DIFFUSION: FROM CLASSICAL TO FRACTIONAL

Theo J. Schep

Technical University of Eindhoven, P.O.Box 513, 5600 MB Eindhoven,
The Netherlands
e-mail: t.schep4@kpnplanet.nl theo.schep@wanadoo.fr

CONTENT

INTRODUCTION

Diffusion can be defined as the motion of some quantity from a region of high concentration to a region of low concentration. It is clear that under this definition diffusion phenomena are ubiquitous in nature and will play a role in many aspects of every day life. Every day life examples are the spreading of odors in air, like the spreading of smoke or the perfume of a lady in a room, the spreading of seasoning in the marinade in your kitchen, or the global spreading of a virus in these times with intensive air traffic. Diffusion denotes the mixing of two or more substances and is the net result on the macro-scale from interactions on the micro-scale with some individual agents, like collisions with molecules. One can say that diffusion is the ensemble average over microscopic interactions. Such processes are studied in numerous sciences and play a role in such diverse areas as chemistry, economy, biology, physics, physiology, etc.. A well known example from sociology is the spreading of information under a population and from biology the spreading of diseases in life stock and bird populations.

In physics, diffusion processes belong to the larger field of transport phenomena. A transport phenomenon is any mechanism by which particles or quantities are moved from one place to another. There exist three main categories in transport theory
. mass transfer
. heat transfer
. momentum transfer
A standard example of the first process is the spreading of heavy particles in a fluid of lighter particles. This is related with Brownian motion. The standard example of the second process is the conduction of heat through a solid. A typical case of momentum diffusion is the loss of momentum of heavy particle or a solid object moving through a fluid due to collisions with the background particles. All three phenomena play a role in transport in fluids and plasmas.

Usually transport phenomena are divided into three main areas: diffusion, convection, and radiation. Quite often, these phenomena occur simultaneously in a process. However, in this course we will only consider the first subject, diffusion processes.

A cornerstone in the history of diffusion theory is the work by Fick, published some 150 years ago. His work was theoretical and can best be characterized as phenomenological. Fifty years later Einstein offered an explanation for Brownian motion in terms of a random walk and gave a molecular basis to Fick's diffusive law. Gaussian probability distribution functions are at the heart of this by now classical theory of diffusion, which imply that the mean square displacement is proportional to time

$$< (\Delta x)^2(t) >= 2Dt,$$

D being the diffusion coefficient. The brackets denote some appropriate statistical averaging.

However, collisions on the micro-scale are not the only cause of diffusion. In magnetized plasmas, like those observed in tokamaks and stellarators, e.g., there is a large discrepancy between the diffusion processes as given by classical or neoclassical theories and the experimentally observed transport phenomena. Such deviations from classical theory are observed in many other fields. This has turned the attention to *anomalous diffusion*, i.e., to transport caused by fluctuating fields. This subject of diffusion in turbulent media is an extremely broad area of research.

Turbulence is a regime that is characterized by chaotic, stochastic changes in the properties of the system under consideration. Examples are
- the streaming of water around the supports of a bridge,
- mixing layers in the atmosphere and in ocean currents,
- smoke rising from a cigarette,
- turbulent flow behind a car and the wings of an airplane.
In a turbulent medium, unsteady regular structures, like vortices and magnetic islands, are generated on many different scale-lengths. In regular fluids, most of the energy input is in the larger scales. This energy cascades to smaller scales where it is dissipated.

In spite of considerable efforts, turbulence is still poorly understood. It has a wide variety of causes and appearances, and contains fluctuations on many scales. This leads to non-standard diffusion processes.

The large variety of forms in which turbulence manifests itself, requires a wide variety of approaches and a wide variety of methods in the analysis of diffusion processes. Not only standard tools like the random walk (Brownian motion), but also the concepts like fractality, scaling laws and the theory of percolation are nowadays applied in diffusion theory.

In addition, often several different types of transport might exist simultaneously in turbulent diffusion. Regions with chaotic magnetic fields may exist next to regions with magnetic islands and regular magnetic surfaces. Just like regions with regular streaming, with sheared velocity fields, and with vortex motions might coexist in fluids and plasmas. Different diffusion processes are often associated with different spatial directions in the system. Therefore it is important to take into account the anisotropy of a configuration. This anisotropy is particularly important in magnetized plasmas.

In the field of controlled thermonuclear fusion, e.g., a major obstacle on the way to the realization of fusion in closed magnetic configuration devices is commonly attributed to the existence of anomalous losses of particles, momentum, and energy across the confining magnetic field.

The diffusion of particles and energy in tokamak devices exceeds by a factor $10 - 10^2$

the (neo)classical values for electrons and by a factor 1-10 the (neo)classical values for ions. The large variety of instabilities in a magnetized plasma will lead to different types of turbulence. The anomalous transport is related to this turbulent character of the plasma behavior. In spite of considerable efforts, this problem is still far from understood.

The classical diffusion models are based upon processes that are local in space and do not have any memory on the history of the system. These properties of spatial locality and absence of memory are often violated in turbulent media. Long temporal and/or spatial correlations might exist due, e.g., to trapping of particles in coherent structures like vortices or islands or to advection by zonal flows.

In the dynamics of continuous fluids, a tracer is a particle that travels with the local fluid velocity, but that has otherwise no influence on the properties of this fluid. It is a particle that is passively advected by the fluid. The motion of passive tracers in fully developed, isotropic and homogeneous turbulence is well described by Brownian motion. Macroscopically it satisfies a Fick's type, local transport equation, and microscopically a random walk with Gaussian statistics. However, in a turbulent system which contains coherent structures like vortices and magnetic islands that may trap particles for long times, and zonal flows that advect tracers over long distances, this theory breaks down. Trapping in coherent structures and the presence of zonal flows will lead to 'memory' effects, to non-Markovian behavior, and imply that the tracer will undergo Lévy flights that will lead to non-Gaussian statistics.

Such processes result into anomalous diffusion laws where the mean square displacement behaves as

$$< (\Delta x)^2(t) > \propto t^{\gamma}.$$

Diffusion processes with $\gamma \neq 1$ are often called 'strange'. They are coined super-diffusive for $\gamma > 1$ and sub-diffusive for $\gamma < 1$. This 'strange' diffusion leads to Fokker-Planck type equations containing fractional derivatives. Transport of passive particles in a system with coherent structures will lead to *strange diffusion* and to diffusion equations that contain *fractional operators*. In general, sub-diffusion occurs in systems with geometric constraints like fractals, doped crystals and magnetic fields, while super-diffusion is encountered in turbulent fluids.

Since we deal with diffusion problems, we are interested in the behavior of systems on long time- and length-scales, actually we want to find out what happens on macroscopic scales. We do not need to know what happens on small scales. From this point of view, the kinetic equations and e.g the master equation for continuous time random walk (CTRW) contain often far too much information for the description of transport on macroscopic scales. We do not require a full kinetic description of the underlying random walk, but we are interested in the continuum (fluid) limit of these equations. The information we are looking for is contained in the large scales i.e. in the tails of the probability distribution functions.

Roughly speaking, one could discern three main directions in diffusion theory.

A. The most fundamental method starts with the kinetic equation for the distribution function $F(x, v, t)$ of a single particle. The dynamics is induced by external random E and B fields and by collisions. To determine macroscopic transport laws, the system has to be averaged over phase-space and over all realizations of the fluctuating fields. This programme seems to be an impossible task.

B. A second direction starts from the Langevin equation for the motion of test particles and use a Fokker-Planck type equations or the hybrid kinetic equation to find the macroscopic solution.

C. Finally, methods exist that have abandoned any deterministic dynamical law in the description of the motion of a test particle. Examples are the application of percolation models and the theory of the continuous time random walk. In both cases, the dynamical laws are replaced by statistical models.

PART ONE: CLASSICAL DIFFUSION AND RANDOM WALKS

I. Classical diffusion

The modern theory of diffusion processes started in the nineteenth century. One of the important building blocks is Fick's law (1855). Another one is the understanding of Brownian motion.

A. Fick's law

In the physics of continuous fluids, Fick's first law (Adolf Fick 1829-1901) is a phenomologically obtained constitutive equation that relates the flux of some quantity with the concentration of that quantity,

$$\mathbf{\Gamma_D} = n\mathbf{v}_D = -D\nabla n, \tag{1}$$

where
$\mathbf{\Gamma}_D$ is the diffusive flux in $m^{-2}s^{-1}$,
$\mathbf{v}_D$ is the diffusion velocity in ms^{-1},
n the concentration or density in m^{-3},
D is the diffusion coefficient in m^2s^{-1}.
This relationship says that differences in concentration of the substance under consideration generate fluxes that transport the substance from places with high concentration to regions of low concentration.

The dynamical equation of particle conservation reads,

$$\frac{\partial n}{\partial t} + \nabla \cdot \mathbf{\Gamma} = 0, \qquad \mathbf{\Gamma} \ = n\mathbf{v}. \tag{2}$$

Here, we are not interested in the streaming of the fluid due to any other cause than diffusion. Thus, assuming that the flux is only due to diffusive processes, $\mathbf{\Gamma} = \mathbf{\Gamma}_D$, one obtains from (2)

$$\frac{\partial n}{\partial t} = D\nabla^2 n. \tag{3}$$

This is Fick's second law and has become known as the *standard diffusion equation*. It is a macroscopic law that does not refer to any microscopic process! This diffusion equation has been applied to model transport processes in numerous fields of the sciences : foods, fluids and plasmas, semiconductor physics, population dynamics, etc..

In an unbounded space, the diffusion equation can be solved as follows. Assume a d-dimensional space and apply to (3) the Fourier transform

$$\hat{n}(\mathbf{k}, t) = \int d^d x \ \hat{n}(\mathbf{x}, t) \exp i\mathbf{k} \cdot \mathbf{x}.$$

The result is

$$\frac{\partial \hat{n}(\mathbf{k},t)}{\partial t} = -k^2 D n(\mathbf{k},t).$$

The solution to this equation is

$$\hat{n}(\mathbf{k},t) = \hat{n}(\mathbf{k},0)\exp -k^2 Dt.$$

Its inverse Fourier transform is

$$n(\mathbf{x},t) = \frac{1}{(2\pi)^d}\int d^d k\ \hat{n}(\mathbf{k},0)\exp\left(-i\mathbf{k}\cdot\mathbf{x} - k^2 Dt\right)$$

$$= \frac{1}{(2\pi)^d}\int d^d k\ \exp(-k^2 Dt)\int d^d x'\ n(\mathbf{x}',0)\exp i\mathbf{k}\cdot(\mathbf{x}'-\mathbf{x}).$$

This leads to the solution of (3)

$$n(\mathbf{x},t) = \int d^d x'\ P(\mathbf{x}-\mathbf{x}';t) n(\mathbf{x}',0), \tag{4}$$

where $P(\mathbf{x}-\mathbf{x}';t)$ is the Gaussian or normal distribution

$$P(\mathbf{x}-\mathbf{x}';t) = \frac{1}{(2\pi Dt)^{d/2}}\exp -\frac{\Sigma_i (x_i - x_i')^2}{4Dt}. \tag{5}$$

This solution suggests the following interpretation. If properly normalized ($\int n d^d x = 1$), we may interpret $n(\mathbf{x},t)d^d x$ as being the probability of finding a particle in a small volume $d^d x$ around the position $\mathbf{x}$ and $P(\mathbf{x}-\mathbf{x}',t)$ as the transition probability that a particle that is at $\mathbf{x}'$ at $t=0$ makes a jump $(\mathbf{x}-\mathbf{x}')$ in a time interval t.

For a particle that starts at $\mathbf{x}'=0$, one has $n(\mathbf{x}',0)=\delta(\mathbf{x}')$. Then,

$$n(\mathbf{x},t) = \frac{1}{(2\pi Dt)^{d/2}}\exp -\frac{\Sigma_i x_i^2}{4Dt}. \tag{6}$$

The length $\sqrt{4Dt}$ is called the diffusion length. *It is clear that (6) is not an equilibrium distribution. The distribution broadens and its maximum value decreases in time.*

The mean square displacement (MSD) of a particle is according to (6)

$$< x_i^2(t) >= \int d^d x\ x_i^2 n(\mathbf{x},t) = 2Dt. \tag{7}$$

The mean square displacement is proportional to the diffusion coefficient and increases linearly with time.

This interpretation is based upon the description of diffusion processes with the Chapman-Kolmogorov functional equation for the probability density of a *stationary and homogeneous Markov process* (see Appendix)

$$n(x, t+\tau) = \int_{-\infty}^{\infty} G(\Delta, \tau) n(x-\Delta, t) d\Delta. \tag{8}$$

Here, $n(x,t)$ is the probability density that the stochastic variable takes the value x at time t and $G(\Delta, \tau)$ is the transition probability that the stochastic variable makes a jump Δ in time interval τ. This equation is *nonlocal in space-time*.

The basic assumption is that the jumps are small as compared with the characteristic scale of $n(\mathbf{x}, t)$, which means that n depends only weakly on Δ. It also implies that $G(\Delta, \tau)$ is a strongly peaked function of Δ. Then, in the limit $t/\tau \to \infty$, (8) can be transformed into a local diffusion equation

$$\frac{\partial n}{\partial t} + V_0 \frac{\partial n}{\partial x} = D \frac{\partial^2 n}{\partial x^2}, \tag{9}$$

where

$$V_0 = \lim \int_{-\infty}^{\infty} d\Delta \frac{\Delta}{\tau} G(\Delta, \tau), \qquad D \equiv \lim \int_{-\infty}^{\infty} d\Delta \, \frac{\Delta^2}{2\tau} G(\Delta, \tau). \tag{10}$$

This is the local version of (8). Equation (9) is a Fokker-Planck type equation and is equivalent to Fick's law. Its solution is given by (4) and (5) with x replaced by $\hat{x} = x - V_0 t$. As will be shown in the next section, (6) is also the pdf for a Gaussian random walk (Brownian motion).

The approach described in this section leads to a diffusion equation that only describes local effects. *In many systems, however, transport phenomena can not be described on the basis of such processes that are local in space and/or time with a constant diffusion coefficient. These more complex diffusion processes are called anomalous. Nevertheless, this local proces still acts as a bench mark for systems which exhibit anomalous transport.*

REMARKS
I. Note that the standard diffusion equation (3) is invariant under the scaling transformation

$$\hat{t} = \alpha t, \quad \hat{x} \to \sqrt{\alpha} x. \tag{11}$$

Requiring mass conservation $n(x,t)dx = n(\hat{x}, \hat{t}) d\hat{x}$ one finds that in addition the relation

$$n(\sqrt{\alpha} x, \alpha t) = \alpha^{-1/2} n(x, t) \tag{12}$$

must hold. This scaling of the density is determined by the Gaussian distribution (5). It means that the distribution of $\sqrt{\alpha} X(\alpha t)$ is the same as that for $X(t)$. Hence, the fluctuations are generated on each scale in a statistically identical manner. If the fluctuations are

known on the interval $[\alpha t,\ t]$, they are also known on the extended interval $[t,\ t/\alpha]$ and on the contracted interval $[\alpha^2 t,\ \alpha t]$.
A distribution that is invariant under a transformation that scales time and space by different factors is called *affine*. A distribution is called *self-similar* if it is invariant under a transformation that scales space and time by the same factor.
II. If the diffusion coefficient is not a constant, then Fick's law reads in 1D

$$\frac{\partial n}{\partial t} = \frac{\partial}{\partial x} D \frac{\partial}{\partial x} n.$$

If $D = D(x)$ depends on x, then there does not exist a simple relation between D and the MSD. This can be seen as follows. Consider an infinite domain $(-\infty, \infty)$ and assume that the density vanishes sufficiently fast at $|x| \to \infty$. Then,

$$\frac{d < x >}{dt} = \frac{d}{dt} \int dx\ xn = \int dx\ x \frac{\partial}{\partial x} D \frac{\partial}{\partial x} n = \int dx\ n \frac{\partial D}{\partial x}$$

and

$$\frac{d < x^2 >}{dt} = \int dx\ x^2 \frac{\partial}{\partial x} D \frac{\partial}{\partial x} n = 2 \int dx\ n(D + x \frac{\partial D}{\partial x}).$$

If $D = constant$, one obtains

$$\frac{d < x >}{dt} = 0, \quad \frac{d}{dt} \left(< x^2 > - < x >^2 \right) = 2D.$$

This latter relationship is lost when $\partial D / \partial x \neq 0$.
III. In case the fluid also has an average flow velocity $\mathbf{V}_0$, then the total flux is given

$$\mathbf{\Gamma} = n\mathbf{V}_0 - D\nabla n,$$

so that we find instead of (3)

$$\frac{\partial n}{\partial t} + \mathbf{V}_0 \cdot \nabla n = D \nabla^2 n.$$

Its solution is given by (4) and (5) with $\mathbf{x}$ replaced by $\hat{\mathbf{x}} = \mathbf{x} - \mathbf{V}_0 t$.
IV. In this section we have assumed the diffusion process to be isotropic. In the general case, the diffusive flux is given by

$$\mathbf{\Gamma}_D = -\mathbf{D} \cdot \nabla n,$$

with $\mathbf{D}$ being a tensor. In many cases this tensor will be diagonal

$$\mathbf{D} = D_1 \mathbf{e}_1 \mathbf{e}_1 + D_2 \mathbf{e}_2 \mathbf{e}_2 + D_3 \mathbf{e}_3 \mathbf{e}_3,$$

where $\mathbf{e}_i$ is the unit vector in the i-direction. In case of a strongly magnetized plasma, the magnetic field introduces a privileged direction and the plasma will have a gyrotropic symmetry,

$$\mathbf{D} = D_\perp (\mathbf{e}_1 \mathbf{e}_1 + \mathbf{e}_2 \mathbf{e}_2) + D_\parallel \mathbf{e}_3 \mathbf{e}_3,$$

the magnetic field being in the 3-direction.
V. Other constitutive relations that are analogous to Fick's law are e.g Ohm's law, Darcy's law, and Fourier's law.
Ohm's law (1827) describes the relation between the current density and the electric field. In the electrostatic case where $\mathbf{E} = -\nabla\phi$, it reads

$$\mathbf{J} = \sigma\mathbf{E} = -\sigma\nabla\phi.$$

Fourier's law gives a relation between the heat flow and the temperature

$$\mathbf{q} = -\kappa\nabla T$$

Actually, Fick formulated his first law in analogy to this relationship.
Darcy's law (1856) is an equation that describes the flow through a porous medium

$$\mathbf{q} = -\frac{\kappa}{\mu}(\nabla P - \rho g\mathbf{e}_z).$$

Here, P is the pressure and $g\mathbf{e}_z$ the gravity force; the constants κ and μ represent the permeability and the coefficient of viscosity, respectively.

B. Brownian motion

Brownian motion is the irregular movement of a heavy particle suspended in a liquid or a gas. It is named after the botanist Robert Brown who studied pollen particles floating in water under a microscope. These observations were made in 1827 and published in 1828. Brown was, however, not the first to observe such motion. The dutch physicist Jan Ingenhousz described the movement of coal dust on the surface of alcohol in 1785. Nevertheless, the phenomenon is traditionally regarded as being discovered by Brown.

The size of a pollen is about $10^3 nm$, while the size of a water molecule is $10^{-1} nm$. Thus a Brownian particle is really a heavy particle in a light fluid. Due to the many collisions with the fluid particles, the velocity of the particle undergoes irregular jumps. At a certain velocity the particle will experience more collisions in front causing an inbalance in force.

Einstein was the first to understand that the quantity of interest is not the average velocity but the mean square displacement of the particle. Looking at the irregular trajectory of a Brownian particle in the figure below, it is clear that the vector parallel to the trajectory is not an useful quantity. Also, experimentally, it is not the velocity but the position of the particle that is observed. The interval between two successive observations is long compared to the correlation time of the velocity. Hence, what is observed is the net displacement of a particle after many variations of the velocity. This displacement is a random variable. Each displacement, is to a good approximation, independent of the

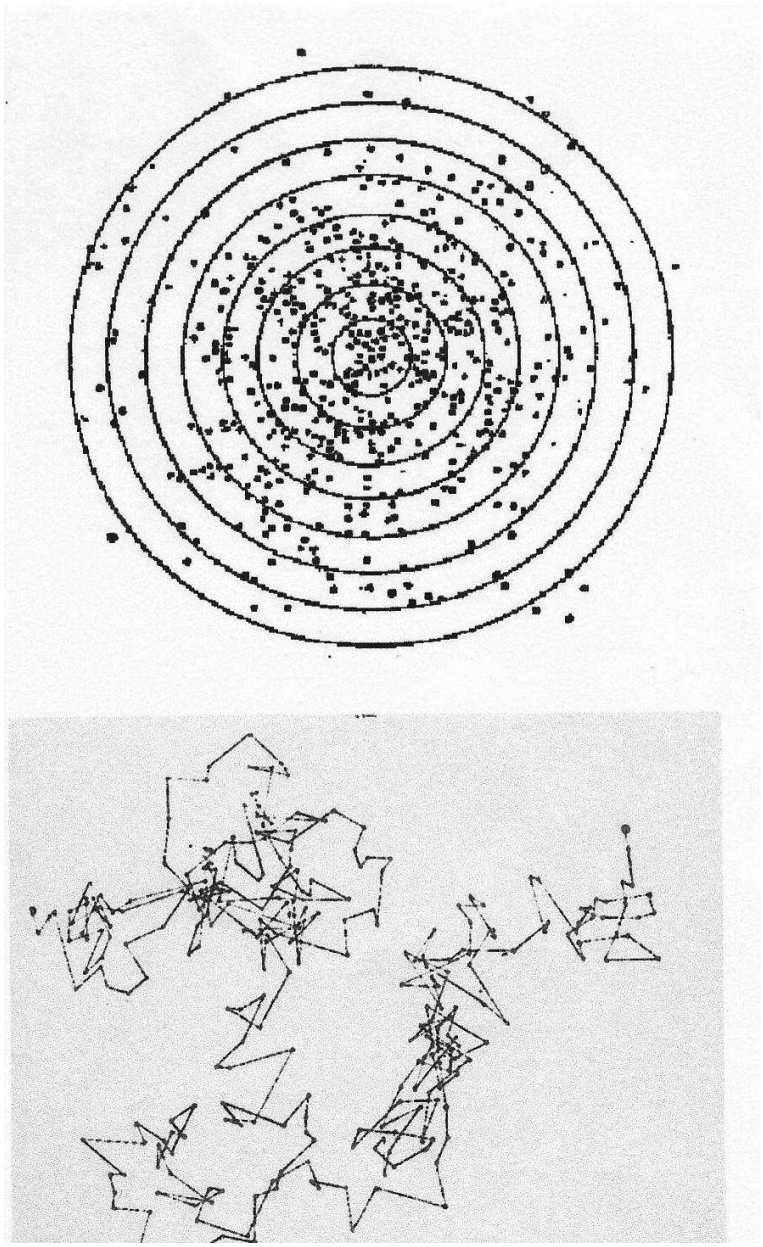

Figure 1: Brownian motion, taken from [1]. The top figure represents the statistical distribution of displacements and the lower figure the trajectory. The circles correspond to fractions and multiples of the mean square displacement $\Sigma(< x_i^2(t) > - < x_i >^2)$.

previous one. This means that *Brownian motion is a Markov process* (see Appendix).

The trajectory of a Brownian particle is represented in Figure 1. This trajectory is a fractal curve. This was already noticed by Jean Perrin [1, 2], who got the Nobel prize in 1926 for his experimental research on the atomic and molecular character of matter. In his words: "If we would plot the particle positions at time intervals hundred times smaller, every linear segment of the trajectory would take a polygonal shape as complicated as the whole trajectory....".

Let us represent the Brownian motion as a random walk. At each time step Δ the walker hops to one of the sites on a d-dimensional lattice. After n steps, the net displacement is $\mathbf{r}(n) = \Sigma_{i=1}^{n} \mathbf{e}_i$, so that

$$r^2(n) = \Sigma_{i=1}^{n} e_i^2 + 2\Sigma_{i>j} \mathbf{e}_i \cdot \mathbf{e}_j. \tag{13}$$

Assume that the steps are independent, and the pdf of each step is such that all steps have zero average $< \mathbf{e}_i >= 0$ and the same variance $< e_i^2 >= l^2$. Then, $< \mathbf{e}_i \cdot \mathbf{e}_j >= 0$ for $i \neq j$, and we have

$$< r^2(n) >= nl^2. \tag{14}$$

Hence, taking the limit of continuous time $n\Delta \to t$, Δ being the size of the time step, one obtains the mean square displacement

$$< r^2(t) >= 2dt \frac{l^2}{2d\Delta} = 2tdD. \tag{15}$$

This expression is equivalent to (7). The diffusion coefficient is

$$D = \frac{l^2}{2d\Delta}. \tag{16}$$

This is Einstein's relation which connects the macroscopic diffusion coefficient D with the microscopic jumps. It is a bridge between the microscopic and macroscopic world. Relation (15) is also frequently called the Einstein-Smoluchovski relation.

In equilibrium situations, the diffusion coefficient (16) can be expressed in terms of macroscopic quantities characterizing the system,

$$D = \frac{T}{\nu M} \tag{17}$$

This is Einstein's second relation. Here, T is the equilibrium temperature, M the mass of the particle and ν is the damping exponent of the fluid. In order to derive this relation, we follow the approach introduced by Langevin.

C. The Langevin approach

The equation of motion of a particle under the action of a random force $f(t)$ and a restoring force $\nu v(t)$ is

$$\frac{dv}{dt} = -\nu v + f(t). \tag{18}$$

This equation is normalized on the mass of the particle. The first term on the right represents viscous damping. The random force $f(t)$ represents the rapid kicks exerted by the fluid particles. This force has zero average and is taken to be δ-correlated in time,

$$< f >= 0, \quad < f(t_1)f(t_2) >= A\delta(t_1 - t_2). \tag{19}$$

The right-hand-side of (18) with the properties specified by (19) is called a Langevin force and (18) the Langevin equation. The random kicks tend to spread the velocity over a wider and wider range, while the damping tries to reduce the velocity to zero.

Equation (18) can easily be solved,

$$v(t) = v(0)\exp -\nu t + \int_0^t d\tau\ f(\tau)\exp \nu(\tau - t). \tag{20}$$

This gives $< v(t) >= v(0)\exp -\nu t$. The correlation function is

$$\begin{aligned} &< v(t_1)v(t_2) >= v^2(0)\exp -\nu(t_1 + t_2) \\ &+ \int_0^{t_1} d\tau \int_0^{t_2} ds\ < f(\tau)f(s) > \exp[-\nu(t_1 + t_2) + \nu(\tau + s)] \\ &= (v^2(0) - \frac{A}{2\nu})\exp -\nu(t_1 + t_2) + \frac{A}{2\nu}\exp -\nu|t_1 - t_2|. \end{aligned} \tag{21}$$

For long times, $\nu t >> 1$, i.e. for times long as compared with the damping time ν^{-1}, the influence of the initial velocity will disappear. Further, the system is assumed to be in equilibrium, so that (22) has to be averaged over a Maxwellian velocity distribution (indicated below by the subscript v_0). Then, the mean square velocity for large times must be

$$< v^2(t) >_{v_0} = \frac{A}{2\nu} = \frac{kT}{M}. \tag{22}$$

The interpretation of this process is that the kicks due to the random force $f(t)$ tend to spread the velocity over a larger range, while the damping brings the velocity back to its equilibrium value.

The mean square displacement is

$$< (x(t) - x(0))^2 >= \int_0^t \int_0^t dt_1 dt_2\ < v(t_1)v(t_2) >$$

$$= \frac{2T}{\nu M}[t - \frac{1}{\nu}(1 - \exp -\nu t)].$$

For short times, inertial effects, that are absent in Brownian motion, will dominate. For large times we find

$$< (x(t) - x(0))^2 > \approx \frac{2T}{\nu M} t. \tag{23}$$

Upon comparing (23) and (15) for $d = 1$ one obtains (17).

Applying (18) to an electron population with particle density n_0, we may interpret $n_0 e^2/m\nu$ as the conductivity of the medium, so that (17) can be written as

$$\Sigma = \frac{n_0 e^2}{T} D. \tag{24}$$

II. The classical random walk

The random walk as a model for a stochastic process occurs not only in numerous branches of physics, but also in many other sciences. In this Section we consider some aspects of the discrete-time random walk.

A. The relationship with the Gaussian and the binomial distribution

It will be shown that in the limit of a large number of steps the pdf of a random walker to be at the position x at time t is a Gaussian. It is shown that this normal distribution follows in the limit from the binomial distribution.

Suppose that the walker takes m steps to the right, each with probability p, and $n - m$ steps to the left, each with probability $q = 1 - p$. For simplicity we take all steps to be of equal length l. The time step is Δ.

At step n, i.e. at time $t = n\Delta$, the walker is at

$$x = [mp - (n - m)q]l = (m - nq)l, \tag{25}$$

The probability $P(m, n)$ to make m steps to the right and $n - m$ steps to the left is the binomial distribution

$$P(m, n) = \frac{n!}{m!(n - m)!} p^m (1 - p)^{n-m}. \tag{26}$$

It follows that

$$\sum_{m=0}^{n} P(m, n) = 1, \quad \sum_{m=0}^{n} mP(m, n) = np, \quad \sum_{m=0}^{n} m^2 P(m, n) = np + n(n-1)p^2, \tag{27}$$

so that the average position and the mean square displacement are

$$< x >= n(p-q)l, \qquad < (x- < x >)^2 >= npql^2. \tag{28}$$

Interpreting n as the time, it is seen that the first term represents a displacement with constant velocity. The second term is the variance, which, as we have seen in the previous section, is related with the diffusion coefficient.

It will be shown that *the normal distribution is obtained from the binomial distribution in the limit of large n at fixed probability p.*

Introduce the variable

$$\lambda = \frac{m- < m >}{\sqrt{npq}} = \frac{x- < x >}{\sqrt{2Dt}}, \qquad D = pq\frac{l^2}{2\Delta}, \tag{29}$$

and use Stirling's formula $\ln n! \approx (n+1/2)\ln n - n + \ln\sqrt{2\pi} + O(n^{-1})$.
Then, from (26) one obtains

$$P(m,n) \approx \frac{1}{\sqrt{2\pi npq}} \exp\{(\frac{1}{2} + np + \lambda\sqrt{npq})\ln(1+\frac{\lambda q}{\sqrt{npq}})^{-1}$$

$$+(\frac{1}{2} + nq - \lambda\sqrt{npq})\ln(1-\frac{\lambda p}{\sqrt{npq}})^{-1}\}.$$

For simplicity we take equal probability for a step to the left and for one to the right $p = q = 1/2$, so that the average displacement vanishes. Using the expansion $\ln(1+x) \approx x - x^2/2$ and $p(\lambda,n)\Delta\lambda = \Sigma_{\Delta m}P(m,n)$, one obtains in the limit $n \to \infty$ the Gaussian distribution

$$P(\lambda,n) = \frac{1}{\sqrt{2\pi}} \exp -\frac{\lambda^2}{2}. \tag{30}$$

Finally, with $p(\lambda,n)d\lambda = P(x,t)dx$, the distribution function is found as a function of space-time

$$P(x,t) = \frac{1}{\sqrt{4\pi Dt}} \exp -\frac{x^2}{4Dt}. \tag{31}$$

This is the pdf to find the random walker at the position x at time t.

Problem

1. Proof the relations (27). Use

$$\sum_{m=0}^{n} P(m,n) = (p+q)^n = 1.$$

B. The standard diffusion equation

The relationship between the standard diffusion equation and the random walk process can be seen as follows.

Since the steps are independent, the probability of the walker being at m at step $n+1$ is equal to the sum of the probability that the walker is a $m-1$ at time n and makes a step to the right at the next time step, and the probability that the walker is at $m+1$ and makes a step to the left,

$$P_{n+1}(m) = pP_n(m-1) + qP_n(m+1) \tag{32}$$

Rewrite this equation as follows

$$P_{n+1}(m) - P_n(m) = p[P_n(m-1) - P_n(m)] + q[P_n(m+1) - P_n(m))].$$

and take the continuous time limit $n\Delta \to t$. This yields,

$$\frac{\partial P(m)}{\partial t} = \frac{p}{\Delta}[P(m-1) - P(m)] + \frac{q}{\Delta}[P(m+1) - P(m)]. \tag{33}$$

This the master equation for the *continuous time random walk*. Taking $ml \to x$ and expanding around x yields

$$\frac{\partial P(x,t)}{\partial t} = -A\frac{\partial P(x,t)}{\partial x} + D\frac{\partial^2 P(x,t)}{\partial x^2} \tag{34}$$

with

$$A = (p-q)\frac{l}{\Delta}, \qquad D = \frac{l^2}{2\Delta}. \tag{35}$$

This is the Fokker-Planck equation. For $A = 0$ it is identical in form to the standard diffusion equation (3). The Fokker-Planck equation is usually obtained from the *master equation* (see Appendix).

C. The central limit theorem

The reason that the Gaussian or normal distribution is dominant in so many fields of statistics is the *central limit theorem*.

The central limit theorem is a statement about the sum of a large number of random variables. It says that if the random variables are
1. independent and identically distributed (IID) and
2. the distribution function of each random variable has a finite variance,
then the sum of the random variables approaches a *Gaussian distribution*.

All other things being equal, a sequence of coin flips and a sequence of dice rolls are IID. The displacements of a heavy particle in Brownian motion may be considered as IID. The sum of the random variables may, e.g., represent the position of a particle as a consequence of the cumulative effect of a large number of collisions. The position of a random walker after a large number of steps is a standard example of such a variable.

Although we take the random walk as a standard example to illustrate the theory of stochastic variables, one should keep in mind that the symbol X may represent any stochastic variable. It may e.g. represent the variation of the electric or magnetic field, of the local velocity of a fluid or of the temperature of some substance.

In many cases of physical interest, $< X^2 >$ represents an energy density. This implies that the variance must be finite so that condition 2. of the *central limit theorem* is satisfied on physical grounds.

Let $X_1, X_2, ..., X_n, ...$ be a set of independent variables, each having the same distribution function with average $< X_i >= \mu$ and variance σ, $< X_i^2 > - < X_i >^2 = \sigma^2$. This means that for the sum of n of such IID random variables one has

$$< \sum_1^n X_i >= n\mu, \qquad < \sum_1^n (X_i - \mu)^2 >= n\sigma^2. \tag{36}$$

Thus, the average value of the sum grows with n and its variance with $\sqrt{n}$.

Introduce the variables y_i and the sums s_n,

$$y_i = \frac{X_i - \mu}{\sigma\sqrt{n}}, \qquad s_n = \sum_n y_i, \tag{37}$$

and write $\lim_{n\to\infty} s_n = s$.
The central limit theorem states that in the limit of large n, the distribution function of s_n approaches the Gaussian distribution

$$\lim_{n\to\infty} P(s_n) = P(s) = \frac{1}{\sqrt{2\pi}} \exp -s^2. \tag{38}$$

This is an asymptotic distribution. A finite number of observations may already lead to a reasonable approximation of the distribution close to its peak. To determine the tails requires a large number of observations. In d-dimensions one also obtains (38), but with $\sqrt{2\pi}$ replaced by $(2\pi)^{d/2}$.

On the basis of the central limit theorem it is clear that the Brownian motion introduced in section B is described by a Gaussian distribution function.

The proof of the theorem is remarkably simple. The characteristic function of the distribution function $p(y_i)$ is

$$\hat{p}_{y_i}(k) = \int_{-\infty}^{+\infty} dy_i \; p(y_i) \exp iky_i. \tag{39}$$

Since $< y_i >= 0$ and $< y_i^2 >= 1/n$, this may be approximated for $k^2/n \to 0$ by

$$\hat{p}_{y_i}(k) = 1 - \frac{1}{2}\frac{k^2}{n} + o(\frac{k^2}{n}), \quad n \to \infty, \tag{40}$$

where $o(k^2/n)$ indicates a function that goes to zero faster than k^2/n. Since all y_i are independent, the probability $P(s_n)$ can be written as

$$P(s_n) = <\delta(s_n - \sum y_i)>$$

$$= \int dy_1 \int dy_2 .... \int dy_n \; P(y_1, y_2, ..., y_n)\delta(s_n - \sum y_i)$$

$$= \int dy_1 .... \int dy_n \; P(y_1, ..., y_n) \int \frac{dk}{2\pi} \exp -ik(s_n - \sum y_i)$$

$$= \int \frac{dk}{2\pi} \; \hat{p}_{y_i}^n(k) \exp(-iks_n). \tag{41}$$

This means that the characteristic function of the sum is equal to the product of the individual characteristic functions, $\hat{P}_{s_n}(k) = \hat{p}_{y_i}^n(k)$.
Since all random variables y_i are identically distributed, the characteristic function of s_n for small values of k^2/n may be approximated by

$$\hat{P}_{s_n}(k) = [1 - \frac{1}{2}\frac{k^2}{n} + o(\frac{k^2}{n})]^n \quad \to \quad e^{-k^2/2}, \quad n \to \infty. \tag{42}$$

The inverse FT (41) of this expression gives the Gaussian (38).

D. Number of returns

A random walker may return many times to any previous position. This will strongly depend on the dimensions of the space. We will analyze this Brownian trajectory with self-intersection with a specific 1D model.

Consider a 1D random walk in a layered medium. A particle diffuses in the x-direction through the layers. The width of a layer is large as compared with the step size. The particle will diffuse through the layer of thickness a in a time τ_D,

$$a = \sqrt{2D\tau_D}. \tag{43}$$

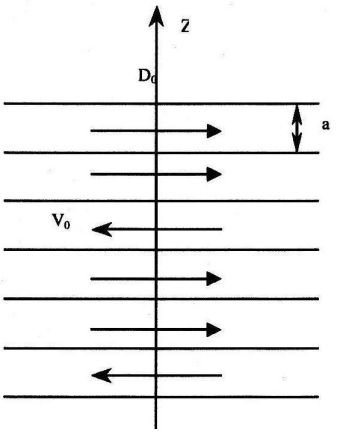

Figure 2:

In a time $t = N\tau_D$, the particle diffuses though N layers. On the other hand, the diffusion length is $l = \sqrt{2DN\tau_D} = a\sqrt{t/\tau_D} = a\sqrt{N}$. This means that the average number of *different* layers visited in a time t is $\sqrt{N}$.
The probability to be in the first layer at time t is obtained from (31)

$$P(0,t)a = \frac{1}{\sqrt{4\pi Dt}}a. \tag{44}$$

This will be equal to the relative number of times that the walker returns to the first layer in a time t. Hence, the number of times the particle returns to the first layer is

$$N_r = \frac{a}{\sqrt{4\pi Dt}}N \approx \frac{N}{\sqrt{t/\tau_D}} \approx \sqrt{N}. \tag{45}$$

III. Lévy flights

In this section we reconsider the random walk. The walker is assumed to take steps at fixed points in time. Again we assume that all steps are independent and identically distributed (IID). However, *we do not assume a constant step size. We do not even assume a priori that the mean value or the variance of the pdf of a step exist.* Thus, the conditions for the validity of the central limit theorem and the occurrence of Gaussian distributions are not necessarily satisfied.

The probability function $P_n(\mathbf{r})$ for the walker to be at the position $\mathbf{r}$ at the nth step can be obtained from

$$P_n(\mathbf{r}) = \int d\mathbf{r}'\ P_{n-1}(\mathbf{r}')p(\mathbf{r}-\mathbf{r}'), \quad P_0(\mathbf{r}) = \delta(\mathbf{r}), \tag{46}$$

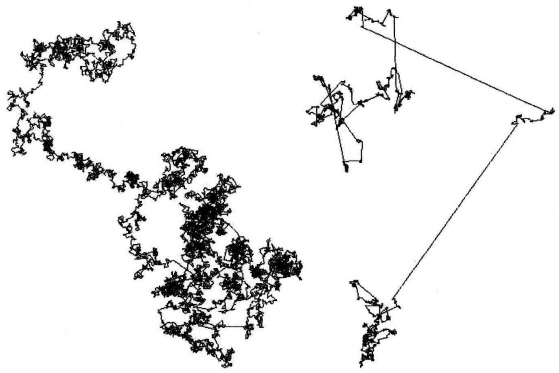

Figure 3: Classical random walk (left) and Lévy flights (right).

where $p(\mathbf{r})$ is the probability density to make a step $\mathbf{r}$. Use the characteristic function

$$\hat{P}_n(\mathbf{k}) = \int d\mathbf{r}\ P_n(\mathbf{r}) \exp i\mathbf{k}\cdot\mathbf{r}. \tag{47}$$

Then, one obtains from (46), $\hat{P}_n(\mathbf{k}) = \hat{P}_{n-1}(\mathbf{k})\hat{p}(\mathbf{k})$, so that

$$\hat{P}_n(\mathbf{k}) = \hat{p}^n(\mathbf{k}), \tag{48}$$

and

$$P_n(\mathbf{r}) = \frac{1}{(2\pi)^d}\int d\mathbf{k}\ \hat{p}^n(\mathbf{k}) \exp(-i\mathbf{k}\cdot\mathbf{r}). \tag{49}$$

This is the solution to the random walk problem. It depends only on the pdf $p(\mathbf{r})$ which is also called the structure function.

A. A structure function with a long tail

Assume that for large values of the (microscopic) step size $\mathbf{y}$, the pdf $p(\mathbf{y})$ is isotropic and decays according to some power of $y = |\mathbf{y}|$,

$$p(\mathbf{y}) = (1-\alpha)\delta(\mathbf{y}) + \frac{A}{y^l}H(y-y_0). \tag{50}$$

The first term represents the bulk of the distribution, which is not of interest in the present discussion. The second term represents the tail, the function H being the Heavyside function. This tail is a *Lévy distribution. It corresponds to large but infrequent steps, so-called rare events*.

From the normalization $\int d^dy\ p(\mathbf{y}) = \int dyd\Omega\ y^{d-1}p(\mathbf{y}) = 1$ follows that

$$A = \frac{\alpha}{S_d} f y_0^f, \qquad f = l - d > 0,$$

with S_d being the solid angle, $S_1 = 1,\ S_2 = 2\pi,\ S_3 = 4\pi$. (This can be generalized to $S_d = 2\pi^{d/2}/\Gamma(d/2)$)). Hence, we have

$$p(\mathbf{y})d^d y = (1-\alpha)\delta(\mathbf{y})d^d y + \frac{\alpha}{S_d}\frac{f y_0^f}{y^{f+1}}H(y-y_0)d\Omega dy. \tag{51}$$

The power f is called the *microscopic step index*.

A. It is seen that for $f > 2$ the mean square step exists,

$$< y^2 >= \int d^d y\ y^2 p(\mathbf{y}) = \alpha y_0^2 \frac{f}{f-2}.$$

As a consequence, the average mean square position after n steps

$$< r^2 >= n < y^2 >= t < y^2 > \tag{52}$$

also exists and the central limit theorem tells us that the distribution $\lim P_n(\mathbf{r})$ in (49) is Gaussian.

B. For $1 < f < 2$ the mean square displacement does not exist, but the mean step is still finite

$$< y >= \int d^d y\ y p(\mathbf{y}) = \alpha y_0 \frac{f}{f-1},$$

so that $< r >= n < y >= t < y >$.

C. In the interval $0 < f < 1$ even the mean step size does not exist.

Next, we consider the probability P_n given by (49) for large values of n. The Fourier transform of the characteristic function $p(\mathbf{y})$ is

$$\hat{p}(\mathbf{k}) = \int d\mathbf{y}\ p(\mathbf{y}) \exp i\mathbf{k}\cdot\mathbf{y} = 1 + \frac{\alpha}{S_d}\int_{y_0}^{\infty} dy \int d\Omega \frac{f y_0^f}{y^{1+f}}[\exp(i\mathbf{k}\cdot\mathbf{y}) - 1] \tag{53}$$

$$\approx 1 - D(ky_0)^f, \quad D = constant, \quad k = |\mathbf{k}|,$$

which gives with $n \to t$

$$\hat{p}^n(\mathbf{k}) \approx [1 - D(ky_0)^f]^n \approx \exp(-Dn(ky_0)^f) = \exp(-Dt(ky_0)^f). \tag{54}$$

Substitution into (49) gives

$$P(\mathbf{r},t) = \lim_{n\to\infty} P_n(\mathbf{r}) = \int \frac{d^d\mathbf{k}}{(2\pi)^d}\exp(-i\mathbf{k}\cdot\mathbf{r} - tDk^\mu y_0^\mu), \tag{55}$$

with - $f > 2$, $< r^2 >$ exists and $\mu = 2$ on the basis of the central limit theorem,

- $f < 2$, $< r^2 >$ does not exist and $\mu = f$.

B. Stable probability distributions functions

Consider the pdf $n(\mathbf{x})$ of a d-dimensional random process. Its characteristic function is

$$\hat{n}(\mathbf{k}) = \int d^d x \; n(\mathbf{x}) exp i\mathbf{k} \cdot \mathbf{x}. \tag{56}$$

Any probability distribution function whose characteristic function satisfies the form-invariance property

$$\hat{n}(a_1\mathbf{k})\hat{n}(a_2\mathbf{k}) = \hat{n}(a\mathbf{k}) \tag{57}$$

is called a **stable law**. For simplicity we will consider only symmetric pdf's that depend on $r = \sqrt{\mathbf{x} \cdot \mathbf{x}}$.

Writing $\hat{n}(a\mathbf{k}) = \exp A(ak)$, (57) gives

$$A(ak) = A(a_1 k) + A(a_2 k).$$

It follows that the general solution is the power law $A(k) = -Ck^\mu$. The constants a, a_1, a_2 must satisfy

$$-C(a_1^\mu + a_2^\mu) = -Ca^\mu.$$

Given a_1 and a_2, this relationship determines a. Hence, the characteristic function of a stable law reads

$$\hat{n}(\mathbf{k}) = \exp(-C|k|^\mu). \tag{58}$$

The inverse transform is

$$n(\mathbf{x}) = (2\pi)^{-d} \int d^d\mathbf{k} \; exp(i\mathbf{k} \cdot \mathbf{x} - C|k|^\mu). \tag{59}$$

Any pdf of this form is called a *symmetric Lévy distribution function*.

For convergence the constant C must be positive. Regularity of $\hat{n}(\mathbf{k})$ at $k = 0$ requires $\mu > 0$. The inverse Fourier transform (59) requires $\mu \leq 2$, since it is not everywhere positive on the domain for $\mu > 2$.

The mean square displacement is

$$< r^2(t) >= -\frac{\partial^2 \hat{n}(\mathbf{k})}{\partial \mathbf{k} \cdot \partial \mathbf{k}}|_{\mathbf{k}=0} = C\mu[(\mu + d - 2)|k|^{\mu-2}]|_{|k|=0}.$$

This limit does not exist for $\mu < 2$. For $\mu = 2$ one finds $< r^2(t) >= 2dC$, which yields the classical result with $C = \sigma^2 t$. For $\mu > 2$ it follows that $< r^2(t) >= 0$, which is only possible if the pdf $n(\mathbf{x}, t)$ is not everywhere positive on the domain. Note hat the distributions (59) are properly normalized

$$\int n(\mathbf{x}) d^d x = \int d^d\mathbf{k} \; \delta(\mathbf{k}) \exp -C|k|^\mu = 1. \tag{60}$$

The Gaussian distribution with $\mu = 2$ corresponds to the solution of the standard diffusion equation,

$$\hat{n}(\mathbf{k}) = e^{-|k|^2}, \qquad n(\mathbf{x}) = \frac{1}{2\pi^{1/2}} e^{-x^2}. \tag{61}$$

For $\mu = 1$ one obtains the Cauchy distribution,

$$\hat{n}(\mathbf{k}) = e^{-|k|}, \qquad n(x) = \frac{1}{\pi(1+x^2)}. \tag{62}$$

For $\mu = 3/2$ one finds the Holtsmark distribution [],

$$\hat{n}(\mathbf{k}) = e^{-|k|^{3/2}}, \qquad n(x) = \frac{x}{(1+x^2)^{3/2}}. \tag{63}$$

Note that all these probability densities, except the Gaussian with $\mu = 2$, have power "tails" for large values of x. In Part V it will be discussed that for $\mu < 2$, the distributions (62) are associated with diffusion equations with *fractional derivatives*.

NOTE
Let $n(\mathbf{x})$ be the pdf of a d-dimensional random variable $\mathbf{x}$. Then, the pdf $n_a(\mathbf{x})$ of a random variable $\mathbf{y} = a^{-1}\mathbf{x}$ is of the form $n(\mathbf{x}) = C_a n_a(a^{-1}\mathbf{x})$. Both pdf's are normalized to unity, so that $C_a = a^{-d}$. Hence we find

$$n_a(\mathbf{x}) = a^{-d} n(a^{-1}\mathbf{x}).$$

In Fourier space this relation reads

$$\hat{n}_a(\mathbf{k}) = \hat{n}(a\mathbf{k}).$$

C. The inverse transform

First, we derive an approximate expression for the integral (59) in 2 dimensions $(d = 2)$. We start from the expression

$$n(\mathbf{x}, t) = \int \frac{d^2\mathbf{k}}{(2\pi)^2}\, exp(-i\mathbf{k}\cdot\mathbf{x} - C|k|^\mu) = \frac{2}{(2\pi)^2}\int_0^\infty \int_0^\pi kdkd\theta\ \exp(ikrcos\theta - Ck^\mu). \tag{64}$$

Use the integral representation for the Bessel function

$$J_0(z) = \frac{1}{\pi}\int_0^\pi d\theta\ e^{iz\cos\theta},$$

and subsequently the representation

$$J_0(z) = \frac{2}{\pi}\int_1^\infty dy \frac{sinzy}{\sqrt{y^2-1}}.$$

Then, (71) becomes

$$n(r,t) = \frac{1}{\pi^2} Re\, i \int_1^\infty \frac{dy}{\sqrt{y^2-1}} \int_0^\infty dk\, k e^{-ikry - Ck^\mu}.$$

Extend this integral into the complex k-plane and close the contour in the fourth quadrant along the negative imaginary axis ($k \to -iq$). Further, we extend the integral up to k_0, where $k_0 >> 1$ is such that $k_0 r >> k_0^\mu$. Then, the integral along the real axis is converted into the asymptotic integral,

$$n(r,t) \approx \frac{1}{\pi^2} Re\, i \int_1^\infty \frac{dy}{\sqrt{y^2-1}} \int_0^\infty dq\, (-q) e^{-qry} (1 - Cq^\mu e^{-i\pi\mu/2}).$$

The first term on the right does not give any contribution, so that

$$n(r,t) \approx \frac{1}{\pi^2} C sin\frac{\pi\mu}{2} \int_1^\infty \frac{dy}{\sqrt{y^2-1}} \int_0^\infty dq\, q^{\mu+1} e^{-qry}$$

$$\approx \frac{1}{\pi^2} sin\frac{\pi\mu}{2} \frac{C}{r^{\mu+2}} \Gamma(\mu+2) \int_1^\infty \frac{dy}{y^{\mu+2}\sqrt{y^2-1}}.$$

The integral can be expressed in terms of the Beta function ($y = v^{-1/2}$),

$$\int_1^\infty \frac{dy}{y^{\mu+2}\sqrt{y^2-1}} = \int_0^1 dv \frac{v^\mu}{\sqrt{1-v}}$$

$$= \frac{1}{2} B(\mu/2+1, 1/2) = \frac{1}{2} \frac{\Gamma(\mu/2+1)\Gamma(1/2)}{\Gamma(\mu/2+3/2)}.$$

Thus, we finally obtain the approximate expression

$$n(r,t) \approx \frac{C}{2\pi^2} sin\frac{\mu\pi}{2} \frac{\Gamma(\mu+2)\Gamma(\mu/2+1)\Gamma(1/2)}{\Gamma(\mu/2+3/2)} r^{-\mu-2}. \tag{65}$$

The coefficient can be further simplified by using relationships between the Gamma functions.

Next we derive an asymptotic expression for large values of x of the integral (62) in the one dimensional case ($d = 1$),

$$n(\mathbf{x},t) = \int \frac{dk}{(2\pi)}\, exp(-ikx - C|k|^\mu) = \frac{1}{\pi} Re \int_0^\infty dk\, \exp(ikx - Ck^\mu). \tag{66}$$

Extend the integral in the complex plane and close the integral in the fourth quadrant along the negative imaginary axis ($k \to -iq$). Further, we extend the integral up to k_0, where $k_0 >> 1$ is such that $k_0 x >> k_0^\mu$,

$$n(\mathbf{x},t) = \frac{-1}{\pi} Re\, i \int_0^\infty dq\, (1 - Cq^\mu \exp i\pi\mu/2) \exp -qx = \frac{1}{\pi} sin\frac{\pi\mu}{2} \Gamma(1+\mu) \frac{1}{x^{1+\mu}}. \tag{67}$$

D. Lévy distributions

Non-local effects can be also described in terms of a random walk model. Such a model leads to a description in terms of an integral equation.

In the case of homogeneous turbulence, the Chapman-Kolmogorov functional equation can be written as

$$\frac{\partial n(\mathbf{x},t)}{\partial t} = \int_{-\infty}^{\infty} G(\mathbf{x}-\mathbf{x}')n(\mathbf{x}',t)d^dx'. \tag{68}$$

This equation is Markovian and cannot describe memory effects. It is non-local in space but local in time. Introducing the Fourier transform with respect to space, this equation becomes,

$$\frac{\partial \hat{n}(\mathbf{k},t)}{\partial t} = \hat{G}(\mathbf{k})\hat{n}(\mathbf{k},t). \tag{69}$$

In case of isotropic turbulence the kernel will be a function $k = |\mathbf{k}|$. If the kernel has the form

$$\hat{G}(k) = -Dk^2, \tag{70}$$

(70) yields the classical diffusion equation after application of the inverse Fourier transform.

This has led to different phenomenological methods to improve the diffusion equation, The discussion in the previous section suggests that the form

$$\hat{G}(k) = -A|k|^{\mu}, \tag{71}$$

where A has the dimensions $cm^{\mu}s^{-1}$, is relevant for the description of turbulent diffusion. Then, we have

$$\frac{\partial \hat{n}(k,t)}{\partial t} = -A|k|^{\mu}\hat{n}(k,t). \tag{72}$$

In Part V it will be shown that the Fourier transform of the right hand side corresponds to a symmetric Riesz fractional derivative.

The solution of (68) is

$$\hat{n}(k,t) = \hat{n}(k,0)\exp(-A|k|^{\mu}t). \tag{73}$$

With $n(x,0) = \delta(x)$, the inverse transform yields the integral (62) with $C = At$.

IV. APPENDIX

A. Stochastic variables

A stochastic or random variable X is defined by a set of values x that it can attain and by a probability distribution $P(x)$ over this set. This set is also called the 'sample space' or 'range'. This space may be either continuous or discrete or a mixture of the two. The total probability has to be equal to unity

$$\int_{range} dx\ P(x) = 1, \qquad \Sigma_n P_n = 1. \tag{74}$$

The discrete case can be recovered from the continuous case with $P(x) = \Sigma_n\ P_n \delta(x - x_n)$.

The nth moment of $P(x)$ is defined as

$$\mu_m =< x^m >= \int dx\ x^m P(x). \tag{75}$$

The first moment is the average or expected value

$$\mu_1 =< x >= \int_{range} dx\ xP(x). \tag{76}$$

The stochastic variable x is called *centered* if $< x >= 0$.

The second moment μ_2 is related with the variance σ

$$\sigma^2 =< (x - < x >)^2 >= \mu_2 - \mu_1^2 = \int_{range} dx\ (x - \mu_1)^2 P(x), \tag{77}$$

The variance will often exist and be finite. E.g. if x represents the velocity of a particle or the magnitude of a fluctuating magnetic field, then the variance is proportional to the square root of an energy or energy density, which has to remain finite.

The skewness and flatness of a centered random variable are defined respectively as

$$S = \frac{< x^3 >}{< x^2 >^{3/2}}, \qquad F = \frac{< x^4 >}{< x^2 >^2}. \tag{78}$$

The Fourier transform of a probability distribution $P(x)$ is called its characteristic function,

$$\hat{P}(z) =< \exp izx >= \int dx\ P(x) \exp izx \tag{79}$$

It is seen that

$$\hat{P}(0) = 1, \quad |\hat{P}(z)| \leq 1, \qquad \hat{P}^*(-z) = \hat{P}(z). \tag{80}$$

The characteristic function generates all the moments of $P(x)$,

$$\hat{P}(z) = \sum_{m=0}^{\infty} \frac{(iz)^m}{m!}\mu_m, \qquad < z^m >= \frac{1}{i^m}\frac{d^m}{dz^m}\hat{P}(z). \tag{81}$$

The mean square displacement i.e. the average of the squared distance $r^2 = \mathbf{x}^2 = \Sigma_1^d x_i^2$ in d-dimensions is

$$< r^2 >= -\frac{\partial}{\partial \mathbf{z}} \cdot \frac{\partial}{\partial \mathbf{z}} \hat{P}(z)|_{z=0}. \tag{82}$$

A d-dimensional, centered random variable is Gaussian if

$$P(\mathbf{x}) = \frac{1}{(2\pi\sigma^2)^{d/2}} \exp -\frac{x^2}{2\sigma^2}. \tag{83}$$

Its characteristic function is

$$\hat{P}(z) =< \exp i\mathbf{x} \cdot \mathbf{z} >= \exp -\frac{1}{2}\sigma^2 z^2, \tag{84}$$

where σ is the variance (77) in d-dimensions,

$$\sigma^2 =< \mathbf{x}^2 >= \int_{range} d^d x\ \mathbf{x}^2 P(\mathbf{x}). \tag{85}$$

All odd moments of a Gaussian distribution vanish, so that it has only even moments $\mu_2,\ \mu_4, ...$ All higher moments of a Gaussian can be expressed in terms of the second moment μ_2.

All quantities Y_X that are defined by some mapping $f(X)$ are also stochastic variables. In particular when such a function also depends on a additional variable t, quite often the time,

$$Y_X(t) = f(x,t), \tag{86}$$

$Y_X(t)$ is called a stochastic process; the sample function $f(x,t)$ is a realization of that process. A stochastic process can be considered as an ensemble of these sample functions.

Consider a stochastic process X_t. The probability density that the stochastic (random) variable takes the value $x_1, x_2, ..., x_n$ at the successive 'times' $t_1 < t_2 < .... < t_n$, is $P_n(x_n, t_n; x_{n-1}, t_{n-1}; ....; x_1, t_1)$. Clearly one must require that

$$\int dx_1 P_n(x_n, t_n; x_{n-1}, t_{n-1}; ....; x_1, t_1) = P_{n-1}(x_n, t_n; x_{n-1}, t_{n-1}; ....; x_2, t_2). \tag{87}$$

The probability P_n may be written as

$$P_n(x_n, t_n; x_{n-1}, t_{n-1}; ....; x_1, t_1) = P_{n-k|k}(x_n, t_n; ...., x_{k+1}, t_{k+1}|x_k, t_k; ....; x_1, t_1)x \tag{88}$$

$$P_k(x_k, t_k; ....; x_1, t_1).$$

where $P_{n-k|k}(...)$ is the conditional probability that the variable takes the value x_n at time t_n etc. and the value x_{k+1} at time t_{k+1}, given that it has the value x_k at time t_k etc. and the value x_1 at time t_1.

In case of a completely random process, all values $\{x_k\}$ are independent, so that one has

$$P_n(x_n, t_n; x_{n-1}, t_{n-1}; ....; x_1, t_1) = \Pi_k P_1(x_k, t) \tag{89}$$

This means that the knowledge of a single distribution function, $P_1(x,t) = P(x,t)$, would be sufficient to determine the process.

The next simplest case is called a Markov process.

B. Markov processes

B..1 The Chapman-Komolgorov equation

For a Markov process, the conditional probability that the stochastic variable takes the value x_n at time t_n only depends on its value x_{n-1} at the previous time t_{n-1} and does not depend on its values at earlier times, i.e. for any n,

$$P_{1|n-1}(x_n, t_n|x_{n-1}, t_{n-1}; ....; x_1, t_1) = P_{1|1}(x_n, t_n|x_{n-1}, t_{n-1}). \tag{90}$$

This means that the transition from (x_{n-1}, t_{n-1}) to (x_n, t_n) does not dependent on earlier transitions. This lack of memory is characteristic for Markov processes. Such a process is completely determined by two probability densities, $P_1(x,t)$ and the transition probability $P_{1|1}(x_2, t_2|x_1, t_1)$. It follows that

$$P_3(x_3, t_3; x_2, t_2; x_1, t_1) = P_{1|2}(x_3, t_3|x_2, t_2; x_1, t_1) P_2(x_2, t_2; x_1, t_1) \tag{91}$$

$$= P_{1|1}(x_3, t_3|x_2, t_2) P_{1|1}(x_2, t_2|x_1, t_1) P_1(x_1, t_1), \qquad t_3 > t_2 > t_1.$$

Integrating over x_2 and dividing by $P_1(x_1, t_1)$ yields the famous *Chapman-Kolmogorov* equation,

$$P_{1|1}(x_3, t_3|x_1, t_1) = \int dx_2 P_{1|1}(x_3, t_3|x_2, t_2) P_{1|1}(x_2, t_2|x_1, t_1). \tag{92}$$

This equation says that the conditional probability to find the value x_3 at t_3 given that it is x_1 at t_1, is equal to the product of the conditional probabilities integrated over all values that the variable could attain at some intermediate time t_2.

Upon multiplying (92) with $P_1(x_1, t_1)$ and integrating over x_1, it is seen that the distribution function $P_1(x, t)$ must satisfy

$$P_1(x,t) = \int dx' P_{1|1}(x,t|x',t') P_1(x',t'), \qquad t' < t. \tag{93}$$

The best known example of a Markov process is Brownian motion. The mathematical model that describes Brownian motion is the continuous time random walk or the *Wiener* process. It can easily be checked that the transition probability

$$P_{1|1}(\mathbf{x}_2, t_2|\mathbf{x}_1, t_1) = \frac{1}{\sqrt{2\pi D(t_2 - t_1)}} \exp -\frac{(x_2 - x_1)^2}{2D(t_2 - t_1)}, \quad t_2 > t_1, \tag{94}$$

satisfies the Chapman-Kolmogorov equation. Together with the initial condition $P(x, 0) = \delta(x)$, this transition probability specifies a non-stationary Markov process.

According to (94) the probability density is

$$P(x,t) = \int dx' P_{1|1}(x,t|x',0) P(x',0) = \frac{1}{\sqrt{2\pi Dt}} \exp -\frac{x^2}{2Dt}. \tag{95}$$

Markov processes that are invariant under a shift in time are of special interest. For these *stationary processes* the conditional probability is a function of the *time difference*

$$P_{1|1}(\mathbf{x}_2, t_2; \mathbf{x}_1, t_1) = P(\mathbf{x}_2|\mathbf{x}_1; t_2 - t_1), \tag{96}$$

while the probability P_1 is independent of time, $P_1(\mathbf{x}, t) = P(\mathbf{x})$. $P(\mathbf{x})$ is the familiar equilibrium distribution as described by statistical equilibrium mechanics.

If the process is homogeneous, then the transition probability will depend only on the difference $\mathbf{x}_2 - \mathbf{x}_1$.

B..2 The master equation

The Chapman-Kolmogorov equation can be rewritten in integro-differential form. Write in (92) $x_3 = x, t_3 = t + \Delta t, x_2 = x', t_2 = t, x_1 = x_0, t_1 = t_0$. Then, the Chapman-Kolmogorov equation reads

$$P_{1|1}(\mathbf{x}, t + \Delta t|\mathbf{x}_0, t_0) = \int d^d x' P_{1|1}(\mathbf{x}, t + \Delta t|\mathbf{x}', t) P_{1|1}(\mathbf{x}', t|\mathbf{x}_0, t_0). \tag{97}$$

In the limit $\Delta t \to 0$ this equation can be written as

$$\frac{\partial P_{1|1}(\mathbf{x}, t|\mathbf{x}_0, t_0)}{\partial t} = \lim_{\Delta t \to 0} \int d^d x' \frac{P_{1|1}(\mathbf{x}, t + \Delta t|\mathbf{x}', t) - \delta(\mathbf{x} - \mathbf{x}')}{\Delta t} P(\mathbf{x}', t|\mathbf{x}_0, t_0).$$

A natural initial condition is $P_{1|1}(\mathbf{x}, t|\mathbf{x}'; t) = \delta(\mathbf{x} - \mathbf{x}')$. Adopt the limit

$$\lim_{\Delta t \to 0} \frac{P_{1|1}(\mathbf{x}, t + \Delta t|\mathbf{x}', t) - \delta(\mathbf{x} - \mathbf{x}')}{\Delta t} = W(\mathbf{x}|\mathbf{x}') - a_0 \delta(\mathbf{x} - \mathbf{x}'), \tag{98}$$

where $W(\mathbf{x}|\mathbf{x}')$ is the *transition probability per unit time* and

$$a_0 = \int d^d x \, W(\mathbf{x}|\mathbf{x}') \tag{99}$$

is determined by the normalization condition $\int d^d x \, P(\mathbf{x}, t + \Delta t|\mathbf{x}', t) = 1$.

In the limit $\Delta t \to 0$, the integral equation (97) becomes the integro-differential equation

$$\frac{\partial P(\mathbf{x}, t|\mathbf{x}_0, t_0)}{\partial t} = \int d^d x' \, [W(\mathbf{x}|\mathbf{x}') P(\mathbf{x}', t|\mathbf{x}_0, t_0) - W(\mathbf{x}'|\mathbf{x}) P(\mathbf{x}, t|\mathbf{x}_0, t_0)]. \tag{100}$$

The indices have been omitted. This equation is known as the *master equation*. It is the differential form of the Chapman-Kolmogorov equation. Upon multiplying this equation with $P(\mathbf{x}_0, t_0)$ and by integrating over $\mathbf{x}_0$, one obtains the same equation with the conditional probability $P(\mathbf{x}, t|\mathbf{x}_0, t_0)$ replaced by the probability density $P(\mathbf{x}, t)$

$$\frac{\partial P(\mathbf{x}, t)}{\partial t} = \int d^d x' \, [W(\mathbf{x}|\mathbf{x}') P(\mathbf{x}', t) - W(\mathbf{x}'|\mathbf{x}) P(\mathbf{x}, t)]. \tag{101}$$

If the range of X is a set of discrete states labelled by n, the master equation reads

$$\frac{dP_n(t)}{dt} = \Sigma_{n'} \, [W_{nn'} P_{n'}(t) - W_{n'n} P_n(t)]. \tag{102}$$

Equations (101) and (102) describe *Gain-Loss* processes. The time rate of change of the probability function in (101) is determined by two processes:
1. the system is in state x and undergoes a transition to a different state x'. This is a loss for state x.
2. the system is in state x' and undergoes a transition to state x. This is a gain for state x.
Replacing x by n and x' by n', these statements also hold for (102).

An important class of processes that can be described by the master equation are one-step or birth-death processes. These are continuous time Markov processes in which only transitions between neighboring states can occur. For such a process the master equation (102) reduces to

$$\frac{dP_n(t)}{dt} = W_{nn+1} P_{n+1}(t) + W_{nn-1} P_{n-1}(t) - W_{n-1n} P_n(t) - W_{n+1n} P_n(t). \tag{103}$$

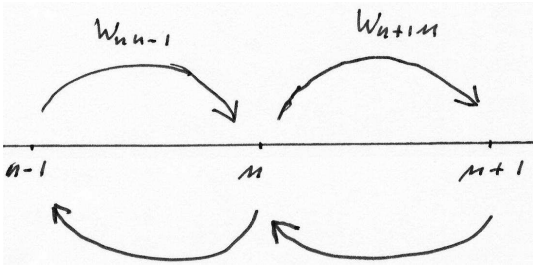

Figure 4: One-step processes

Examples

I. An interesting example is the *Poisson process* in which only transitions in one 'direction' take place with constant transition probabilities, $W_{nn+1} = W_{n-1n} = 0,\ W_{n+1n} = W_{nn-1} = p$. The steps to the right with probability p occur at random times.
The master equation is

$$\frac{dP_n(t)}{dt} = pP_{n-1}(t) - pP_n(t), \quad P_n(0) = \delta_{n0}. \tag{104}$$

The solution is the probability that the particle is at position n at time t

$$P_n(t) = \frac{(pt)^n}{n!} \exp -pt. \tag{105}$$

II. Another well-known continuous time process is $W_{nn-1} = W_{n+1n} = p$ and $W_{n-1n} = W_{nn+1} = q$. This represents a continuous time random walk with transition probability p to make a step to the right and q to make a step to the left. The master equation reads,

$$\frac{dP_n(t)}{dt} = qP_{n+1}(t) + pP_{n-1}(t) - (q+p)P_n(t), \qquad P_n(0) = \delta_{n0}. \tag{106}$$

REMARK
The result (49) can also be obtained as follows. The position of the walker at time t after n steps is

$$\mathbf{r}(t) = \sum_1^n \mathbf{y}_i, \tag{107}$$

$\mathbf{y}_i$ being the ith step. The pdf for the walker being at $\mathbf{r}$ at time t is

$$P(\mathbf{r},t) = \int ... \int d\mathbf{y}_1 d\mathbf{y}_2 .....d\mathbf{y}_n P(\mathbf{y}_n, t_n; \mathbf{y}_{n-1}, t_{n-1}; .....; \mathbf{y}_1, t_1)\delta(\mathbf{r} - \mathbf{r}(t)), \tag{108}$$

this means

$$P(\mathbf{r},t) = <\delta[\mathbf{r} - \mathbf{r}(t)]>$$

$$= \int \frac{d\mathbf{k}}{(2\pi)^d} \exp -i\mathbf{k}\cdot(\mathbf{r}-\mathbf{r}(t)) \int ... \int d\mathbf{y}_1 d\mathbf{y}_2 ..... d\mathbf{y}_n P(\mathbf{y}_n, t_n; \mathbf{y}_{n-1}, t_{n-1}; .....; \mathbf{y}_1, t_1). \tag{109}$$

All steps are IID, so that

$$P(\mathbf{r}, t) = \int \frac{d\mathbf{k}}{(2\pi)^d} \exp(-i\mathbf{k}\cdot\mathbf{r})[\int d\mathbf{y}\ p(\mathbf{y}) \exp i\mathbf{k}\cdot\mathbf{y}]^n.$$

$$= \frac{1}{(2\pi)^d} \int d\mathbf{k}\ \hat{p}^n(\mathbf{k}) \exp(-i\mathbf{k}\cdot\mathbf{r}). \tag{110}$$

References

[1] Jean Perrin, *Les Atoms*, 1ère éd Paris (1913), reprint Flammarion, Paris (1991).

[2] Jean Philibert, *One and a Half Century of Diffusion: Fick, Einstein, Before and Beyond*, The Open-Access Journal for the Basic Principles of Diffusion Theory, Experiemnt and Application.