Bloch-Torrey Equation & Diffusion Imaging (DWI, DTI, q-space Imaging)

Expanded lecture notes for ISMRM Weekend Educational Course *MR Physics for Physicists*, Toronto, Canada, May 30, 2015, 12:00

Dmitry S. Novikov*
Bernard and Irene Schwartz Center for Biomedical Imaging,
Department of Radiology, NYU School of Medicine, New York, NY 10016, USA

CONTENTS

* dima@alum.mit.edu

I. INTRODUCTION: THE MESOSCOPIC SCALE

A. NMR of simple liquids: The molecular scale

In the first few decades after the discovery of NMR, its main applications focused on the structure of molecules. A macroscopic sample, often a chemical solution in an NMR tube of a few cubic millimeters volume, was uniform down to molecular dimensions. This well-mixed solution ensured that the NMR measurement was a giant ensemble average, recording average NMR properties of a proton at a given molecular site. Equivalently, one could think about studying one molecule many times and averaging over the outcomes. All relevant interactions affecting a given nuclear spin played out at the molecular level, in very close proximity: dipole-dipole interactions between adjacent spins, fluctuating in time due to a stochastic nature of the molecular environment, resulted in the chemical shift and in BPP relaxation [1]. These phenomena occurred at the *molecular scale* of about a nanometer.

Quantifying these microscopic processes and explaining NMR line shapes for different nuclei presented a theoretical challenge. The relevant microscopic, i.e. atomic or molecular, scale dictated the development of the necessary theoretical framework for quantifying the observed effects, which took a couple of decades to build. The framework was based on the quantum-mechanical evolution of a few coupled spins, with the solutions typically averaged over thermal fluctuations. Hence, the relevant mathematics [2–4] was based on the Schrödinger equation, the evolution equation for the density matrix, and the chemical exchange models [5]. As a particular outcome, the relaxation phenomena were understood in terms of spin-spin couplings and the parameters of the environment and the chemical exchange rates.

With the invention of MRI, and practically from the 1980s onwards, biomedical applications [6] came to the forefront. At first, it was exciting enough to recognize the correlations of empirical NMR parameters, such as T_1, T_2, and the diffusion coefficient D, with anatomy and (patho)physiology.

Empirical correlations between MRI parameters and pathology have proven to be extremely useful, even though often times their biophysical origins are yet to be clarified. The famous example is an almost twofold drop of the diffusion coefficient in acute ischemia [7], an effect discovered in 1990, utilized clinically for over a decade, and whose microstructural origins are under intense debate even today.

However, collecting empirical correlations cannot be the end goal of scientific exploration. Eventually, we are faced with hard questions — what is it that we measure, and what does it really mean at the cellular level. Addressing them is

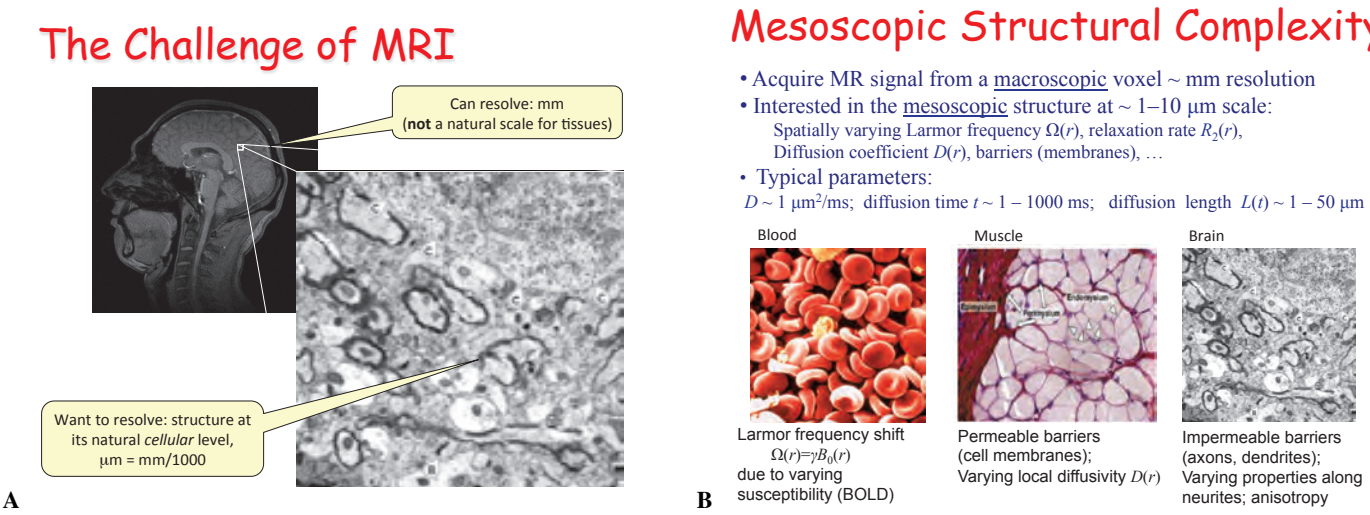

FIG. 1. (A) The challenge of the biomedical MRI is to quantify the structural complexity at the mesoscopic scale, orders of magnitude below the nominal imaging resolution. (B) Examples of tissue architecture at mesoscopic scale, which greatly exceeds the molecular dimensions (nanometers), yet it is still far finer than the macroscopic imaging resolution of millimeters.

important not just to satisfy our curiosity, but is crucial in order to develop markers which are not just empirically sensitive, but which are specific to particular pathological processes.

Here is where the *mesoscopic scale* comes into play.

B. Tissue architecture: The mesoscopic scale

The chief difference between the "chemical" NMR of the 1950s–1970s and the "biological" NMR and MRI of the modern days is the presence of immense structural complexity at the mesoscopic scale, Fig. 1. This scale, ranging between a fraction of a μm to tens of μm, is intermediate (hence "meso") between the molecular scale, and the macroscopic scale of the MRI resolution. This is the scale of the cellular tissue architecture, which makes tissues so specific, complex, and radically different from a mere solution of proteins in water.

In connection to the mesoscopic scale, the most important number to keep in mind is the water diffusion coefficient value. Between 10°C and body temperature, it ranges

$$D_{\rm water} \sim 1 - 3\,\mu{\rm m}^2/{\rm ms}\,. \qquad (1.1)$$

Everywhere here we will use "natural" units, μm and ms, for temporal and spatial scales, as it is in these units the diffusion coefficient is of the order unity. Reported in the standard SI units, m^2/s, the above value does not look very user friendly.

The diffusion time t, which is set by the echo time or the mixing time (depending on the NMR sequence), can be practically varied between a few ms and about 1 s, limited by the performance of diffusion gradients from below, and by T_1 from above. This readily yields the following range for the *diffusion length* in water under our control:

$$L(t) = \sqrt{2Dt} \sim 1 - 50\,\mu{\rm m}\,. \qquad (1.2)$$

The diffusion length, to remind, is the standard deviation for the molecules to spread over the time t:

$$L(t) = \Big\langle \big(x(t) - x(0)\big)^2 \Big\rangle^{1/2}\,. \qquad (1.3)$$

We should now appreciate that the value (1.1) of the water diffusion coefficient is a blessing. Indeed, based on this value, we are essentially given a length scale (1.2), under our control by varying t, that is far smaller than the nominal imaging resolution. Furthermore, we are immensely lucky that $L(t)$ is generally *commensurate with cell dimensions*. Hence, by varying the diffusion time t, we can probe the most biologically relevant length scales!

But what does it mean, to probe the mesoscopic structure via the varying $L(t)$? Practically, it means that we study *transient* processes, as opposed to the relatively simple stationary macroscopic parameters. Thus we expect that our observed relaxation rates and diffusion metrics would in general *depend on the echo time*, evolving as the properties of the tissue architecture are being sampled by water molecules traveling over the gradually increasing diffusion length (1.3). This time dependence, a necessary counterpart of generally non-exponential relaxation and non-Gaussian diffusion, is the footprint of tissue complexity at the mesoscopic scale.

The new physics emerging at the mesoscopic scale requires development of a corresponding theoretical apparatus, which differs from that used for NMR properties of liquids [2–4]. Instead of the quantum-mechanical description of coupled spins, and the site-exchange models, the relevant theoretical framework at the mesoscale involves *purely classical evolution*, albeit in a structurally complex and randomly looking environment at the scale of the tissue building blocks. Technically, the tools include averaging over the Brownian paths of the spins, and over the stationary *structural disorder* representing the mesoscopic tissue architecture, which restricts and modifies molecular motion. These tools may be borrowed from

condensed matter physics and statistical physics, which have faced similar challenges for over half a century in a different context, describing transport in environments which are heterogeneous at the mesoscopic scale [8–15].

The future of advanced diffusion MRI relies on building the necessary framework for quantifying tissue properties at the mesoscopic scale. The purpose of these notes is to provide a foundation of the relation between measurable diffusion metrics and the mesoscopic structural complexity.

II. THE DIFFUSION EQUATION

A. The conserved current

The diffusion equation consists of *two* equations with very different physical meanings. The first one describes the conservation of the number of molecules:

$$\partial_t \psi(t,\mathbf{r}) = -\partial_{\mathbf{r}} \cdot \mathbf{j}(t,\mathbf{r})\,. \tag{2.1}$$

It is equivalent to saying that the number of particles $Q_V = \int_V \mathrm{d}\mathbf{r}\, \psi(t,\mathbf{r})$ in a given volume V can only change by means of the particle flow in or out of this volume:

$$\partial_t Q_V = -\int_{\partial V} \mathrm{d}\mathbf{s}\cdot\mathbf{j}\,, \tag{2.2}$$

where the right-hand side is the net particle current flowing through the volume's boundary ∂V.

Equation (2.1) is *exact*, although not particularly useful yet, as long as the current is left unspecified. In order to write down the partial differential equation containing only $\psi(t,\mathbf{r})$, one needs to relate the current $\mathbf{j}$ to the density ψ. The latter relation is always *approximate*, and depends on the microscopic properties of the medium.

B. Fick's law

Let us guess the relation between $\mathbf{j}$ and ψ, based on the symmetry arguments, in the simplest case of a uniform medium. Imagine that we labeled certain water molecules, with their physical properties otherwise being the same as those of other water molecules. Obviously, if the density $\psi(t,\mathbf{r})$ of the labeled molecules is uniform, there is no net current of them, $\mathbf{j} \equiv 0$, as their fluxes in each direction statistically compensate each other. Hence, the current only occurs when the density is inhomogeneous, and should increase with the inhomogeneity. It should also be directed from the excess to the shortage of the labeled molecules. Finally, the current $\mathbf{j}$ is a vector which should change sign if the density gradient $\partial_{\mathbf{r}}\psi(t,\mathbf{r})$ (also a vector!) is reflected, i.e. when the excess and shortage are swapped places. These considerations naturally lead to the following approximation:

$$\mathbf{j}(t,\mathbf{r}) \approx -D\,\partial_{\mathbf{r}}\psi(t,\mathbf{r})\,, \tag{2.3}$$

i.e. the particle current is proportional to the density gradient and is directed against the latter. This is Adolf Fick's first law, reported in 1855. Note that this law is similar to Ohm's law (electric current is proportional to the gradient of the voltage), and to the Fourier law (heat flux is proportional to the gradient of temperature).

The relation (2.3) is the simplest linear response relation, appropriate for small variations in ψ. However, it is not the most general linear relation: As we discuss in detail in Sec. IV below, in principle, symmetry considerations allow one to write down an expansion in the odd powers of the gradients of ψ, such as $\partial_{\mathbf{r}}\nabla^2\psi$, etc. Also, such relation may in general be *retarded* rather than instantaneous as implied by Eq. (2.3), i.e. dependent on the values of $\psi(t',\mathbf{r})$ at the preceding moments of time $t' < t$. These generalizations of the Fick's law emerge as a result of averaging over a spatially heterogeneous sample, as we will describe in Sec. IV B 3 below.

Practically, the simple Fick's approximation (2.3) works remarkably well in simple liquids. Combining Eqs. (2.1) and (2.3), we obtain the diffusion equation

$$\partial_t \psi = D\,\partial_{\mathbf{r}}^2\psi\,, \tag{2.4}$$

where the single parameter characterizing the molecular motion is the diffusion constant D.

For a heterogeneous medium, characterized by spatially varying diffusion coefficient $D(\mathbf{r})$, the Fick's law can be written *locally* for the vicinity of each point $\mathbf{r}$,

$$\mathbf{j}(t,\mathbf{r}) \approx -D(\mathbf{r})\,\partial_{\mathbf{r}}\psi(t,\mathbf{r})\,. \tag{2.5}$$

Combining Eqs. (2.1) and (2.5), we obtain the general form of the diffusion equation in a medium with the spatially varying diffusion properties:

$$\boxed{\partial_t\psi = \partial_{\mathbf{r}}\big(D(\mathbf{r})\partial_{\mathbf{r}}\psi\big)\,.} \tag{2.6}$$

Note again that the current conservation requires that the right-hand side of Eq. (2.6) be a full divergence. Therefore writing down the diffusion equation in a different form, e.g. $\partial_t\psi = D(\mathbf{r})\partial_{\mathbf{r}}^2\psi$, would result in the net flow (drift).

C. A permeable barrier (membrane)

The spatially varying $D(\mathbf{r})$ in Eq. (2.6) is in principle allowed to vary strongly, so that this equation can be thought of as containing the effect of *permeable membranes, or barriers*, in addition to "smooth" variations of $D(\mathbf{r})$. Indeed, permeability microscopically stems from a spatially varying $D(\mathbf{r})$, since a permeable barrier can be defined as a thin sheet in which both the local diffusion coefficient D_m and the thickness l_m approach zero, while their ratio

$$\kappa = \frac{D_m}{l_m} \tag{2.7}$$

remains finite. This ratio κ is the *permeability* of the barrier.

As the barrier is thin, we can assume that, upon applying a density mismatch $\Delta\psi = \psi|_{x+0} - \psi_{x-0}$ across a barrier at point x, the corresponding current $j = -D_m\cdot(\Delta\psi/l_m) \equiv$

$-\kappa\Delta\psi$ and the density gradient $\Delta\psi/l_m$ become uniform inside the barrier practically instantaneously, and the current inside the barrier instantly equilibrates with the current at both sides of the barrier. Hence, the effect of the barrier can also be formulated as a *boundary condition* on ψ, with the normal component of the current at the position of the barrier is proportional to the discontinuity in the density across it,

$$-\mathbf{n}\cdot\mathbf{j}|_{\mathbf{r}_m} = \kappa\left[\psi_{\mathbf{r}_m+\mathbf{n}0} - \psi_{\mathbf{r}_m-\mathbf{n}0}\right]. \qquad (2.8)$$

Here, $\mathbf{n}$ is the normal to the barrier at the position $\mathbf{r}_m$.

Alternatively, the effect of the barrier can be represented as a singular term, proportional to the derivative δ' of the delta-function along the normal $\mathbf{n}$ at $\mathbf{r}=\mathbf{r}_m$, in the right-hand side of Eq. (2.6), as described in Ref. [16], such that the diffusion equation in a uniform medium with diffusion constant D_0 and a barrier at the origin reads

$$\partial_t\psi = D_0\partial_x^2\psi - \frac{D_0^2}{\kappa}\delta'(x)[\partial_x\psi]_{x=0}. \qquad (2.9)$$

Both ways of incorporating the effect of a barrier essentially stem from the spatially varying $D(\mathbf{r})$ at the scale of the barrier thickness. Hence, in a way, Eq. (2.6) can be thought of as incorporating both smooth variations of $D(\mathbf{r})$ and the abrupt drops in $D(\mathbf{r})$ within barriers. Understood in this general way, Eq. (2.6) can in principle be specified so as to contain all the mesoscopic structural complexity; its solution, *if only we could find it*, would be as much as we could possibly wish for. For tissues, of course, finding this solution is impossible.

In biological tissues, cf. Fig. 1, the value $D(\mathbf{r})$ is wildly fluctuating in space, and is difficult to specify (much like it is practically impossible to precisely specify the positions and parameters of all cells and cell membranes in a voxel). Therefore, one seeks to have a more tractable relation between the density and the current, with both quantities *averaged* over random placements of cells in a voxel, which retains measurable structural information about the underlying tissue building blocks at the mesoscopic scale. This relation emerges in the effective medium theory (EMT) treatment, as described in Ref. [17] and below in Sec. IV.

But before developing parsimonious approaches to dealing with Eq. (2.6), let us turn to the measurement.

III. WHAT DO WE MEASURE?

A. The Bloch-Torrey equation

With NMR, after an excitation, one measures the transverse magnetization $M(t,\mathbf{r})$, which is a complex-valued quantity (a two-dimensional vector in the plane transverse to the B_0 field). Its evolution is governed by the Bloch-Torrey equation [18], which in the rotating frame reads

$$\partial_t M = \partial_{\mathbf{r}}\big(D(\mathbf{r})\partial_{\mathbf{r}}M\big) - i\Omega(\mathbf{r})M - R_2(\mathbf{r})M. \qquad (3.1)$$

The NMR-specific additional terms, compared to the pure diffusion equation (2.6), are the locally varying Larmor frequency offset $\Omega(\mathbf{r})$ and transverse relaxation rate $R_2(\mathbf{r})$. Crucially, their presence result in the right-hand side of Eq. (3.1) *not* being the divergence of any "current".

The absence of the conservation law of the form (2.1) for $M(t,\mathbf{r})$ results in the *transverse relaxation* (decay) of the net magnetization $\int d\mathbf{r}\, M(t,\mathbf{r})$ acquired within a voxel. Simply put, the number of water molecules and protons is conserved, but their magnetization is not.

When both Ω and R_2 are uniform across a sample (voxel), the substitution $M = \psi\, e^{-i\Omega t - R_2 t}$ returns us back to the diffusion equation (2.6). When a sample has nonuniform magnetic properties, this factorization does not work, and the relaxation effects bias the diffusion metrics [19, 20].

In what follows, we will not consider the relaxation effects, and will focus solely on the conserving diffusional dynamics of the water molecules, Eq. (2.6).

B. The diffusion propagator and q-space imaging

The diffusion-weighted measurement is really a transverse relaxation measurement in disguise. After all, we only measure the time evolution of the net transverse magnetization $\propto \int_V d\mathbf{r} M(t,\mathbf{r})$, where V is our macroscopic volume (e.g. an imaging voxel). The trick, due to Stejskal and Tanner [21], is to apply a known external $\Omega(t,\mathbf{r})$, such that the transverse relaxation of the observed "diffusion-weighted signal" $S(t) \propto \int_V d\mathbf{r} M(t,\mathbf{r})$ would have a footprint of the diffusive properties of the medium.

For simplicity, let us assume an ideal case of a balanced narrow-pulse gradient of the Larmor frequency,

$$\mathbf{g}(\tau) = \mathbf{q}\left[\delta(\tau - t) - \delta(\tau)\right]. \qquad (3.2)$$

In other words, we have a narrow gradient pulse at $\tau = 0$ and an opposite pulse of the same magnitude at $\tau = t$. Here, $\mathbf{q}$ is by construction a vector in the direction of $\mathbf{g}$ whose magnitude is given by the integral under the gradient pulse. The parameters $\mathbf{q}$ and t are under our experimental control.

The Bloch-Torrey evolution of unit magnetization $M_{t;\mathbf{r}_t,\mathbf{r}_0}$, which initially were concentrated around $\mathbf{r}_0$, $M_{t;\mathbf{r}_t,\mathbf{r}_0}|_{t=0} = \delta(\mathbf{r}_t - \mathbf{r}_0)$, is then described, according to Eq. (3.1) with $\Omega(\tau,\mathbf{r}) = \mathbf{g}(\tau)\mathbf{r}$, as

$$M_{t;\mathbf{r}_t,\mathbf{r}_0} = e^{-i\mathbf{q}\mathbf{r}_t}\mathcal{G}_{t;\mathbf{r}_t,\mathbf{r}_0}e^{i\mathbf{q}\mathbf{r}_0}. \qquad (3.3)$$

Here, the diffusion *propagator* $\mathcal{G}_{t;\mathbf{r}_t,\mathbf{r}_0}$ is by definition the probability density function (PDF) of molecular displacements from point $\mathbf{r}_0$ to point $\mathbf{r}_t$ over time t. It is the most basic characteristic of diffusion in a given medium (tissue). Formally, it is the fundamental solution of the diffusion equation (2.6) and it will be discussed in detail in the following Section IV, cf. Eq. (4.2) below. Qualitatively, $\mathcal{G}_{t;\mathbf{r}_t,\mathbf{r}_0}$ governs the purely diffusive spreading of a "packet" of magnetization carried by water molecules in-between the two pulses.

The measurement effectively averages over the initial points $\mathbf{r}_0$ (as the signal is acquired from all protons in a voxel), and sums over all the possible finite points $\mathbf{r}_t$ (since the probabilities of mutually excluding events of taking different Brow-

nian paths emanating from $\mathbf{r}_0$ add up). Introducing the displacement $\mathbf{r} = \mathbf{r}_t - \mathbf{r}_0$, we obtain the key relation between the diffusion-weighted signal $S_\mathbf{q}(t)$ and the voxel-averaged propagator [22]:

$$S_\mathbf{q}(t) = \int_V \mathrm{d}\mathbf{r}\, \frac{\mathrm{d}\mathbf{r}_0}{V}\, e^{-i\mathbf{q}\mathbf{r}} \mathcal{G}_{t;\mathbf{r}_0+\mathbf{r},\mathbf{r}_0} \equiv G_{t,\mathbf{q}}\,. \tag{3.4}$$

Defined in this way, the diffusion-weighted signal is normalized to be $S_\mathbf{q}(t)|_{t=+0} \equiv 1$. Eq. (3.4) says that we are directly measuring the spatial Fourier transform of the voxel-averaged displacement PDF, discussed in greater detail around Eq. (4.6) of the following Section. But already at this point it is intuitively clear that (i) the diffusion-weighted measurement contains a lot of information about tissue structure and (ii) this information is "hidden" due to the averaging over all spins and their Brownian paths within a voxel.

Sampling the q-space, i.e. the space of all vectors $\mathbf{q}$ defined via Eq. (3.2), by applying balanced diffusion gradients $\mathbf{g}(\tau)$ in three dimensions, has been introduced under the term "q-space imaging" by Callaghan *et al.* [23] in 1988. Currently, there are numerous schemes of sampling the q-space. The total number and configuration of points (Cartesian, spherical shells, etc), and the maximal q depend on the gradient performance and on the limitations on the scan time. In general, the choice of an acquisition scheme should be tied to the model of the mesoscopic tissue structure which we would like to probe; there is certainly no single "best" q-space sampling scheme.

C. DWI signal representation: Cumulant expansion

1. To model or to represent?

The fundamental relation (3.4) is completely general, i.e. sample- and model-independent. The goal of this subsection is to outline the common way to *represent* the measured propagator $G_{t,\mathbf{q}}$. Note that *representing the propagator is not the same as modeling diffusion*; this is the point where literature is sometimes misleading, when using the term "modeling" for both genuine models and signal representations.

A representation has little or no assumptions and has a very broad scope of applicability. In other words, representing a general $G_{t,\mathbf{q}}$ is akin to expanding it in a certain basis. Think of the Taylor expansion, or the Laplace transform (a multiexponential fit) as examples. As there are infinite ways to represent any continuous function, the choice of representation is often dictated by convenience or tradition, and arguments about what is the best representation sometimes resemble religious disputes.

A model of diffusion has concrete assumptions and tissue parameters as inputs, and a very particular form for the diffusion propagator $G_{t,\mathbf{q}}$ or its characteristics (such as the diffusion coefficient, as described below).

The basis coefficients of a given representation can serve as empirical parameters which may e.g. correlate with pathology, i.e. they can be *sensitive*. It is often useful to use commonly accepted signal representations, such as the cumulant expansion described below, when we are not yet sure about the relevant tissue structure affecting our signal, but we still would like to present our experimental results in the way that can be useful for the future analysis.

However, representations by construction are *not specific* — i.e. they cannot provide information about particular tissue changes at a cellular level. To become specific (which is usually much more difficult), one needs to develop mesoscopic models of diffusion based on our knowledge of tissue structure at the level of the diffusion length (1.3), examples of which will be considered further, in Section V.

Both models and representations have their roles in diffusion MRI data analysis, and it is very important to distinguish between them.

2. The simplest case: Gaussian diffusion

Consider a uniform medium (e.g. a simple liquid), with $D(\mathbf{r}) = D_0 = \mathrm{const}$, so that the diffusion equation is given by Eq. (2.4). Let us find its *fundamental solution* $G^{(0)}_{t,\mathbf{r}}$, i.e. the one corresponding to a unit "packet" of random walkers spreading from the origin for time $t > 0$. It is governed by Eq. (2.4) with a unit source term

$$\partial_t G^{(0)}_{t,\mathbf{r}} = D_0 \partial^2_\mathbf{r} G^{(0)}_{t,\mathbf{r}} + \delta(t)\delta(\mathbf{r})\,. \tag{3.5}$$

The evolution of the corresponding spin magnetization at $\mathbf{r} = \mathbf{r}_0$ would in this case be governed by Eq. (3.3) with $\mathcal{G}_{t;\mathbf{r}_t,\mathbf{r}_0} \to G^{(0)}_{t,\mathbf{r}-\mathbf{r}_0}$.

As Eq. (3.5) is the partial differential equation with constant coefficients, it is readily solved by the Fourier transform

$$G^{(0)}_{t,\mathbf{r}} = \int \frac{\mathrm{d}\omega}{2\pi} \frac{\mathrm{d}^d\mathbf{q}}{(2\pi)^d}\, e^{-i\omega t + i\mathbf{q}\mathbf{r}}\, G^{(0)}_{\omega,\mathbf{q}} \tag{3.6}$$

in d spatial dimensions. In the Fourier domain, Eq. (3.5) turns to the algebraic equation

$$-i\omega G^{(0)}_{\omega,\mathbf{q}} = -D_0 q^2 G^{(0)}_{\omega,\mathbf{q}} + 1\,. \tag{3.7}$$

Its solution is given by the Lorentzian spectral line shape

$$G^{(0)}_{\omega,\mathbf{q}} = \frac{1}{-i\omega + D_0 q^2} \tag{3.8}$$

(recall, we are measuring the transverse relaxation in the applied gradient!). The inverse temporal Fourier transform of the Lorentzian, calculated by closing the integration contour in the lower half plane of the complex plane of ω, yields the exponentially decaying signal

$$S^{(0)}_\mathbf{q}(t) = G^{(0)}_{t,\mathbf{q}} \equiv \theta(t)\, e^{-D_0 q^2 t}\,, \tag{3.9}$$

where $\theta(t > 0) = 1$ and $\theta(t < 0) = 0$ is a unit step function. A Gaussian propagator in $\mathbf{q}$ results in a Gaussian propagator in $\mathbf{r}$:

$$G^{(0)}_{t,\mathbf{r}} = \frac{\theta(t)}{(4\pi D_0 t)^{d/2}}\, e^{-r^2/(4D_0 t)}\,. \tag{3.10}$$

3. Cumulant expansion

Let us now go back to the general case of diffusion governed by Eq. (2.6). The propagator will then not be a simple Gaussian; how should we characterize it in the general case? Intuitively, we expect the packet spreading as some kind of a bell-shaped curve; it turns out that there exists a century-old general way to characterize such curves, often emerging in statistics as probability density functions (PDFs), that deviate from the Gaussian shape [24]. This representation for an arbitrary PDF is called *cumulant expansion* [25, 26].

Consider the *characteristic function* (the Fourier transform)

$$\tilde{p}(k) \equiv \left\langle e^{ikw} \right\rangle = \int \mathrm{d}w\, p(w)\, e^{ikw} \tag{3.11}$$

for any PDF $p(w)$. Its Taylor expansion in k generates the *moments* $m_n \equiv \langle w^n \rangle$:

$$\tilde{p}(k) = 1 + ik\,\langle w\rangle + \frac{(ik)^2}{2!}\left\langle w^2\right\rangle + \dots = \sum_{n=0}^{\infty} \frac{(ik)^n m_n}{n!}\,. \tag{3.12}$$

The *cumulants* are defined as the Taylor expansion coefficients of $\ln \tilde{p}(k)$:

$$\ln \tilde{p}(k) = ik\,\langle w\rangle + \frac{(ik)^2}{2!}\left[\left\langle w^2\right\rangle - \langle w\rangle^2\right] + \dots = \sum_{n=1}^{\infty} \frac{(ik)^n c_n}{n!}\,. \tag{3.13}$$

The first cumulant $c_1 \equiv m_1$ is the *mean*, the second $c_2 = m_2 - m_1^2$ is *variance*, the dimensionless ratio $c_3/c_2^{3/2}$ is called *skewness*, the dimensionless ratio c_4/c_2^2 is called *kurtosis*, etc. We can see that the Gaussian distribution is characterized by the first two nonzero cumulants and vanishing $c_n = 0$, $n > 2$; higher order cumulants effectively tell us how far our distribution is from the Gaussian shape. Skewness is the measure of its asymmetry relative to the mean, and kurtosis tells how dominant are its tails relative to the tails of the Gaussian distribution.

Intuitively, the cumulants are more convenient than the moments, in that the higher-order terms improve our accuracy of describing the bell-shaped PDF around its peak. Formally, c_n is obtained from m_n by subtracting the "trivial" part, corresponding to combinations of the lower-order moments, such as the term $-m_1^2$ in the definition of variance, or $c_4 = \left\langle (w - m_1)^4 \right\rangle - 3c_2^2$ for the fourth-order cumulant.

A beautiful combinatorial relation expressing a moment m_n through cumulants of up to order n, called *linked cluster expansion* derived by Mayer in the statistical physics context [27] in 1941, and whose generalization is the basis of the modern field theoretical description of condensed matter systems [28], is proven in ref. [26]. It generalizes relations $m_2 = c_2 + c_1^2$, $m_3 = c_3 + 3c_2c_1 + c_1^3$, $m_4 = c_4 + 4c_3c_1 + 6c_2c_1^2 + 3c_2^2 + c_1^4$, by connecting the combinatorial coefficients to the number of possibilities of linking n points into all possible clusters.

4. Cumulant expansion for the diffusion-weighted signal

To make connection between diffusion MRI and the general theory [24, 25], we note that the fundamental relation (3.4) between the diffusion propagator and the measurement can be recast as an average, cf. Eq. (3.11), where the random variable is the spin precession phase ϕ, and the averaging $\langle \dots \rangle$ is performed both over the Brownian paths of a given spin and over all spins in a voxel (the double average, cf. also Section IV below):

$$S_{\mathbf{q}}(t) = \left\langle e^{i\phi} \right\rangle, \tag{3.14a}$$

$$\phi = -\int_0^t \mathbf{g}(\tau)\mathbf{r}(\tau)\,\mathrm{d}\tau = \int_0^t \mathbf{q}(\tau)\mathbf{v}(\tau)\,\mathrm{d}\tau, \tag{3.14b}$$

$$\mathbf{q}(t) = \int_0^t \mathbf{g}(\tau)\,\mathrm{d}\tau\,. \tag{3.14c}$$

Here the surface terms in the integration by parts in Eq. (3.14b) vanish under the balanced gradient condition $\int_0^t \mathbf{g}(\tau)\,\mathrm{d}\tau \equiv 0$, and $\mathbf{v} = \mathrm{d}\mathbf{r}/\mathrm{d}t$ is the molecular velocity.

The form (3.14) generalizes the narrow-pulse limit described in Sec. III B above. In the case of arbitrary gradient train $\mathbf{g}(\tau)$, the diffusion-weighted signal $S_{\mathbf{q}}(t)$ is *not* equal to the diffusion propagator for a given $\mathbf{q}$ and t. In general, the diffusion-weighted signal is a *functional* of the gradient form $\mathbf{g}(\tau)$, or, equivalently, of its "anti-derivative" $\mathbf{q}(\tau)$: $S \equiv S[\mathbf{q}(\tau)]$.

Consider now the case of a sufficiently small diffusion weighting parameter $q = |\mathbf{q}|$. This is especially relevant for clinical MRI, where diffusion gradients are limited. In this case, the signal (3.4) over the time interval t can be found within the Gaussian phase approximation [22, 26] which amounts to keeping only first two cumulants:

$$S = \left\langle e^{i\phi} \right\rangle \approx e^{-\left\langle \phi^2 \right\rangle/2}, \quad \langle \phi \rangle \equiv 0\,. \tag{3.15}$$

The sample average of the phase vanishes, $\langle \phi \rangle \equiv 0$, due to time reversal invariance in the absence of net flow, $\langle \mathbf{v} \rangle = 0$. (Time reversal invariance actually implies that all *odd* moments of the phase vanish, $\langle \phi^n \rangle = 0$, $n = 1, 3, 5, \dots$.) As a result, Eq. (3.15) yields

$$-\ln S(t) \simeq \frac{1}{2}\int_0^t \mathrm{d}\tau_1 \mathrm{d}\tau_2\, q_i(\tau_1)\,\langle v_i(\tau_1) v_j(\tau_2)\rangle\, q_j(\tau_2)\,. \tag{3.16}$$

Here the Einstein's convention of summing over repeating indices $i, j = 1, 2, 3$ is implied. The signal depends on the total duration t of the gradient train $g(\tau)$, and is a functional of the diffusion-weighting (3.14c).

For uniform (but generally anisotropic) media,

$$\langle v_i(\tau_1) v_j(\tau_2)\rangle = 2D_{ij}\delta(\tau_1 - \tau_2)\,, \tag{3.17}$$

where D_{ij} are by definition the components of the symmetric *diffusion tensor*. This leads to the anisotropic generalization of the standard Gaussian expression (3.9) obtained above:

$$\ln S \simeq -b_{ij} D_{ij}\,, \tag{3.18}$$

where the so-called b-matrix, defined as

$$b_{ij} = \int_0^t q_i(\tau) q_j(\tau)\, \mathrm{d}\tau\,, \tag{3.19}$$

with $b = \mathrm{tr}\, b_{ij}$ called the b-value, is the commonly accepted parameter characterizing the strength and the spatial arrangement of the diffusion weighting.

When the diffusion is non-Gaussian, but we are using narrow pulses (3.2), we may represent Eq. (3.16) as

$$-\ln S(t) \simeq q_i q_j t \cdot D_{ij}(t)\,, \quad D_{ij}(t) = \frac{\langle \delta x_i(t) \delta x_j(t)\rangle}{2t}\,, \tag{3.20}$$

where q_i is the amplitude of the ith component of the vector $\mathbf{q}$ defined after Eq. (3.2), $\delta x_i(t) = x_i(t) - x_i(0)$ is the displacement along the ith coordinate axis, and $D_{ij}(t)$ are the time-dependent components of the diffusion tensor.

The linear estimation problem (3.18) or (3.20) for the diffusion tensor, referred to as the diffusion tensor imaging (DTI), has been solved by Basser *et al.*[29]. It requires measurement along at least 6 non-collinear diffusion directions in addition to the $b = 0$ (unweighted) image. The parameter $\bar\lambda \equiv \frac{1}{3}\,\mathrm{tr}\, D_{ij}$ is called mean diffusivity; the fractional anisotropy

$$\mathrm{FA} = \sqrt{\frac{3}{2}}\sqrt{\frac{(\lambda_1-\bar\lambda)^2 + (\lambda_2-\bar\lambda)^2 + (\lambda_3-\bar\lambda)^2}{\lambda_1^2+\lambda_2^2+\lambda_3^2}} \tag{3.21}$$

characterizes imbalance between (in general, time-dependent) diffusion tensor eigenvalues $\lambda_{1,2,3}$.

Likewise, the mapping of the kurtosis tensor, via the expansion up to q^4,

$$\ln S \simeq -b \hat g_i \hat g_j D_{ij} + \frac{b^2 \bar\lambda^2}{6} \hat g_i \hat g_j \hat g_k \hat g_l W_{ijkl} \tag{3.22}$$

called diffusion kurtosis imaging (DKI), has been introduced by Jensen *et al.*[30]. Above, $\hat g_i$ are the components of the unit vector in the direction of the gradient $\mathbf{g}$. The weights for unbiased estimation of diffusion and kurtosis tensors for non-Gaussian NMR noise were suggested recently [31].

We emphasize that the cumulants $D_{ij}, W_{ijkl}, \ldots$ of the signal obtained via Taylor-expanding its logarithm in the (even) powers of q_i, or equivalently, in the powers of b, correspond to the cumulants of the genuine molecular displacements $\mathbf{r} = \mathbf{r}_t - \mathbf{r}_0$ only in the narrow pulse limit (3.2), whenever the precession phase $\phi = -\mathbf{qr}$. When the finite gradient pulse duration δ is comparable to the time scale of the transient processes (in the case of time-dependent diffusion coefficient and higher-order cumulants), the measurement acts as a low-pulse filter with a cutoff frequency $\sim 1/\delta$.

Even in the narrow-pulse limit (3.2), the phase diagram of diffusion weighted imaging is two-dimensional (Fig. 2), characterized both by the magnitude q and the diffusion time t between the pulses. Generally, the b-value alone does not uniquely characterize the measurement — unless the corresponding cumulants are time-independent. Physically, this corresponds to Gaussian diffusion in all non-exchanging tissue compartments.

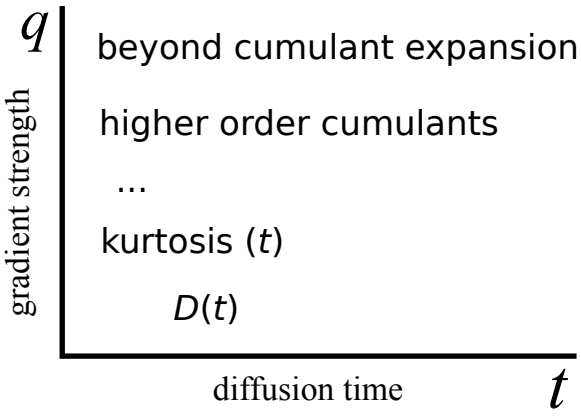

FIG. 2. The phase diagram of diffusion MRI is two-dimensional: By increasing q one accesses the progressively higher-order diffusion cumulants $\langle(\delta x)^2\rangle$, $\langle(\delta x)^4\rangle - 3\langle(\delta x)^2\rangle^2$, ..., whereas the dependence along the t-axis reflects their evolution over an increasing diffusion length scale $L \sim \sqrt{tD(t)}$. The b-matrix (3.19) alone does not uniquely describe the diffusion measurement, unless diffusion in all tissue compartments is Gaussian.

An interesting measurement scheme involving *pairs* of narrow pulses, has been proposed by Mitra in 1995 [32]. In this case, the measurement contains information beyond the single-particle diffusion propagator, since $\langle e^{-i\mathbf{q}_1\mathbf{r}_1 - i\mathbf{q}_2\mathbf{r}_2}\rangle \neq \langle e^{-i\mathbf{q}_1\mathbf{r}_1}\rangle\langle e^{-i\mathbf{q}_2\mathbf{r}_2}\rangle$. However, multiple pulse measurements are equivalent to a combination of single pulse measurements up to $\mathcal{O}(q^2)$, and provide unique contrast only at the level $\mathcal{O}(q^4)$ and beyond [33].

How many parameters do the successive cumulant tensors have? A term of rank l_c is a fully symmetric tensor, which can be represented in terms of the so-called symmetric trace-free (STF) tensors of rank $l_c, l_c-2, \ldots, 2, 0$ [34]. Each set of $2l+1$ STF tensors of rank l realizes an irreducible representation of the SO(3) group of rotations, equivalent to the set of $2l+1$ spherical harmonics Y_{lm} [34]. Hence, the total number of nonequivalent components in the rank-l_c cumulant is

$$n_c(l_c) = \sum_{l=0,2,\ldots}^{l_c} (2l+1) = \frac{1}{2}(l_c+1)(l_c+2)\,. \tag{3.23}$$

Suppose we truncate the cumulant series at the (even) term of the rank l_c. Thereby we determine all the parameters of the cumulant tensors (diffusion, kurtosis, ...) of the ranks $2,\ 4,\ \ldots,\ l_c$. This total number of parameters in the truncated series is equal to

$$N_c(l_c) = \sum_{l=2,4,\ldots}^{l_c} n_c(l) = \frac{l_c}{3}\left[\frac{l_c}{2}+1\right]\left[\frac{l_c}{2}+2\right] + \frac{l_c}{4}\left[\frac{l_c}{2}+3\right] \tag{3.24}$$

corresponding to $N_c = 6,\ 21,\ 49, \ldots$ for $l_c = 2,\ 4,\ 6, \ldots$ (we did not include the proton density $S|_{b=0}$ in our counting).

IV. DIFFUSION CHARACTERISTICS OF A MACROSCOPIC SAMPLE. EFFECTIVE MEDIUM THEORY

In this Section, we take a step back from specifics of diffusion MRI measurements, and focus on a connection between diffusion propagator acquired from a large volume (e.g. voxel) and the statistical properties of the medium (tissue).

Consider a macroscopic sample with mesoscopic structure (e.g. a tissue voxel), where diffusion is governed by our main equation, Eq. (2.6). Consider for simplicity a statistically isotropic situation (i.e. low FA). Our macroscopic spatial resolution by definition does not allow us to directly resolve the contribution of water molecules from a particular part of this sample to the overall NMR signal. Hence, the measurement (3.4) effectively *averages* over the structural contributions "seen" by all the spins (molecules) in all parts of the sample.

Our technical task is to relate the measurement to the parameters of the original Eq. (2.6). Loosely speaking, we want to "average", in some sense, this equation over the sample. Of course, it is not Eq. (2.6), but its *solution*, the measured signal (3.4), that is really being averaged. This is good news — otherwise it would only be the sample average $\langle D(\mathbf{r})\rangle$ that would matter. This is clearly not the case, and some mesoscopic features of $D(\mathbf{r})$ do manifest themselves even after acquiring the signal from a macroscopic sample. Unfortunately, most of the potentially interesting mesoscopic structural information gets washed out for good. This is what makes the recovering, or quantifying, the mesoscopic structural information a very difficult ill-posed problem.

The purpose of this Section is to establish a number of general relations between the macroscopic diffusion metrics irrespective of a particular model of the mesoscopic structure, and to provide connections with dispersive quantities which emerge in the effective medium description of transport, dielectric, magnetic and optical properties of structurally complex media. This will be the most technical Section of these notes. For the first reading, it would be possible to understand most of the material in the subsequent Sections, at least on the qualitative level, without going through all technical details escalating towards the end of this Section.

A. Macroscopic measurement: The double average

Let us now look into the origins of a measurement outcome from a macroscopic sample. The most commonly reported metric is the diffusion coefficient defined via the mean squared molecular displacement in a particular direction $\hat{\mathbf{x}}$ (cf. Eq. (3.20)):

$$D(t) = \frac{\langle \delta x^2(t)\rangle}{2t}\,. \tag{4.1}$$

Here $\delta x(t) \equiv x(t) - x(t)|_{t=0}$. For the general case of Eq. (2.6), both the propagator $\mathcal{G}_{t;\mathbf{r},\mathbf{r}_0}$ and all of its cumulants would depend on the initial point $\mathbf{r}_0$ [17]. In particular, in the continuous limit, this propagator satisfies Eq. (2.6)

$$\partial_t \mathcal{G}_{t;\mathbf{r},\mathbf{r}_0} = \partial_{\mathbf{r}}\left(D(\mathbf{r})\partial_{\mathbf{r}}\mathcal{G}_{t;\mathbf{r},\mathbf{r}_0}\right) + \delta(t)\delta(\mathbf{r}-\mathbf{r}_0) \tag{4.2}$$

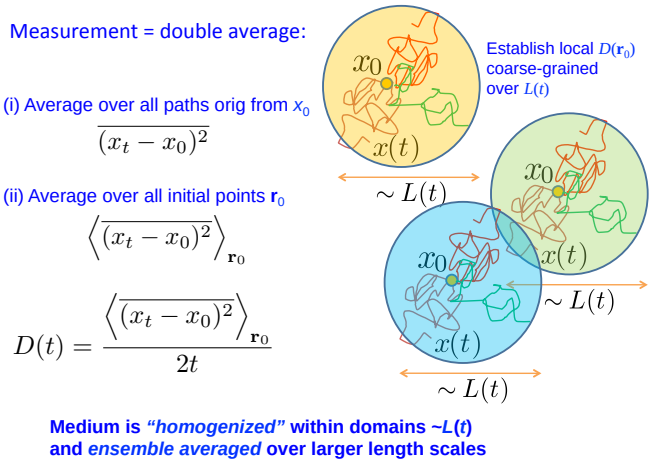

FIG. 3. The double average: (i) over the Brownian paths, establishing the local value of $D(\mathbf{r})|_{L(t)}$ coarse-grained over the diffusion length $L(t)$; (ii) over the ensemble, yielding the macroscopic diffusion metrics, such as Eq. (4.1).

with a source term corresponding to $\mathcal{G}_{t;\mathbf{r},\mathbf{r}_0}|_{t=+0} = \delta(\mathbf{r}-\mathbf{r}_0)$ and $\mathcal{G}_{t;\mathbf{r},\mathbf{r}_0}|_{t<0} \equiv 0$, cf. Eq. (3.5). The current conservation (2.1) ensures that the probability is conserved,

$$\int d\mathbf{r}\, \mathcal{G}_{t;\mathbf{r},\mathbf{r}_0} \equiv 1 \tag{4.3}$$

irrespective of the initial point $\mathbf{r}_0$, for all $t > 0$.

The lack of translation invariance due to the presence of the mesoscopic structure leads us to a subtle point regarding taking the average $\langle\ldots\rangle$ in the definition of the diffusion coefficient (4.1) and other diffusion metrics acquired from a macroscopic volume (voxel). Let us step back and realize that, in principle, there are different equivalent ways of measuring the diffusion coefficient. For instance, one can directly record [35] many paths taken sequentially by a given optically tagged molecule. In the case of NMR, we happen to acquire the information about molecular displacements from a macroscopic number of molecules in a given sample (voxel) simultaneously. But in any case, to obtain the sample diffusion coefficient, we need to subject the squared displacement to the *two averaging procedures* (Fig. 3):

• (i) Averaging

$$\overline{(x_t - x_0)^2} \equiv \int d\mathbf{r}\, \mathcal{G}_{t;\mathbf{r},\mathbf{r}_0}\,(x - x_0)^2 \tag{4.4}$$

over the Brownian paths $\mathbf{r}_{t'}$, $0 < t' < t$, originating from a given point $\mathbf{r}_0$, with $x_t = \mathbf{r}_t \cdot \hat{\mathbf{x}}$ being the displacement along chosen direction $\hat{\mathbf{x}}$. Most of those paths will be contained within a domain of size $\sim L(t)$, the (local) diffusion length. As Fig. 4 demonstrates (also cf. Sec. V below), this effectively *homogenizes* the sample's properties over a window of size $\sim L(t)$, establishing a local "coarse-grained" value of the diffusion coefficient $D(\mathbf{r}_0)|_{L(t)} = \overline{(x_t - x_0)^2}/2t$ which

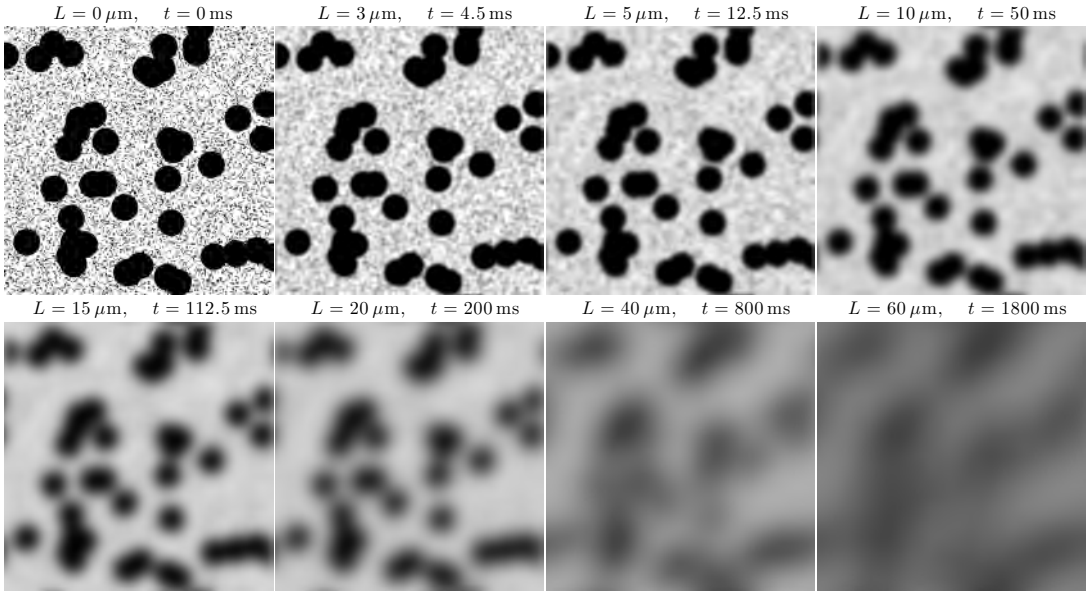

FIG. 4. Diffusion as a homogenization process, as a result of the double average, cf. Fig. 3. An example of a medium with mesoscopic structure was created by randomly placing black regions (disks) of two different radii, $R_{\rm small} = 1\,\mu$m and $R_{\rm large} = 20\,\mu$m, top left panel. To obtain the snapshots of the medium as "seen" by the diffusing molecules, we used a Gaussian filter with width $L/2$, where $L(t) = \sqrt{2Dt}$, and ignored the time dependence of D in the definition of diffusion length, setting it to a typical value $D = 1\,\mu$m^2/ms for simplicity.

is smoother than the sharply varying one, $D(\mathbf{r})$, that enters the original Eq. (2.6).

• (ii) Subsequent ensemble averaging

$$\langle \delta x^2(t)\rangle \equiv \langle \overline{(x_t - x_0)^2}\rangle_{\mathbf{r}_0} = \int d\mathbf{r} \frac{d\mathbf{r}_0}{V}\, \mathcal{G}_{t;\mathbf{r},\mathbf{r}_0}\, (x - x_0)^2 \tag{4.5}$$

over all initial points $\mathbf{r}_0$ within a macroscopic sample. As we discuss below, given a large enough sample, this effectively averages over all realizations of the structural disorder.

Generally, in a macroscopic measurement, it is the above double average that is always present in any global diffusion metric, e.g. the diffusion coefficient (4.1), or the signal (3.4) and (3.14a). From now on, we will drop the averaging bar and $\mathbf{r}_0$ and will use the usual $\langle \dots \rangle$ notation for the double average for brevity.

To illustrate the role of both averages (i) and (ii), consider Fig. 4. A sample medium was created by randomly placing black regions (disks) of two different radii. For the illustration purposes, we can consider the situation when molecules can travel everywhere, but the black regions are made of a material with diffusion coefficient different from that of the white background. As the diffusion time t evolves, the increasing diffusion length $L(t)$ acts as a window for the (approximately) Gaussian filter, which homogenizes the properties over the scale L as a result of the averaging (i), producing the increasingly smoothly varying coarse-grained local diffusion coefficient $D(\mathbf{r})|_{L(t)}$. Hence, from the point of the molecules, in this particular example, for time $t \gtrsim 50$ ms the small disks are practically invisible, i.e. their effect is indistinguishable from a uniform "gray" medium with some effective diffusion coefficient filling in the space between the large disks. Eventually, as $t \approx 2$ s, the medium becomes almost completely uniform, at which point the macroscopic (bulk) diffusion coefficient D_∞ is approached. At each time instant, the averaging (ii) in the definition (4.1) implies that the observed diffusion coefficient is the sample average of the local $D(\mathbf{r})|_{L(t)}$. In Section V, we will relate the time dependence of the observed diffusion coefficient to the decreasing variance of the local $D(\mathbf{r})|_{L(t)}$.

To summarize: The measured diffusion characteristics, such as Eq. (4.1), describe a macroscopic sample *as a whole*. They do not belong to any given Brownian path, but rather emerge as a result of averaging (i) over all possible Brownian paths that could be taken by a given molecule, and (ii) over the initial positions of all molecules in a sample.

B. Diffusion equation in a macroscopic sample

In Sec. II, based on the conservation law and symmetry considerations, we have derived Eq. (2.6) for the general case of a spatially varying local diffusion coefficient. This equation is valid at the mesoscopic scale, where the local value of $D(\mathbf{r})$ is already well defined, and onwards.

Here, we ask a different question: **How would a diffusion equation look at the level of a macroscopic sample (voxel)?**

At first, this sounds like a dumb question: The correct equation (2.6) is already written, what else can one ask for? Well, in reality, we want to know the *solution* to Eq. (2.6), and better analytically than numerically. But it is *truly difficult*. In fact, it's even harder than it looks because one cannot in all honesty write this equation down, i.e. precisely specify the local values of $D(\mathbf{r})$ at all points in space in a voxel at a microme-

ter scale. It's just impossible at the moment — we simply do not know tissue properties so well! And, besides, cells are all somewhat different and a voxel is so huge compared to a single cell, that we would need literally millions of parameters to specify the diffusion equation. Hence, Eqs. (2.6) and (4.2) for a macroscopic tissue sample (voxel) are a complete mess.

Then, one can reasonably say, solving these equations is a losing proposition, so we should all pack up and go home.

1. *Disorder averaging and translation invariance*

But suppose we are not in the position of giving up. (And also, otherwise we wouldn't have asked the above question to begin with. For those who are impatient to see the answer, see Eq. (4.19) below.)

What can make our lives easier? Fortunately, it is quite obvious that, in order to interpret a macroscopic measurement, we never need the solution of Eqs. (2.6) and (4.2) involving all those millions of hard-to-specify parameters. Indeed, in agreement with the averaging (ii) described above, any metric acquired over the sample, including the full diffusion propagator (the displacement PDF), cf. Eq. (3.4),

$$G_{t,\mathbf{r}} \equiv \langle \mathcal{G}_{t;\mathbf{r}_0+\mathbf{r},\mathbf{r}_0} \rangle_{\mathbf{r}_0} \tag{4.6}$$

gets effectively averaged over all the initial points $\mathbf{r}_0$.

Remarkably, the ensemble-averaged propagator (4.6) is much simpler than the unknown $\mathcal{G}_{t;\mathbf{r},\mathbf{r}_0}$, which is to say that one difficulty (of performing ensemble average over all parts of the voxel, item (ii) from Sec. IV A) mostly cancels out the other one (the need to specify Eq. (4.2) precisely).

In particular, one notes that the propagator (4.6) has become *translation invariant* after averaging over the initial point $\mathbf{r}_0$, i.e. it depends only the displacement but not on the start and end points separately (as does $\mathcal{G}_{t;\mathbf{r},\mathbf{r}_0}$). This is the main signature of the averaging over $\mathbf{r}_0$, or, as physicists sometimes say, of the *disorder averaging*. This is quite a general point, valid not just for diffusion. One can imagine a typical disordered sample with an irregular mesoscopic structure of the same "origin", or "kind". Basically, think about particular spatial configurations of this structure in different parts of the voxel as taken from some probability distribution. If a voxel is large enough, one can mentally split it into many smaller parts; then, these different parts of the voxel would contain different *disorder realizations*. Hence, averaging over the initial point $\mathbf{r}_0$ corresponds to averaging over all possible realizations (configurations) of the disorder. In fact, this is why the averaging (ii) is called the ensemble averaging — we average over the ensemble of the disorder configurations of the same general origin. One then says that disorder averaging restores translation invariance. This is a somewhat different, but often helpful, interpretation of the averaging (ii) and of Eq. (4.6). Concrete examples of restored translation invariance and calculation of the disorder-averaged propagator (4.6) can be found in [16, 17, 36].

In terms of the disorder-averaged propagator (4.6), the net result of the double average (i) and (ii) follows from Eqs. (4.5) and (4.6) after the change of the integration measure $\mathbf{d}\mathbf{r}\mathbf{d}\mathbf{r}_0 \to \mathrm{d}(\delta\mathbf{r})\mathbf{d}\mathbf{r}_0$, where $\delta\mathbf{r} = \mathbf{r} - \mathbf{r}_0$:

$$\langle \delta x^2(t) \rangle = \int \mathrm{d}(\delta\mathbf{r})\, G_{t;\delta\mathbf{r}}\, (\delta x)^2 \,. \tag{4.7}$$

The double-average of the higher-order displacement cumulants is obtained in the same way. As a result, acquisition over a macroscopic sample effectively means measuring the disorder-averaged PDF (4.6) or, practically, its first few displacement cumulants.

2. *Dispersion due to heterogeneity*

In Physics, we deal with disorder-averaged properties all the time. Recall, for instance, electrodynamics, or optics, in continuous media. We generally operate with the material parameters, such as the dielectric function ε, or the magnetic permittivity μ, or the electrical conductivity σ, or the refraction index n. The most important feature of all those quantities is that they are *dispersive*, i.e. they depend on the *frequency*.

Probably the first known example of dispersion has been spotted by Newton in his experiments with prisms. As understood centuries later, the observation of a rainbow after a ray of white light passes through a prism (a signature of the frequency-dependent refraction index), is a consequence of a complicated atomic and mesoscopic structure of glass. In other words, **dispersion is a sign of "invisible" structural complexity**.

Consider a different well-known example: suppose we drive an electric current through a material. In general, its relation to the applied voltage, or to the electric field $\mathbf{E}$, will be dispersive [37–39]:

$$\mathbf{J}_\omega = \sigma(\omega)\mathbf{E}_\omega \,. \tag{4.8}$$

This is the familiar differential form of the Ohm's law [38]. A simpler example of the same phenomenon is an RLC circuit where the admittance (reciprocal of the impedance) depends on the frequency. A similar relation, $\mathbf{D}_\omega = \varepsilon(\omega)\mathbf{E}_\omega$, exists between the electrical induction $\mathbf{D}$ and field in a dielectric [38], in an ionic solution, or in plasma [39]; or between magnetic induction and field in a para- or diamagnetic material [38], etc.

In the time domain, the dispersive relation such as (4.8) signifies the delayed, or *retarded* response [40]:

$$\mathbf{J}_t = \int_{-\infty}^{t} \mathrm{d}t'\, \sigma(t-t')\, \mathbf{E}_{t'} \,, \quad \sigma(\tau) = \int \frac{\mathrm{d}\omega}{2\pi}\, \sigma(\omega) e^{-i\omega\tau} \,. \tag{4.9}$$

This means that the current at time t is determined by the *history* of the applied voltage or field at earlier moments $t' \le t$.

Clearly, the dispersion of material properties arise from the "structure". This structure is generally at the mesoscopic scale (complex and spatially varying molecular properties causing light refraction in glass, dielectric screeining, etc). Much like for the case of diffusion described above, these dispersive metrics characterize a sufficiently macroscopic sample (there is

no dielectric constant or refraction index for a single Silicon atom but there is one for a piece of silicon).

Besides the temporal dispersion, there exists in general also the *spatial dispersion*, i.e. a linear relation separately for each spatial Fourier harmonic of, say, the current $\mathbf{J}_{\omega,\mathbf{q}}$, or the electrical induction $\mathbf{D}_{\omega,\mathbf{q}}$, and of the electric field, so the material characteristics depend both on ω and $\mathbf{q}$, such as [37–39]

$$\mathbf{J}_{\omega,\mathbf{q}} = -\sigma(\omega,\mathbf{q})\, i\mathbf{q}\Phi_{\omega,\mathbf{q}}\,, \tag{4.10}$$

where the electric field $\mathbf{E} = -\partial_{\mathbf{r}}\Phi \to -i\mathbf{q}\Phi$ is given by the gradient of the potential Φ.[1] For a macroscopic sample, $q \to 0$, and the relation (4.10) can be expanded in the (odd) powers of q, with $\sigma(\omega) \equiv \sigma(\omega,\mathbf{q})|_{q=0}$ in Eq. (4.8).

All these dispersive material characteristics arise as a result of averaging over the mesoscopic structure during a macroscopic measurement. They represent essentially what survives this (double) averaging, since almost all the details of the atomic and molecular environment are washed out.

Note that, following the arguments of Sec. IV B 1, the notation in Eq. (4.10) assumes that the *space translation invariance is restored*, such that in real d-dimensional space this corresponds to

$$\mathbf{J}_{t,\mathbf{r}} = \int_{-\infty}^{t} \mathrm{d}t' \int \mathrm{d}^d\mathbf{r}\, \sigma(t-t',\mathbf{r}-\mathbf{r}')\, \mathbf{E}_{t',\mathbf{r}'}\,, \tag{4.11}$$

$$\sigma(\tau,\mathbf{r}) = \int \frac{\mathrm{d}\omega}{2\pi}\frac{\mathrm{d}^d\mathbf{q}}{(2\pi)^d}\, \sigma(\omega,\mathbf{q})\, e^{-i\omega\tau+i\mathbf{q}\mathbf{r}}\,. \tag{4.12}$$

As the properties become homogeneous after the ensemble averaging, one may assume the following point of view: let us assume translation invariance in all our equations such as Eq. (4.11), keeping in mind that all the (invisible) heterogeneity is what causes the spatial and temporal dispersion, absent in a truly uniform sample. In effect, we supplemented the real, horribly microscopically complex material by a fictitious one which looks completely uniform, albeit having dispersive dielectric and conductive properties, refraction index, dispersive diffusivity, etc. This is the *effective medium theory* (EMT) way of thinking [9, 13, 16, 17, 37].

We can see that our problems in tissue biophysics are not that different from the ones studied in condensed matter physics, electrodynamics and optics. Similarly, we expect a general dispersive dependence between the (diffusive) current and the particle density; and similarly, most of the interesting short-distance details will disappear for good.

So the question we need be asking, while modeling and interpreting MRI metrics in tissues, is really the following one: What part of the tissue complexity survives the averaging due to the measurement in a macroscopic sample (voxel), and what is its contribution to a few macroscopic dispersive diffusion metrics which we can realistically measure?

[1] The part of the current related to the contribution to $\mathbf{E}$ coming from the vector potential $\mathbf{A}$ is not essential for the spatial dispersion, hence we omitted it in Eq. (4.10).

3. *Conserved current and generalized Fick's law*

In Sec. II, we introduced microscopic density ψ and current $\mathbf{j}$. Now let us introduce the *disorder-averaged* quantities

$$\Psi(t,\mathbf{r}) = \langle\psi(t,\mathbf{r})\rangle\,, \quad \mathbf{J}(t,\mathbf{r}) = \langle\mathbf{j}(t,\mathbf{r})\rangle\,. \tag{4.13}$$

These are the quantities averaged over the ensemble of disorder realizations: Suppose we create a lump of density of "labeled" particles and observe its diffusive evolution hindered by the mesoscopic structure. Now let us repeat the same observation many times, each time having a different realization of the underlying mesoscopic "disorder" in $D(\mathbf{r})$, such that averaging over all of those realizations in Eq. (4.13) represents the averaging over our macroscopic sample.

The question we asked in the very beginning of Sec. IV B is equivalent to the following one: What is the equation governing the evolution of $\Psi(t,\mathbf{r})$?

The current conservation (2.1) being an exact property, it still holds for the averaged quantities, since the averaging cannot possibly create or destroy molecules or spins:

$$\partial_t\Psi = -\partial_{\mathbf{r}}\cdot\mathbf{J}\,, \tag{4.14}$$

which for the Fourier components reads as

$$\omega\Psi_{\omega,\mathbf{q}} = \mathbf{q}\cdot\mathbf{J}_{\omega,\mathbf{q}}\,. \tag{4.15}$$

It turns out [17] that it is the Fick's law (2.3) that gets modified, becoming a generally dispersive relation (both in time and in space) between the current and the density gradients:

$$\mathbf{J}_{\omega,\mathbf{r}} = -\mathcal{D}(\omega)\partial_{\mathbf{r}}\Psi_{\omega,\mathbf{r}} + \mathcal{O}(\partial_{\mathbf{r}}\nabla^2\Psi_{\omega,\mathbf{r}})\,, \tag{4.16}$$

where the right-hand side contains all the odd-order gradients of density. Equivalently, in both ω and $\mathbf{q}$ space,

$$\mathbf{J}_{\omega,\mathbf{q}} = -\mathcal{D}(\omega)i\mathbf{q}\Psi_{\omega,\mathbf{q}} + \mathcal{O}(\mathbf{q}q^2\Psi_{\omega,\mathbf{q}})\,. \tag{4.17}$$

Note a similarity with Eqs. (4.10) and (4.11).

The general form of the Fick's law (4.16) and (4.16) emerging after the disorder averaging makes $\mathcal{D}(\omega)$ (together with higher-order dispersive parameters in the above equation) a central object in the effective medium description of diffusion in disordered media [16, 17, 36, 41]. Indeed, combined with the conservation law (4.15), it defines the disorder-averaged diffusion equation

$$-i\omega\Psi_{\omega,\mathbf{q}} = -\mathcal{D}(\omega)q^2\Psi_{\omega,\mathbf{q}} + \mathcal{O}\left(q^4\Psi_{\omega,\mathbf{q}}\right) \tag{4.18}$$

in ω, $\mathbf{q}$ domain (cf. Eq. (3.7)), or equivalently

$$-i\omega\Psi_{\omega,\mathbf{r}} = \mathcal{D}(\omega)\nabla^2_{\mathbf{r}}\Psi_{\omega,\mathbf{r}} + \mathcal{O}\left(\nabla^4_{\mathbf{r}}\Psi_{\omega,\mathbf{r}}\right) \tag{4.19}$$

in ω, $\mathbf{r}$ domain, which incorporates the characteristics of the mesoscopic restrictions that can be still quantified with a bulk measurement. This is the answer to the question "How would a diffusion equation look at the level of a macroscopic sample?" posed above. While Eq. (4.19) may look fairly abstract, in order to obtain it, we have realized how the macroscopic diffusion metrics emerge as a result of the signal acquisition

over a sample (voxel), and have drawn parallels with previously studied physics paradigms. We also note that a similar dispersive equation can be written for the sample's magnetization when the local Larmor frequency $\Omega(\mathbf{r})$ entering the Bloch-Torrey equation (3.1) is varying at the mesoscopic scale [42]. Hence, the EMT approach naturally unifies the diffusive and relaxational phenomena.

4. *The self-energy part*

Similar to Eqs. (4.18) and (4.19), one may ask for the corresponding equation for the disorder-averaged propagator (4.6). This is the solution for Eq. (4.18) with the unit source (corresponding to $1_{\omega,\mathbf{q}} \leftrightarrow \delta(t)\delta(\mathbf{r})$, cf. Eqs. (3.7) and (4.2)),

$$-i\omega G_{\omega,\mathbf{q}} = \mathcal{D}(\omega)q^2 G_{\omega,\mathbf{q}} + \mathcal{O}\left(q^4 G_{\omega,\mathbf{q}}\right) + 1\,. \qquad (4.20)$$

We can rewrite this equation in the following form:

$$G_{\omega,\mathbf{q}} = \frac{1}{-i\omega + D_\infty q^2 - \Sigma(\omega,\mathbf{q})}\,, \qquad (4.21)$$

where the *self-energy part* $\Sigma(\omega,\mathbf{q})$ describes all non-Gaussian diffusion effects, i.e. deviations from the Gaussian bulk diffusion propagator of the form (3.8)

$$G^{(0)}_{\omega,\mathbf{q}} = \frac{1}{-i\omega + D_\infty q^2} \qquad (4.22)$$

in a fully homogenized sample (cf. the longest-time snapshot in Fig. 4). Technically, it is the expansion of the self-energy part in the powers of q^2,

$$\Sigma(\omega,\mathbf{q}) = [D_\infty - \mathcal{D}(\omega)]\,q^2 + \Sigma_4(\omega)q^4 + \dots\,, \qquad (4.23)$$

that yields the dispersive corrections to the diffusion equation (4.19) and the Fick's law (4.17), Ref. [17].

5. *The dispersive diffusivity*

Let us for simplicity drop the higher-order terms in Eq. (4.19), and focus on $\mathcal{D}(\omega)$. It must provide a retarded response,

$$\mathbf{J}_{t,\mathbf{r}} = -\int \mathrm{d}t'\,\mathcal{D}(t-t')\nabla_{\mathbf{r}}\Psi_{t',\mathbf{r}} \qquad (4.24)$$

in analogy with the conductivity in Eqs. (4.8) and (4.9). This means that its temporal Fourier transform $\mathcal{D}(t)$, defined as

$$\mathcal{D}(t) = \int \frac{\mathrm{d}\omega}{2\pi}\, e^{-i\omega t}\mathcal{D}(\omega)\,, \qquad (4.25)$$

vanishes for $t<0$, since the current $\mathbf{J}_{t,\mathbf{r}}$ cannot emerge before the lump of density $\Psi_{t',\mathbf{r}}$ has been created. Indeed, one can show [17, 37] that $\mathcal{D}(t)$ is the Fourier transform of the retarded velocity autocorrelation function:

$$\mathcal{D}(t) \equiv \theta(t)\,\langle v(t)v(0)\rangle\,, \qquad (4.26)$$

with $\theta(t)$ a unit step function, cf. Fig. 5, and $\langle\dots\rangle$ is the double average described above in Sec. IV A. This also means that $\mathcal{D}(\omega)$ is analytic (has no singularities) in the upper half plane of the complex variable ω [17].

As a result of the averaging, therefore, the macroscopic sample looks *as if* it were uniform but with the velocity autocorrelator having *memory*,

$$\langle v(t_0+\tau)v(t_0)\rangle = \mathcal{D}(|\tau|)\,, \qquad (4.27)$$

$$\langle v_{-\omega}v_\omega\rangle \equiv \int \mathrm{d}\tau e^{i\omega\tau}\,\langle v(t_0+\tau)v(t_0)\rangle = 2\,\mathrm{Re}\,\mathcal{D}(\omega)\,. \qquad (4.28)$$

The above autocorrelator is independent of t_0 due to time translation invariance.

C. Back to diffusion MRI: The second-order cumulant

Let us now utilize an equivalent, and often more convenient way to represent Eq. (3.16), in terms of Fourier transformed quantities introduced above, such as $q_\omega = \int_0^T \mathrm{d}t\, e^{i\omega t}q(t)$:

$$-\ln S(T) \simeq \frac{1}{2}\int \frac{\mathrm{d}\omega}{2\pi}\, q_{-\omega}\,\langle v_{-\omega}v_\omega\rangle\, q_\omega\,. \qquad (4.29)$$

(Here we dropped spatial indices focusing on the isotropic case.) The velocity autocorrelator in the frequency representation is defined as in Eq. (4.28). The representation (4.29) underscores that, knowing the correlator $\langle v_{-\omega}v_\omega\rangle$, one can evaluate the diffusion-weighted signal (3.16) up to $\mathcal{O}(q^2)$ for any gradient waveform $g(t)$. Conversely, by selecting a particular form of $q(t)$ according to its Fourier representation q_ω, one effectively allocates a larger or a smaller weight to particular Fourier harmonics $\langle v_{-\omega}v_\omega\rangle$ contributing to the measured signal (4.29), Ref. [41, 43].

There are two advantages of working in the frequency representation (4.29). From the practical standpoint, a single integral in ω is simpler than a double integral in t. This reduction is due to the time translation invariance not explicitly utilized in Eq. (3.16). From the fundamental standpoint, $\langle v_{-\omega}v_\omega\rangle$ is directly related, via Eq. (4.28), to the dispersive diffusivity $\mathcal{D}(\omega)$ entering the Fick's law (4.16) discussed above, cf. Fig. 5.

As a result, the knowledge of $\mathcal{D}(\omega)$ allows one to find the second cumulant contribution to the signal attenuation for any pulse sequence $g(t)$ via Eq. (4.29):

$$-\ln S(T) \simeq \int \frac{\mathrm{d}\omega}{2\pi}\, q_{-\omega}\mathcal{D}(\omega)q_\omega\,. \qquad (4.30)$$

Here, only $\mathrm{Re}\,\mathcal{D}(\omega)$ contributes, as $\mathrm{Im}\,\mathcal{D}(\omega)$, odd in ω, yields zero after being integrated with an even function $q_{-\omega}q_\omega = |q_\omega|^2$. Fortunately, $\mathrm{Im}\,\mathcal{D}(\omega)$ does not contain additional information as it can be restored using the Kramers – Kronig relations [40]. As we show below, it may be useful to work with the analytic function $\mathcal{D}(\omega)$ (in the sense of the complex variable ω), rather than with its non-analytic real part.

The dispersive diffusivity can be obtained exactly from the narrow-pulse PFG diffusion coefficient (4.1) as

$$\mathcal{D}(\omega) = D_0 + \int_0^\infty \mathrm{d}t\, e^{i\omega t}\partial_t^2\,[tD(t)]\,, \qquad (4.31)$$

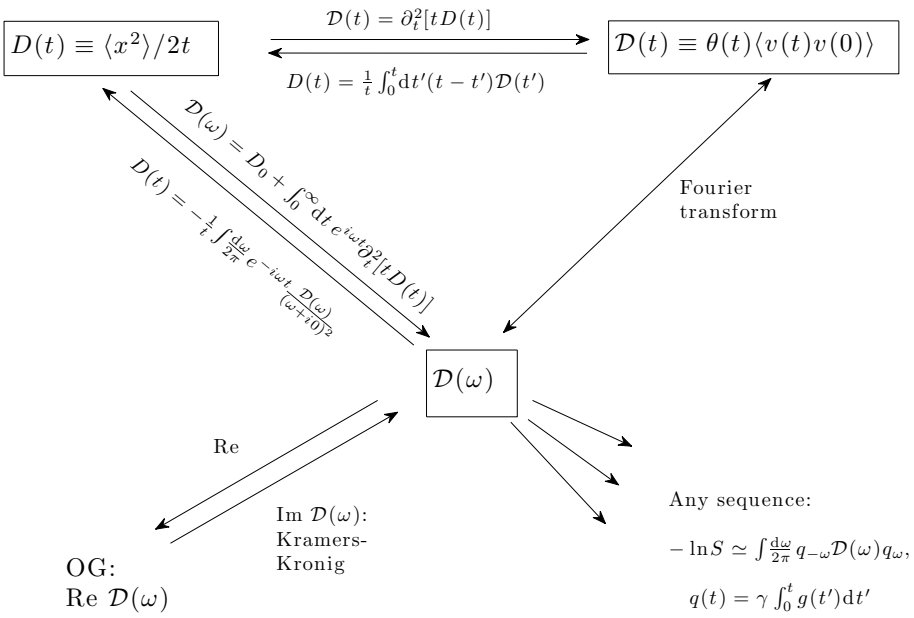

FIG. 5. General relations between the three diffusion metrics: $\mathcal{D}(\omega)$, $\mathcal{D}(t)$ and $D(t)$, and the signal attenuation, from Ref. [41].

where $D_0 \equiv D(t)|_{t=0}$ (cf. Eq. (D3) in Appendix D of Ref. [17]).

The three diffusion metrics: the dispersive diffusivity $\mathcal{D}(\omega)$; the retarded velocity autocorrelator $\mathcal{D}(t)$; and the time-dependent diffusion coefficient $D(t)$ **contain the same amount of information about the mesoscopic structure**, and thus can be expressed via each other [17, 41], as illustrated schematically in Fig. 5.

D. Oscillating gradients

A comprehensive diffusion-weighted measurement must provide a way to obtain the diffusivity $\mathcal{D}(\omega)$, or the correlator $\langle v_{-\omega}v_\omega\rangle$, for all ω. From this point of view, the oscillating gradients (OG) method, with $g(t) = g_0\cos\omega_0 t$, is the easiest one to interpret, as in the limit of the large number $N = \omega_0 T/2\pi \gg 1$ of oscillations,

$$q_\omega = \frac{i\pi\gamma g_0}{\omega_0}\left[\delta(\omega-\omega_0) - \delta(\omega+\omega_0)\right]$$

effectively selects the ω_0 component $\langle v_{-\omega_0}v_{\omega_0}\rangle$, so that

$$-\ln S(T)|_{g(t)=g_0\cos\omega_0 t} \simeq \frac{(\gamma g_0)^2 T}{2\omega_0^2}\cdot \mathrm{Re}\,\mathcal{D}(\omega_0)\,. \quad (4.32)$$

Here we used $\delta(\omega)|_{\omega=0} = T/2\pi$ from the Fourier representation of $\delta(\omega)$. As a result, it is $\mathrm{Re}\,\mathcal{D}(\omega)$ that is measured via the OG techniques [17, 41]. In the above equation, the attenuation over each oscillation period is accumulated, such that the signal (4.32) can be represented as

$$S = e^{-b\cdot \mathrm{Re}\,\mathcal{D}(\omega_0)} \quad \text{with} \quad b = Nb_1\,, \quad b_1 = \frac{\pi(\gamma g_0)^2}{\omega_0^3}\,. \quad (4.33)$$

For the dispersive $\mathcal{D}(\omega)$, the b-value alone does not define the measurement: the same value, achieved with different oscillation frequencies ω_0, will yield different results for S.

Remarkably, the signal S is also sensitive to the initial phase φ of the oscillation $g_\varphi(t) = g_0\cos(\omega_0 t - \varphi)$, yielding

$$-\ln S(T)|_{g_\varphi(t)} \simeq \frac{(\gamma g_0)^2 T}{\omega_0^2}\cdot\left[\frac{1}{2}\,\mathrm{Re}\,\mathcal{D}(\omega_0) + \sin^2\varphi\cdot D(T)\right], \quad (4.34)$$

where $D(T) \simeq D_\infty \equiv \mathcal{D}(\omega)|_{\omega\to 0} = D(t)|_{t\to\infty}$ practically is the tortuosity asymptote, since the latter is typically reached over the sufficiently long total measurement time T. Physically, the initial phase φ leads to the admixture of the PFG attenuation over the time T due to the nonzero value of $q_\omega|_{\omega\to 0} \propto \sin\varphi$, cf. Ref. [44] for $\varphi = \frac{\pi}{2}$.

Equation (4.32), as well as its more general counterpart (4.34), link the diffusive response function $\mathcal{D}(\omega)$ of any medium, which enters the generalized Fick's law (4.16), to the OG attenuation with $N \gg 1$ oscillations.

In Ref. [41], the above relations were used to relate the universal short-time behavior [45] of the diffusion coefficient $D(t)$ (discussed below in Sec. V), to its high-frequency counterpart, $\mathcal{D}(\omega)$, measurable with OG. The links between the different equivalent diffusion metrics outlined in Fig. 5 were explicitly demonstrated for this particular problem.

V. TIME-DEPENDENT DIFFUSION

Let us now demonstrate on a few examples how the diffusion metrics of a macroscopic sample acquire time-dependence as a consequence of the mesoscopic structure, Fig. 1, exemplified by Eq. (2.6) containing the effects of both

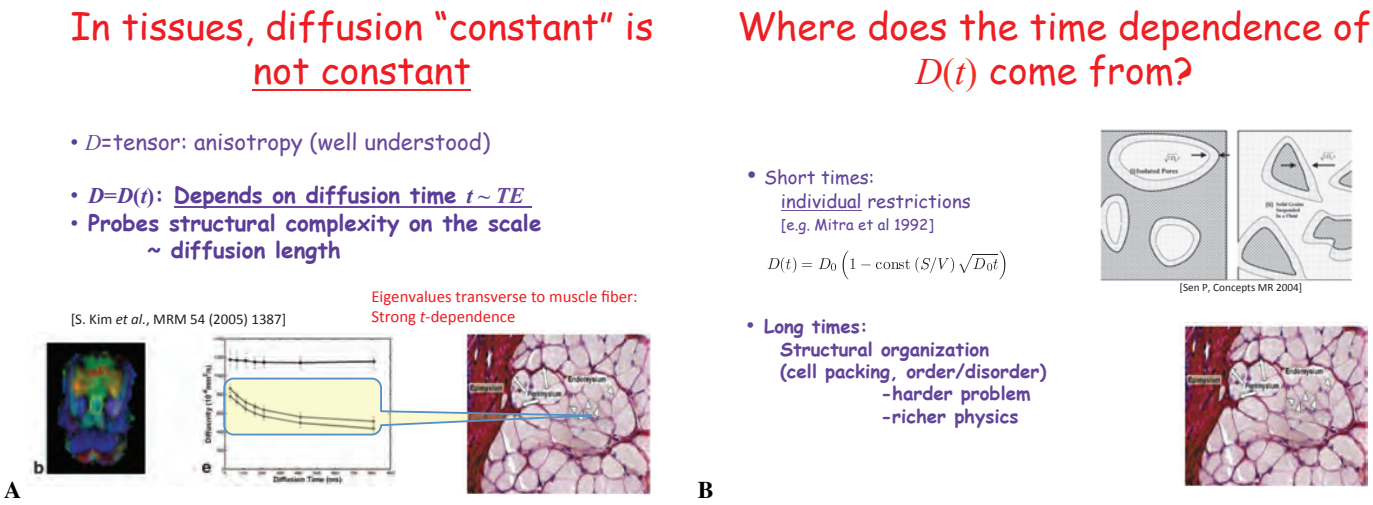

FIG. 6. (A) In general, the diffusion tensor eigenvalues depend on time as a result of the mesoscopic tissue structure. As an example, the pronounced time dependence of the eigenvalues transverse to muscle fibers between $t = 30$ ms and 800 ms has been observed in Ref. [46]. (B) The time dependence at short and long times is due to qualitatively different reasons.

slowly varying local diffusivity (e.g. different intrinsic diffusivity in cytoplasm and in the extracellular fluid), as well as the sharp variations corresponding to cell membranes.

Increasing the time t amounts to exploring sample's complexity over a set of increasing length scales $L(t) = \langle \delta x^2(t)\rangle^{1/2}$. The presence of the mesoscopic structure slows down the diffusion, with qualitatively different physical reasons behind the short- and long-time behavior of the diffusion metrics, Fig. 6.

A. Short times: Net amount of the restrictions

At $t = 0$, each random walker only senses its own microenvironment with a strongly varying local diffusion coefficient $D(\mathbf{r})$. The Brownian path averaging (i) of Sec. IV A is then trivial, and the ensemble averaging (ii) yields a sample average

$$D_0 \equiv D(t)|_{t=0} = \langle D(\mathbf{r}_0)\rangle_{\mathbf{r}_0}\,. \tag{5.1}$$

At short times, the decrease in $D(t)$ is caused by the local restrictions. Suppose a sample is filled with randomly oriented walls with the surface-to-volume ratio S/V. Since, after the averaging (i), the diffusion effectively vanishes inside a layer of thickness $\sim L(t)$ along the walls, the ensemble averaging (ii) yields the $-\sqrt{t}$ correction [45]

$$D(t) \simeq D_0\left(1 - \frac{4}{3d\sqrt{\pi}}\cdot\frac{S}{V}\sqrt{D_0 t}\right). \tag{5.2}$$

The $\sqrt{t}$ term, which counts net amount of the restrictions *irrespective of their positions*, has been used to quantify the ratio S/V in samples filled with NMR-visible molecules, such as water and other liquids [47–49], and polarized ^{129}Xe gas [50] in porous media, as well as water in packed erythrocytes [51].

B. Long times: Structural correlations

Over time, the random walkers probe the spatial organization of the sample's microstructure. The time-dependence of the diffusion metrics becomes intricately tied to an increasingly large number of structural characteristics. Finding, say, $D(t)$ or $\mathcal{D}(\omega)$ analytically in a realistic complex sample is an intractable problem. Technically, it amounts to including the contributions from the spatial correlations of the local diffusion coefficient $D(\mathbf{r})$ and of the positions of all restrictions up to an infinitely high order.

As it has been recently shown [36], this problem greatly simplifies in the long time limit, when the diffusion coefficient approaches its macroscopic value

$$D_\infty \equiv D(t)|_{t\to\infty} = \mathcal{D}(\omega)|_{\omega\to 0}\,. \tag{5.3}$$

In this case, almost all microscopic complexity can be lumped into the value of D_∞, which we assume to be empirically known from a long-t measurement. *The only relevant structural details will be those responsible for the large scale sample heterogeneity, or long range structural correlations.* These correlations will manifest themselves in the specific values of the exponent ϑ in the *power law tail* of the molecular velocity autocorrelation function (4.26)

$$\mathcal{D}(t) \sim t^{-(1+\vartheta)}\,, \quad \vartheta > 0\,. \tag{5.4}$$

Practically, the tail (5.4) can be identified in the way the time-dependent *instantaneous* diffusion coefficient

$$D_{\rm inst}(t) \equiv \frac{\partial}{\partial t}\frac{\langle \delta x^2\rangle}{2} = \int_0^t \mathrm{d}t'\,\mathcal{D}(t') \tag{5.5}$$

approaches the finite bulk diffusion constant D_∞,

$$D_{\rm inst}(t) \simeq D_\infty + \text{const}\cdot t^{-\vartheta}\,. \tag{5.6}$$

The instantaneous diffusion coefficient is related to the standard diffusion coefficient (4.1) measured with narrow pulses as follows:

$$D_{\rm inst}(t) = \partial_t \left[tD(t) \right] . \tag{5.7}$$

1. *Time-dependent diffusion at long times: A simple example*

To understand where the long range structural correlations enter the picture, consider a simple example of a (disordered) mesoscopic structure in $d = 1$ dimension, Fig. 7: randomly positioned identical barriers with permeability κ, Eq. (2.7), and mean density $\bar{n}$, which restrict the diffusion with a free diffusion constant D_0.

First let us consider the simplest problem, the tortuosity limit $t \to \infty$. In this limit, the diffusion becomes Gaussian with macroscopic diffusion constant [16]

$$D_\infty = \frac{D_0}{1 + \frac{\bar{n} D_0}{\kappa}} . \tag{5.8}$$

This one-dimensional result is a particular case of a more general exact one-dimensional relation

$$\frac{1}{D_\infty} = \left\langle \frac{1}{D(x)} \right\rangle \tag{5.9}$$

between the tortuosity limit D_∞ and *any* spatially varying local diffusion coefficient $D(x)$. One can derive Eq. (5.9) directly, but it is instructive to understand it using the Einstein relation [9] between dc conductivity σ and the diffusion constant, $\sigma \propto D$. A segment of length l with diffusion constant D is analogous to a resistor with resistivity $\rho \propto 1/D$, corresponding to the resistance $R = l/D$ (we dropped the inessential proportionality coefficient in the Einstein relation as it will cancel out anyway). As resistances add up in one dimension, $R = \int \mathrm{d}x \, \rho(x)$, it is the resistivity, as well as the inverse diffusivity, that gets averaged, yielding Eq. (5.9).

Now, to obtain Eq. (5.8), we note that based on the definition (2.7), a barirer corresponds to a resistor with dc resistance $l_m/D_m = 1/\kappa$ [16]. Hence, the resistance of a segment of length L with N barriers in series is given by $L/D_0 + N/\kappa$, yielding the macroscopic diffusion constant (5.8) in the limit of $N \to \infty$, $L \to \infty$, and $N/L = \bar{n}$.

Now let us consider long, but finite times t, which correspond to the diffusion length $L(t) \simeq \sqrt{2D_\infty t} \gg \bar{a}$, where $\bar{a} = 1/\bar{n}$ is the mean barrier spacing. We will apply the double average of Sec. IV A implied in a macroscopic measurement.

The local averaging (i) results in the diffusing molecules homogenizing, or coarse-graining the line into domains of size $\sim L(t)$, similar to the two-dimensional example of Fig. 4. Crudely speaking, this can be represented as splitting the line into segments of length $L(t)$. Hence, the averaging (i) establishes the local "average" (or, more precisely, coarse-grained) diffusion coefficient $D(x)|_{L(t)} \approx D_j$ of each segment, Fig. 7. The problem now looks as if, instead of the original barriers, we have segments with slightly different local diffusivities. It is important to note that they are slightly different, because the

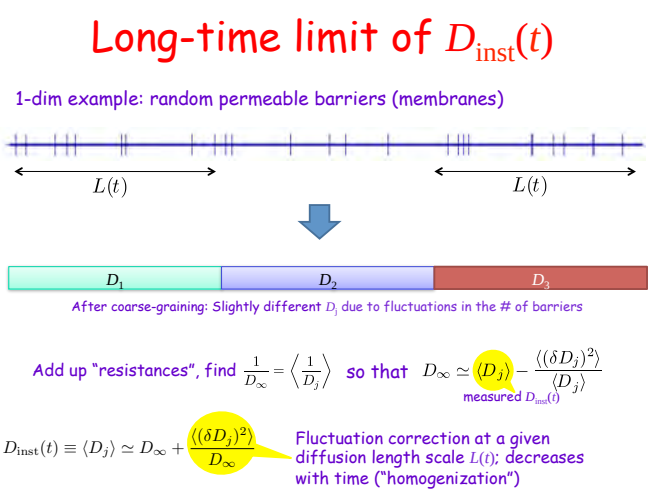

FIG. 7. The qualitative picture for the origin of the time dependence of the diffusion coefficient at long t

numbers of barriers in each segment is slightly different from $\bar{n}L(t)$ due to the *structural fluctuations* (disorder in barrier positions). This is where these fluctuations, and the amount of disorder in a sample, begin to matter.

Let us now apply the ensemble average (ii). The tortuosity asymptote is still given by Eq. (5.9), albeit now it must be calculated using the coarse-grained $D(x)|_{L(t)}$ (involving, approximately, the average $\langle 1/D_j \rangle$). What is essential is that the tortuosity limit D_∞ is *independent of our mental coarse-graining procedure*. Therefore, the left-hand side of Eq. (5.9) does not depend on t. Hence, the right-hand side must be t-independent, too! Let us find it. Expanding up to the second order in the local deviations $\delta D(x)|_{L(t)} = D(x)|_{L(t)} - \langle D \rangle$,

$$\frac{1}{D_\infty} = \left\langle \frac{1}{D(x)} \right\rangle = \left\langle \frac{1}{\langle D \rangle + \delta D(x)} \right\rangle \simeq \frac{1}{\langle D_j \rangle} - \frac{\langle (\delta D_j)^2 \rangle}{\langle D_j \rangle^3} ,$$

where all the terms correspond to a particular coarse-graining scale $L(t)$ (segment size), and $\langle \delta D_j \rangle \equiv 0$. Note that both the mean $\langle D_j \rangle$ over the segments and their variance $\langle (\delta D_j)^2 \rangle$ both depend on t, but the above combination is time independent, as it corresponds to $1/D_\infty$.

What do we measure? A rigorous calculation [36] shows that it is the mean

$$D_{\rm inst}(t) \equiv \langle D(x) \rangle \, |_{L(t)} = \langle D_j \rangle \tag{5.10}$$

that corresponds to the macroscopic *instantaneous* time-dependent diffusion coefficient (5.5). Expressing $D_{\rm inst}(t)$ via D_∞ and the variance $\langle (\delta D(x))^2 \rangle \, |_{L(t)} = \langle (\delta D_j)^2 \rangle$, we obtain

$$D_{\rm inst}(t) \simeq D_\infty + \frac{\langle (\delta D(x))^2 \rangle \, |_{L(t)}}{D_\infty} . \tag{5.11}$$

Above, we discarded terms higher-order in the variance $\langle (\delta D(x))^2 \rangle \, |_{L(t)}$, considering the latter to be small at sufficiently long t (corresponding to sufficiently large coarse-graining window $L(t)$).

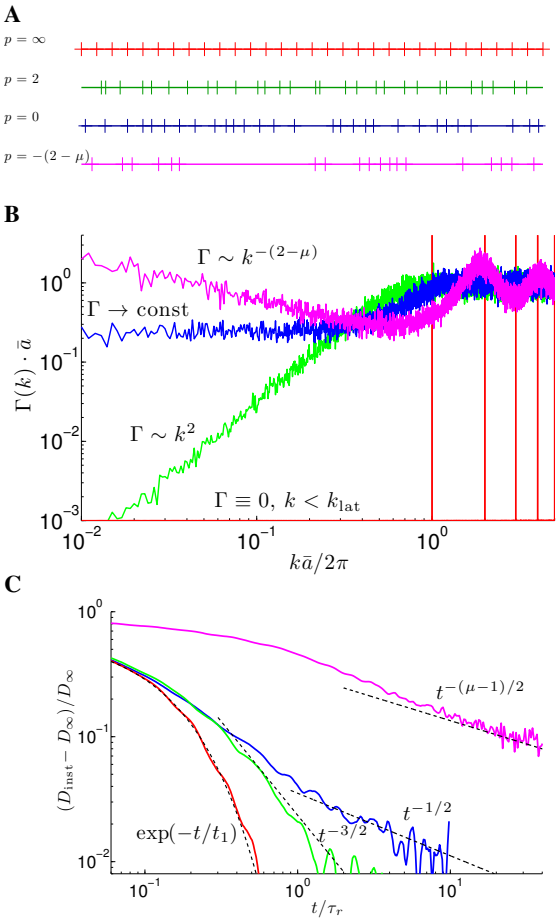

FIG. 8. Time-dependent diffusion distinguishes between structural complexity classes in one dimension, represented by the placement of identical barriers with permeability κ and the same mean density $n = 1/\bar{a}$, from Ref. [36]. **(A)** *Order* (red), *hyperuniform disorder* (green), *short-range disorder* (blue), and *strong disorder* (magenta). **(B)** The barrier densities have qualitatively different large-scale fluctuations, reflected in the small-k behavior of their density correlator $\Gamma(k) \sim k^p$ (see text). **(C)** Numerical results confirming the relation (5.13). The time-dependence (5.5) clearly distinguishes between the four arrangements, while the value D_∞ is the same for all of them. The dashed lines are the exact power laws, and the exponential decrease is from the exact solution; $\tau_r = \bar{a}/2\kappa$.

We observe that the deviation of $D_{\text{inst}}(t)$ from its $t \to \infty$ limit D_∞ is given by the variance of the distribution of coarse-grained diffusivities homogenized at the scale $L(t)$. As time increases, this variance decreases due to self-averaging, and the above procedure becomes asymptotically exact. The result (5.11) is corroborated by a rigorous calculation, and is qualitatively similar in dimensions $d > 1$.

How do we relate the time dependence (5.11) to structure? Let us recall that the coarse-grained $D(x)|_{L(t)}$ depends on the coarse-grained local barrier density $n(x)$. Hence, its variance is proportional, for small deviations $n(x)-n$, to a typical density fluctuation $\langle (\delta n)^2 \rangle |_{L(t)}$ of the restrictions in a segment of size $L(t)$. The latter fluctuation, in the case of Poissonian disorder (uncorrelated barriers) scales as $1/L(t) \sim t^{-1/2}$ according to the central limit theorem. As a result, we obtain that when barrier positions are uncorrelated, the instantaneous diffusion coefficient approaches its macroscopic limit as $t^{-1/2}$:

$$D_{\text{inst}}(t) \simeq D_\infty + \text{const} \cdot t^{-1/2}\,. \tag{5.12}$$

We can see that for Poissonian (uncorrelated) arrangement of the structure (barriers), the dynamical exponent entering Eq. (5.6) is $\vartheta = 1/2$.

2. *Classification of structure based on long-range correlations*

Equation (5.12) provides an intuition behind a fairly general statement [36]: the dynamical exponent ϑ in Eq. (5.6)

$$\vartheta = (p+d)/2 \tag{5.13}$$

is related to the statistics of large scale structural fluctuations. This statistics is described in terms of the *structural exponent* p characterizing global structural organization in d spatial dimensions. The structural exponent p determines the $\Gamma(k)|_{k\to 0} \sim k^p$ behavior of the Fourier transform of the correlation function $\Gamma(\mathbf{r}) = \langle n(\mathbf{r}+\mathbf{r}_0)n(\mathbf{r}_0)\rangle_{\mathbf{r}_0}$ for the underlying microstructure. Hence, p characterizes global structural complexity, taking discrete values robust to local perturbations. This enables the classification of the types and topologies of the mesoscopic disorder, such as shown in Fig. 8 in dimension $d = 1$. In particular, the dependence (5.12) indeed follows from the structural exponent $p = 0$ (short-range disorder) in $d = 1$, blue curves in Fig. 8.

Relation (5.13) provides a way to determine the exponent p and, thereby, the structural complexity class, using any type of bulk diffusion measurement. Local properties affect the coefficients, e.g. the values of D_∞ and the prefactor of $t^{-\vartheta}$ in (5.5), but not the exponent ϑ. The latter exponent is *universal*, i.e. is independent of microscopic details, and is robust with respect to variations between samples of a similar origin.

VI. OUTLOOK

The overarching goal of these notes was to present diffusion MRI as a vibrant scientific field which is actively developing by drawing on remarkably deep connections to more established branches of physics, such as statistical and condensed matter physics. It is fascinating to realize that the methods of studying transport in complex systems in a completely different context [8–15] appear to be increasingly useful to quanfity mesoscopic tissue properties noninvasively. In this respect, we are witnessing a kind of a "phase transition" in our field, where the paradigm is shifting from empirical correlations to biophysical parameters characterizing mesoscopic structure of tissues, such as spatial arrangement and correlations of cells, membrane permeability and other parameters. While the nominal spatial resolution of human MRI is unlikely to enable direct imaging of at the cellular level, the future innovation will be in many ways determined by deepening of our fundamental understanding of the link between diffusion measurements and tissue structure.

[1] N. Bloembergen, "Relaxation effects in nuclear magnetic resonance absorption," Physical Review **73**, 679 (1948).

[2] A. Abragam, *Principles of Nuclear Magnetism* (Oxford University Press, New York, 1961).

[3] C. P. Slichter, *Principles of Magnetic Resonance*, 3rd ed. (Springer, New York, 1990).

[4] R. R. Ernst, G. Bodenhausen, and A. Wokaun, *Principles of Nuclear Magnetic Resonance in One and Two Dimensions* (Oxford University Press, New York, 1987).

[5] J. R. Zimmerman and W. E. Brittin, "Nuclear magnetic resonance studies in multiple phase systems: Lifetime of a water molecule in an adsorbing phase on silica gel," The Journal of Physical Chemistry **61**, 1328 (1957).

[6] D. Le Bihan, ed., *Diffusion and Perfusion Magnetic Resonance Imaging* (Raven Press, New York, 1995).

[7] M. E. Moseley, Y. Cohen, J. Mintorovitch, L. Chileuitt, H. Shimizu, J. Kucharczyk, M. F. Wendland, and P. R. Weinstein, "Early detection of regional cerebral-ischemia in cats – comparison of diffusion-weighted and T2-weighted MRI and spectroscopy," Magnetic Resonance In Medicine **14**, 330 (1990).

[8] P. W. Anderson, "Absence of diffusion in certain random lattices," Physical Review **109**, 1492 (1958).

[9] M. Lax, "Fluctuations from the nonequilibrium steady state," Reviews of Modern Physics **32**, 25 (1960).

[10] N. F. Mott and E. A. Davies, *Electronic processes in non-crystalline materials* (Oxford University Press, New York, 1971).

[11] H. Scher and E. W. Montroll, "Anomalous transit-time dispersion in amorphous solids," Physical Review B **12**, 2455 (1975).

[12] B. I. Shklovskii and A. L. Efros, *Electronic properties of doped semiconductors* (Springer, Heidelberg, 1984).

[13] P. A. Lee and T. V. Ramakrishnan, "Disordered electronic systems," Reviews of Modern Physics **57**, 287 (1985).

[14] Y. Imry, *Introduction to mesoscopic physics* (Oxford University Press, New York, 1997).

[15] A. Kamenev, *Field Theory of Non-Equilibrium Systems* (Cambridge University Press, New York, 2011).

[16] D. S. Novikov, E. Fieremans, J. H. Jensen, and J. A. Helpern, "Random walks with barriers," Nature Physics **7**, 508 (2011).

[17] D. S. Novikov and V. G. Kiselev, "Effective medium theory of a diffusion-weighted signal," NMR Biomed **23**, 682 (2010).

[18] H. C. Torrey, "Bloch equations with diffusion terms," Physical Review **104**, 563 (1956).

[19] J. Zhong, R. P. Kennan, and J. C. Gore, "Effects of susceptibility variations on {NMR} measurements of diffusion," Journal of Magnetic Resonance (1969) **95**, 267 (1991).

[20] V. G. Kiselev, "Effect of magnetic field gradients induced by microvasculature on nmr measurements of molecular self-diffusion in biological tissues," J Magn Reson **170**, 228 (2004).

[21] E. O. Stejskal and J. E. Tanner, "Spin diffusion measurements: Spin echoes in the presence of a time dependent field gradient," Journal of Chemical Physics **42**, 288 (1965).

[22] P. T. Callaghan, *Principles of Nuclear Magnetic Resonance Microscopy* (Clarendon, Oxford, 1991).

[23] P. Callaghan, C. Eccles, and Y. Xia, "Nmr microscopy of dynamic displacements: k-space and q-space imaging," Journal of Physics E: Scientific Instruments **21**, 820 (1988).

[24] R. A. Fisher and J. Wishart, "The derivation of the pattern formulae of two-way partitions from those of simpler patterns," Proc. London Math. Soc. (1) **s2-33**, 195 (1932).

[25] N. G. van Kampen, *Stochastic Processes in Physics and Chemistry*, 1st ed. (Elsevier, Oxford, 1981).

[26] V. G. Kiselev, "Diffusion MRI: Theory, methods and applications," (Ed. Jones, D. K., Oxford University Press, New York, 2010) Chap. 10. The cumulant expansion: An overarching mathematical framework for understanding diffusion NMR.

[27] J. E. Mayer and E. Montroll, "Molecular distribution," Journal of Chemical Physics **9**, 2 (1941).

[28] A. A. Abrikosov, L. P. Gorkov, and I. E. Dzyaloshinskii, *Methods of Quantum Field Theory in Statistical Physics* (Dover, 1975).

[29] P. Basser, J. Mattiello, and D. LeBihan, "Estimation of the effective self-diffusion tensor from the nmr spin-echo," Journal of Magnetic Resonance Series B **103**, 247 (1994).

[30] J. H. Jensen, J. A. Helpern, A. Ramani, H. Lu, and K. Kaczynski, "Diffusional kurtosis imaging: the quantification of non-gaussian water diffusion by means of magnetic resonance imaging," Magn Reson Med **53**, 1432 (2005).

[31] J. Veraart, J. Rajan, R. R. Peeters, A. Leemans, S. Sunaert, and J. Sijbers, "Comprehensive framework for accurate diffusion mri parameter estimation," Magn Reson Med **70**, 972 (2013).

[32] Mitra, "Multiple wave-vector extensions of the nmr pulsed-field-gradient spin-echo diffusion measurement," Phys Rev B Condens Matter **51**, 15074 (1995).

[33] S. N. Jespersen, "Equivalence of double and single wave vector diffusion contrast at low diffusion weighting," NMR Biomed **25**, 813 (2012).

[34] K. S. Thorne, "Multipole expansions of gravitational radiation," Reviews of Modern Physics **52**, 299 (1980).

[35] A. Kusumi, C. Nakada, K. Ritchie, K. Murase, K. Suzuki, H. Murakoshi, R. S. Kasai, J. Kondo, and T. Fujiwara, "Paradigm shift of the plasma membrane concept from the two-dimensional continuum fluid to the partitioned fluid: High-speed single-molecule tracking of membrane molecules," Annual Review of Biophysics and Biomolecular Structure **34**, 351 (2005).

[36] D. S. Novikov, J. H. Jensen, J. A. Helpern, and E. Fieremans, "Revealing mesoscopic structural universality with diffusion," Proc Natl Acad Sci U S A **111**, 5088 (2014).

[37] R. Kubo, "The fluctuation-dissipation theorem," Rep. Prog. Phys. **29**, 255 (1966).

[38] L. D. Landau, L. P. Pitaevskii, and E. M. Lifshitz, *Electrodynamics of Continuous Media (Course of Theoretical Physics, Vol. 8)*, 2nd ed. (Butterworth–Heinemann, Oxford, 2004).

[39] E. M. Lifshitz and L. P. Pitaevskii, *Physical Kinetics (Course of Theoretical Physics, Vol. 10)* (Pergamon Press, Oxford, 1981).

[40] L. D. Landau and E. M. Lifshitz, *Statistical Physics, Part I (Course of Theoretical Physics, Vol. 5)* (Pergamon Press, Oxford, 1969).

[41] D. S. Novikov and V. G. Kiselev, "Surface-to-volume ratio with oscillating gradients," Journal of Magnetic Resonance **210**, 141 (2011).

[42] D. S. Novikov and V. G. Kiselev, "Transverse NMR relaxation in magnetically heterogeneous media," J Magn Reson **195**, 33 (2008).

[43] S. N. Jespersen, "Equivalence of double and single wave vector diffusion contrast at low diffusion weighting," NMR Biomed **25**, 813 (2012).

[44] M. D. Does, E. C. Parsons, and J. C. Gore, "Oscillating gradient measurements of water diffusion in normal and globally ischemic rat brain," Magn Reson Med **49**, 206 (2003).

[45] P. P. Mitra, P. N. Sen, L. M. Schwartz, and P. Le Doussal, "Diffusion propagator as a probe of the structure of porous media," Physical Review Letters **68**, 3555 (1992).

[46] S. Kim, G. Chi-Fishman, A. S. Barnett, and C. Pierpaoli, "Dependence on diffusion time of apparent diffusion tensor of ex vivo calf tongue and heart," Magnetic Resonance In Medicine **54**, 1387 (2005).

[47] L. L. Latour, P. P. Mitra, R. L. Kleinberg, and C. H. Sotak, "Time-dependent diffusion coefficient of fluids in porous media as a probe of surface-to-volume ratio," J. Magn. Reson. Ser A **101**, 342 (1993).

[48] L. J. Zielinski and M. D. Hürlimann, "Probing short length scales with restricted diffusion in a static gradient using the cpmg sequence," J Magn Reson **172**, 161 (2005).

[49] J. Stepisnik, S. Lasic, A. Mohoric, I. Sersa, and A. Sepe, "Velocity autocorrelation spectra of fluid in porous media measured by the cpmg sequence and constant magnetic field gradient," Magnetic Resonance Imaging **25**, 517 (2007).

[50] R. W. Mair, G. P. Wong, D. Hoffmann, M. D. Hurlimann, S. Patz, L. M. Schwartz, and R. L. Walsworth, "Probing porous media with gas diffusion NMR," Physical Review Letters **83**, 3324 (1999).

[51] L. L. Latour, K. Svoboda, P. P. Mitra, and C. H. Sotak, "Time-dependent diffusion of water in a biological model system," Proceedings of the National Academy of Sciences of the United States of America **91**, 1229 (1994).