_1$ (red arrows) and the measured vector $\boldsymbol{P}_1$ (green arrows) for specimens $S_1$, $S_2$, $S_4$, $S_5$, $S_{10}$, and $S_{11}$ on the anterior region of the femur and of $S_{15}$ to $S_{18}$ on the posterior region.

and Allena 2018] we have identified the projection $\boldsymbol{P}_1$ on the femur surface of the main direction of orthotropy. Then we have estimated the error between $\boldsymbol{P}_1$ and the numerical longitudinal vector $\boldsymbol{i}_1$ (Table 1).

For the comparison, we only consider the available cortical specimens from [Cluzel and Allena 2018, Table 1]: $S_1$, $S_2$, $S_4$, $S_5$, $S_8$, $S_{10}$, $S_{11}$, $S_{15}$, $S_{16}$, $S_{17}$, and $S_{18}$. In Figure 3, we show in red the vector $\boldsymbol{i}_1$ and in green the vector $\boldsymbol{P}_1$, except for $S_8$, which is on the other side of the femur.

Overall, the error goes from a minimal value of about 3.7° for $S_{16}$ in the posterior diaphysis to a maximal value of about 62.8° for $S_8$ in the less trochanter where a high gradient of the $\boldsymbol{i}_1$ is observed. However, in the anterior and posterior diaphysis, in the greater trochanter, and in the neck the error varies between 3.7° and 26.5°.

## 4. Discussion

For an orthotropic material such as bone or a woven composite, the degradation mechanisms are oriented along the principal directions of the mesostructure. Therefore, in order to be able to correctly describe the mechanical as well as the damage behavior of such materials, the associated models need to be built with respect to the local systems of orthotropic coordinates. Nonetheless, if the orthotropic directions constantly change across the 3D geometry, their description becomes even more difficult. Then, being able to obtain an approximation of the orthotropic system of coordinates may constitute a numerical challenge, but it allows the simulation of the mechanical response of those structures for which the orthotropy plays a critical role.

In this article, we have proposed an FE technique based on the assumption that, for many heterogeneous structures, the main directions of the orthotropic behavior are determined by the external surface that shapes the geometry, except near the loading regions. This aspect can be easily observed in the case of a woven composite material where each wire is made of a large number of continuous fibers. However, it becomes more complex for biological materials (i.e., cortical and spongy bone), especially when the boundary conditions are applied over a large region of the structure [Duda et al. 1998].

Here we have employed our method to describe the field of orthotropic directions in a left human femur. The 3D geometry is obtained through a stack of CT images. Once the outer surface of the femur is accurately defined, the radial vector $\boldsymbol{i}_3$ across the thickness of the structure is computed using an appropriate diffusion equation. To obtain the longitudinal vector $\boldsymbol{i}_{10}$, a second diffusion equation is used which must take into account specific boundary conditions. As in [Huiskes et al. 1987; Doblaré and García 2001; Hambli et al. 2012] and for the sake of simplicity, such boundary conditions coincide with the main muscular anchoring regions (i.e., the head, the condyles, and the greater and the lesser trochanter). A sensibility study has been performed in order to confirm that the variation in intensity of the boundary conditions does not affect the final results. In fact, by increasing or decreasing by $\pm 0.1$ the values of $u_{10}$ in (5), successively, we found a variation of about $0.01^\circ$ in the main direction of orthotropy $\boldsymbol{i}_1$, which can be neglected. Such an approach provides a description of the orthonormal systems of orthotropic coordinates across the cortical bone. Nevertheless, in those regions where the thickness of the cortical bone is very thin (i.e., $h_c < 0.5$ mm) or where the boundary conditions are applied, if $\boldsymbol{i}_3$ is properly determined, the longitudinal ($\boldsymbol{i}_1$) and the circumferential ($\boldsymbol{i}_2$) directions have no particular physical meaning.

In the literature, most of the numerical models describe cortical bone as an isotropic material due to several technical issues that one may encounter. A few orthotropic models have been proposed which include an orthotropic macroscale description of the bone behavior [Martínez-Reina et al. 2014; Taylor et al. 2002; Doblaré and García 2001; Gómez-Benito et al. 2005; Peng et al. 2006; Báča et al. 2008; Hambli et al. 2012; Ün and Çalık 2016]. Our approach allows a global and continuous representation of the orthotropic directions throughout the femur to be obtained while taking into account the local variations in specific regions of interest such as the neck and the lesser and greater trochanter.

In order to ensure the quality of the description of the orthotropic directions, some authors have measured in vitro the orthotropic field which has been manually implemented in the numerical models over about 20 regions in [Báča et al. 2007] and by cubes 2 mm in length in [Wirtz et al. 2003], for instance. In our approach, we rather estimate a posteriori the error between the numerical longitudinal vector $\boldsymbol{i}_1$ and the direction $\boldsymbol{P}_1$ identified via the $\mu$CT images [Cluzel and Allena 2018].

To conclude, the FE technique that we propose is globally consistent with the experimental data and allows one to obtain a proper approximation of the orthotropic field of coordinates. As in [Doblaré and García 2001], here we have assumed that the orthotropy directions are determined by both the applied boundary conditions and the external geometry of the femur. Additionally, only the boundary conditions at the extremities of the structure (i.e., the head, the greater and lesser trochanter, and the condyles) have been considered at this stage. Nonetheless, further muscular anchoring surfaces should be taken into account in the thicker diaphysis region [Duda et al. 1998], which may lead to slight rotations of the osteons with respect to the longitudinal axis, as observed in [Petrtýl et al. 1996; Báča et al. 2007]. To detect such variations, we are currently acquiring additional measurements along the diaphysis in order to obtain a more complete map of the osteons' orientation. Then we will build corrective functions to adjust both the numerical axial ($\boldsymbol{i}_1$) and circumferential ($\boldsymbol{i}_2$) unit vectors in the diaphysis via a rotation around $\boldsymbol{i}_3$. We envisage undertaking a series of simulations for different types of loading on the femur to quantify the precision needed for positioning the coordinates systems. We expect to find some differences in the failure mechanisms of the orthotropic damage model rather than on the global displacements. Taking into account the exact anisotropic directions of the cortical bone microstructure will allow one to rigorously describe the damage model. This constitutes a fundamental element to describe the overall remodeling process, including the evolution in time of the anisotropic directions [Placidi et al. 2004], and more specifically the interplay between the biological and the mechanical processes involved [Frame et al. 2017; Schmitt et al. 2016].

## 5. Acknowledgement

The authors are grateful to the ParisTech BiomecAM chair program on subject-specific musculoskeletal modeling for partial funding of the work.

## References

[Allena and Aubry 2011] R. Allena and D. Aubry, "A novel technique to parametrize shell-like deformations inside biological membranes", *Comput. Mech.* **47**:4 (2011), 409–423.

[Báča et al. 2007] V. Báča, D. Kachlík, Z. Horák, and J. Stingl, "The course of osteons in the compact bone of the human proximal femur with clinical and biomechanical significance", *Surg. Radiol. Anat.* **29**:3 (2007), 201–207.

[Báča et al. 2008] V. Báča, Z. Horák, P. Mikulenka, and V. Džupa, "Comparison of an inhomogeneous orthotropic and isotropic material models used for FE analyses", *Med. Eng. Phys.* **30**:7 (2008), 924–930.

[Bernard et al. 2013] S. Bernard, Q. Grimal, and P. Laugier, "Accurate measurement of cortical bone elasticity tensor with resonant ultrasound spectrosco", *J. Mech. Behav. Biomed.* **18** (2013), 12–19.

[Bessho et al. 2009] M. Bessho, I. Ohnishi, T. Matsumoto, S. Ohashi, J. Matsuyama, K. Tobita, M. Kaneko, and K. Nakamura, "Prediction of proximal femur strength using a CT-based nonlinear finite element method: differences in predicted fracture load and site with changing load and boundary conditions", *Bone* **45**:2 (2009), 226–231.

[Cluzel and Allena 2015] C. Cluzel and R. Allena, "Modelling of anisotropic cortical bone based on degradation mechanism", *Comput. Method. Biomech.* **18**:S1 (2015), 1914–1915.

[Cluzel and Allena 2018] C. Cluzel and R. Allena, "Heterogeneous directions of orthotropy in three dimensional structures: geometrical identification from $\mu$CT images", preprint, 2018.

[Courant 1962] R. Courant, *Methods of mathematical physics*, vol. II: Partial differential equations, Wiley-VCH, 1962.

[Cowin and He 2005] S. C. Cowin and Q.-C. He, "Tensile and compressive stress yield criteria for cancellous bone", *J. Biomech.* **38**:1 (2005), 141–144.

[Doblaré and García 2001] M. Doblaré and J. M. García, "Application of an anisotropic bone-remodelling model based on a damage-repair theory to the analysis of the proximal femur before and after total hip replacement", *J. Biomech.* **34**:9 (2001), 1157–1170.

[Doblaré et al. 2004] M. Doblaré, J. M. García, and M. J. Gómez, "Modelling bone tissue fracture and healing: a review", *Eng. Fract. Mech.* **71**:13–14 (2004), 1809–1840.

[Duchemin et al. 2008] L. Duchemin, V. Bousson, C. Raossanaly, C. Bergot, J. D. Laredo, W. Skalli, and D. Mitton, "Prediction of mechanical properties of cortical bone by quantitative computed tomography", *Med. Eng. Phys.* **30**:3 (2008), 321–328.

[Duda et al. 1998] G. N. Duda, M. Heller, J. Albinger, O. Schulz, E. Schneider, and L. Claes, "Influence of muscle forces on femoral strain distribution", *J. Biomech.* **31**:9 (1998), 841–846.

[Frame et al. 2017] J. Frame, P.-Y. Rohan, L. Corté, and R. Allena, "A mechano-biological model of multi-tissue evolution in bone", *Continuum Mech. Therm.* (online publication December 2017).

[García et al. 2001] J. M. García, M. A. Martinez, and M. Doblaré, "An anisotropic internal-external bone adaptation model based on a combination of CAO and continuum damage mechanics technologies", *Comput. Method. Biomech.* **4**:4 (2001), 355–377.

[Gómez-Benito et al. 2005] M. J. Gómez-Benito, J. M. García-Aznar, and M. Doblaré, "Finite element prediction of proximal femoral fracture patterns under different loads", *J. Biomech. Eng.* **127**:1 (2005), 9–14.

[Hambli et al. 2012] R. Hambli, A. Bettamer, and S. Allaoui, "Finite element prediction of proximal femur fracture pattern based on orthotropic behaviour law coupled to quasi-brittle damage", *Med. Eng. Phys.* **34**:2 (2012), 202–210.

[Herman et al. 2010] B. C. Herman, L. Cardoso, R. J. Majeska, K. J. Jepsen, and M. B. Schaffler, "Activation of bone remodeling after fatigue: differential response to linear microcracks and diffuse damage", *Bone* **47**:4 (2010), 766–772.

[Huiskes et al. 1987] R. Huiskes, H. Weinans, H. J. Grootenboer, M. Dalstra, B. Fudala, and T. J. Slooff, "Adaptive bone-remodeling theory applied to prosthetic-design analysis", *J. Biomech.* **20**:11–12 (1987), 1135–1150.

[Martínez-Reina et al. 2014] J. Martínez-Reina, I. Reina, J. Domínguez, and J. M. García-Aznar, "A bone remodelling model including the effect of damage on the steering of BMUs", *J. Mech. Behav. Biomed.* **32** (2014), 99–112.

[O'Brien et al. 2007] F. J. O'Brien, D. Taylor, and T. C. Lee, "Bone as a composite material: the role of osteons as barriers to crack growth in compact bone", *Int. J. Fatigue* **29**:6 (2007), 1051–1056.

[Peng et al. 2006] L. Peng, J. Bai, X. Zeng, and Y. Zhou, "Comparison of isotropic and orthotropic material property assignments on femoral finite element models under two loading conditions", *Med. Eng. Phys.* **28**:3 (2006), 227–233.

[Petrtýl et al. 1996] M. Petrtýl, J. Heřt, and P. Fiala, "Spatial organization of the haversian bone in man", *J. Biomech.* **29**:2 (1996), 161–169.

[Placidi et al. 2004] L. Placidi, S. H. Faria, and K. Hutter, "On the role of grain growth, recrystallization and polygonization in a continuum theory for anisotropic ice sheets", *Ann. Glaciol.* **39** (2004), 49–52.

[Rémond et al. 2016] Y. Rémond, S. Ahzi, M. Baniassadi, and H. Garmestani, *Applied RVE reconstruction and homogenization of heterogeneous materials*, Wiley, 2016.

[Rho 1996] J.-Y. Rho, "An ultrasonic method for measuring the elastic properties of human tibial cortical and cancellous bone", *Ultrasonics* **34**:8 (1996), 777–783.

[Rho et al. 1998] J.-Y. Rho, L. Kuhn-Spearing, and P. Zioupos, "Mechanical properties and the hierarchical structure of bone", *Med. Eng. Phys.* **20**:2 (1998), 92–102.

[Schmitt et al. 2016] M. Schmitt, R. Allena, T. Schouman, S. Frasca, J. M. Collombet, X. Holy, and P. Rouch, "Diffusion model to describe osteogenesis within a porous titanium scaffold", *Comput. Method. Biomech.* **19**:2 (2016), 171–179.

[Seref-Ferlengez et al. 2015] Z. Seref-Ferlengez, O. D. Kennedy, and M. B. Schaffler, "Bone microdamage, remodeling and bone fragility: how much damage is too much damage?", *BoneKEy Rep.* **4** (2015), 644.

[Spingarn et al. 2017] C. Spingarn, D. Wagner, Y. Rémond, and D. George, "Multiphysics of bone remodeling: a 2D mesoscale activation simulation", *Med. Eng. Phys.* **28**:S1 (2017), 153–158.

[Taylor et al. 2002] W. R. Taylor, E. Roland, H. Ploeg, D. Hertig, R. Klabunde, M. D. Warner, M. C. Hobatho, L. Rakotomanana, and S. E. Clift, "Determination of orthotropic bone elastic constants using FEA and modal analysis", *J. Biomech.* **35**:6 (2002), 767–773.

[Tsai and Wu 1971] S. W. Tsai and E. M. Wu, "A general theory of strength for anisotropic materials", *J. Compos. Mater.* **5**:1 (1971), 58–80.

[Ün and Çalık 2016] K. Ün and A. Çalık, "Relevance of inhomogeneous-anisotropic models of human cortical bone: a tibia study using the finite element method", *Biotechnol. Biotec. Eq.* **30**:3 (2016), 538–547.

[Ural and Vashishth 2007] A. Ural and D. Vashishth, "Anisotropy of age-related toughness loss in human cortical bone: a finite element study", *J. Biomech.* **40**:7 (2007), 1606–1614.

[Vashishth 2007] D. Vashishth, "Hierarchy of bone microdamage at multiple length scales", *Int. J. Fatigue* **29**:6 (2007), 1024–1033.

[Wasserman et al. 2008] N. Wasserman, B. Brydges, S. Searles, and O. Akkus, "In vivo linear microcracks of human femoral cortical bone remain parallel to osteons during aging", *Bone* **43**:5 (2008), 856–861.

[Wirtz et al. 2003] D. C. Wirtz, T. Pandorf, F. Portheine, K. Radermacher, N. Schiffers, A. Prescher, D. Weichert, and F. U. Niethard, "Concept and development of an orthotropic FE model of the proximal femur", *J. Biomech.* **36**:2 (2003), 289–293.

[Wolff 1892] J. Wolff, *Das Gesetz der Transformation der Knochen*, Hirschwald, 1892.

Received 9 Apr 2018. Revised 14 Jun 2018. Accepted 25 Jun 2018.

RACHELE ALLENA: rachele.allena@ensam.eu
*Institute de Biomécanique Humaine George Charpak, Arts et Métiers ParisTech, Paris, France*

CHRISTOPHE CLUZEL: cluzel@lmt.ens-cachan.fr
*Laboratoire de Mécanique et Technologie, Ecole Normale Supérieure Paris Saclay, Cachan, France*

and

*Département Science et Genie des Materiaux, Institut Universitaire de Technologie d'Evry Val d'Essonne, Every, France*

MATHEMATICS AND MECHANICS OF COMPLEX SYSTEMS
Vol. 6, No. 4, 2018
dx.doi.org/10.2140/memocs.2018.6.353

# A GENERAL METHOD FOR THE DETERMINATION OF THE LOCAL ORTHOTROPIC DIRECTIONS OF HETEROGENEOUS MATERIALS: APPLICATION TO BONE STRUCTURES USING $\mu$CT IMAGES

CHRISTOPHE CLUZEL AND RACHELE ALLENA

To assess the degree (i.e., isotropy, transverse isotropy, or orthotropy) and the directions of anisotropy of a three-dimensional structure, information about its mesostructure is necessary. Usually, a topological analysis of computed tomography or microcomputed tomography images is performed and requires an interpretation of the constitutive elements of the three-dimensional structure, which may lead to a simplistic description of the geometry. In this paper we propose an alternative technique based on a geometric tensor and we use it to analyze 38 representative elementary volumes extracted from 24 specimens of cortical bone in a human femur whose geometries have been reconstructed via microcomputed tomography images.

## 1. Introduction

Computed tomography (CT) and microcomputed tomography ($\mu$CT) are powerful imaging tools allowing the visualization of three-dimensional (3D) geometries which can be used to simulate the global and personalized response of the mechanical structure [Rémond et al. 2016]. If such geometries are constituted of heterogeneous materials like bone or composites [Placidi et al. 2017; Giorgio et al. 2017], one needs to describe their constitutive behavior as a function of the local systems of anisotropy. Then, additional information is required at the scale of their mesostructure to identify the anisotropic field.

Cortical bone is constituted of several elements oriented in space leading to a very significant anisotropy at different levels, from the nanoscale (i.e., collagen fibers) to the mesoscale (i.e., osteons) [Rho et al. 1998]. As a consequence, the elastic behavior at the macroscale is highly anisotropic and more specifically orthotropic as has been quantified in [Rho 1996; Bernard et al. 2013].

The identification of the directions of orthotropy may be straightforward and given by the direct observation of the Haversian canals. For instance, in [Heřt

**Communicated by Francesco dell'Isola**.
*Keywords:* orthotropic materials, cortical bone, $\mu$CT images, geometric tensor.

et al. 1994; Petrtýl et al. 1996], the canals are previously ink-soaked and then developed by successive polishing. A similar technique has also been adopted by Báča et al. [2007] to describe the directions of the canals on the bone surface.

Alternatively, a topological analysis of CT or $\mu$CT images can be employed to identify the degree (i.e., isotropy, transverse isotropy, or orthotropy) and the main directions of anisotropy after a 3D skeletonization as in [Pothuaud et al. 2000] or a 3D finite element (FE) simulation as in [Nazemi et al. 2016], both applied on trabecular bone. Nonetheless, such an approach requires a complex interpretation of the constitutive elements of the 3D structure. Therefore, in this paper we propose an alternative technique based on a geometric tensor and we use it to analyze a series of representative elementary volumes (REVs) extracted from cortical bone specimens and whose 3D geometries are obtained via $\mu$CT images. Assuming an orthotropic elastic behavior for the cortical bone, the average directions of the mesostructure are computed.

In the following sections we describe the experimental approach used to identify the main directions of orthotropy of the cortical bone mesostructure. This includes the specimen extraction (Section 2.1.1), the $\mu$CT imaging (Section 2.1.2), and the computation of the geometrical tensor associated with the femur mesostructure (Section 2.1.3). In Section 3, we first show the consistency of the technique to identify the directions of orthotropy through simple geometric configurations (Section 3.1) and second we apply our approach on the bone specimens (Section 3.2). Finally, in Section 4, the results are discussed and some limitations and perspectives of the work are considered.

## 2. Material and methods

### 2.1. *Experimental analysis of the orthotropic field.*

**2.1.1.** *CT-scanning and specimen extraction.* A left human male femur (91 years old) was collected and frozen at $-20\,^{\circ}$C in a plastic bag. Once defrosted, the femur was cleaned by a clinician to remove soft tissues around it and dried with ethanol.

A total of 24 specimens $S_i$, with $i$ from 1 to 24, were extracted at different regions of the proximal side of the femur as follows (Figure 1):

- 3 along the upper anterior diaphysis (AD) ($S_1$ to $S_3$),
- 2 in the greater trochanter (GT) ($S_4$ and $S_5$),
- 4 around and on top of the lesser trochanter (LT) ($S_6$ to $S_9$),
- 3 along the femoral neck (N) ($S_{10}$ to $S_{12}$),
- 2 in the femoral head (H) ($S_{13}$ and $S_{14}$),
- 4 in the upper posterior diaphysis (PD) ($S_{15}$ to $S_{18}$), and
- 6 around the diaphysis (D) ($S_{19}$ to $S_{24}$).

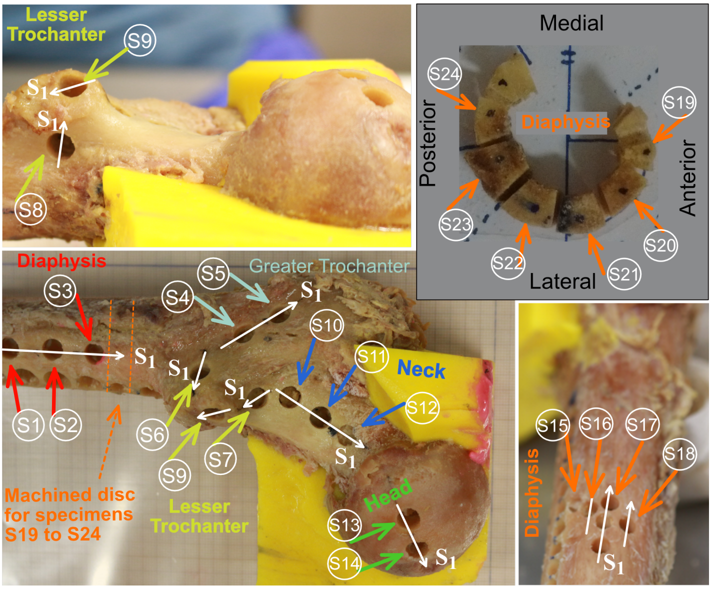


**Figure 1.** Extraction and position of the 24 specimens from the human left femur.

Diamond-tipped drills were used to machine specimens $S_1$ to $S_{18}$, which have a cylindrical shape with diameter 6 mm and height equal to the thickness of the cortical bone. Specimens $S_{19}$ to $S_{24}$ were manually cut and show a trapezoidal shape. During the cutting, water was used in order to reduce both friction and temperature rise. Before the extraction, an easily identifiable mark in the direction $\boldsymbol{S}_1$ has been carved on the external surface of each specimen in order to orient it with respect to the femur (Figures 1 and 2). The direction $\boldsymbol{S}_1$ is used to locate each specimen in the femur when the 3D microstructure is reconstructed from $\mu$CT images. Thus, it could be any direction. Here, for the sake of simplicity, we have

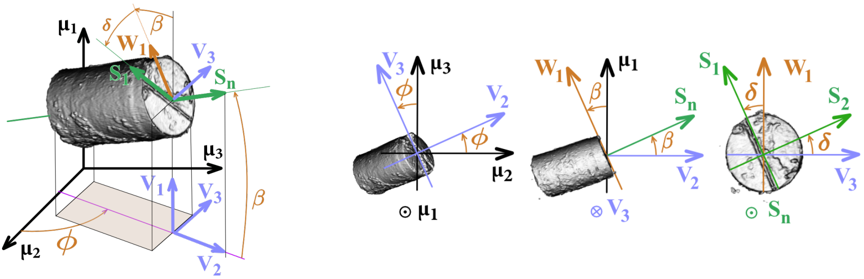


**Figure 2.** Position of a specimen with respect to the $\mu$CT system of coordinates $R_\mu$.

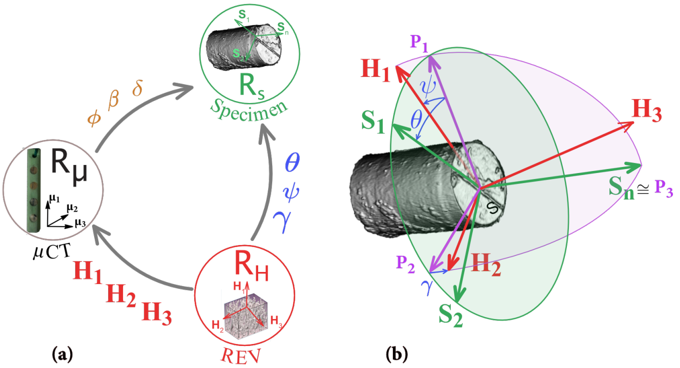


**Figure 3.** (a) Main steps to acquire the directions of orthotropy of the Haversian canals. (b) Angles defining the projection of $R_H$ on the external surface of the specimen.

decided to let it coincide with the middle line of the femur in each region of interest (Figure 1). Specimens were immersed in a solution of zinc iodide for 24 hours to stain the Haversian canals inside the osteons.

**2.1.2.** *$\mu$CT imaging.* The specimens were placed on a shelf trying to align the direction $\boldsymbol{S}_1$ with the vertical axis $\boldsymbol{\mu}_1$ of the $\mu$CT scanner in the best possible way (Figure 3b). They were scanned using a $\mu$CT scanner (Scanco Medical XtremeCT with voxel size 7.4 $\mu$m). It consisted of a microfocus X-ray source, a rotating specimen holder, and a detector system, with a $2048 \times 2048$ pixel CCDD camera. The images were acquired using the following protocol: 90 kVp, 155 $\mu$A, 0.5 mm aluminum filter, and integration time 200 ms per slice.

After acquisition, a stack of about 800 cross-sectional images stored in DICOM format was obtained and the 3D reconstruction was made using software from FEI (Hillsboro, Oregon, USA). First, we built the 3D volume of the specimens in order to compute the outward normal vector $\boldsymbol{S}_n$ to its external surface (Figure 3b). Second, by defining a specific threshold and by extracting one or more representative elementary volumes (REVs) for each specimen, we were able to obtain the 3D network of the Haversian canals. It is worth noting that an REV includes a sufficient number of osteons (i.e., at least 10, which corresponds to 3 to 4 osteons per side) and does not present any porosity which could trigger artifacts. In both cases (whole specimen and REV), the final 3D geometry was stored as an STL file constituted of a large number of facets $N_f$ ($150000 < N_f < 200000$) providing a uniform and smooth surface.

**2.1.3.** *Identification of the main directions of the Haversian canals.* In this section we detail the successive steps used to acquire the main directions of orthotropy associated with the Haversian canals (Figure 3a). Each system of reference used in this section is a direct orthonormal system of coordinates.

First, each specimen $S_i$ is defined by its proper system of reference $R_S = \{\boldsymbol{S}_1, \boldsymbol{S}_2, \boldsymbol{S}_n\}$, where the subscript $S$ stands for specimen, $\boldsymbol{S}_1$ and $\boldsymbol{S}_n$ were previously defined (Section 2.1.1), and $\boldsymbol{S}_2$ is obtained via a vector product between $\boldsymbol{S}_1$ and $\boldsymbol{S}_n$ (Figure 3b). In order to determine the position of the specimen with respect to the $\mu$CT system of reference $R_\mu = \{\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \boldsymbol{\mu}_3\}$ (the subscript $\mu$ stands for $\mu$CT), three angles are measured between $R_s$ and $R_\mu$: $\phi$, $\beta$, and $\delta$ (Appendix A1).

Second, once an REV is extracted from a specimen and using the geometrical information included in the STL files previously obtained (Section 2.1.2), the system of reference $R_H = \{\boldsymbol{H}_1, \boldsymbol{H}_2, \boldsymbol{H}_3\}$ (the subscript $H$ stands for Haversian canals) can be computed. To do so, rather than performing a topological analysis [Boyle and Kim 2011] of the REV surface mesh which would require an approximation of each Haversian canal by a regular geometry, we propose an approach which only takes into account the external surface of each Haversian canal while maintaining the precision, as demonstrated via simple illustrative examples in Section 3.1.

Each facet of the REV surface mesh is identified by its proper outward normal vector $\boldsymbol{n}_j$. Since for each REV the mesh facets have mostly the same area and their total number $N_f$ is high, no weighting has been applied. The product $\boldsymbol{n}_j\boldsymbol{n}_j^T$ enables one to obtain a tensorial form of $\boldsymbol{n}_j$, which includes more information than the vector itself (i.e., eigenvalues and eigenvectors). Then, by summing all these tensors, the global tensor $\boldsymbol{G}$ can be computed as

$$\boldsymbol{G} = \sum_{j=1}^{N_f} \boldsymbol{n}_j\boldsymbol{n}_j^T. \tag{1}$$

To quantify the morphology and the geometrical effects, we use the normalized eigenvalues $0 \le \lambda_k \le 3$, with $k \in [1, 2, 3]$, of $\boldsymbol{G}$, which are obtained from the eigenvalues $\lambda_{10} \le \lambda_{20} \le \lambda_{30}$ as

$$\lambda_k = \frac{3\lambda_{k0}}{\lambda_{10} + \lambda_{20} + \lambda_{30}}. \tag{2}$$

For each normalized eigenvalue $\lambda_k$, the associated eigenvector $\boldsymbol{H}_k$ can be calculated.

Finally, the projection of $R_H$ on the external surface of the specimen is computed to obtain the system of reference $R_P = \{\boldsymbol{P}_1, \boldsymbol{P}_2, \boldsymbol{P}_3\}$, where the subscript $P$ stands for projection (Appendix A2). The vectors $\boldsymbol{P}_1$, $\boldsymbol{P}_2$, and $\boldsymbol{P}_3$ are the projections of $\boldsymbol{H}_1$, $\boldsymbol{H}_2$, and $\boldsymbol{H}_3$, respectively (Figure 3b). Then, the position of $R_H$ for each REV with respect to $R_S$ can be found through three angles: $\psi$, $\gamma$, and $\theta$ (Figures 3b and 4). The vector $\boldsymbol{P}_1$ will be directly compared to the corresponding numerical vector, which is obtained as described in the following sections.

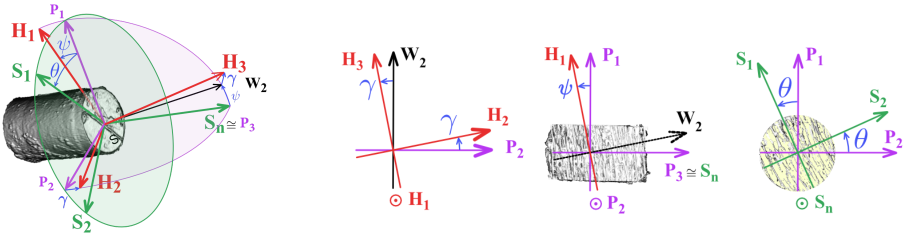


**Figure 4.** Angles defining the projection of $R_H$ on the external surface of the specimen ($\boldsymbol{W}_2$ is an intermediate unit vector).

## 3. Results

**3.1.** ***Validation of the technique to identify the directions of orthotropy.*** To validate and illustrate our approach presented in Section 2.1.3, five simple examples, whose average direction $\boldsymbol{V}$ is known, are proposed as shown in Figure 5.

It has to be noticed that the cutting sections of the tubes are not proper surfaces of the tubes themselves but rather fictive ones obtained through the REV extraction. Therefore, for configurations (a) to (d), the upper and lower cutting planes are not taken into account. However, for the sake of practicality, for configuration (e) the extremities are included in the analysis.

For each configuration, the tubes are characterized by their direction $\boldsymbol{V}_{t0}$ ($t$ being the number of the tube in the specific configuration and going from 1 to $N_t$, the total number of tubes), which is defined in a spherical system of coordinates as $\boldsymbol{V}_{t0} = \{\cos\alpha_t \cos\gamma_t, \cos\alpha_t \sin\gamma_t, \sin\alpha_t\}$.

The average direction $\boldsymbol{V}$ of a configuration is then defined as

$$\boldsymbol{V} = \frac{\sum_{l=1}^{N_t} \boldsymbol{V}_{t0}}{\left\| \sum_{l=1}^{N_t} \boldsymbol{V}_{t0} \right\|} \tag{3}$$

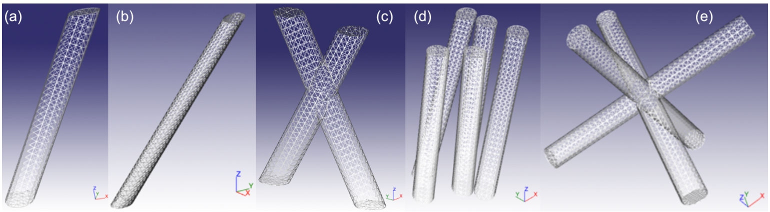


**Figure 5.** Simple examples to illustrate our approach to compute the geometric tensor $\boldsymbol{G}$. (a) Single tube with circular section. (b) Single tube with elliptical section. (c) Two crossed tubes. (d) Five noncrossed and randomly oriented tubes with circular section. (e) Three orthogonally crossed tubes.

| | (a) | (b) | (c) | (d) | (e) |
|---|---|---|---|---|---|
| $\alpha_t/\gamma_t$ | 60°/30° | 60°/30° | 60°/90°<br>60°/−90° | 50°/20°<br>55°/25°<br>60°/30°<br>65°/35°<br>70°/40° | 0°/0°<br>0°/90°<br>90°/0° |
| $\boldsymbol{V}$ | 0.433 | 0.433 | 0 | 0.437 | |
| | 0.250 | 0.250 | 0 | 0.237 | |
| | 0.866 | 0.866 | 1 | 0.868 | |
| $\lambda_1$ | 0.010 | 0.000 | 0.320 | 0.03 | 0.998 |
| $\lambda_2$ | 1.480 | 0.990 | 0.930 | 1.420 | 1.001 |
| $\lambda_3$ | 1.510 | 2.010 | 1.750 | 1.540 | 1.001 |
| $\boldsymbol{H}_1$ | 0.432 | 0.433 | −0.001 | 0.408 | 0.577 |
| | 0.251 | 0.250 | 0.001 | 0.240 | 0.577 |
| | 0.866 | 0.866 | 1.000 | 0.881 | 0.577 |
| $\boldsymbol{H}_2$ | 0.517 | −0.501 | −0.013 | 0.598 | −0.305 |
| | 0.718 | 0.865 | 0.999 | 0.659 | −0.503 |
| | −0.466 | 0.001 | −0.001 | −0.456 | 0.808 |
| Anisotropy mode | TI | O | O | TI | I |

**Table 1.** Overall results for each configuration (a) to (e) (TI = transverse isotropy, O = orthotropy, and I = isotropy).

with $\|\cdot\|$ the Euclidean norm of a vector. The results associated with each configuration are reported in Table 1. The direction $\boldsymbol{V}$ can be alternatively computed using the approach presented in Section 2.1.3. In fact, the first eigenvector $\boldsymbol{H}_1$ of the geometrical tensor $\boldsymbol{G}$ corresponds to $\boldsymbol{V}$. It is interesting to notice that an eigenvector $\boldsymbol{H}_k$ ($k$ going from 1 to 3) is correctly computed only if the associated eigenvalue $\lambda_k$ is notably different from the other two. Thus, the direction $\boldsymbol{V}$ is properly estimated by the proposed approach for cases (a) to (d). However, for case (e), the eigenvalues $\lambda_k$ of $\boldsymbol{G}$ being identical and close to 1, the overall geometry is isotropic and the eigenvectors do not provide any further information. Similarly, the eigenvector $\boldsymbol{H}_2$ is indicative only if the eigenvalues $\lambda_2$ and $\lambda_3$ are different. This is not the case for configurations (a) and (d) where there is not a preferential transversal direction (i.e., transverse isotropy). Finally, for cases (b) and (c), the eigenvector $\boldsymbol{H}_2$ clearly shows the geometrical orthotropy.

The consistency of our technique has then been shown, and we can now apply it to the cortical bone.

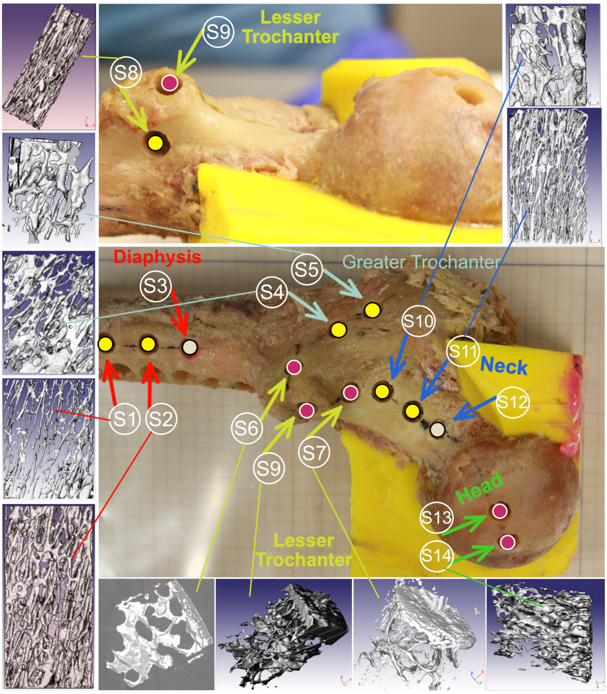


**Figure 6.** 3D reconstruction of the Haversian canals network for the REVs extracted from specimens $S_1$, $S_2$, $S_4$, $S_5$, $S_8$, $S_{10}$, and $S_{11}$ (yellow circles). For the other specimens (pink or gray circles), the reconstruction was not possible due to the very thin cortical thickness or some technical issues.

**3.2.** ***Measurement of the directions of orthotropy.*** For each of the 24 specimens the 3D geometry was reconstructed and its position with respect to the $\mu$CT system of coordinates $R_\mu$ was identified. In order to obtain the main direction $\boldsymbol{H}_1$ of the Haversian canals, at least one REV for each specimen was extracted where only Haversian canals (i.e., no large porosities) could be observed. Nevertheless, for some specimens (i.e., $S_6$, $S_7$, $S_9$, $S_{13}$, and $S_{14}$) only a few Haversian canals (i.e., fewer than five) could be segmented due to the very thin thickness of the cortical domain. Then, the orthotropic system of coordinates $R_H$ could not be computed, but we rather estimated the eigenvectors $\boldsymbol{H}_k$ of the geometric tensor $\boldsymbol{G}$ associated with the spongy trabeculae. Additionally, specimens $S_3$, $S_6$, and $S_{12}$ could not be retrieved due to some technical issues. In Figure 6, the REVs for specimens $S_1$, $S_2$, $S_4$, $S_5$, $S_8$, $S_{10}$, and $S_{11}$ are presented. In summary, we distinguish between the specimens with a cortical thickness $h_C$ greater (19, named cortical specimens) and less (5, named trabecular specimens) than 0.5 mm. In the following sections, for each specimen we identify the eigenvalues ($\lambda_1$, $\lambda_2$, $\lambda_3$) of the geometric tensor $\boldsymbol{G}$,

| $S_i$ | REV | region | s.t. | $\lambda_1$ | $\lambda_2$ | $\lambda_3$ | $\theta$ | $\psi$ | $\gamma$ | $h_c$ | a.m. |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | ° ±1 | ° ±1 | ° ±1 | mm ± 0.4 | |
| $S_1$ | $S_{1,1}$ | AD | C | 0.39 | 0.93 | 1.67 | −20.1 | 11.2 | −3.8 | | O |
| $S_2$ | $S_{2,1}$ | AD | C | 0.46 | 1.12 | 1.42 | −9.9 | 2.0 | −35.5 | 4.1 | O/TI |
| $S_3$ | $S_{3,1}$ | AD | C | | | | | | | | LD |
| $S_4$ | $S_{4,1}$ | GT | C | 0.42 | 1.06 | 1.52 | 31.3 | 1.1 | 0.9 | 0.8 | O |
| $S_5$ | $S_{5,1}$ | GT | C | 0.38 | 0.76 | 1.86 | 20.5 | 3.0 | 10.4 | 0.8 | O |
| $S_6$ | $S_{6,1}$ | LT | T | | | | | | | 0.5 | NVM |
| $S_7$ | $S_{7,1}$ | LT | T | 0.43 | 0.72 | 1.80 | 33.0 | 2.9 | 5.7 | 0.4 | O |
| $S_8$ | $S_{8,1}$ | LT | C | 0.35 | 1.01 | 1.64 | 87.5 | 0.2 | −3.5 | 2.5 | O |
| $S_9$ | $S_{9,1}$ | LT | T | 0.62 | 0.99 | 1.39 | 70.8 | 9.8 | 83.9 | 0.5 | O |
| $S_{10}$ | $S_{10,1}$ | N | C | 0.37 | 0.74 | 1.90 | 20.5 | 7.3 | 3.9 | 0.7 | O |
| $S_{11}$ | $S_{11,1}$ | N | C | 0.37 | 1.14 | 1.48 | −2.3 | 5.2 | −0.2 | | O |
| $S_{12}$ | $S_{12,1}$ | N | C | | | | | | | | LD |
| $S_{13}$ | $S_{13,1}$ | H | T | 0.43 | 1.04 | 1.53 | −96.1 | 80.5 | −69.1 | | O |
| $S_{14}$ | $S_{14,1}$ | H | T | 0.43 | 1.08 | 1.49 | −89.2 | 79.7 | −60.9 | 0.5 | O |
| $S_{15}$ | $S_{15,1}$ | PD | C | 0.76 | 1.11 | 1.14 | −7.5 | 7.9 | −53.2 | 5.3 | TI |
| $S_{16}$ | $S_{16,1}$ | PD | C | 0.63 | 1.10 | 1.28 | −4.2 | 6.2 | 87.4 | 5.7 | O/TI |
| $S_{17}$ | $S_{17,1}$ | PD | C | 0.58 | 1.11 | 1.31 | 10.6 | 4.4 | −89.1 | 6.1 | O/TI |
| $S_{18}$ | $S_{18,1}$ | PD | C | 0.39 | 1.20 | 1.40 | 9.6 | 1.7 | −28.3 | 6.0 | O/TI |

**Table 2.** Overall results for specimens $S_1$ to $S_{18}$. Under s.t. (specimen type): C = cortical, T = trabecular. Under a.m. (anisotropy mode): TI = transverse isotropy, O = orthotropy, I = isotropy, LD = lost data, NVM = no visible mark.

the angles $\theta$, $\psi$, and $\gamma$, and the cortical thickness $h_C$. In Tables 2 and 3 all the results are reported. For the angles, the uncertainty of 1° is due to the manual measurement. For the cortical thickness, an uncertainty of 0.4 mm corresponds to the largest transition region between the cortical and the spongy bone.

**3.2.1.** *Cortical specimens.* Although in reality there is no sharp transition between isotropy (I), transverse isotropy (TI), and orthotropy (O), here, in order to classify the degree of anisotropy for each specimen $S_i$, we have defined the intervals

- for I $|\lambda_i - \lambda_j| \leq 0.05$,
- between I and TI $0.05 < |\lambda_i - \lambda_j| < 0.37$,
- for TI with respect to $\boldsymbol{H}_1$ $|\lambda_2 - \lambda_1| \geq 0.37$ and $|\lambda_3 - \lambda_2| \leq 0.05$,
- between TI and O $|\lambda_2 - \lambda_1| \geq 0.37$ and $0.05 < |\lambda_3 - \lambda_2| < 0.37$, and
- for O $|\lambda_i - \lambda_j| \geq 0.37$,

where $i, j \in \{1, 2, 3\}$ and $i \neq j$.

Overall, for each cortical specimen, $\lambda_1$ is much smaller than $\lambda_2$ and, more specifically, we found that specimens $S_2$, $S_{15}$, $S_{16}$, $S_{17}$, and $S_{18}$ show a transverse isotropy.

| $S_i$ | REV | $\chi$ <br> $^\circ\pm1$ | region | s.t. | $\lambda_1$ | $\lambda_2$ | $\lambda_3$ | $\theta$ <br> $^\circ\pm1$ | $\psi$ <br> $^\circ\pm1$ | $\gamma$ <br> $^\circ\pm1$ | a.m. |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $S_{19}$ | $S_{19,1}$ | −11.3 | D | C | 0.62 | 1.06 | 1.30 | 8.4 | 1.8 | −2.5 | O/TI |
| | $S_{19,2}$ | −4.6 | D | C | 0.76 | 1.04 | 1.20 | 2.1 | −2.1 | −16.3 | TI/I |
| | $S_{19,3}$ | 5.4 | D | C | 0.68 | 0.97 | 1.35 | 3.6 | 1.6 | −4.7 | O |
| $S_{20}$ | $S_{20,1}$ | 33.4 | D | C | 0.64 | 1.03 | 1.33 | 16.3 | 7.3 | 12.0 | O |
| | $S_{20,2}$ | 42.4 | D | C | 0.68 | 1.03 | 1.29 | 19.4 | 4.4 | −7.5 | O/TI |
| | $S_{20,3}$ | 51.8 | D | C | 0.59 | 1.10 | 1.31 | 16.8 | 1.8 | −26.3 | O/TI |
| $S_{21}$ | $S_{21,1}$ | 70.0 | D | C | 0.62 | 1.08 | 1.30 | 11.7 | −0.6 | −36.1 | O/TI |
| | $S_{21,2}$ | 80.3 | D | C | 0.63 | 1.10 | 1.27 | 10.2 | −1.6 | −34.2 | O/TI |
| | $S_{21,3}$ | 87.7 | D | C | 0.53 | 1.21 | 1.26 | 8.6 | −1.9 | −71.3 | TI |
| $S_{22}$ | $S_{22,1}$ | 100.9 | D | C | 0.62 | 1.05 | 1.33 | 9.3 | 0.6 | −26.1 | O/TI |
| | $S_{22,2}$ | 118.2 | D | C | 0.61 | 1.13 | 1.26 | 9.5 | −0.4 | −5.6 | O/TI |
| | $S_{22,3}$ | 122.3 | D | C | 0.51 | 1.05 | 1.44 | 9.2 | 1.1 | −18.9 | O |
| | $S_{22,4}$ | 126.9 | D | C | 0.68 | 1.05 | 1.27 | 18.0 | −1.1 | −17.9 | O/TI |
| | $S_{22,5}$ | 130.0 | D | C | 0.74 | 1.05 | 1.21 | 16.6 | −1.1 | −5.1 | TI/I |
| $S_{23}$ | $S_{23,1}$ | 147.2 | D | C | 0.65 | 1.07 | 1.28 | 4.7 | −0.7 | −13.1 | O/TI |
| | $S_{23,2}$ | 160.2 | D | C | 0.68 | 0.99 | 1.32 | 1.3 | 0.6 | −4.8 | O/TI |
| | $S_{23,3}$ | 171.1 | D | C | 0.71 | 1.01 | 1.28 | −7.9 | 2.7 | −7.9 | O/TI |
| $S_{24}$ | $S_{24,1}$ | 181.0 | D | C | 0.79 | 1.09 | 1.12 | −23.1 | −1.7 | −17.2 | TI |
| | $S_{24,2}$ | 187.1 | D | C | 0.67 | 0.99 | 1.34 | −35.9 | −5.7 | −1.4 | O |
| | $S_{24,3}$ | 197.6 | D | C | 0.68 | 1.06 | 1.26 | −28.6 | −16.1 | −22.9 | O/TI |

**Table 3.** Overall results for specimens $S_{19}$ to $S_{24}$. Under s.t. (specimen type): C = cortical. Under a.m. (anisotropy mode): TI = transverse isotropy, O = orthotropy, I = isotropy.

Then, in these cases, the angle $\gamma$, which generally provides the circumferential direction, is not relevant since $\lambda_2 \approx \lambda_3$. The remaining cortical specimens (i.e., $S_1$, $S_4$, $S_5$, $S_8$, $S_{10}$, and $S_{11}$) are orthotropic. Around the diaphysis (i.e., from $S_{19}$ to $S_{24}$), some REVs clearly show an orthotropic behavior, whereas others are between orthotropy and transverse isotropy (Table 3).

It is interesting to focus on the specimens in the AD ($S_1$ and $S_2$), in the PD (from $S_{15}$ to $S_{18}$), and around the diaphysis (from $S_{19}$ to $S_{24}$) for which $h_c > 0.5$ and the Haversian canals are uniformly oriented. We found that the angle $\theta$ between the drawn mark $\boldsymbol{S}_1$ on each specimen and $\boldsymbol{P}_1$, the projection of the principal direction of the Haversian canals $\boldsymbol{H}_1$ on the external surface of the femur, varies between −35.86° for $S_{24,2}$ and 19.44° for $S_{20,2}$.

We analyzed the evolution of $\theta$ along the radial direction (Figure 7). To do so, we split the reconstructed $S_1$ specimen into seven successive slices with a spacing of 0.5 mm (Figure 7) and we found $\theta$ equal to (a) −43°, (b) −22°, (c) −15°, (d)

| Image | $\lambda_1$ | $\lambda_2$ | $\lambda_3$ | $\theta$ (°) |
|---|---|---|---|---|
| Figure 7a | 0.39 | 0.55 | 1.09 | −42.7 |
| Figure 7b | 0.44 | 0.99 | 1.56 | −21.9 |
| Figure 7c | 0.40 | 0.98 | 1.61 | −15.0 |
| Figure 7d | 0.37 | 1.15 | 1.48 | −16.1 |
| Figure 7e | 0.39 | 1.22 | 1.39 | −16.6 |
| Figure 7f | 0.44 | 1.15 | 1.40 | −17.6 |
| Figure 7g | 0.44 | 1.18 | 1.38 | −14.9 |

**Table 4.** Results for the extracted slices from specimen $S_1$.

−16°, (e) −17°, (f) −18°, and (g) −15°. If we consider that in the outermost slices (parts (a) and (b) of Figure 7) the presence of the external surface distorts the final outcome, one may conclude that $\theta$ does not change significantly when going from the outer to the inner cortical domain. The computed angles are reported in Table 4.

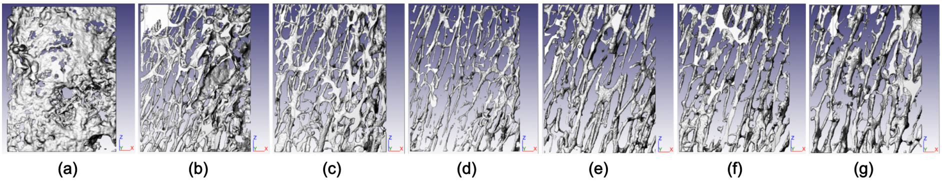


**Figure 7.** Successive slices from the exterior (a) to the interior (g) of an REV of specimen $S_1$.

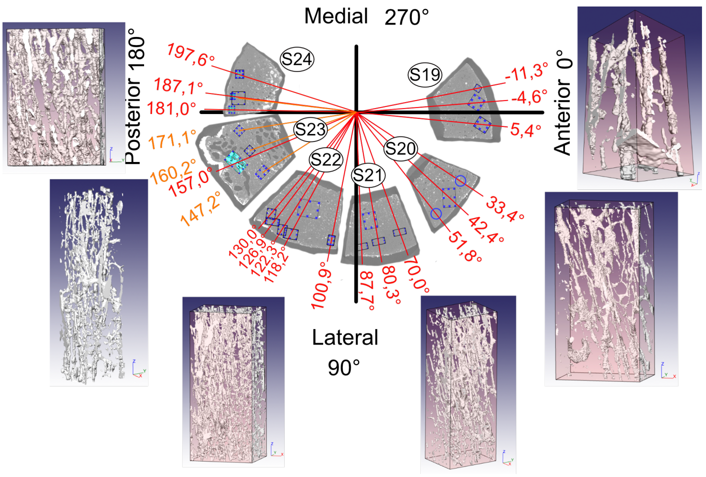


**Figure 8.** Specimens $S_{19}$ to $S_{24}$ with the relative circumferential position $\chi$ and some of the extracted REVs.

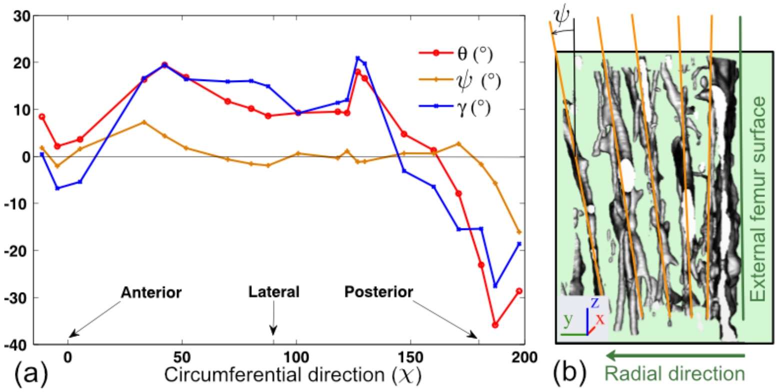


**Figure 9.** (a) Evolution of $\theta$, $\psi$, and $\gamma$ with respect to the circumferential position $\chi$ of the REVs in specimens $S_{19}$ to $S_{24}$. (b) Orientation of $\psi$ close to the external surface of the femur.

For specimens $S_{19}$ to $S_{24}$, twenty REVs were extracted, whose position around the diaphysis circumference is defined by the angle $\chi$, varying between $-11.3°$ and $197.6°$ (Figure 8 and Table 3). In Figure 9a, the evolutions of the angles $\theta$ (red line), $\psi$ (brown line), and $\gamma$ (blue line) around the upper diaphysis (i.e., from the anterior to the posterior region passing by the lateral region) are shown. We can notice that $\theta$ and $\gamma$ vary as functions of the circumferential coordinate $\chi$, whereas $\psi$ does not change significantly.

Finally, we investigated the variation of $\psi$, the angle between $\boldsymbol{P}_1$ and $\boldsymbol{H}_1$, across the cortical thickness along the radial direction for the specimen $S_1$. We found that, closer to the outer surface, the Haversian canals are nearly parallel to the surface (Figure 9b).

**3.2.2.** *Trabecular specimens.* As mentioned earlier, for some specimens ($S_6$, $S_7$, $S_9$, $S_{13}$, and $S_{14}$) the cortical thickness is very thin. Thus, it is difficult to extract an REV with isolated Haversian canals and identify the average direction $\boldsymbol{H}_1$.

Although in the present work we focus on the orthotropy of the cortical bone, it is interesting to quantify the geometric tensor $\boldsymbol{G}$ and the associated variables for such specimens (Table 2). To do so, the parameters used for the reconstruction of the 3D geometries from the $\mu$CT images have been adapted in order to detect the trabeculae and larger REVs have been extracted.

Some remarks can be drawn. For specimens $S_{13}$ and $S_{14}$ extracted in the head, the angle $\psi$ is equal to 80.5° and 79.7°, respectively. This means that the spongy trabeculae are mainly oriented perpendicular to the outer surface of the femur. On the contrary, for specimens $S_7$ and $S_9$ extracted in the neck, $\psi$ is found equal to 2.9° and 9.8°, respectively, which implies that the trabeculae are almost parallel to the external femoral surface (Table 2). Such outcomes are in agreement with the literature and more specifically with the works of Jacobs et al. [1995] and Tsubota et al. [2009].

## 4. Discussion

Our approach is based on experimental data and more specifically on $\mu$CT images of the Haversian network, which provide the same trends found in [Petrtýl et al. 1996; Wirtz et al. 2003; Báča et al. 2007]. We have defined a geometric tensor $\boldsymbol{G}$, via the outward normal vectors to the facets of an REV, and we have computed its eigenvalues and eigenvectors to estimate the degree and the directions of orthotropy, respectively.

Despite the interesting results, such a technique shows few main drawbacks. First, the position of each specimen with respect to the $\mu$CT system of coordinates $R_\mu$ needs to be carefully retrieved in order to obtain consistent results. Second, it is a discrete approach since the measurements are obtained via a series of specimens in specific regions of the 3D structure. In order to circumvent this issue, we propose in a further work a numerical technique based on diffusion equations and on the experimental data obtained in this paper. This approach will allow us to get a continuum description of the field of orthotropic directions in a 3D structure such a femur.

## Appendix

**A1.** ***Position of $R_S$ with respect to $R_\mu$.*** As explained in Section 2.1.3, it is necessary to determine the position of a specimen $S_i$ with respect to the $\mu$CT. Each specimen is identified by the mark $\boldsymbol{S}_1$ and the outward normal vector $\boldsymbol{S}_n$, which constitute with $\boldsymbol{S}_2$ the system of coordinates $R_S = \{\boldsymbol{S}_1, \boldsymbol{S}_2, \boldsymbol{S}_n\}$. During the image acquisition, the position of a specimen with respect to the $\mu$CT system of coordinates $R_\mu = \{\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \boldsymbol{\mu}_3\}$ is defined by three angles (Figure 2):

- $\phi$, the angle between $\boldsymbol{\mu}_2$ and $\boldsymbol{V}_2$,
- $\beta$, the angle between $\boldsymbol{V}_2$ and $\boldsymbol{S}_n$, and
- $\delta$, the angle between $\boldsymbol{W}_1$ and $\boldsymbol{S}_1$,

where $R_v = \{\boldsymbol{V}_1, \boldsymbol{V}_2, \boldsymbol{V}_3\}$ is an intermediate system of reference obtained by projecting the specimen on the plane $(\boldsymbol{\mu}_2, \boldsymbol{\mu}_3)$. Then, $\boldsymbol{V}_1 = \boldsymbol{\mu}_1$, $\boldsymbol{V}_2$ is the normalized projection of $\boldsymbol{S}_n$ on the plane $(\boldsymbol{\mu}_2, \boldsymbol{\mu}_3)$, and $\boldsymbol{V}_3$ is the cross product between $\boldsymbol{V}_1$ and $\boldsymbol{V}_2$. Finally, the vector $\boldsymbol{W}_1$ is the normalized projection of $\boldsymbol{V}_1$ on the specimen external surface.

**A2.** ***Projection of $R_H$ on the external surface of a specimen.*** As described in Section 2.1.3, it is necessary to identify the position of the Haversian canals with respect to the specimen. When an REV is extracted from a specimen, the system of orthotropic coordinates $R_H = \{\boldsymbol{H}_1, \boldsymbol{H}_2, \boldsymbol{H}_3\}$ of the Haversian canals can be determined. The projection of $R_H$ on the external surface of the specimen is computed and the system of reference $R_P$ is obtained. Then, three angles define the position of $R_H$ with respect to $R_S$ (Figure 4):

- $\theta$, the angle between $\boldsymbol{P}_1$ and $\boldsymbol{S}_1$, expressed as

$$\theta = -\arcsin(\boldsymbol{P}_1^T \boldsymbol{S}_2), \tag{4}$$

- $\psi$, the angle between $\boldsymbol{H}_1$ and $\boldsymbol{P}_1$, which reads

$$\psi = -\arcsin(\boldsymbol{S}_n^T \boldsymbol{H}_1), \tag{5}$$

  and

- $\gamma$, the angle between $\boldsymbol{H}_2$ and $\boldsymbol{P}_2$, defined as

$$\gamma = -\arcsin(\boldsymbol{P}_2^T \boldsymbol{H}_3), \tag{6}$$

where $(\boldsymbol{c}^T \boldsymbol{d})$ indicates the dot product, the superscript $T$ the transpose of a vector, and

$$\boldsymbol{P}_2 = \frac{\boldsymbol{H}_1 \wedge \boldsymbol{S}_n}{\|\boldsymbol{H}_1 \wedge \boldsymbol{S}n\|}, \tag{7}$$

$$\boldsymbol{P}_1 = \frac{\boldsymbol{S}_n \wedge \boldsymbol{P}_2}{\|\boldsymbol{S}_n \wedge \boldsymbol{P}_2\|} \tag{8}$$

with $(\boldsymbol{c} \wedge \boldsymbol{d})$ the vector product. It is worth noting that in (4), (5), and (6), rather than simply computing the arc cosine between the two involved vectors, the arc sine is used in order to detect the sign of the angle of interest.

## Acknowledgements

The authors are grateful to the ParisTech BiomecAM chair program on subject-specific musculoskeletal modeling for partial funding of the work and to Sonia Simoes for her support during $\mu$CT imaging.

## References

[Báča et al. 2007] V. Báča, D. Kachlík, Z. Horák, and J. Stingl, "The course of osteons in the compact bone of the human proximal femur with clinical and biomechanical significance", *Surg. Radiol. Anat.* **29**:3 (2007), 201–207.

[Bernard et al. 2013] S. Bernard, Q. Grimal, and P. Laugier, "Accurate measurement of cortical bone elasticity tensor with resonant ultrasound spectroscopy", *J. Mech. Behav. Biomed.* **18** (2013), 12–19.

[Boyle and Kim 2011] C. Boyle and I. Y. Kim, "Three-dimensional micro-level computational study of Wolff's law via trabecular bone remodeling in the human proximal femur using design space topology optimization", *J. Biomech.* **44**:5 (2011), 935–942.

[Giorgio et al. 2017] I. Giorgio, U. Andreaus, F. dell'Isola, and T. Lekszycki, "Viscous second gradient porous materials for bones reconstructed with bio-resorbable grafts", *Extreme Mech. Lett.* **13** (2017), 141–147.

[Heřt et al. 1994] J. Heřt, P. Fiala, and M. Petrtýl, "Osteon orientation of the diaphysis of the long bones in man", *Bone* **15**:3 (1994), 269–277.

[Jacobs et al. 1995] C. R. Jacobs, M. E. Levenston, G. S. Beaupré, J. C. Simo, and D. R. Carter, "Numerical instabilities in bone remodeling simulations: the advantages of a node-based finite element approach", *J. Biomech.* **28**:4 (1995), 449–451.

[Nazemi et al. 2016] S. M. Nazemi, D. M. L. Cooper, and J. D. Johnston, "Quantifying trabecular bone material anisotropy and orientation using low resolution clinical CT images: a feasibility study", *Med. Eng. Phys.* **38**:9 (2016), 978–987.

[Petrtýl et al. 1996] M. Petrtýl, J. Heřt, and P. Fiala, "Spatial organization of the haversian bone in man", *J. Biomech.* **29**:2 (1996), 161–169.

[Placidi et al. 2017] L. Placidi, U. Andreaus, and I. Giorgio, "Identification of two-dimensional pantographic structure via a linear D4 orthotropic second gradient elastic model", *J. Eng. Math.* **103** (2017), 1–21.

[Pothuaud et al. 2000] L. Pothuaud, P. Porion, E. Lespessailles, C. L. Benhamou, and P. Levitz, "A new method for three-dimensional skeleton graph analysis of porous media: application to trabecular bone microarchitecture", *J. Microsc.* **199**:2 (2000), 149–161.

[Rémond et al. 2016] Y. Rémond, S. Ahzi, M. Baniassadi, and H. Garmestani, *Applied RVE reconstruction and homogenization of heterogeneous materials*, Wiley, 2016.

[Rho 1996] J.-Y. Rho, "An ultrasonic method for measuring the elastic properties of human tibial cortical and cancellous bone", *Ultrasonics* **34**:8 (1996), 777–783.

[Rho et al. 1998] J.-Y. Rho, L. Kuhn-Spearing, and P. Zioupos, "Mechanical properties and the hierarchical structure of bone", *Med. Eng. Phys.* **20**:2 (1998), 92–102.

[Tsubota et al. 2009] K.-i. Tsubota, Y. Suzuki, T. Yamada, M. Hojo, A. Makinouchi, and T. Adachi, "Computer simulation of trabecular remodeling in human proximal femur using large-scale voxel FE models: approach to understanding Wolff's law", *J. Biomech.* **42**:8 (2009), 1088–1094.

[Wirtz et al. 2003] D. C. Wirtz, T. Pandorf, F. Portheine, K. Radermacher, N. Schiffers, A. Prescher, D. Weichert, and F. U. Niethard, "Concept and development of an orthotropic FE model of the proximal femur", *J. Biomech.* **36**:2 (2003), 289–293.

Received 9 Apr 2018. Revised 13 Jun 2018. Accepted 29 Jul 2018.

CHRISTOPHE CLUZEL: cluzel@lmt.ens-cachan.fr
*Laboratoire de Mécanique et Technologie, Ecole Normale Supérieure Paris Saclay, Cachan, France*

RACHELE ALLENA: rachele.allena@ensam.eu
*Institute de Biomécanique Humaine George Charpak, Arts et Métiers ParisTech, Paris, France*

# Guidelines for Authors

Authors may submit manuscripts in PDF format on-line at the submission page.

**Originality.** Submission of a manuscript acknowledges that the manuscript is original and and is not, in whole or in part, published or under consideration for publication elsewhere. It is understood also that the manuscript will not be submitted elsewhere while under consideration for publication in this journal.

**Language.** Articles in MEMOCS are usually in English, but articles written in other languages are welcome.

**Required items.** A brief abstract of about 150 words or less must be included. It should be self-contained and not make any reference to the bibliography. If the article is not in English, two versions of the abstract must be included, one in the language of the article and one in English. Also required are keywords and a Mathematics Subject Classification or a Physics and Astronomy Classification Scheme code 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 amsart class, but submissions in other varieties of TeX, and exceptionally in other formats, are acceptable. Initial uploads should normally be in PDF format; after the refereeing process we will ask you to submit all source material.

**References.** Bibliographical references should be complete, including article titles and page ranges. 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, however, for submission you may use the format of your choice. Links will be provided to all literature with known web locations and authors are encouraged to provide their own links in addition to those supplied in the editorial process.

**Figures.** Figures must be of publication quality. After acceptance, you will need to submit the original source files in vector graphics format for all diagrams in your manuscript: vector EPS or vector PDF files are the most useful.

Most drawing and graphing packages — Mathematica, Adobe Illustrator, Corel Draw, MATLAB, etc. — allow the user to save files in one of these formats. Make sure that what you are saving is vector graphics and not a bitmap. If you need help, please write to graphics@msp.org with as many details as you can about how your graphics were generated.

Bundle your figure files into a single archive (using zip, tar, rar or other format of your choice) and upload on the link you been provided at acceptance time. 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.

MATHEMATICS AND MECHANICS OF COMPLEX SYSTEMS
vol. 6 no. 4 2018

*MEMOCS* is a journal of the International Research Center for
the Mathematics and Mechanics of Complex Systems
at the Università dell'Aquila, Italy.