Technische Universität München
Physik-Department, T30g

Ageing and memory effects in anomalous diffusion processes

Johannes H. P. Schulz

Vollständiger Abdruck der von der Fakultät für Physik der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften (Dr. rer. nat.)

genehmigten Dissertation.

Vorsitzende(r):	Univ.-Prof. Dr. Friedrich C. Simmel
Prüfer der Dissertation:	
	1. Univ.-Prof. Dr. Ralf Metzler, Universität Potsdam
	2. Univ.-Prof. Dr. Martin Zacharias

Die Dissertation wurde am 09.01.2014 bei der Technischen Universität München eingereicht und durch die Fakultät für Physik am 17.02.2014 angenommen.

Summary

The use of probabilistic methods has a long tradition in the physical sciences. The idea is to describe high-dimensional, fast dynamics not by deterministic, overwhelmingly complex equations of motion, but instead in terms of simple statistical models. The prototypical example is the collective behaviour of molecules of an ideal gas, which was the main object of studies for the pioneers of the *molecular theory of heat*. Their combined efforts built the fundament on which rests now the discipline of Statistical Mechanics. It provides wide access and powerful tools for theories of all forms of matter, capable of explaining phase transitions, thermodynamic laws and material properties.

In the same spirit, the theory of stochastic processes tries to predict the behaviour of individuals which are in part affected by deterministic forces, but also coupled to the hectic, complex dynamics of a surrounding environment. Thus, probabilistic aspects are introduced into the classical, continuum theories of transport and diffusion. The paradigmatic role that the ideal gas plays in Statistical Mechanics, is taken by *Brownian motion* amongst diffusion models. The theory of Brownian motion has been studied and extended continuously during the past century. It turned out to provide an extremely versatile approach to model all kinds of fluctuating systems. Today, the applications span all fields of sciences, including biology, chemistry, engineering, geology, finance, and the computer and social sciences.

However, within roughly the last ninety years, it became clear that the standard models of diffusion fail to describe the phenomena observed in highly disordered environments, like amorphous semi-conductors, turbulent fluids, biological cells, glasses, porous surfaces or stock markets. The alternative models invented to replace the standard techniques fundamentally depart from the standard Brownian motion assumptions, and are therefore collected under the umbrella term *anomalous diffusion* processes.

The analysis of such processes on a theoretical basis is the major scope of this thesis, which is devised into four parts. Chapter 1 outlines the historical development of stochastic process theory. Previous knowledge on stochastic theory is helpful, but by no means necessary.Several standard methods and concepts are being introduced along the way, such as random walks, Langevin equations, Fokker-Planck equations and central limit theorems.

In chapter 2, three paradigmatic anomalous diffusion models are presented. The *Lévy flights* are distinct by their scale-free, discontinuous trajectories and fractal exploration of space. *Continuous time random walks* are characterised by scale-free sticking or caging times, resulting in several peculiar phenomena such as weak-ergodicity breaking or ageing. The *fractional Brownian motion* is a model for long-ranged memory, its behaviour ranging from vivid, anti-persistent jiggling to seemingly biased motion going in cycles of arbitrary duration. The distinct properties of these anomalous processes are

highlighted and contrasted to the standard Brownian motion.

An attempt to unify these anomalous diffusion processes is made in chapter 3. The study focuses on the interplay of the various "anomalies" and on ways to distinguish and analyse them. Moreover, a specific model ingredient is added, which has not been considered in previous studies, namely both scale-free and long-range correlated relaxation dynamics. It turns out that memory effects in temporal and spatial relaxation dynamics can mingle and mask each other. Thus the analysis in terms of simple time scalings or propagator functions can be misleading.

Complimentarily, chapter 4 is concentrated on a single anomalous diffusion process, the continuous time random walk. Its ubiquitous ageing features are emphasised. The term ageing is used whenever the properties of a process depend on the time of its observation. Continuous time random walks exhibit remarkable ageing effects, such as the splitting of random walker populations in a mobile and a seemingly immobilised fraction. At the core of these peculiar phenomena lies the scale-free *ageing renewal process*. Reviewing the anomalous diffusion in the light of the more fundamental renewal theory will be crucial to extending the methods and results beyond the class of diffusion processes, such as quantum mechanical two-state systems.

Contents

1 Stochastic modelling of diffusion phenomena – a brief history

1.1 The random walk's walk to fame

In 1905, Nature published a letter from the English mathematician Karl Pearson, who addressed "The Problem of the Random Walk" [1] to the journal's audience:

> "A man starts from a point O and walks l yards in a straight line; he then turns through any angle whatever and walks another l yards in a second straight line. He repeats this process n times. I require the probability that after these n stretches he is at a distance between r and $r + dr$ from his starting point, O.
>
> The problem is one of considerable interest, but I have only succeeded in obtaining an integrated solution for two stretches. I think, however, that a solution ought to be found, if only in the form of a series in powers of $1/n$, when n is large."

The *random walk* was thus introduced by Pearson to the physics community as an abstract gedankenexperiment. The application Pearson immediately had in mind was a random migration problem. During his long and fruitful collaboration with zoologist Walter F. R. Weldon, he got interested in quantitatively analysing the spreading of a mosquito infestation in a forest [2].

But the theoretical concept turned out to be extremely versatile and was adopted to model various phenomena throughout different scientific fields. It proved to be "of considerable interest" indeed.[1] Thus, the desired solution in the limit of large n was provided by no other than Nobel laureate-to-be John William Strutt, Lord Rayleigh, who identified the analogy with a subject from a seemingly unrelated discipline. In the course of his own acoustics studies, he considered [4] the superposition of elementary sound waves $x_j(t)$, with common frequency ω and amplitude a, but varying phases ϕ_j. Let $x_j(t) = \mathrm{Re}[z_j e^{i\omega t}]$ represent the time evolution of an individual wave, where the complex number $z_j = a e^{i\phi_j}$ incorporates both the (real valued) wave amplitude and phase shift. Adding n such independent waves produces a signal $X_n(t) = \sum_{j=1}^{n} x_j(t) = \mathrm{Re}[Z_n e^{i\omega t}]$, with $Z_n = \sum_{j=1}^{n} z_j = A_n e^{i\Phi_n}$. If the phases ϕ_j are uniformly random and independent, then Z_n is exactly Pearson's random walk in the complex z-plane. The single-wave amplitude a corresponds to the walking distance l, random phases ϕ_j are the random turning angles of the walker, and the total wave amplitude A_n is the distance r from the

[1] The list of applications of random walk theory presented here is far from complete. For further reading, consult the excellent introductory chapter of Hughes' book [3], from which many of the following examples and references are borrowed.

origin after covering n stretches. Rayleigh was able to derive the desired asymptotics, and answered Pearson's inquiry only one week after its publication [5]:

"If n be very great, the probability sought is

$$\frac{2}{n}\exp\left(\frac{r^2}{n}\right)\, r\, dr.\text{ "} \tag{1.1}$$

Similar ideas later entered the analysis of waves of different physical origin, including oscillatory electrical signals [6], and coherent laser light [7, 8, 9, 10] or radar waves [11] scattered at irregular plane surfaces. Several modifications of the model were introduced as needed, such as replacing the fixed walk distance l by a random one, in order to describe microwave radiation in the sea [12]. Concurrently, the accuracy of Eq. (1.1) was improved in terms of power series for moderately large n [13].

Further examples in the biological field, from which Pearson originally drew his motivation from, are nowadays to be found also in the behaviour of microscopic living organisms. Some bacteria use *flagella* to swim through a fluid medium. Their motion pattern was observed to be piecewise straight [14], and thus can be approximately described by a three-dimensional random walk, or *random flight*, with random travel lengths. However, all turning angles should not be considered equal, when the organism can feel chemical or temperature gradients. In this case, the random walk becomes locally biased [15]. The migration patterns of larger species are studied mainly in the context of foraging behaviour, and several modifications and refinements of the original random walk idea were made to model, compare and predict animal search strategies [16, 17, 18, 19].

The distance r to the origin of the random walk is not always the prime quantity of interest. The *ideal chain* (or *freely-jointed chain*) is the simplest model to describe synthetic or natural polymers [20, 21]. A polymer is a large molecule that consists of a large number of aligned, identical elements, the *monomers*. The ideal chain model assumes that monomers are approximately line-shaped, rigid objects. The joints between consecutive monomers are taken to allow for force- and torque-free rotation. Hence, a single conformation of the ideal chain, linearly composed of n monomers of length l, corresponds to a single random flight path generated from n flights of distance l. When the polymer is placed in a solvent, the end-to-end distance of the polymer (as represented by the random flight coordinate r) is a relatively uninteresting quantity, as it is rather difficult to access experimentally. Rather, a statistical physics analysis naturally brings up the task of counting the number of conformations (i.e. the number of random flight paths) which are compatible with certain geometric constraints (like selected bonds sticking to a surface or spatial confinement). Another statistical quantity of interest is the *radius of gyration*, that is the root mean squared distance of joints from the common center of mass. In order to give quantitatively accurate results, the relatively crude ideal chain approximation has to be modified in various ways, but the basic idea to model polymer conformations in terms of (refined) random walks proved to be an excellent approach to the problem [22].

1.2 Probability theory in physics beyond estimation of errors

Mathematically, the random walk idea was not exactly invented by Pearson. Uncoupled from any real world application, the addition of independent and identically distributed two-dimensional random vectors appears to have been considered before (see citation of Ch.M. Schols in Ref. [23]. The random vectors can be interpreted as the directed walking distances during a random walk. At least in its one-dimensional form, the random walk had been (and still is) studied extensively in probability theory, where it is commonly related to games of chance [24]. The random steps – which in one dimension simply go forwards or backwards with equal probability – are losses or gains of a gambler during a fair game.

While such parallels might seem obvious from a modern perspective – standard probabilistic and combinatorial methods are commonly taught in elementary mathematics and physics classes, as they form the basics of statistical physics – the discipline of "applied probability theory" at the beginning of the 20th century was developed poorly, to say the least. There was nothing like a lively, networking statistics community, at a time when several competing languages were dividing the readers of scientific publications anyway. Considering, in addition, the fast pace at which new ideas and concepts were emerging, it does not come as a surprise that scientists from separate fields were partially studying the same topics in parallel, without taking notice of one another. Pearson apologises [25]:

> "I ought to have known it, but my reading of late years has drifted into other channels, and one does not expect to find the first stage in a biometric problem provided in a memoir on sound."

For sure, he did not expect to read relevant information on this issue in a doctoral thesis on finance either. Thus, the innovative work of the French mathematician Louis Bachelier from 1900 completely escaped Pearson's attention. Likewise, his physicist contemporaries Rayleigh, Einstein and Smoluchowski (more on them below) apparently did not take notice. Several aspects included in the thesis certainly would have been valuable to any of them. Bachelier modeled the rise and fall of price options at the French stock market by a simple (one-dimensional) random walk (although he did not use that term). Even more, he provided probabilistic arguments to demonstrate that the long time statistics are described in terms of a diffusion equation (to be re-derived by different means five years later by Einstein and Smoluchowski) and showed the Gaussian solution (which implies a one-dimensional analogue of Rayleigh's asymptotics Eq. (1.1)).

Karl Pearson in the context of random walk theory should be credited for presenting and popularising a probabilistic model to a physics community that was largely either not aware of it or simply not interested in it. The name "random walk" is colourful and the associated picture is intuitive (as compared to e.g. the addition of sound waves in the complex plane). Pearson later further improved the pictorial quality by making the walking protagonist "a drunken man who is at all capable of keeping on his feet" [25]. The

notion of a *drunkard's walk* or *drunk sailor's walk* can frequently be found in scientific texts to this day and even triggered empirical studies on the topic itself [26].

Pearson and Rayleigh, among others, deserve critical acclaim for promoting the use of probabilistic methods to describe phenomena which are at the heart deterministic in nature. The foundations of what we call now statistical physics were laid by a series of works of James C. Maxwell, Ludwig Boltzmann, Josiah Gibbs and others dating back to 1860 [27]. But the physical theory they aimed to achieve, the *kinetic theory of heat*, was far from complete and subject to extensive debates. Its critics' reluctance to admit to such a concept was not only based on physical objections, but in part originated in rather philosophical concerns. On the one hand, a positivistic attitude lead some to refute the necessity for postulating the existence of atoms or molecules in the first place. At a time when a successful theory of thermodynamics, resting on firm axiomatic grounds, already existed, why hypothesise an object which can not actually be "seen" nor its direct properties be measured otherwise? On the other hand, some may have felt uncomfortable with the implied statistical doctrine, as being used to approach physical problems with a stringent, deterministic way of thinking. The traditional realm of probability theory in physics was error estimation – from where actually Maxwell borrowed the basic formulae.

The discoveries at the beginning of the 20th century should drastically change the physicists' scientific and philosophical perspective. The need to develop advanced statistical tools and mathematical foundations became immanent when quantum theory appeared, mathematical finance emerged, and both the atomic theory and the kinetic theory of heat were fully established and confirmed. The latter experienced a decisive breakthrough, when Maxwell's and Boltzmann's ideas were extended to theoretically describe, in an experimentally verifiable manner, the phenomenon of *Brownian motion*. From a history point of view, the contributions that random walk theory made in this context, are maybe the most valuable ones.

1.3 Molecules and heat: probably related

The prime assumption of the kinetic theory of heat is that matter is not continuous, but composed of small elemental particles, which constantly exert an agitated motion. The properties of a macroscopic substance are then interpreted as the collective effect of microscopic particle dynamics: the pressure onto a wall of a confined volume stems from constant particle bombardment, and any transfer of heat is actually a transfer of kinetic energies, realised by mixing or collisions of particles. In this respect, the kinetic theory competed with alternative approaches such as the *caloric theory of heat*. The latter suggests that heat is being transferred by means of a *caloric*, a fluid substance flowing from hotter to colder bodies. At the same time, the idea of an aether transmitting electromagnetic waves, was not completely dead. By the turn of the 19th century, the existence of molecules and atoms was considered by most an established fact, or at least

a very practicable working hypothesis.[2] Several physical and chemical studies as the likes of Robert Boyle, Joseph Gay-Lussac, John Dalton and Amedeo Avogadro indicated that any substance indeed consists of an integer amount of molecules, defining the general properties of the substance. Molecules in turn are built from chemically indivisible atoms, and any chemical transition of substances implies the decomposition or merging of atoms into other types of molecules.

However, it was not clear in which exact form energy is stored within and distributed among molecules, how they interact or how large or heavy they actually are. In fact, the atomic theory was, to some extent, compatible with ideas of a caloric, filling the space between molecules as a mean to transfer energies, and thus heat, among them. The kinetic theory for (ideal) gases as developed by Maxwell and Boltzmann made very specific assumptions on the nature and dynamics of molecules: their size is very small as compared to their average distance; interactions between the molecules themselves and between molecules and a confining wall can be described in terms of instantaneous, fully elastic collisions; except during collisions, interactions among molecules are negligible (ruling out the existence of a caloric); due to the very large number of molecules contained within a macroscopic amount of gas, their dynamics are to be treated statistically (implying in particular a purely statistical interpretation of *thermodynamic equilibrium*, the *equipartition theorem* and the *second law of thermodynamics*). Such precise and extensive presumptions were of course difficult to justify without concise experimental evidence. But as Joseph Loschmidt pointed out in 1865, an estimation of the size of a molecule compliant with kinetic theory can be obtained from density measurements on the same same substance in liquid and gaseous state. He arrived at an estimate of the size of an average "air molecule" of the order 10^{-7} cm. Direct observation of the trajectory of an non-isolated object that small and fast is still beyond practical reach to this day.

Yet, there was hope to find a nearly direct experimental validation in a phenomenon signalled by Scottish botanist Robert Brown back in 1828 [29]. Brown examined pollen grains suspended in water under the microscope. Inside, he spotted minute particles which, when ejected from the grain, started to exert a jiggering motion inside the solution. Figure 1.1 provides an image of the *Clarkia pulchella* pollen grain that Brown was studying. A similar behaviour had been reported back in 1785 by Dutch physician Jan Ingenhousz for coal dust particles on the surface of alcohol.[3] Brown carried out his studies further with much care and detail. He asserted that the particle agitation is not specific to the plant's species. Even more, sufficiently fine powders from non-organic matter initiate the same type of movement in water. This ruled out any live-related

[2] Ref. [28] provides an extensive outline of the history of the kinetic theory of heat from a modern perspective. Its author's main views are adopted in the present introduction.

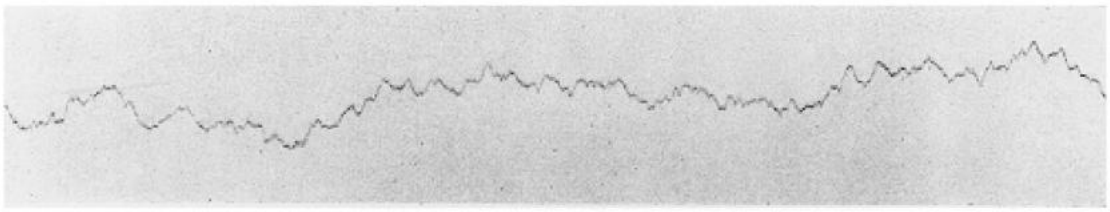
[3] From a historical perspective, a bit of chance appears to be involved in the naming of the "Brownian" motion. Albert Einstein, the man who triggered the breakthrough of the theories of Brownian motion, was working at the time in patent office and was thus, scientifically, relatively isolated. He simply did not know about Ingenhousz' earlier work, and therefore chose to coin the phenomenon "Brownian" motion – although he reportedly even did not have access to Brown's original work.

Figure 1.1: A *Clarkia pulchella* pollen grain bursts open and releases its content into the surrounding water. The larger cell organelles are detectable under the microscope. When entering the solution, they begin an erratic motion, which was first observed by Robert Brown [29]. The image is from a modern high resolution microscopy camera used in a recent experiment [32].

explanation. In the years to come, several scientists contributed to experimentally characterising the phenomenon, including Christian Wiener (1863 [30]) and Louis Georges Gouy (1888). The erratic movement indeed persisted, regardless of how long the time of observation, how low the liquid temperature, how intense the incident light, how long its wave length, and how careful the shielding from any thinkable external source of perturbation. An example of an experimental observation of Brownian motion can be studied in Fig. 1.2, which was recorded roughly 100 years after Brown's original discovery by Eugen Kappler [31].

Such findings rendered the Brownian motion (a term to be coined later by Einstein) very pre-dispositioned towards a kinetic theory interpretation. Water molecules may be too small to be traced individually and thus appear as a fluid continuum under the microscope. But an immersed micrometer-sized particle is large enough to be identified, still might just be small enough to be visibly impacted by the immense number of surrounding water molecules. Such an explanation seemed satisfying and elegant – unfortunately, the first attempts to make quantitative predictions failed to describe the observed reality. This was mainly due to inappropriate theoretical reasoning. For instance, Swiss cytologist Karl von Nägeli – who is commonly credited for considering the first collision model of Brownian motion – argued that the suspended particle has a mass many orders of magnitude larger than a water molecule, and thus each displacement induced by a single collision is way too small to be visible [33, 34]. But the opposing statistical argument for this collective effect, in the words of Belgian Jesuit Ignace Carbonelle, is the following (see quotation in [35]):

> "In the case of surface having a certain area, the molecular collisions of



Registrieraufnahme der Brownschen Bewegung (natürliche Größe).
Direktionskraft $9{,}428 \cdot 10^{-9}$ abs. Einh. Trägheitsmoment: $1 \cdot 10^{-7}$ abs. Einh. Abstand Spiegel–Kamera: 72,1 cm.
Zeitmarke: 30 sec $dx = 1$ mm. a) Atmosphärendruck. Temperatur 13° C

Fig. 5a

Registrieraufnahme der Brownschen Bewegung (natürliche Größe).
Direktionskraft $9{,}428 \cdot 10^{-9}$ abs. Einh. Trägheitsmoment $1 \cdot 10^{-7}$ abs. Einh. Abstand Spiegel–Kamera: 72,1 cm.
Zeitmarke: 30 sec $dx = 1$ mm. b) $1 \cdot 10^{-3}$ mm Hg. Temperatur 13° C

Fig. 5b

Figure 1.2: Trajectories of the Brownian motion of a small mirror, suspended from a fine quartz thread. The erratic angular twist of the mirror is measured by following the motion of a reflected light beam. The high precision data allowed Kappler [31] to obtain the Avogadro number $N = 60.5910^{22} \pm 1\%$. Since the fibre exerts a torsional torque on the mirror, the motion is harmonically bound Brownian. For the two examples shown here, only the pressure conditions are different. In both cases, the motion is overdamped.

> the liquid, which cause the pressure, would not produce any perturbation of the suspended particles, because these, as a whole, urge the particles equally in all directions. But if the surface is of area less than necessary to insure the compensation of irregularities, there is no longer any ground for considering the mean pressure; the inequal pressure, continually varying from place to place, must be recognised, as the law of large numbers no longer leads to uniformity; and the resultant will become more and more apparent the smaller the body is supposed to be, and in consequence the oscillations will at the same time become more and more brisk."

The controversy is exemplary to the case. The traditional physicist's line of argument focused on averaged microscopic quantities such as the mean displacement per collision. The open question was how to properly construct a macroscopic picture, if frequent deviations from the average were to be taken into account. In particular, the precise mathematical mechanism of the inferred *law of large numbers* and its connection to the

central limit theorem (to be detailed below) was not yet part common knowledge. [4] The celebrated efforts of Albert Einstein (1905 [36]) and Marian Smoluchowski (1906 [37]) finally lead to a satisfying probabilistic and quantitative theory of Brownian motion.

1.3.1 Albert Einstein and the diffusion equation

Einstein combined two parallel perceptions of a diffusion phenomenon. First, a large collection of small particles immersed in a fluid can be described in terms of a concentration gradient $f(x,t)$, that is $f(x,t)\,dx$ gives the number of particles contained in a small volume around position[5] x at time t. Upon invoking the principles of kinetic theory, Einstein establishes a law of osmotic pressure for the suspended substance, which is essentially identical to the ideal gas law. Then, along the lines of Fick, Thomson and Fourier, he argues that $f(x,t)$ should be subject to the classical laws of diffusion. In particular, in a dynamical equilibrium, any current induced by global external forces, such as gravity or electric fields, is counterbalanced by a diffusional current $-D\,(\partial f/\partial x)$, expressing the tendency of the substance to fill regions of lower concentration. The *diffusion constant* D has units of $\mathrm{m}^2/\mathrm{sec}$ and indicates the strength of the diffusion current. Einstein further assumes that particles are spherical subject to Stokes friction and uses above-mentioned osmotic pressure law to arrive at the relation

$$D = \frac{R\mathcal{T}}{6\pi\eta r N}, \tag{1.2}$$

where R is the ideal gas constant, $\mathcal{T}$ is the temperature of the fluid, η its friction coefficient, r gives the particle radius, and N is Avogadro's number.

In a second step, Einstein takes an alternative view on the process. The Brownian particles are large enough (order of micrometers) to be studied individually. Einstein now makes the following crucial assumption [36]:

> "It must clearly be assumed that each individual particle executes a motion which is independent of the motions of all other particles; it will also be considered that the movements of one and the same particle in different time intervals are independent processes, as long as these time intervals are not chosen too small.
>
> We introduce a time interval τ into consideration, which is very small compared to the observable time intervals, but nevertheless so large that in two successive time intervals τ, the motions executed by the particle can be thought of as events which are independent of each other."

[4] The first profound mathematical formulation of a central limit theorem for the binomial distribution was by Pierre-Simon Laplace in 1811. Still [38], "It was not until the nineteenth century was at an end that the importance of the central limit theorem was discerned, when, in 1901, Russian mathematician Aleksandr Lyapunov defined it in general terms and proved precisely how it worked mathematically. Nowadays, the central limit theorem is considered to be the unofficial sovereign of probability theory."

[5] In his 1905 paper, Einstein derives essentially one-dimensional formulae. Any external force is assumed to act along the x-axis and displacements are measured in x-direction.

Einstein then considers the probability[6] $\phi(\Delta)$ for any Brownian particle to be displaced by a distance Δ between times t and $t+\tau$. In order to accord with above assumptions, $\phi(\Delta)$ is taken to be the same law for all particles, and to be is symmetric in Δ and independent of time t and past movements. This implies the number of particles in a small volume around x evolves with time according to

$$f(x, t+\tau)\, dx = dx \int_{-\infty}^{\infty} f(x+\Delta, t)\phi(\Delta)\, d\Delta. \tag{1.3}$$

Taylor-expanding both sides with respect to small τ and Δ yields

$$\frac{\partial f}{\partial t} = D\, \frac{\partial^2 f}{\partial x^2}, \tag{1.4}$$

where

$$D = \frac{1}{2\tau} \int_{-\infty}^{\infty} \Delta^2 \phi(\Delta)\, d\Delta. \tag{1.5}$$

Equation (1.4) affirms the original assumption that the concentration gradient indeed evolves according to the classical laws of (Fickean) diffusion. However, in this case the interpretation is based on a single particle picture. The diffusion constant D is related via Eq. (1.5) to the transition probability $\phi(\Delta)$, governing the microscopic particle displacements. Einstein proposes the following method to measure this quantity. After the recording of particle motions, on can spatially shift all trajectories such that the origin of the coordinate system $x = 0$ corresponds to the origin of motion at $t = 0$. By virtue of spatial homogeneity, the transformed particle concentration should still obey the diffusion equation. The initial condition is then[7] $f(x, t=0) = n\,\delta(x)$, where $n = \int_{-\infty}^{\infty} f(x,t)\, dx$ is the total number of suspended Brownian particles. The corresponding solution to Eq. (1.4) is the famous Gaussian distribution, that is

$$\frac{1}{n}\, f(x,t) = \frac{1}{\sqrt{4\pi D t}} \exp\!\left(-\frac{x^2}{4Dt}\right). \tag{1.6}$$

Hence, the diffusion constant can be obtained from empirically establishing the distribution of total particle displacements $f(x,t)$. A more simple and direct approach, as Einstein points out, is to calculate the *mean squared displacement* per particle in the direction of x, i.e. the ensemble average

$$\left\langle x^2(t) \right\rangle = \frac{1}{n} \int_{-\infty}^{\infty} x^2 f(x,t)\, dx = 2Dt. \tag{1.7}$$

[6] Einstein actually avoids the use of the term "probability" (german: "Wahrscheinlichkeit") completely in his 1905 paper. He rather speaks of a "frequency distribution" ("Häufigkeitsverteilung"), that is, in modern terminology, a histogram. Still, he clearly uses the concept of a properly normalised, continuous probability density, and calculates its first and second moment, and he writes down (Eq. (1.3)) what today is known as a Chapman-Kolmogorov equation for a Markov process. His nomenclature might be coincidence or common habit of the time – or it might reflect the author's attempt to appeal to a broad audience who is not completely familiar with probabilistic ideas.

[7] Here, $\delta(x)$ is the Dirac delta function, i.e. a generalised function defined by $\int_{-\infty}^{\infty} g(x)\delta(x)\, dx = g(0)$ for any well-behaved test function $g(x)$. Einstein did not use the delta function, but essentially did use the concept.

The diffusion constant can thus be found by averaging square displacements over a multitude of trajectories of Brownian particles. In combination with Eq. (1.2), this method allows to determine experimentally Avogadro's number N. The latter constituted an essential aspect of the atomic theory (see Perrin's work below).

The originality of Einstein's method lies in his dual approach to the problem: On the one hand, he argues on the level of continuous matter, in terms of concentration gradients, pressure laws and the classical diffusion laws. On the other hand, he uses a probabilistic particle description, analysing the properties of trajectories of individual Brownian bodies. By focusing directly on the position coordinate x at times large as compared to τ, Einstein elegantly circumvented several difficulties of earlier statistical arguments, such as inconsistent estimations of the instantaneous velocity of a Brownian particle or its average displacement per microscopic collision. These delicate issues were later attacked directly by Smoluchowski.

1.3.2 Marian von Smoluchowski and the random walk

Einstein defines the statistical properties of Brownian particle dynamics in terms of the displacement probability $\phi(\Delta)$. He makes certain reasonable assumptions (such as independence of x and t), but he does not exactly specify its functional form, nor does he relate it directly to the dynamics of the surrounding medium. In this sense, the approach presented in 1906 by Smoluchowski [37] is more microscopic, as it explicitly models the effect of single collisions with the invisible fluid or gas molecules. His basic method of choice is the random walk.

Along his derivation, he clears out several precedent misconceptions. For instance, he addresses von Nägeli's concern on the tininess of displacements. The velocity increase of a Brownian particle with a diameter of 10^{-4} cm by a single shock from a light water molecule should be of the order of a mere 10^{-6} cm/sec. This is way below the observed (and observable) Brownian movement characteristics. Although there are many such shocks per second, the net displacement should be zero, according to von Nägeli, due to their unbiased directionality. Smoluchowski argues:[8]

> "This is the same error in reasoning as if a hazard[9] player believed he could never loose an amount larger than the bet of a single throw."

Within n throws of dice, the probability $p(m;n)$ of winning m times is given by

$$p(m;n) = \frac{n!}{2^n\, m!(n-m)!} = \frac{1}{2^n}\binom{n}{m}.$$

[8] English translation and footnote by J. S..

[9] *Hazard* was a popular game of dice in the 17th and 18th centuries. Its rules were relatively complicated, but Smoluchowski uses effectively the simple mechanism of the fair gambling game (and thus of the one-dimensional random walk): the player places a unit bet, and at each throw of the dice there is an equal probability for either winning or loosing the bet.

Having won m out of n rounds, the total earnings are $v = 2m - n$. Since the game is fair, $p(m;n) = p(n-m;n)$. Consequently, the average gain $\langle v \rangle$ is zero, but the average (positive or negative) deviation from zero is

$$\langle |v| \rangle = \sum_{m=0}^{n} |2m-n|\, p(m;n) = 2 \sum_{m=n/2}^{n} \binom{n}{m} \frac{2m-n}{2^n} = \frac{n}{2^n} \binom{n}{n/2},$$

which for large n becomes[10]

$$\langle |v| \rangle \sim \sqrt{\frac{2n}{\pi}}. \tag{1.8}$$

Hence, a gambler who intents to play n rounds should expect to loose or win an amount of the order of $\sqrt{n}$ and should have an appropriate budget at his disposition. Following the random walk analogy, the velocity component v_x in x-direction of a Brownian particle is assumably either increased (gambler wins) or decreased (gambler looses) by a fixed amount at each collision. Then the average (signed) velocity vanishes, $\langle v_x \rangle = 0$, but the average (absolute) speed $\langle |v_x| \rangle$ increases like the square root of n. Within a single second, a suspended Brownian particle experiences about $n = 10^{16}$ shocks (in gaseous air) or $n = 10^{20}$ shocks (in fluid water). In this most simplistic picture, the speed increase per second can hence be of the order of 10^2 or 10^4 cm/sec, respectively. Smoluchowski thus gives a comprehensible demonstration of how an accumulation of random effects, while unbiased, yet can increasingly deviate from the expected zero net effect. In fact, according to the central limit theorem (discussed below), the observed square root dependence on the number of random effects n, if n be large, is something quite general.

While such considerations resolve von Nägeli's concern, this simple random walk must still be rejected as a complete model for Brownian motion. As Smoluchowski points out, Eq. (1.8) indicates an unbounded increase of speed, and thus, an unbounded increase of kinetic energy. The essential flaw of the above theory is that it does not account for the velocity dependence of collision probabilities. In reality, the faster the Brownian particle moves in one direction, the more frequent become collisions with solvent molecules on that side. Hence, collisions drive the Brownian motion, yet they also induce a friction mechanism. Ultimately, the velocity distribution in general and the average speed in particular will attain a certain limiting value. According to Maxwell's idea of a kinetic theory, this is exactly what defines the thermodynamic limit. More precisely, in an equilibrated system, all particles, regardless of size, shape or weight, should bear the same mean kinetic energy. Assuming that the Brownian particles are spherical with mass M and velocity[11] $\boldsymbol{V}$, and the water molecules are point-like with mass m and velocity $\boldsymbol{v}$, this implies

$$\left\langle V^2 \right\rangle = \frac{m}{M} \left\langle v^2 \right\rangle. \tag{1.9}$$

[10] To show this, use Stirling's formula, i.e. $n! \sim (n/e)^n \sqrt{2\pi n}$ for large numbers n.

[11] Bold mathematical symbols are used here to indicate vectors in Euklidean space, e.g. $\boldsymbol{v} = (v_x, v_y, v_z)^T$. Equivalent non-bold characters stand for respective Eukidean norms, $v = |\boldsymbol{v}| = \sqrt{v_x^2 + v_y^2 + v_z^2}$.

Based on this relation, Smoluchowski roughly estimates the average speed of a Brownian particle suspended in water under equilibrium conditions as $\langle V\rangle \approx \sqrt{\langle V^2\rangle} \approx 0.4\,\mathrm{cm/sec}$. Such value, he notes, is much faster than any velocity measured in experiment; a fact that unfortunately brought initial scepticism over the kinetic theory approach. The error is, again, an error in interpretation. Smoluchowski explains:[12]

> "What we observe is just the average position of a particle moving with such speed $[V]$, but changing its direction 10^{16} to 10^{20} times per second. Its centre of mass follows an immensely more complicated, zig-zag shaped path, whose straight line segments are much smaller in magnitude than the dimensions of the particle itself [...]."

The straight point-to-point displacement distance $\delta X = |\boldsymbol{X}(t+\delta t) - \boldsymbol{X}(t)|$ is not even an approximate measure for the real length of the trajectory, as it utterly underestimates its complexity. Any attempt to calculate the instantaneous speed V from $\delta X/\delta t$, with experimental time resolution δt being way beyond $10^{-16}\,\mathrm{sec}$, is bound to fail.

Smoluchowski substantiates his argument by introducing a more elaborate random walk scheme. The position of the Brownian particle after the nth collision with a surrounding fluid or gas molecule is labelled $\boldsymbol{X}(n)$. Motion starts at $X(0) = 0$ in a random direction. For the model to be analytically tractable, Smoluchowski makes several simplifying assumptions. Some are obvious (mostly justified by the asymmetric ratio of masses, $M \gg m$); others are rather subtle, see the original text [37] for details. In short, the model definition is as follows. (i) Collisions with solvent molecules occur at fixed time intervals, any randomness of impact times is neglected. (ii) In between any two collisions n and $n+1$, the motion is straight with constant velocity $\boldsymbol{V}(n)$. The speed of the Brownian particle is a priori dictated by the equipartition theorem (1.9), that is $V(n) \equiv V = \langle V\rangle = \sqrt{\langle v^2\rangle\, m/M}$. In principle, the equipartition theorem is a statistical one and allows for fluctuations around the average, but the latter are neglected. It follows from (i) and (ii) that the displacement vector is given by $\boldsymbol{\delta X}(n) = \boldsymbol{X}(n+1) - \boldsymbol{X}(n) = \boldsymbol{V}(n)/\nu$, where ν is the collision rate. Note that down to this point, the model is a three-dimensional version of Pearson's random walk, since the Brownian particle covers fixed, nonrandom distances $l = \delta X = V/\nu$ between any two consecutive collisions. However, in Smoluchowski's random walk the new direction of movement after each step (collision) is not entirely random. Instead, (iii) the turning angle $\epsilon(n)$, defined through $\cos[\epsilon(n)] = [\boldsymbol{V}(n-1)\,\boldsymbol{V}(n)]/V^2$, is again assumed to be constant and nonrandom, $\epsilon(n) \equiv \epsilon$. Following von Nägeli's argument, the effect of an individual collision should indeed assumed to be small, since $M \ll m$. Consequently, the turning angle is small and to first order proportional to the ratio of momenta $(mv)/(MV)$. Estimating $v \approx \sqrt{\langle v^2\rangle}$ and using V as defined in (ii), this yields $\epsilon = \mathcal{O}\Big(\sqrt{m/M}\Big)$.

Summarising (i) through (iii), after each collision, the tip of the displacement vector $\boldsymbol{\delta X}(n)$ is required to land on a definite small circle centred around $\boldsymbol{\delta X}(n-1)$, see Fig. 1.3.

[12] Translation, addition and omission by J.S.

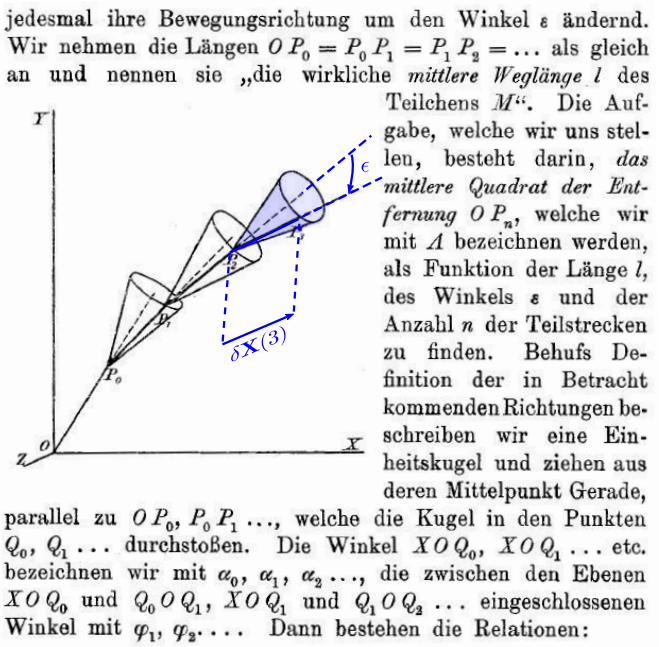


Figure 1.3: Schematic of Smoluchowski's persistent random walk in three dimensions. Starting at $\boldsymbol{X}(0) = 0$, the distance length of displacements is fixed, $\delta X(n) = l$ for all n. The first travel direction is fully random. Consequent turning directions are constrained by a fixed, constant polar angle ϵ. The azimutal angle is uniformly random, and independent from step to step. Figure taken from Ref. [37], with amendments by J. S. marked in blue.

In contrast to the fully random zig-zag pattern of the Pearson walk, Smoluchowski introduces an element of *persistence*: a Brownian particle tends to keep its direction of motion after collisions, which is a manifest of its heavy inertial mass. If ϵ is very small, then even after a moderately large number of collisions, the observed path will appear to be approximately straight. Thus, Smoluchowski argues, emerges the *apparent mean free path* λ, which can be several orders of magnitude larger than the *real mean free path* l.

Smoluchowski gives a quantitative example by considering the mean squared displacement $\left\langle \boldsymbol{X}^2(n) \right\rangle$ of a Brownian body suspended in a gas. Its spatial dimensions in this case are smaller than the mean free path of a gas molecule. Hence, following Boltzmann's original kinetic theory ideas, the effects of consecutive collisions can be assumed to be statistically independent. This completes the definition of the Smoluchowski walk: (iv) Randomness of collisions enters through the azimuthal angles of $\boldsymbol{\delta X}(n)$ along the circle around $\boldsymbol{\delta X}(n-1)$. These angles are drawn independently from a uniform distribution on $[0; 2\pi]$. Identifying the time coordinate as $t = n/\nu$, the results of Smoluchowski's calculations [37] can be written as

$$\left\langle \boldsymbol{X}^2(t) \right\rangle = \begin{cases} (Vt)^2, & \text{for } \nu^{-1} \ll t \ll \tau, \\ 6Dt, & \text{for } t \gg \tau, \end{cases} \tag{1.10}$$

where

$$\begin{aligned} \tau &= \frac{1}{2\nu[1-\cos(\epsilon)]} \approx \frac{1}{\nu\epsilon^2} = \mathcal{O}\left(\frac{M}{\nu m}\right), \\ \lambda &= V\tau = \lambda_x\sqrt{3}, \\ D &= \frac{\lambda_x^2}{2\tau} = \mathcal{O}\left(\frac{\left\langle v^2 \right\rangle}{\nu}\right). \end{aligned} \tag{1.11}$$

These findings are remarkable in several ways. First, Eq. (1.10) underpins Smoluchowski's scale separation argument. Motion appears to be linear, $\sqrt{\langle X^2 \rangle} = Vt$, on temporal scales smaller than τ. Conversely, the chaotic nature of the process becomes noticeable on large scales. This regime is a diffusive one: the mean squared displacement equals the behaviour described by Einstein's Eq. (1.5).[13] But Einstein did not include persistence of motion: Eq. (1.3) explicitly states the independence of past and future displacement vectors. In fact, at large times, Smoluchowski's random walk mimics a non-persistent Pearson walk with effective path lengths $l = \lambda = V\tau$. Such analogy comes by virtue of the central limit theorem (explained below, but in this form not available to Einstein nor Smoluchowski at the time).

Thus, Smoluchowski establishes a connection between the experimentally unobservable small-scale motion characteristics (collision rate ν, real free path V/ν) and the measurable diffusion properties (such as the diffusion constant D). For instance, Eq. (1.11) predicts that D should be independent of the mass ratio m/M. As Smoluchowski explains, a Brownian particle with a heavy mass maintains a low average speed at thermal equilibrium, but is highly persistent in keeping its direction of motion; these two factors apparently balance out each other.

1.3.3 Jean Babtiste Perrin and Avogadro's number

Einstein had presented the satisfactory probabilistic approach to the modelling of particle diffusion. Smoluchowski had demonstrated that random walks can provide the missing link to kinetic collision models. This established a largely consistent theory of Brownian motion, which had yet to be validated experimentally. The outcome of such experiments was published by French physicist Jean Babtiste Perrin [35] in 1909. Together with his students Chaudesaigues and Dabrowski, Perrin carried out several extensive tests to verify the kinetic theory in general, and Einstein's considerations in particular, by study of Brownian motion.

The major focus of Perrin's 1909 paper lied on the determination of Avogadro's number N. The latter is defined as the number of molecules contained within two grams of

[13] Note that under isotropic conditions, $\left\langle \boldsymbol{X}^2(t) \right\rangle = \left\langle X^2(t) \right\rangle + \left\langle Y^2(t) \right\rangle + \left\langle Z^2(t) \right\rangle = 3\left\langle X^2(t) \right\rangle$.

hydrogen.[14] According to Avogadro's principle, this number should be the same for any other gas occupying the same volume as the hydrogen standard, and being maintained in a steady state under the same pressure and temperature conditions. Such principle arises naturally in the kinetic theory of heat, and consequently, N appears as a universal constant in various formulae from atomistic theories.[15] Several estimations of its value were already available, as calculated from, e.g., the ideal gas law, counting of α-decays, elastic scattering of light, electric mobility measurements and black-body radiation. However, the object was to confirm the constant's universality, which required a certain amount of precision. Perrin achieved in calculating N with unprecedented accuracy (roughly 19% off from modern measurements).

Both his 1909 paper and his 1913 book "Les Atomes" [39] are still recommended reading today, as they provide interesting historical insights, but also impart a profound understanding of the theory. Perrin exhibits the basic principles and implications of the kinetic theory up to this point, collects the various experimental methods to measure related physical quantities (e.g. indirect measuring of the atomic radius or elementary electric charge, and existing estimations for Avogadro's number N), discusses recent experimental approaches (a first series of experiments had seemed to disprove Einstein's formulae), and finally he describes in much detail his own careful and profound experimental procedures. Beyond following Einstein's suggestion to calculate N, Perrin indeed confirms the Gaussian nature of the displacement distribution (1.6), verifies experimentally the applicability of Stokes' law of friction for such small particles, establishes a barometric formula for Brownian particles subject to gravitation, and experimentally confirms equipartition of energies (including even rotational contributions); he uses particles of different materials, ranging in diameter from tenths to tens of micrometers, thus indeed showing universality. Perrins complete and detailed studies won him the Nobel prize in 1926 and managed to convert even the harshest critics of kinetic theory [40].

1.4 Crafting the tools of stochastic modelling

The introduction of these novel probabilistic concepts at the beginning of the 20th century sparked rapid developments of analytic methods on several grounds. Such research was finally conducted in a parallel, yet orchestrated way. Various scenarios raising from physics, mathematics and, somewhat delayed, also finance questions were now put on the common stage of *stochastic process theory*.

A stochastic process should be seen as a collection of random variables, $\{X(t)\}_{t\in I}$,

[14] Perrin's definition of Avogadro's number N differs from the modern definition of the *Avogadro constant* N_A. The latter is the number of constituent particles in one *mole* of a given substance. In turn, 12 grams of pure carbon-12 are defined to contain exactly one mole (1 mol). Hence, N and N_A slightly differ in their numerical value. Furthermore, N is dimensionless, while N_A has a base unit mol^{-1}.

[15] Some of these theories were still "under construction" at the time of 1909, in particular the quantum theory and the corpuscle theory of light. Hence, the partial results and formulae Perrin quotes are not presented here. Also note that, when the latter appear in their modern form, Avogadro's constant usually enters indirectly via the Boltzmann factor $k_B = R/N_A$.

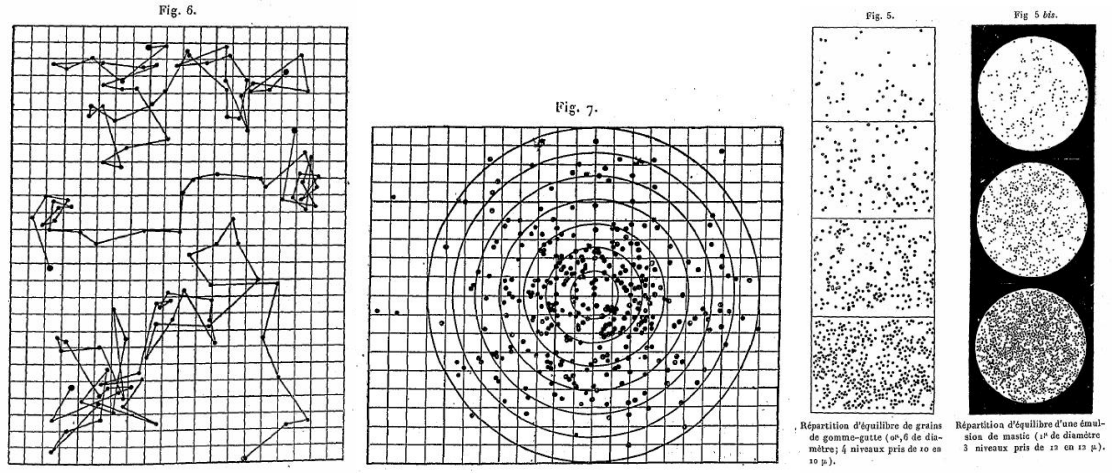


Figure 1.4: Studies on Brownian motion by Jean-Babtiste Perrin [35]. *Left:* Trajectories of a Brownian particle (projected onto a plane), as measured in time intervals of 30 sec. *Center:* All displacements (straight line segments of the rectilinear trajectories) of all trajectories are shifted to the origin. The figure illustrates the Gaussian nature of the distribution of travel lengths, predicted by Einstein. *Right:* Two snapshot pictures of a dispersion of Brownian particles under the influence of the gravitational field along the vertical axis. Due to thermal agitation, Perrin argues, the particles do not sediment, but attain a stationary, barometric (i.e. exponentially decaying) spatial distribution.

indexed by what is commonly referred to as *time parameter* t. The index set can either be discrete (e.g. for a random walk, the obvious choice are the natural numbers, $I = \mathbb{N}_0$) or continuous (for Brownian motion, set $I = \mathbb{R}$). The state of the process at fixed time t is represented by the random variable $X(t)$. The sample space of the latter can be defined as appropriate; for instance, for three-dimensional Brownian motion one should actually consider random vectors $\boldsymbol{X}(t) \in \mathbb{R}^3$. For any finite set of measurement times $(t_1, ..., t_d) \in I^d$, $d \in \mathbb{N}$, one can then draw samples $X(t_1) = x_1, ..., X(t_d) = x_d$, which constitutes a set of *sample points* of a *realisation* or *path* or *trajectory* of the stochastic process $\{X(t)\}$. It should be noted that such definition is straight-forward only for a finite set of sample points. The question of how to find a proper notion of the "probability" of a full path $t \mapsto x(t)$, or of how to probabilistically describe analytic path properties, is highly nontrivial.

1.4.1 The Wiener process

On the mathematics side hence arises naturally the task to consistently and uniquely define specific processes, to classify them, and to define and study properties of their sample paths. Such efforts culminated in the works of Norbert Wiener [41].He studied

a stochastic process $\{B(t)\}_{t\in\mathbb{R}_0^+}$, today named in his honour the *Wiener process*, which is uniquely defined by the following three properties: (i) $B(0) = 0$; (ii) a sample trajectory $B(t)$ is almost surely[16] continuous everywhere; (iii) increments $B(t_2) - B(t_1)$ have a Gaussian distribution with mean 0 and variance $|t_2 - t_1|$, and they are mutually independent for any non-overlapping time intervals $[t_1, t_2]$, $[t_1', t_2']$, etc. This is in fact a mathematical treatment of the (one-dimensional) Brownian motion described by Einstein. (The Wiener process uses arbitrary space and time units, standardised such that $D = 1/2$. To recover Einstein's equations, use a scaled, non-standard Brownian motion $\sqrt{2D}\,B(t)$.) Point (iii) reflects Einstein's requirement that "in two successive time intervals τ, the motions [...] are independent of each other" [36]. Gaussianity is implied by Eq. (1.6), and the variance property is Eq. (1.7). More generally, the position coordinate scales[17] with time as $X(t) \sim t^{1/2}$.

Wiener proved the existence of this process in that he constructed an appropriate space of sample paths, equipped with a proper probability measure, satisfying (i)-(iii). On this basis he derived several path properties; compare to the sample trajectories generated numerically in Fig. 1.5. Arguably the most important feature is that any path of a Wiener process, while being continuous by definition, is non-differentiable everywhere with probability 1. Such is the mathematical precise way of expressing the hopelessness in determining velocities of Brownian particles. Conversely, attempts were made to give a senseful meaning to the integral of a stochastic process, which ultimately led to the development of *stochastic calculus* [42] by Kiyoshi Itō [43, 44], and later Stratonovich's contributions, which have become specifically important for physics applications [51].

1.4.2 The Langevin equation

The desire to invent such mathematical machinery arose with the concurrent discussion of *stochastic differential equations* in the physics community. The first to propose such concept was French physicist Paul Langevin [45], who described Brownian motion in

[16] An event A is said to happen *almost surely* if $\Pr\{A\} = 1$. This does not imply that the event must always happen, when performing a certain number of trails. For example, draw some random integer number and let A be the event that the number is different from 5. Although drawing 5 is possible in principle, the probability for A is one, since there is an indefinite amount of different outcomes. The concept becomes particularly relevant when studying asymptotic convergence of a property A_n, referring, for example to a characteristic of a stochastic path $\{X(n')\}_{n'\leq n}$. One says property A_n holds *asymptotically almost surely* for large n, if $\lim_{n\to\infty}\Pr\{A_n\} = 1$.

[17] Let $\{X(t)\}_{t\in\mathbb{R}^+}$ be some stochastic process and let $H > 0$ be a constant. The notation $X(t) \sim t^H$ indicates a scaling relation with respect to time: the random variable $X(t)$ has the same distribution as $t^H X(1)$. For the associated PDF $p(x;t)$, this equivalent to writing $p(x;t) \equiv t^{-H}p(xt^{-H};1)$. It should be noted however, that for most cases considered in this thesis, an actually stronger statement holds: the motions in question are self-similar with index H, that is, for any constant $c > 0$, the processes $\{X(t)\}_{t\in\mathbb{R}^+}$ and $\{c^{-H}X(ct)\}_{t\in\mathbb{R}^+}$ have the same finite-dimensional distributions.

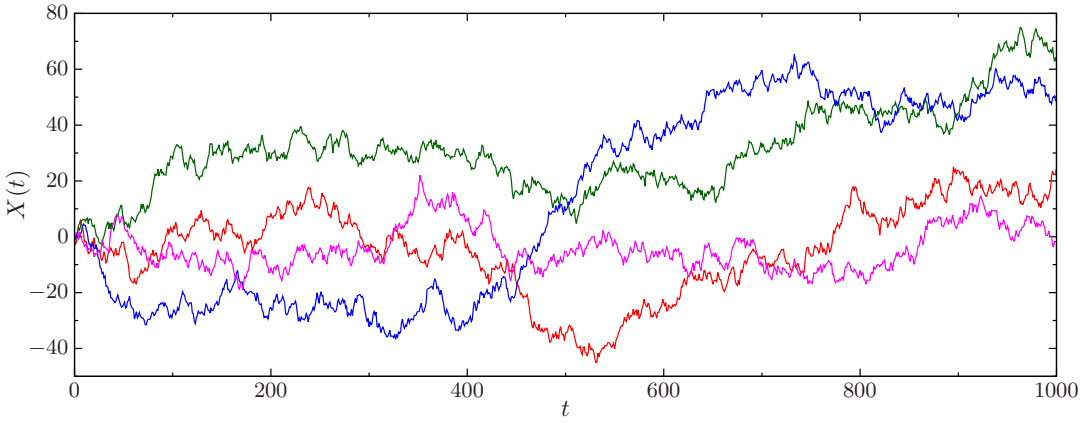


Figure 1.5: Sample paths $X(t)$ of a Wiener process, that is, the mathematical model of ordinary *Brownian motion*, generated by numeric simulation. The sample paths are erratic but continuous and at no time one can determine a favoured direction. Locally, the simulated trajectories agree remarkably well with the real Brownian motion as observed, e.g., by Kappler, see Fig. 1.2. However, due to the torsional stalling force, Kappler's trajectories globally tend to drift towards the equilibrium position.

terms of the equation of motion

$$\begin{aligned} \frac{dX}{dt} &= V(t), \\ m\frac{dV}{dt} &= -\gamma V(t) + \xi(t), \end{aligned} \tag{1.12}$$

where now m denotes the mass of the Brownian particle and γ the linear friction coefficient with respect to the surrounding medium. At first glance this equation appears as ordinary manifest of Newton's second law, with the familiar inertia term on the left hand side and the dissipative friction force on the right. The time-dependent force $\xi(t)$ is meant to represent the momentum transfer by collisions with external fluid or gas molecules. The collision rate is again assumed to be much faster than the typical inverse "time scale dt" defining the scope of Eq. (1.12). The definition of $\xi(t)$ therefore comes quite out-of-the-ordinary as a stochastic process, rather than a deterministic function of time. Such randomness is of course passed on to the motion $X(t)$ (respectively $V(t)$), which should be characterised by its probabilistic traits. Langevin's equation thus evaded classical methods of analysis, and engaged physicists and mathematicians alike in the course of the coming decades [43, 46, 47, 48].

The following treatment is akin to presentations found in modern textbooks [49, 50, 51]. First, one needs to specify the collision process $\xi(t)$. If the environment is being helt at thermal equilibrium conditions, one can readily invoke the principles of kinetic

theory which already inspired Einstein's and Smoluchowski's theories. The surrounding molecules then obey the Maxwell-Boltzmann distribution laws. It follows in particular that momentum transfer is unbiased on average; the Brownian particle is equally likely to be kicked in one direction or the other. Moreover, equilibrium laws are stationary, implying that the processes $\xi(t)$ and $\xi(t+t_0)$ are statistically equivalent for any time shift t_0. Finally, any two collisions are assumed to be statistically independent events. These considerations amount to requiring[18]

$$\begin{aligned} \langle \xi(t) \rangle &= 0 \\ \langle \xi(t)\xi(t') \rangle &= 2D\gamma^2\,\delta(t-t'), \end{aligned} \tag{1.13}$$

for all times t, t'. The occurrence of a diffusion constant D will become clear shortly; for now, it is merely a positive constant whose magnitude determines the amplitude of fluctuations in the collision or *noise process* $\xi(t)$. Later in 1945, Wang and Ornstein [47] demonstrated that it is highly reasonable (today, it is common practice), to define $\xi(t)$ as a *Gaussian process*. This is consistent with the Gaussian nature of the Maxwell-Boltzmann momentum distributions of bombarding molecules, but also has further important implications.[19] The distributional properties of the *Gaussian white noise*[20] $\xi(t)$ are thus uniquely defined in terms of the drift and autocorrelation function[21] in Eq. (1.13).

Next, the stochastic differential equation (1.12) should be provided a reasonable interpretation. One way to do this is to first solve it while treating the noise term formally as an analytic function. The general solution thus obtained can then be used [49] to calculate first and second moments, autocorrelation functions or cross-correlators of $X(t)$ and $V(t)$. Relations (1.13) have to be applied at this point. The solution depends on initial values $X(0)$ and $V(0)$, which are also allowed to be random. Of particular interest is the

[18] Let A and B be two continuous random variables with probability density functions $p_A(x)$ and $p_B(x)$, respectively. Also define their joint probability density function $p_{AB}(x,y)$. A and B are said to be *(statistically) independent*, if $p_{AB} \equiv p_A\, p_B$. Two independent random variables are also uncorrelated, meaning that $\langle AB \rangle = \int\int xy\, p_{AB}(x,y)\, dx\, dy = \int x\, p_A(x)\, dx \int y\, p_B(y)\, dy = \langle A \rangle \langle B \rangle$. The converse holds if, in addition, A and B are jointly Gaussian.

[19] $\{X(t)\}_{t\in I}$ is called a *Gaussian process*, if the distribution of any finite-dimensional random vector $\boldsymbol{X} = (X(t_1), ..., X(t_d))$, $d \in \mathbb{N}$, $t_i \in I$, is *multivariate Gaussian* [52]. A Gaussian process is uniquely defined in terms of its mean drift $\langle X(t) \rangle$ and its correlation function $\langle X(t_1)X(t_2) \rangle$. This is due to *Isserli's theorem*. Defining the zero-drift Gaussian process $\{Y(t)\} = \{X(t) - \langle X(t) \rangle\}$, one has odd moments $\langle Y(t_1) \dots Y(t_{2d-1}) \rangle = 0$, and even moments $\langle Y(t_1) \dots Y(t_{2d}) \rangle$ can be calculated recursively from the covariance function (see e.g. [49]).

[20] "White" noise refers to the constant power spectral density $S(\omega) = 2\int_{-\infty}^{\infty} e^{-i\omega t} \langle \xi(0)\xi(t) \rangle\, dt = 2\gamma k_B \mathcal{T}$. This is analogous to an ideal "white" light, where all wave lengths have equal contributions. Because of the Dirac δ-function in the correlation function, the noise is also referred to as δ-correlated noise.

[21] Mathematically, the *covariance* of two random variables A and B is defined as $\mathrm{Cov}[A,B] = \langle AB \rangle - \langle A \rangle \langle B \rangle$. It is related to the *correlation* via $\mathrm{Corr}[A,B] = \mathrm{Cov}[A,B]/\sqrt{\sigma_A^2 \sigma_B^2}$ in terms of the variances $\sigma_A^2 = \langle A^2 \rangle - \langle A \rangle^2$ and σ_B respectively. Confusingly, the common physicist's habit its to call $\mathrm{Cov}[\cdot]$ a correlation and refer to $\mathrm{Corr}[\cdot]$ as *normalised correlation*. The present work uses the physics terminology. In any case, the terms *covariance/correlation function* of a stochastic process $X(t)$ refer to respective functions of two times t_1, t_2, constructed by setting $A = X(t_1)$ and $B(X(t_2))$.

equilibrium solution, which is characterised by a time independent (Maxwell-Boltzmann) distribution of velocities. This is realised by the initial condition

$$\langle V(0)\rangle = 0 \equiv \langle V(t)\rangle\,, \qquad \left\langle V^2(0)\right\rangle = \frac{\gamma D}{m} \equiv \left\langle V^2(t)\right\rangle. \tag{1.14}$$

Notice that for a Brownian particle with one mechanical degree of freedom, equipartition of energies requires that $\left\langle V^2(0)\right\rangle = (R\mathcal{T})/(Nm)$, where $\mathcal{T}$ is the temperature of the surrounding medium. This defines the constant $D = (R\mathcal{T})/(N\gamma)$. In addition, for a spherical particle of radius r, the Stokes friction coefficient is $\gamma = 6\pi\eta r$, which exactly reproduces Eq. (1.2) found by Einstein by a completely different reasoning. But for the Langevin equation at hand, is D really a coefficient of diffusion? The corresponding mean squared displacement is found to be (set $X(0) = 0$ without loss of generality)

$$\begin{aligned} \langle X(t)\rangle &= 0, \\ \left\langle X^2(t)\right\rangle &= \frac{2D}{\gamma}\left[t/\tau - \left(1 - e^{-t/\tau}\right)\right] \\ &= \begin{cases} \left\langle V^2(0)\right\rangle\, t^2, & \text{for } t \ll \tau, \\ 2Dt, & \text{for } t \gg \tau. \end{cases} \end{aligned} \tag{1.15}$$

The time scale $\tau = m/\gamma$ obviously measures a competition between friction and inertia forces. It separates two temporal regimes. For very small times (Langevin estimates $\tau = 10^{-8}\,\mathrm{sec}$ for a typical Brownian particle in fluid suspension), the mean squared displacement increases with time squared. The regime is called *ballistic*. The motivation for such a terminology becomes evident when comparing to Smoluchowski's Eq. (1.10): a particle travelling at deterministic speed $V_{\mathrm{det}} = \sqrt{\langle V^2(0)\rangle}$ shows the exact same behaviour. However, in contrast to Smoluchowski's ansatz, particle speeds and velocities are here a priori random and, by virtue of Eq. (1.14), are distributed in accordance with the Maxwell-Boltzmann laws.

The large time regime is said to be a *diffusive* one, as the behaviour of the mean squared displacement is exactly congruent with the result (1.5) of Einstein's diffusion model. Moreover, the Gaussian nature of the noise process $\xi(t)$ is passed on to $X(t)$ and $V(t)$ [47]. The position coordinate $X(t)$ in particular is distributed as in Eq. (1.6). In fact, the full behaviour of $X(t)$ is, in an asymptotic sense, essentially that of a Wiener process $B(t)$, for as long as all measurement times and time intervals are large as compared to τ.

An interesting analogy is now found by direct approximation of the Langevin equation (1.12). With $\tau = m/\gamma$ a tiny parameter, one can formally neglect the inertia term (proportional to m) as being small compared to the friction term (proportional to γ). The equation of motion then reads

$$V(t) = \frac{dX}{dt} = \frac{\xi(t)}{\gamma}. \tag{1.16}$$

If indeed the process $X(t)$ in this approximation is literally taken to be a Wiener process, then the noise $\xi(t)$ appears as (being proportional to) the derivative of a Wiener process.

Hence, while sample trajectories of an exact Wiener process are non-differentiable, the derivative process can now be interpreted in a limiting sense. The Langevin equation motion has by definition a well-defined velocity. But on time scales where inertia is negligible, the effects of shocks obtained from $\xi(t)$ are virtually instantaneous displacements $X(t+dt) - X(t) = \gamma^{-1}\xi(t)\,dt$. Consequently, the particle's velocity becomes as sharply peaked and wildly fluctuating as the fuelling collision process.

The mathematical difficulties with such an identifications is being circumvented by using Doob's interpretation [48] of the Langevin equation. Its original form (1.12) is of course highly attractive from the physicist's point of view, since it provides a natural and intuitive approach to translating familiar deterministic mechanical problems to the realm of stochastic motions. Mathematicians instead prefer the (formally) integrated version of such a differential equation. The involved integrals with respect to $\xi(t')\,dt'$ are then interpreted as $dB(t')$, that is an integral with respect to a Wiener process. The latter in turn can be treated conveniently by Itō's stochastic calculus.

On such profound mathematical grounds, one can start to approach more complex problems. For instance, the motion under the influence of an additional external potential $U(x,t)$ is treated by introducing to the Langevin equation (1.12) a force term $-\partial U/\partial x$. This was done by Wang, Ornstein and Uhlenbeck[22] [46, 47] for the harmonic potential $U(x) = m\omega^2/2$. Up to this day, formulating non-deterministic problems in terms of stochastic differential equations is a highly favoured practice. Itō's stochastic calculus can in principle treat any equation of the form

$$\frac{d\boldsymbol{X}}{dt} = \boldsymbol{a}(\boldsymbol{X}(t),t) + \hat{\boldsymbol{g}}(\boldsymbol{X}(t),t)\,\boldsymbol{\xi}(t). \tag{1.17}$$

Here, $\boldsymbol{a}(\boldsymbol{x},t)$ is a vector-valued, $\hat{\boldsymbol{g}}(\boldsymbol{x},t)$ a matrix-valued function, and $\boldsymbol{\xi}(t)$ is a stochastic process whose components ξ_i are mutually independent Gaussian white noises. The vector $\boldsymbol{X}$ collects all variables needed to completely determine the state of the system at a given point in time. For instance, for a d-dimensional Wiener process, $\boldsymbol{X} = (X_1,\ldots,X_d)^T$. To recover Langevin's original equation (1.12) for one-dimensional Brownian motion, set $\boldsymbol{X} = (X,V)^T$. But the components of $\boldsymbol{X}$ can also represent angles, an angular moment or other physical quantities. Thus, the effect of thermally agitated charge carriers (electrons or ions) can be considered be using state variables of the form $\boldsymbol{X} = (\{U_i\},\{I_i\})$, where the U_i and I_i are the various voltages and currents, respectively, measured in a particular electric circuit. Further examples are be given in section 1.5.1 below. Despite the generalised and abstract formulation, equations of the form (1.17) are commonly referred to as Langevin equations in the physics literature.

1.4.3 The Fokker-Planck equation

Bachelier (in the context of stock prices) and Einstein (by study of Brownian motion) showed that the probability density function for the state of a Wiener process $B(t)$ at

[22] Remarkably, at these scientists discussed the harmonically bound Brownian particle, the mathematical principles of stochastic calculus had not yet been developed. Even the question of how to interpret the noise term $\xi(t)$ was subject to debates.

time t obeys a linear, partial differential equation, namely the diffusion equation (1.4). Now, can one find a similar equation for more elaborate dynamics such as the Langevin equation motion?

Considerable efforts in this direction were made in the years which followed Einstein's original publication on Brownian motion. It eventually turned out that any kind of stochastic differential equation (1.17) has associated to it a partial differential equation of the form [50]

$$\frac{\partial}{\partial t}p(\boldsymbol{x};t) = -\sum_{i=1}^{d}\frac{\partial}{\partial x_i}\left[a_i(\boldsymbol{x},t)\,p(\boldsymbol{x};t)\right] + \frac{1}{2}\sum_{i,j=1}^{d}\frac{\partial^2}{\partial x_i \partial x_j}\left[b_{ij}(\boldsymbol{x},t)\,p(\boldsymbol{x};t)\right]. \tag{1.18}$$

The joint PDF for the state variables $\boldsymbol{x}$ is $p(\boldsymbol{x};t)$. The real-valued functions $a_i(\boldsymbol{x};t)$ are the components of $\boldsymbol{a}(x,t)$ appearing in Eq. (1.17). They are called *drift coefficients*. Likewise, the positive functions $b_{ij}(\boldsymbol{x};t)$ are the components of $\hat{\boldsymbol{g}}(\boldsymbol{x};t)\hat{\boldsymbol{g}}(\boldsymbol{x};t)^T$, called *diffusion coefficients*. It is important to contrast the deterministic Fokker-Planck equation for the real-valued probability densities, to the stochastic Langevin-equation, describing a random process evolution.

The complete equation (1.18) was introduced by Andrey Kolmogorov in 1931 [53] and today is frequently found in texts on probability theory as *Kolmogorov forward equation*. Einstein and Bachelier's original one-dimensional diffusion equation is recovered by setting $a(x;t) = 0$, $b(x;t) = 2D$. Adriaan Fokker, during his Ph.D. years with Max Planck (finished 1913) discovered a very general form of Eq. (1.18) in the course of their studies on rotating dipoles [54, 55, 50]. In the physics's literature, they are thus commonly referred to as *Fokker-Planck equations*. In 1915, Smoluchowski generalised the diffusion equation [56] to the now-called *Smoluchowski equation*, which incorporates external forces and operates in the overdamped limit of negligible inertia forces. Conversely, a differential equation in velocity space in the absence of external forces was considered by Lord Rayleigh as early as in 1891 [57]. Finally, the Klein-Kramers equation, named after Oskar Klein [58] and Hans Kramers [59], accounts for both external and inertial forces. It consequently provides a description in complete phase space $\boldsymbol{x} = (x, v)$.

To solve of Fokker-Planck equations (1.18) is complementary to Langevin's method. The Langevin equation treatment is, in a way, more complete, as it fully defines the stochastic process to be studied. The Fokker-Planck equation describes only the behaviour of the associated PDF of the single-time state variable, $p(\boldsymbol{x};t)$. It is thus incapable of predicting, for example, two-time correlations $\langle \boldsymbol{X}(t_1)\boldsymbol{X}(t_2)\rangle$. Yet, in the analysis of the specific PDF the Fokker-Planck equation focuses on, it does its job exceptionally well. Being a deterministic differential equation, several solving and approximation techniques from standard analysis are available such as eigenfunction expansions, separation ansatz or elaborate operator methods developed in the quantum mechanics field [50]. Plus, boundary conditions are particularly easy to implement by using the method of images [60].

1.5 Gaussian white noise: universal approach for modelling fluctuations

The generality of the Langevin equation (1.17) and the associated Fokker-Planck equation (1.18) is impressive. The range of applications is vast, and goes far beyond the analysis of Brownian particle diffusion. The concept is to devise all kinds of influences onto a dynamical variable $\boldsymbol{X}(t)$ under study into deterministic and fluctuating contributions. The deterministic aspects then define the overall structure of the Langevin equation of motion (1.17). Fluctuating components essentially enter through the white Gaussian noise(s) $\boldsymbol{\xi}(t)$.

1.5.1 Applications

Several physical applications of Langevin equation motion in periodic potentials are listed in Risken's 1989 book [50], including the following: synchronisation of an oscillator, locking of two laser modes, quantum noise in ring laser gyros, charge carrier transport in superionic conductors, current through a Josephson Tunneling Junction (two superconductors which are separated by a thin oxide layer), noise in phase locked loops and noise in a laser. Moreover, rotational Brownian motion is central to the Debye theory of dielectric relaxation [49, 50], allowing to calculate susceptibilities in linear response for rotation of dipoles in an external field [50].

W. T. Coffey (2004 [49]) adds dynamic light scattering and Kramer's escape rate theory [50, 59]. The latter was originally developed to explain the breaking of a chemical bond induced by thermal agitation. In this context, the variable $X(t)$ is not a classical position, but rather a reaction coordinate, i.e. the distance between two fragments of a dissociated molecule [49]. The theory yields explicit formulas for the escape rates in any physical system in which there is noise-activated escape from a potential well. Among its diverse applications are [49] the dielectric relaxation of nematic liquid crystals, magnetic relaxation of fine ferromagnetic particles, laser physics, and the dynamics of Josephson junctions.

Concersely, if the typical time scale of a chemical reaction is relatively small, then its progress is predominantly determined by the time it takes the reaction partners to find each other. In this regime, called *diffusion-limited* or *diffusion-controlled* reaction [61], Brownian motion naturally governs statistics and transition rates.

In electrical circuits, Johnson-Nyquist noise [51, 62, 63] is the electronic noise generated by the thermal agitation of the charge carriers (usually the electrons) inside an electrical conductor at equilibrium, which happens regardless of any applied voltage. Thermal noise in an idealistic resistor is approximately white and has a nearly Gaussian amplitude distribution [64]. This highly motivates the study of a Langevin equation for the involved currents and voltages, whenever thermal noise effects are to be taken into account.

The 2002 book of R. M. Mazo [65] contributes by discussion of fluorescence depolarisation of molecules in solutions. The author elaborates on a toy model for diffusion of polymers and provides a large list of references for further reading.

Active Brownian motion (that is, a Langevin equation motion with nonlinear friction term $\boldsymbol{a}(\boldsymbol{X})$) has found biological application in the description of self-propelled motion patterns, such as directed transport in living cells or collective swarming behaviour. [66]

With the first studies of a statistically motivated diffusion equation being conducted by Bachelier in the context of stock markets, it does not come as a surprise that Langevin equations also play an important role in modern mathematical finance. The Black-Scholes model describes the evolution of an option price $X(t)$ under the influence of market fluctuations [67]. It is formulated in terms of a Langevin equation with *multiplicative noise*: the noise strength g appearing in Eq. (1.17) is not constant in this case, but proportional to $X(t)$. This effectively generates a *geometric Brownian motion*, which is essentially and exponentiated Wiener process with drift. Hence, $X(t)$ is always positive, and characterised by independent multiplicative increments, and a log-normal distribution.

1.5.2 Fast versus slow variables

Langevin's object of theoretical study was the Brownian motion. In order to overcome the burden of treating the complex interplay with the dynamics of the ambient medium, Langevin suggested to use his stochastic equations of motion (1.12). There, all interaction with the medium is packed into a linear friction term $-\gamma V$ and the Gaussian white noise $\xi(t)$. This approach is for sure a very attractive one, since it provides intuitive insight and allows for a natural extension to other problems (like the ones cited above). However, it should be stressed that such an ansatz is purely phenomenological. A brief heuristic argument motivating his method can be given in terms of a separation of time scales as follows.

The systems listed in the previous section, albeit being extremely diverse, have the following common trait. On the one hand, there is an individual (or a finite set of individuals) whose dynamics are followed and studied in detail via the variable $\boldsymbol{X}(t)$. On the other hand, the behaviour is coupled to a noisy environment, or, by its physical term, a heat bath. The latter constitutes itself a dynamical system with an immense number of degrees of freedom, but made up of constituents who, in some sense, all act very fast in comparison to the observed individual: the position of a Brownian particle is measured, say, every second, which corresponds to 10^{20} sec collisions with one of the water molecules in the bath. A share price might be looked at once per month, during which several thousands of shares may change owner. The "slow" variable $\boldsymbol{X}(t)$ is not notably affected by single interaction events, but rather by large accumulations thereof. This justifies the concentration on (time) averaged statistical effects. Moreover, correlations between successive interactions should decay fast so as to be negligible for the observed motion. For instance, immediately after the collision between a fluid molecule and the Brownian particle, their motions are correlated – but by the time the two meet again, the fluid molecule has undergone such a large amount of collisions with other fluid molecules that the correlation with the Brownian particle has almost vanished. This explains the use of a white (i.e. δ-correlated) noise in the Langevin approach.

One can conduct these considerations in a much more careful and quantitative way by using elaborate tools such as Mori-Zwanzig projection operators [68, 69]. The procedure is outlined in Ref. [65] for the case of a heavy Brownian particle in a gas or fluid. Starting from a deterministic Liouville equation, the fast variables are eliminated in favour of the slow ones, eventually reproducing the stochastic Langevin equation of motion. This way one even finds the microscopic connection between friction and diffusion coefficient.[23] This mathematical treatment however is far from trivial. The extension to more complex systems poses a considerable challenge, and for many problems, an ultimate resolution is still missing. In fact, the Mori-Zwanzig method is not a prominent approach to legitimate Brownian motion models – it is meant to find the limit where the Brownian theory breaks down. A careful analysis in some cases reveals a need to generalise Langevin's treatment, questioning in particular the uncorrelated noise and the linear friction (see for example the discussion on Brownian motion in a dense fluid [65] and in a lattice of harmonic springs [65, 70]).

1.5.3 Attractiveness matters

Given such reservations, one cannot help but wonder why so many profoundly different phenomena as the ones compiled in section 1.5.1 are still successfully tackled by the same stochastic equations (1.17). How come that fluctuation mechanisms which are microscopically so distinct – a laser beam's disordered scattering sites, the thermal agitation of a charge carrier, or a hard sphere or a magnetic moment, the transactions on a stock market – can universally be captured by the same mathematical noise model $\xi(t)$? In particular, why are the noise statistics Gaussian?

Here, the statistician can give a very clear answer: Gaussianity comes by virtue of the *central limit theorem*. A very large class of motions are, to use the mathematical phrase, *attracted* towards a Gaussian process at large times. The most common form of the theorem is as follows. Suppose one is interested in the statistics of a sum

$$X(n) = \sum_{j=1}^{n} \delta X_j \tag{1.19}$$

of random variables δX_j which are statistically independent, but characterised by identical probability distributions. Let specifically $\mu = \langle \delta X_j \rangle$ and $\sigma^2 = \left\langle \delta X_j^2 \right\rangle - \mu^2$ denote their common average and variance, respectively. Consider now the dimensionless, centred, rescaled sum

$$X'(n) = \frac{X(n) - \mu n}{\sqrt{\sigma^2 n}}. \tag{1.20}$$

and denote with $p'(x'; n)$ its probability density function. The central limit theorem now says that for large n, this probability density converges point-wise to a standard

[23] Notice that Langevin establishes the relation between friction and diffusion constant through an additional equation, independent of his equations of motion. Namely, he requires equipartition of energies at equilibrium.

Gaussian, i.e.

$$p'(x';n) \to p'(x') = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x'^2}{2}\right) \qquad (n \to \infty). \tag{1.21}$$

A basic proof is provided in section 2.1.1. For now, investigate the consequences for stochastic process theory. In this section, two view-points on the central limit theorem will be presented. On the one hand the sum $X(n)$ may be interpreted as a one-dimensional random walk, that is, a step-wise linear motion with random step lengths. This perspective gives additional insight into the classical Brownian motion theories and helps detecting their connection. On the other hand, the δX_j may represent small scale distortions in a noisy environment. If, on an observational scale, only the effect of large accumulations of such microscopic effects are resolvable, then one needs to study increments of $X(n)$. Such an analysis showcases Gaussian noises as a tool of outstanding importance in stochastic theory, but also reveals their limitations.

The central limit theorem for random walks. First, pursue the random walk picture. In this interpretation, δX_j is the distance of the jth step. The random step lengths are assumed to be independent and identically distributed. For example, a one-dimensional Pearson walk has discrete step distances: either $\delta X_j = +l$ or $\delta X_j = -l$ with equal probabilities. Then, $\mu = 0$ and $\sigma^2 = l^2$. In a two-dimensional Pearson walk, the x-coordinate has a continuous distribution: for a single step, $\delta X_j = l\cos(\theta_j)$, where each θ_j is distributed uniformly on $[0; 2\pi)$. In that case, $\mu = 0$ and $\sigma^2 = l^2 \left\langle \cos^2(\theta_j) \right\rangle = l^2/2$. A random walk with $\mu \neq 0$ is said to be *biased*.

Now let n be large, and study the probability density function $p(x;n)$ of the random walk position coordinate $X(n)$. It directly follows from Eqs. (1.20) and (1.21) that $p(x;n)$ is approximately Gaussian with mean μn and variance $\sigma^2 n$. In order to interpret the random walk as a real process, define a physical time coordinate $t = n\tau$, that is, τ is the time passing between consecutive steps. In this form, the central limit theorem reads

$$p(x;t) \sim \frac{1}{\sqrt{4\pi D t}} \exp\left[-\frac{(x-vt)^2}{4Dt}\right], \qquad \text{for } t \gg x^2/D. \tag{1.22}$$

where the drift $v = \mu/\tau$ and diffusion constant $D = \sigma^2/(2\tau)$ are introduced as actual, physically measurable parameters.

The point to stress here is the universality of the result. Independent of the detailed step length statistics, the distribution of the position coordinate is approximately Gaussian; there is a constant drift $\langle X(t) \rangle = vt$, and the square width increases linearly in time, $\left\langle X^2(t) \right\rangle = 2Dt$. For example, Einstein arrives at Eq. (1.6), which corresponds to Eq. (1.22) with $v = 0$. In fact, he implicitly uses a random walk approach: his step lengths are defined in terms of the symmetric function $\phi(\Delta)$. A more specific example is Pearson's random walk in the plane. Both the x-coordinate and the y-coordinate exhibit unbiased random walks with $\sigma_x^2 = \sigma_y^2 = l^2/2$. While the x- and y-motions are not statistically independent (since they are coupled through the turning angles θ_j),

they are uncorrelated: $\langle \delta X_j \delta Y_j \rangle = l^2 \langle \cos(\theta_j) \sin(\theta_j) \rangle = 0$. In such a case, a straightforward generalisation of the central limit theorem to sums of random vectors yields $p(x, y; t) = p(x; t)\, p(y; t)$, where each of the marginal distributions $p(x; t)$ and $p(y; t)$ is again the Gaussian (1.22). For the radial coordinate $r = \sqrt{x^2 + y^2}$, $dx\, dy = r\, d\theta\, dr$, one therefore finds $p(r; t) = [r/(Dt)] \exp\bigl[-r^2/(2Dt)\bigr]$, with two-dimensional diffusion coefficient $D = D_x + D_y = l^2/(2\tau)$. This perfectly coincides with the asymptotics (1.1) provided by Lord Rayleigh (who uses standardised coordinates, $n = t/\tau$ and $r' = r/l$).

To be precise, the Gaussian approximation (1.22) becomes better the larger the ratio $x^2/(Dt)$. Inaccuracies are potentially found in the far tails of the distribution, hence the name "central" limit theorem. From now on, assume that the expression in (1.22) is an *exact* feature of the process $X(t)$. Formally one can achieve this by sending τ and σ^2 to zero, while keeping D fixed. $X(t)$ is then said to be the *diffusion* or *scaling limit* of the random walk $X(n)$, due to the involved rescaling procedure (1.20). Notice that, by its very construction in terms of independent, identically distributed step lengths, the process' increments $X(t_2) - X(t_1)$ are independent if time intervals do not overlap. They are also Gaussian with variance $2D|t_2 - t_1|$, provided that one can only resolve time differences beyond $|t_2 - t_1| \gg \tau$. These considerations highly suggest that the scaling limit of the random walk $X(n)$ is a scaled Wiener process with constant drift, $X(t) = vt + \sqrt{2D}\, B(t)$. Indeed, this statement can be formulated more precisely and proved with mathematical rigour in terms of a theorem by Monroe D. Donsker.

The central limit theorem can be generalised in various ways. Alexandr Lyapunov considered a sequence δX_j of independent step lengths which are allowed to vary in distribution. Given that the higher moments of the step length distribution do not tend to dominate for increasing j, there is again convergence to a Gaussian law. Lyapunov gave a precise necessary condition [71] for this behaviour, which was later refined by Jarl Waldemar Lindeberg [72]. In such a case, the scaling limit $X(t)$ has the Gaussian distribution (1.22), but the coefficients become effectively time dependent, $v = v(t)$ and $D = D(t)$.

Conversely, one may ask whether a stationary sequence δX_j with some amount of inherent dependence is still attracted to the Gaussian limit. Unfortunately, this is a rather subtle question. The answer is affirmative, if the δX_j are *mixing* in time. This means, loosely speaking, that step distances temporally far apart from one another are almost independent. However, the definitions of what exactly is mixing are diverse. Several notions exist in the literature, some intuitive, others rather abstract; for a mathematical survey see Ref. [73]. Determining whether or not a given, dependent sequence of step lengths ultimately sums up to a Gaussian random variable is therefore an intricate task. In addition, even if such a convergence happens on the distribution level, it is not evident that the scaling limit $X(t)$ is clear of any residual interdependence; see the discussion on the fast decay of correlations in the preceding section 1.5.2 and on short- and long-range dependence in section 2.3.1. Two Brownian motion related examples which are indeed attracted to a Wiener process are the Smoluchowski random walk (section 1.3.2) and the Langevin equation motion (section 1.4.2). In both cases, the process increments are initially highly correlated: due to the inertia of the particle, the motion is persistent.

But at times large as compared to the correlation time scale τ, the Wiener process approximation holds.

The central limit theorem for random noises. The appearance of the Wiener processes in stochastic process theory is hence ubiquitous. The Gaussian white noise is of course intimately related as being, in a loose sense, the derivative of a Wiener process. The process shall be revisited now in the light of the central limit theorem. For this, think of the variables δX_j as effects of interactions with a strongly fluctuating environment. For instance, they might model the momentum transfer induced by single or moderately few particle collisions. The object under study might also be subject to other influences, but the δX_j are assumed to be the only random contribution.

In short, the variables δX_j are the only sources of stochastic noise. Now assume that the object being affected by the noise reacts only slowly to those stimuli. The reason might be the object being highly inertial itself, or an additional external, strong but slowly varying influence. A single δX_j then does not have a notable effect, and it is useful to define an *effective noise*

$$\xi_{\bar\tau}(t) = \frac{\gamma}{\bar\tau} \sum_{j=t/\tau}^{(t+\bar\tau)/\tau} \delta X_j = \frac{\gamma}{\bar\tau}[X(t+\bar\tau) - X(t)]. \tag{1.23}$$

Here, the time τ passing between any two consecutive noise shocks is assumed small as compared to $\bar\tau$. The latter measures the typical time it takes to accumulate a sufficient number of noise impacts δX_j in order to create a recognisable effect for the slow object. (The additional prefactors are introduced for consistency with the formulae for Brownian motion. Their meaning and units of course depend on the setting.)

In this context, the statement of the central limit theorem can be rephrased as follows. First, under suitable conditions on the microscopic noise δX_j (Lyapunov or Lindeberg condition, mixing), the distribution of the effective noise $\xi_{\bar\tau}(t)$ becomes approximately Gaussian. Second, from the detailed statistics of the microscopic noise δX_j, only the means μ_j and variances σ_j are relevant, determining its center and width, respectively. The effective noise can be split into two parts, $\xi_{\bar\tau}(t) = \xi_{\bar\tau}^{\mathrm{drift}}(t) + \xi_{\bar\tau}^{\mathrm{fluct}}(t)$. The first is a drift term and accounts for the accumulated means μ_j. It is *deterministic* and defines the average evolution of the noise: $\langle \xi_{\bar\tau}(t)\rangle = \langle \xi_{\bar\tau}^{\mathrm{drift}}(t)\rangle \equiv \xi_{\bar\tau}^{\mathrm{drift}}(t)$. The second term emerges from the microscopic fluctuations σ_j. It is zero on average and gives the *random* deviations from the deterministic drift behaviour: $\langle [\xi_{\bar\tau}(t)]^2\rangle - \langle \xi_{\bar\tau}(t)\rangle^2 = \langle [\xi_{\bar\tau}^{\mathrm{fluct}}(t)]^2\rangle$. It is therefore reasonable to treat drift effects as if they were another external, deterministic influence. For example, according to the Langevin equation (1.12), a Brownian particle at a given velocity v experiences two kinds of forces. Both originate in interactions with fluid molecules. On the one hand, the frequency of particle collisions is spatially asymmetric for $v \neq 0$. This results on average in a net deceleration $-\gamma v$. On the other hand, the randomness of particle collisions enters through the separate noise $\xi(t)$, which is unbiased.

Finally, consider the case where the fluctuating contributions are stationary, $\sigma_j \equiv \sigma$, and where the temporal "range" of dependency[24] in the microscopic noise sequence $\{\delta X_j\}$ is small as compared to $\bar{\tau}$. In this case, the scaling limit becomes $\xi_{\bar{\tau}}^{\mathrm{fluct}}(t) = \gamma\bar{\tau}^{-1}\sqrt{2D}\,[W(t+\bar{\tau}) - W(t)]$ with diffusion constant[25] $D > 0$. In other words, $\xi_{\bar{\tau}}^{\mathrm{fluct}}(t)$ is a Gaussian process with

$$\left\langle \xi_{\bar{\tau}}^{\mathrm{fluct}}(t) \right\rangle = 0,$$

$$\left\langle \xi_{\bar{\tau}}^{\mathrm{fluct}}(t)\, \xi_{\bar{\tau}}^{\mathrm{fluct}}(t') \right\rangle = \frac{2D\gamma^2}{\bar{\tau}} \begin{cases} 0 & \text{for } |t-t'| \geq \bar{\tau}, \\ \left(1 - \dfrac{|t-t'|}{\bar{\tau}}\right) & \text{for } 0 < |t-t'| < \bar{\tau}, \\ 1 & \text{for } t = t'. \end{cases} \tag{1.24}$$

Here, both the independence and stationarity of Wiener process increments came to use. It follows from direct comparison with Eq. (1.13) that for all practical purposes, the effective noise $\xi_{\bar{\tau}}^{\mathrm{fluct}}(t)$ behaves like a Gaussian white noise. It is not perfectly white, but this is only relevant on time scales of the order of $\bar{\tau}$. The formal limit $\bar{\tau} \to 0$ conveniently eliminates this regime, and the noise becomes effectively δ-correlated.

Recall that the statistics of the discrete, microscopic noises δX_j are not necessarily Gaussian. They are even allowed (to a limited extent) to vary with j and be to mutually dependent. By virtue of the central limit theorem, their accumulated effect is then approximated well by a Gaussian white noise process at sufficiently large times. This justifies their ubiquitous use in stochastic modelling.

The history of the central limit theorem dates back to the studies on the Bernoulli distribution of de Moivre (1733) and Laplace (1812). The first precise formulation of the theorem in a general setting was given by Lyapunov in 1901 [38]. As with other fields of probability theory, the development of the central limit theorem in its various forms started with the beginning of the 20th century. Important contributions attribute to von Mises, Pòlya, Lindeberg, Lévy, and Cramér. To complete the historical record, the reader is referred to the thorough accounts in Refs. [38, 74].

In retrospective, the early works by the visionaries of the kinetic theory of heat in the mid 19th century are thus indeed impressive. At their time, the ample scope of the central limit theorem was not yet apparent. Appreciated by many merely as a law of errors, it revealed its beauty only to the very enthusiasts of statistical methods. One such convinced individual was Sir Francis Galton,[26] who in 1889 praised the central limit theorem with the following words [75]:

[24] The precise definition of short-range dependence is provided in section 2.3.1. Here, the condition makes certain that the noise $\xi_{\bar{\tau}}(t)$ can be approximated in terms of a Wiener process increment.

[25] If the microscopic noise impacts δX_j are mutually independent, then $D = \sigma^2/(2\tau)$. If not, the value of D is determined by the exact nature of noise correlations, see section 2.3.1.

[26] Galton is also credited for being the inventor of linear regression analysis and popularised the concept of correlation. He was mentor of Karl Pearson, who some consider his "statistical heir".

> "I know of scarcely anything so apt to impress the imagination as the wonderful form of cosmic order expressed by the 'Law of Frequency of Error'. The law would have been personified by the Greeks and deified, if they had known of it. It reigns with serenity and in complete self-effacement, amidst the wildest confusion. The huger the mob, and the greater the apparent anarchy, the more perfect is its sway. It is the supreme law of Unreason. Whenever a large sample of chaotic elements are taken in hand and marshaled in the order of their magnitude, an unsuspected and most beautiful form of regularity proves to have been latent all along."

1.6 Anomalous diffusion and noise: not Gaussian, not white

The elimination of fast variables technique and the central limit theorem analysis lay the fundament for theories based on Langevin stochastic differential equations, Eq. (1.17), and Fokker-Planck diffusion equations, Eq. (1.18). But a careful treatment also indicates the limitations of these approaches. In general, a breakdown of the classical noise and diffusion theories is to be expected in the context of complex, disordered environments. It became common habit to refer to such a behaviour as *anomalous* dynamics. This term of course requires further classification.

The most common form to characterise an *anomalous diffusion* process $X(t)$ is in terms of the mean squared displacement (assume $X(0) = 0$)

$$\left\langle X^2(t) \right\rangle \simeq t^{2H}. \tag{1.25}$$

In the absence of any external binding or drift forces, and at large times, the mean squared displacement of a classical diffusion processes – as treated by Einstein, Smoluchowski or Langevin – is Eq. (1.25) with $H = 1/2$. The linear dependence is thus indication of *normal diffusion*. Processes with a *Hurst exponent* H ranging within $0 < H < 1/2$ are coined *subdiffusive*. The case $H > 1/2$ is referred to as *superdiffusion*, and $H = 1$ is the limiting *ballistic* or *wavelike* regime.

Examples for anomalous diffusion of the form (1.25) range from the motion of charge carriers in amorphous semiconductors [76, 77] over the diffusion of submicron tracers in living biological cells [78] or the dynamics of small particles in weakly chaotic flows [79] to the dispersion of chemical tracers in the groundwater [80] or the dynamics of stock markets [81], just to name a few [82, 83, 84, 85].

In general anomalous diffusion processes are not universal and thus their definition through equation (1.25) is not unique. Instead, this form may be caused by multiple physical mechanisms, some of which are very distinct conceptually. Several pathways to anomalous diffusion have been discussed. The object of the chapter 2 is to introduce and characterise in detail three prominent paradigmatic diffusion processes used in modern stochastic modelling.

Lévy flights. Under suitable conditions, the noise fluctuations can take on extreme values. The central limit theorem requires that the variance of individual noise shocks

σ^2 be finite. If such a preliminary is violated, convergence to a non-Gaussian limit is consequential. Lévy flights are instead characterised by *heavy-tailed* statistics. Moreover, the motion is highly discontinuous and superdiffusive. The field of application for Lévy flights, or their close relative, the Lévy walks, are dynamic motions in fractal embedding spaces or in highly turbulent environments. A detailed discussion on Lévy flights, including several examples of applications, is given in section 2.1.

Continuous time random walks. The pathway to normal diffusion generally goes along one or several time scale separation arguments. Continuous time random walks tackle systems where in fact the internal relaxation time scales are subject to a considerable amount of randomness. When the concept of a "typical" noise time scale has to be rejected, then also the central limit theorem needs to be reconsidered. The situation may be encountered when diffusion takes place in a highly heterogeneous medium, involving multi-scale binding, trapping or reaction dynamics. In the most extreme case, the expectation of the relaxation time is infinite, which leads to ubiquitous phenomena such as subdiffusion, ageing and weak ergodicity breaking. An extensive treatment of this topic is provided in sections 2.2 and 4.

Fractional Brownian motion and noise. Elimination of fast variables is a formidable technique to construct Langevin or Fokker-Planck equations directly from the underlying Hamiltonian dynamics. However, identifying the slow and fast variables in a system more complex than the Brownian motion in a dilute gas is not straight-forward. It turns that slow modes in the dynamics of dense or crowded environments can make significant contributions to the large scale behaviour. Effectively, the simple white noise and linear drift terms in the associated Langevin equation are then to be endowed with a power-law memory component. Such a noise is called fractional Brownian (or Gaussian) noise and the integrated diffusion process is a fractional Brownian motion. The latter, while still Gaussian, feature persistent or anti-persistent behaviour and may lead to both to sub- or superdiffusion. These processes are the focus of section 2.3.

The range of distinct anomalous diffusion processes is wide, and listing them in encyclopedic completeness goes beyond the scope of this work. For extensive accounts on the topic, including additional information on self-avoiding walks, obstructed motion, diffusion on fractals, single-file diffusion and models of quenched disorder, see Refs. [3, 82, 83, 84, 85]. The program of the following chapter is rather to present the three models introduced above as paradigmatic case studies. Their distinct anomalous diffusion properties are highlighted and contrasted. Section 3 is devoted to a unified model, and the interplay of anomalous diffusion phenomena is investigated.

2 Paradigms of anomalous diffusion

This chapter highlights three paradigmatic models which are commonly put under the large umbrella of *anomalous diffusion* processes: *Lévy flights*, *continuous time random walks* (CTRWs) and *fractional Brownian motion* (FBM). They are studied here first on an abstract theoretical basis. Each defines a class of stochastic processes which somehow depart from the laws of normal diffusion. The latter refers to the statistical interpretation of the Brownian motion as described in the introductory chapter. To be specific, from here on throughout the rest of this work, the term "Brownian motion" is being identified with the mathematical notion of a "Wiener process" (see section 1.4.1).[1] The ubiquitous properties of each anomalous diffusion paradigm naturally define the field of typical real world applications, a collection of which is given at the end of each subsection.

2.1 Lévy flights

Lévy flights can be defined in terms of a random walk process. Let $X(n)$ label the position of a random walker along a continuous line after making a discrete number of $n \in \mathbb{N}_0$ steps. Let any step length $\delta X_n = X(n) - X(n-1)$, $n \geq 1$, be random, but independent from all previous step lengths. Moreover, all step lengths share the same probability density function (PDF) $\lambda(\delta x)$, i. e.

$$\Pr\{\delta X_n \leq a\} = \int_{-\infty}^{a} \lambda(\delta x)\, d(\delta x) \tag{2.1}$$

for all $n \in \mathbb{N}_0$ and $a \in \mathbb{R}$. In short: step lengths are assumed to be independent, identically distributed (IID) random variables.

A specific choice of step length characteristics $\lambda(\delta x)$ implies specific random walk statistics. As outlined in section 1.5, the focus of the applied theory often does not rest on the detailed, step-by-step random walk features, but on the approximated behaviour as measured on very large spatial and temporal scales. Now, the powerful central limit theorem and Donsker's theorem dictate an ultimate convergence in distribution of a random walk to a Brownian motion process under very general conditions. To escape the realm of normal diffusion, one therefore needs to target at least one of the theorem's requirements.

In the case of Lévy flights, the requirement in question is the finiteness of the second moment of step lengths. From a probabilistic point of view, such an assumption is not

[1] This may be considered as an abuse of language, as it confuses the physical observation with its mathematical model. Still, the nomenclature is common especially in theoretically oriented texts. For instance, what Mandelbrot's calls fractional Brownian motion is a stochastic integral with respect to a Wiener process, see below.

self-evident. For example, a Lévy flight is typically generated from a *heavy-tailed* PDF, that is

$$\lambda(\delta x) \sim A_\mu \sigma^\mu \, |\delta x|^{-1-\mu}, \qquad 0 < \mu < 2, \tag{2.2}$$

for large distances $|\delta x| \gg \sigma > 0$. The constant $A_\mu = \sin(\pi\mu/2)\Gamma(1+\mu)/\pi$ is introduced for convenience, see the calculations of the following sections. Notice that such a $\lambda(\delta x)$ is perfectly integrable at $\pm\infty$, but its key feature is the diverging second moment, $\left\langle (\delta X_n)^2 \right\rangle = \int_{-\infty}^{\infty} (\delta x)^2 \lambda(\delta x)\, d(\delta x) = \infty$. In fact, any moment $\langle |\delta X_n|^q \rangle$ with $q \geq \mu$ is infinite. The distribution is hence, in a sense to be made more specific below, very broad. The parameter $\sigma > 0$ here plays the role of a typical microscopic scale for individual step distances. But it should *not* be confused with the statistical notion of the standard deviation of steps, since the latter, again, is infinite.

Now, with this peculiar choice of step length distributions, will the dynamics at large times still be Brownian-motion-like? This question is the focus of the following section.

2.1.1 Generalized central limit theorem and stable random variables

Consider a random walk $X(n)$ as introduced above. For simplicity, assume the IID step lengths δX_n are symmetric, $\lambda(\delta x) \equiv \lambda(|\delta x|)$. Apart from that, $\lambda(\delta x)$ is arbitrary for now. The random walk is hence not biased globally or locally, and without loss of generality, one can assume $X(0) = 0$. Let $p_\mu(x; n)$ be the PDF for the random walker position $X(n)$. Its asymptotic form can be derived in terms of a generalised central limit theorem, due to Gnedenko and Kolmogorov [87]. It is instructive to go into the basic reasoning of this theorem in some detail.

Since $X(n)$ is a sum of independent, identically distributed random variables, the analysis can be conducted most conveniently by use of characteristic functions:

$$\begin{aligned} p_\mu(k; n) &= \langle \exp(ikX(n)) \rangle = \left\langle \exp\left(ik \sum_{j=1}^{n} \delta X_j \right) \right\rangle \\ &= \prod_{j=1}^{n} \langle \exp(ik\, \delta X_j) \rangle = [\lambda(k)]^n. \end{aligned} \tag{2.3}$$

The factorisation of averages from the first to the second line is allowed due to the independence of step lengths. $\lambda(k)$ is the characteristic function, or *Fourier transform*[2] of the probability density $\lambda(\delta x)$.

At this point, one has to distinguish two cases. First assume that $\lambda(\delta x)$ decays so fast for large $|\delta x|$ that all of its moments are finite. Then the latter are encoded in a Taylor

[2] The Fourier transform $f(k) = \mathscr{F}_{t\to k}\{f(t)\} = \int_{-\infty}^{\infty} f(x) \exp(ikx)\, dx$ of a function $f(x)$ is expressed, throughout this work, by explicit dependence on the Fourier variable k. Fourier inversion is occasionally indicated explicitly as $f(x) = (2\pi)^{-1} \int_{-\infty}^{\infty} f(k) \exp(-ikx)\, dk = \mathscr{L}^{-1}_{k\to x}\{f(k)\}$.

expansion of $\lambda(k)$:

$$\begin{aligned}\lambda(k) = \langle \exp(ik\,\delta X_j)\rangle &= \int_{-\infty}^{\infty} e^{ik\,\delta x}\,\lambda(\delta x)\,d(\delta x) \\ &= \int_{-\infty}^{\infty}\left[1 + ik\,\delta x - \frac{(k\,\delta x)^2}{2} + \dots\right]\lambda(\delta x)\,d(\delta x) \\ &= 1 - \sigma^2 k^2 + \mathcal{O}\big(k^4\big)\,, \end{aligned} \tag{2.4}$$

where specifically $2\sigma^2 = \big\langle (\delta X_j)^2 \big\rangle$ denotes the second moment. The first moment vanishes for symmetric steps. Furthermore, the asymptotics on the last line of Eq. (2.4) can be extended to include cases where σ^2 is finite, but some higher moments diverge. Then, the full Taylor series becomes inaccurate only in the higher order terms.

The second case are the heavy-tailed distributions (2.2), where the second moment is infinite. Here, the Taylor expansion fails and one has to invoke on Tauberian theorems [24] to find

$$\lambda(k) \sim 1 - \sigma^\mu |k|^\mu. \tag{2.5}$$

Again, the parameter σ does not represent the standard deviation here, since the latter is infinite. Still, notice that the cases of non-heavy-tailed distributions with finite variance can actually be included into the asymptotics (2.5) by allowing the value $\mu = 2$.

In order to find a large-n approximation for the random walk statistics $p_\mu(x;n)$, define a rescaled, dimensionless coordinate

$$X'(n) = \frac{X(n)}{\sigma n^{1/\mu}}. \tag{2.6}$$

The limit of the respective characteristic function $p'_\mu(k;n)$ can be found by combining Eqs. (2.3) and (2.5):

$$\begin{aligned} p'_\mu(k;n) = \big\langle \exp\big[ikX'(n)\big]\big\rangle &= \left\langle \exp\left[ik\frac{X(n)}{\sigma n^{1/\mu}}\right]\right\rangle \\ &= p(k/(\sigma n^{1/\mu});n) \\ &\to \left(1 - \frac{|k|^\mu}{n}\right)^n \\ &\to \exp(-|k|^\mu)\,, \qquad (n\to\infty). \end{aligned} \tag{2.7}$$

To interpret the random walk as a real process, define a physical time coordinate $t = n\tau$, that is, τ is the time passing between consecutive steps. For the implied random motion $X(t)$ follows the large-time approximation

$$p_\mu(k;t) = \exp(-K_\mu |k|^\mu t)\,, \tag{2.8}$$

where the effectively observable parameter $K_\mu = \sigma^\mu/\tau$ is a *generalised diffusion constant*, bearing units of $\mathrm{cm}^\mu/\mathrm{sec}$.

In the scaling limit of large times, the random variable $X(t)$ is uniquely defined in terms of the characteristic function $p_\mu(k;t)$, as specified in Eq. (2.8). Two important properties can be derived from this representation by merely looking at it. Arguably the most remarkable observation is: the distribution is Gaussian for $\mu = 2$, but it is not Gaussian for any $0 < \mu < 2$. Hence, indeed Lévy flights qualify as anomalous diffusion processes. Second, without carrying out the Fourier inversion of Eq. (2.8) explicitly, one can infer the scaling form

$$p_\mu(x;t) = (K_\mu t)^{-1/\mu}\, \ell_\mu\Big[x(K_\mu t)^{-1/\mu}\Big]\,. \tag{2.9}$$

In other words, the position coordinate evolves with time as $X(t) \sim t^{1/\mu}$, that is faster than in the Brownian case ($\sim t^{1/2}$), which is why Lévy flights are referred to as *superdiffusive* processes.

The scaling function ℓ_μ is a standard[3] *symmetric μ-stable law*, uniquely defined by its characteristic function

$$\int_{-\infty}^{\infty} e^{ikx}\ell_\mu(x)\mathrm{d}x = \exp(-|k|^\mu)\,. \tag{2.10}$$

Stable distributions [24, 88] were studied extensively by French mathematician Paul Lévy – who also lent his name to the Lévy flight – and play a crucial role in the context of generalised central limit theorems [87]. In fact, Lévy showed that stable random variables (i.e. random variables with stable distributions) form the basin of attraction for *any* sum of IID random variables. The Gaussian random variables, $\mu = 2$, are a particular subclass which are relevant in the case of finite variance (see section 1.5.3). Symmetric stable laws (2.10) are the proper candidates for symmetric random variables with heavy-tailed distributions (2.2). A third subclass are the *one-sided* stable distributions, which will be the topic of section 2.2.

Stable random variables are central to the analysis of both Lévy flights and continuous time random walks. Two very important features are to be listed here. The statements come without mathematical proof; see for instance the book of Samorodnitzky and Taqqu [88] for a complete discussion.

First, the attribute "stable" refers to the following property. Let X_1 and X_2 be two IID stable random variables. Then the linear combination $aX_1 + bX_2$ with real coefficients a and b is again stable. More precisely, the latter has the distribution of $(|a|^\mu + |b|^\mu)^{1/\mu}X_1 + d$, where d is a suitable real constant and $0 < \mu \leq 2$. If, in addition, X is symmetric then $d = 0$, and the exponent μ is the same as in the characteristic function (2.10). This property can of course be recursively extended to include any finite linear combinations $\sum_{i=1}^n a_i X_i$.

[3] A non-standard symmetric μ-stable law has characteristic function $\exp(-\sigma^\mu|k|^\mu)$ with some $\sigma > 0$. The "standard" of Eq. (2.10) thus lies in the special choice $\sigma = 1$. The parameter σ is called scaling parameter, as it tunes the scale of the random variable: if X is standard symmetric μ-stable, then σX is (non-standard) symmetric μ-stable with scale parameter σ. Note that a standard symmetric 2-stable distribution is actually a Gaussian with a non-standard variance of 2. Still, the common definition of physical parameters is such that the generalised diffusion constant K_μ is related to the ordinary diffusion constant D of Brownian motion via $K_2 = D$.

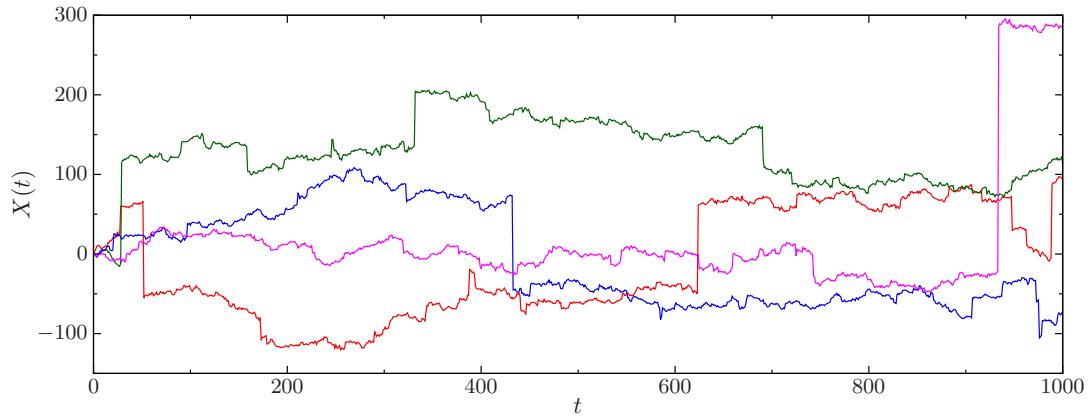


Figure 2.1: Sample paths $X(t)$ of *Lévy flights*. The motion is uncorrelated and unbiased, but characterised by large-scale, discontinuous jumps. The tail index is $\mu = 1.5$.

Second, except for the Gaussian case $\mu = 2$, any μ-stable distribution is heavy-tailed with tail exponent μ. For instance, in the symmetric case, the characteristic function (2.10) has small-k expansion (2.5) with $\sigma = 1$. Therefore, $\ell_\mu(x) \sim A_\mu x^{-1-\mu}$ for $x \gg 1$. This also implies divergence of any moments $\langle |X|^q \rangle$ with $q \geq \mu$.

To conclude, the distributional properties of Lévy flights strongly contrast the Gaussian nature of ordinary Brownian motion, even in the scaling limit of large times. However, the latter is an effect which is to be observed in the ensemble statistics of many independent random walkers. Can even an individual Lévy flight trajectory be discerned from its Brownian motion counterpart? The answer is given in the following section.

2.1.2 A scaling limit free of scales

An ordinary random walk with finite variance displacements converges, according to Donsker's theorem, to an ordinary Brownian motion. Several sample trajectories are provided in section 1.4.1, Fig. 1.5. Lévy flights are fundamentally different processes, as already indicated by the distributional and scaling analysis. But their distinctiveness actually is obvious at first glance when looking at an individual trajectory, such as provided in Fig. 2.1.

While locally the motion might be mistaken for an ordinary Brownian motion by a naked eye inspection, it is interrupted by occasional but remarkably far displacements. This definitely contrasts the continuous Brownian behaviour (compare with Fig. 1.5). The Lévy stable motions shown in Fig. 2.1 are generated from a huge amount of Lévy flight steps. The coordinate system used in the figure is rescaled appropriately, so as to capture the full trajectories. On these macroscopic scales, the scale parameter σ, which determines the "typical scale" of individual displacements (via Eq. (2.2)), is infinitely

small. Still, displacements of observable size do occur regularly in all four trajectories.

The following mathematical analysis helps to understand the origin of this phenomenon. For a discrete random walk $X(n) = \sum_{j=1}^{n} \delta X_j$, consider the random variable

$$L(n) = \max\{|\delta X_j| : j \leq n\}, \tag{2.11}$$

that is, $L(n)$ denotes the largest displacement distance in a single step among the first n steps. What is the distribution of this random variable, and how does it evolve with n? For displacements δX_n which are IID with common PDF $\lambda(\delta x)$, the calculation is simple: since the maximum displacement $L(n)$ falls below some value l if and only if all the displacements up to this point do, one has

$$\Pr\{L(n) > l\} = 1 - \Pr\{L(n) \leq l\} = 1 - \left[\int_{-l}^{l} \lambda(\delta x)\, d(\delta x)\right]^n. \tag{2.12}$$

As n increases, for any fixed value of l, this probability tends to one. This is natural, since the random walker gets more and more chances to draw at least one step distance larger than l. So, in short, the largest single step distance $L(n)$ grows with n. But so does the overall random walk coordinate $X(n)$. Thus the critical question is: do the maximum displacements $L(n)$ increase fast enough as to be visible on spatial scales dictated by the full motion $X(n)$?

It turns out the question is affirmative for Lévy flights, that is, for heavy-tailed step length distributions as in Eq. (2.2). To see this, define $L'(n)$ analogous to $X'(n)$ as

$$L'(n) = \frac{L(n)}{n^{1/\mu}}. \tag{2.13}$$

Now calculate

$$\begin{aligned}
\Pr\left\{L'(n) > l'\right\} &= \Pr\left\{L(n) > l' n^{1/\mu}\right\} \\
&= 1 - \left[\int_{-l'n^{1/\mu}}^{l'n^{1/\mu}} \lambda(\delta x)\, d(\delta x)\right]^n \\
&\to 1 - \left[1 - 2\int_{l'n^{1/\mu}}^{\infty} A_\mu \sigma^\mu (\delta x)^{-1-\mu}\, d(\delta x)\right]^n \\
&= 1 - \left[1 - \frac{1}{n}\frac{2A_\mu}{\mu}\left(\frac{\sigma}{l'}\right)^\mu\right]^n \\
&\to 1 - \exp\left[-\frac{2A_\mu}{\mu}\left(\frac{\sigma}{l'}\right)^\mu\right] \qquad (n \to \infty).
\end{aligned} \tag{2.14}$$

Therefore, the large scale time evolution of the distribution of the maximum step length, expressed in terms of $t = n\tau$, is

$$\Pr\{L(t) > l\} = 1 - \exp\left[-\frac{2A_\mu}{\mu}\frac{K_\mu t}{l^\mu}\right]. \tag{2.15}$$

The most important aspect of this result is the scaling relation $L(t) \sim t^{1/\mu}$, which is the exact scaling behaviour of the Lévy flight itself. This implies that within any section of the trajectory one may indeed observe a single step distance which is of the order of the observable spatial scales. Yet, the temporal scale τ cannot be resolved in this limit, so these displacements appear to be instantaneous. This is why they are also commonly referred to as *jumps*, and one speaks rather of a Lévy *flight* than a Lévy *walk* (although the latter notion also exists, see section 2.1.4). In the same spirit, Lévy flights are said to exhibit *scale-free* spatial dynamics. From the exact jump length distribution all details besides the tail behaviour are irrelevant to the motion beyond microscopic scales, $x \gg \sigma$, $t \gg \tau$. But the fundamental characteristic of jump-like dynamics persists on even the largest of scales.

Notice that this line of reasoning must be restricted to the case of heavy-tailed jump length distributions. As a counterexample, consider a power-law distribution of jump distances as in Eq. (2.2), but with $\mu > 2$. Then the scaling relation $L(t) \sim t^{1/\mu}$ still holds, since the derivation of Eq. (2.15) is independent of μ (provided μ is positive). But since in this case the second moment of jump lengths is finite, the process falls in the scope of the (standard) central limit theorem. $X(t)$ is attracted to a Brownian motion with $X(t) \sim t^{1/2}$, which is now *faster* than $t^{1/\mu}$. Hence, the larger the observational scales, the smaller in comparison become jump distances. Ultimately, the trajectory will thus appear continuous. (Although, if μ is close to 2, the convergence to this continuous limit can actually be very slow.)

The large-time scaling limit of a (symmetric) Lévy flight is called a *(symmetric) Lévy stable motion*. This is analogous to interpreting Brownian motion as the scaling limit of a random walk with finite-variance jump lengths. As Wiener showed (see section 1.4.1), there is an equivalent way of defining Brownian motion directly without invoking a limit theorem. The same can be done for a Lévy μ-stable motion $L_\mu(t)$, in terms of the following properties (see e.g. Ref. [88]): (i) $L_\mu(0) = 0$; (ii) increments $L_\mu(t_2) - L_\mu(t_1)$ are mutually independent for non-overlapping time intervals; (iii) increments are equal in distribution to $|t_2 - t_1|^{1/\mu} \cdot L_\mu(1)$, and $L_\mu(1)$ obeys a standard μ-stable law. (In the symmetric case, the standard μ-stable law is given by Eq. (2.10). Notice that above definition uses the same mathematical reductionism as the standard Wiener process: coordinates are rescaled such that $K_\mu = 1$. This can be reversed by considering $(K_\mu)^{1/\mu} \cdot L_\mu(t)$ instead.)

For $\mu = 2$, the definition is effectively equivalent the definition of an ordinary Brownian motion.[4] While the above definition is very useful in calculating ensemble properties, such as distributions, moments, dependence structures, etc., it hides the scale-free character of jumps. It also emphasizes an important parallel between the symmetric Lévy

[4] This is not strictly correct. Notice that (i)-(iii) only specify the properties all finite dimensional distributions of the process. Path properties are not included. In particular, $L_2(t)$ is not required to have continuous sample paths, so it is not necessarily a Brownian motion. However, the two processes are identical in all finite dimensional distributions, and sample path properties are not the scope of the present thesis. To be perfectly concise, standard Lévy 2-stable motion is equal in distribution to a standard Brownian motion multiplied by a scaling factor of 2 (compare footnote on page 36).

stable motion and Brownian motion: the increments of both processes are stationary (i.e. invariant under time shifts) in distribution and statistically independent. This is in fact the crucial defining property of any Lévy process. It represents the continuous time analogue to constructing random walks or Lévy flights from IID jump distances.

2.1.3 First passage statistics, leap-over lengths and other properties

The fingerprint characteristics of Lévy flights as presented in the previous sections are: superdiffusive scaling, heavy-tailed, stable distributions and scale-free jump-like spatial motion. Several other interesting properties are to be subsummized in the following. Some might be considered as straight-forward implications, others are maybe less expected. The list is far from exhaustive, but rather meant to give a feeling for the special character of Lévy flights. For proofs or mathematical details, the interested reader is referred to the cited literature.

A μ-stable PDF (2.9) with $0 < \mu < 2$ does not solve the ordinary diffusion equation (1.4). Standard methods like the analysis of the Fokker-Planck equation (1.18) therefore have to be modified to capture general stable dynamics. This can be done by using *fractional derivatives* [83, 85, 86, 89]. They appear in a Fourier space representation as algebraic power-law factors $|k|^\mu$, while in real space, diffusion equations are turned into integro-differential equations with power-law memory kernels.

But also the standard treatment of such diffusion equations should be reconsidered. For instance, the method of images fails to correctly predict the effect of an absorbing boundary [90]. The implied Dirichlet condition $p(a;t) = 0$ in fact only introduces an absorbing *point* at position $x = a$. In contrast to Brownian motion, Lévy flights exhibit jumps on all scales and therefore the point a can be passed by without terminating the motion. The proper way of treating boundary conditions for fractional diffusion equations is rather to use non-local operators [91] or to consider directly first passage statistics (see below).

The behaviour of Lévy flights in the presence of confining boundaries is indeed curious. The time evolution of a radial Lévy flight starting in the center of a periodic box is studied extensively in Ref. [92]. On the one hand, the initial motion for times $t \ll a^\mu/K_\mu$ (a is half the edge length of the box, K_μ is the diffusion constant) appears to be unbounded: the PDF $p_\mu(x;t)$ is well-fitted by a stable distribution, and the analysis of moments $\langle |X(t)|^q \rangle \simeq t^{q/\mu}$, with $q < \mu$, is consistent with the unbounded motion's scaling law, $L_\mu(t) \sim t^{1/\mu}$. On the other hand, due to the capability of covering huge distances in only a few steps, the particle motion is affected by the boundary from the very beginning of the motion, no matter how large the confining box. Therefore, bounded Lévy flights exhibit an initial hybrid behaviour: moments with $q \geq \mu$ do not diverge; the second moment in particular is finite and evolves linearily in time, just like for ordinary Brownian motion; the assotiated diffusion constant depends on a (the larger the box, the faster the diffusion); an analysis of the fractal box counting dimension [93] reveals a bifractal behaviour (part Lévy flight, part Brownian like).

Another interesting aspect, are first passage statistics and leap-over lengths. It is

worthwhile to go into these topics in some detail, since they are of particular importance to the discussion on continuous time random walks and ageing renewal theory.

For one-dimensional motions considered here, the *first passage time* $T(x)$ measures how long it takes a process $X(t)$, starting at $X(0) = 0$, to cross a given boundary $x > 0$ for the first time. Mathematically, the random variable $T(x)$ can be defined as

$$T(x) = \inf \{t > 0 : X(t) > x\} \tag{2.16}$$

Depending on the scenario in question, it can be thought of as the time needed for a share at the market to pass a certain threshold price, for an animal to find a food location, or for an enzyme to reach a target polymer cite.

For a Brownian motion $X(t) = B(t)$, the PDF $p_f(t;x)$ of the first passage time can be calculated as [60]

$$p_f(t;x) = \frac{x}{\sqrt{4\pi D t^3}} \exp\left(-\frac{x^2}{4Dt}\right). \tag{2.17}$$

This implies a scaling law $T(x) \sim x^2$, as may be anticipated from the "inverse" scaling law $X(t) \sim t^{1/2}$. Moreover, the asymptotics of the first passage time PDF are $p_f(t;x) \simeq t^{-3/2}$ as $t \gg x^2/(2D)$. It therefore gives a fine example of a heavy-tailed PDF, and consequently, the average first passage time diverges, $\langle T(x) \rangle = \infty$.

According to Sparre-Anderson's theorem [94, 95] this is in fact a universal property: for any random walk based process characterized by IID jump lengths with some continuous, symmetric distribution,[5] the first passage time density asymptotically decays like $t^{-3/2}$. This includes in particular the symmetric Lévy flights $X(t) = L_\mu(t)$, for which one finds [96]:

$$p_f(t;x) \simeq \sqrt{\frac{x^\mu}{K_\mu}} t^{-3/2}, \tag{2.18}$$

as $t \gg x^\mu/K_\mu$. The omitted numerical prefactors depend on μ only. Here, the scaling is $T(x) \sim x^\mu$, consistent with $X(t) \sim t^{1/\mu}$.

But especially in the case of Lévy flights, the analysis of first passage problems does not close with the successful calculation of $p_f(t;x)$. Depending on the application in mind, it might not be enough to know when the dynamical variable $X(t)$ *passes* the boundary x; it can be crucial to be able to estimate how close it *lands*. For instance, an animal which cannot scan for food "on the fly", needs to arrive on a target site within the local field of vision. The stock market analyst is certainly interested in knowing how far a stock share is expected to fall below a given lower bound. This motivates introducing the concept of a *leap-over length* $Z(x)$. It gives the distance to the arrival position directly after first passage at x. Mathematically,

$$Z(x) = |X(T(x)) - x|. \tag{2.19}$$

[5] The situation is of course different, if either a drift or a reflective boundary biases the motion towards the boundary x. In such a case, the average first passage time is finite [60].

If $X(t)$ is a Brownian motion, $\mu = 2$, then clearly $Z(x) = 0$ almost surely. For continuous paths, the concepts of first passage and first arrival are equivalent. In contrast, the jumps of Lévy flights, $\mu < 2$, are scale-free, so that leap-overs do occur. The respective PDF $p_l(z; x)$ is given by [96]

$$p_l(z; x) = \frac{\sin(\pi\mu/2)}{\pi} \frac{|x|^{\mu/2}}{z^{\mu/2}(|x| + z)}. \tag{2.20}$$

Observe that this is independent of K_μ. While the diffusion constant does set the scale for the typical first passage time, it does not influence the distance by which the motion overshoots the target bound. Instead, there is a scaling relation $Z(x) \sim x$. The further away the bound of first passage from the origin of motion, the longer the leap-over distance. Finally, notice that also this PDF is heavy-tailed, $p_l(z; x) \simeq z^{-\mu/2}$ ($z \gg |x|$), albeit with a tail exponent which is even "heavier" than the tail exponent μ of the distribution of individual jump distances.

2.1.4 Applications

When trying to survey even a small range of Lévy flight applications, it is almost unavoidable to also touch the issue of *Lévy walks*. Hence, this closely related anomalous diffusion process shall be briefly introduced here. The spatial displacement statistics of a Lévy walk are, much like for the Lévy flights, governed by heavy-tailed, that is scale-free distributions of the form (2.2). Yet, displacements do not occur instantaneously, but any transition is penalised by a certain amount of time required to get from one point to the other. The travel distances and travel times may be connected simply through a fixed, constant velocity, but also more complicated relations have been considered [97]. In effect, the coupling between flight length and duration makes the trajectories smooth in comparison to the discontinuous Lévy flight paths. In addition, the mean squared displacement is finite for Lévy walks, despite the superdiffusive scaling. Especially this last property is why Lévy walks are often considered conceptually superior over Lévy flights. Indeed, scale-free, instantaneous jumps imply arbitrarily large velocities, which is for sure physically unreasonable for a lot of systems. Still, in many respects the two processes are very similar, and insights or results gained from one of them can readily be transferred to the other.

Sometimes, the distinction is even irrelevant. For instance, the authors of Ref. [98] describe how to prepare a highly disordered optical medium. When an incident light beam hits the surface, it is scattered very irregularly. The microscopic structure of the surface can even be designed to be *fractal*, i.e. self-similar in a statistical sense, and scale-free. Figure 2.2 provides a schematic of the surface landscape in the left panel and includes also a hypothetical "trajectory" of a scattered light beam. The scale-free distribution of scattering sites, gives rise to a scale-free distribution of straight line segment lengths of the light path. With the individual light beam "travelling" at the speed of light, studying the temporal evolution of the "diffusion" process clearly cannot be the experimental object. Instead, the experiments here are static: e.g. one can

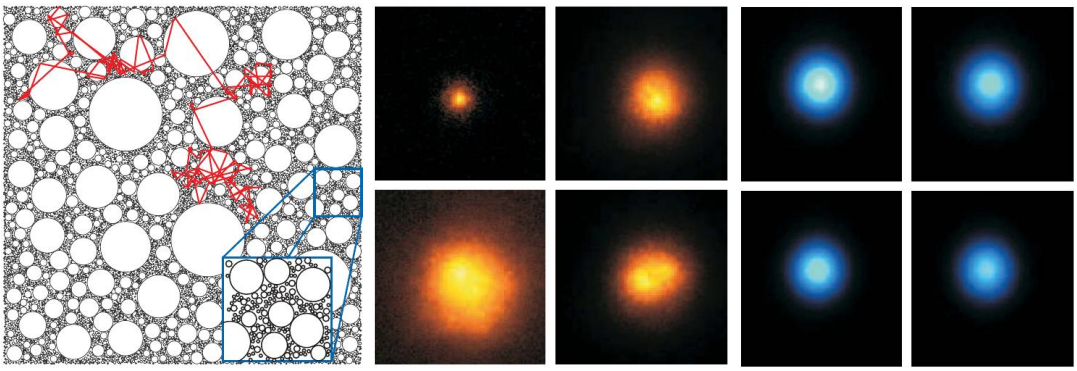


Figure 2.2: Light scattering on a fractal surface [98]. *Left panel*: The disordered optical medium has a fractal, i.e. statistically self-similar structure, see the magnifying inset. In addition, the distribution of scattering sites is scale-free. This feature is carried over to the trajectory of a singly light beam (red). *Centre panel*: Transmission intensity of an incident light bundle. Since scale-free scatter distributions are dominated by exceptionally large local structures, the intensity profiles vary largely from one surface section to the next. *Right panel*: In contrast, a optical medium lacking the self-similarity property generates a reproducible, Gaussian intensity profile throughout the surface.

measure the transmission intensity of an incident bundle of light beams, see the middle panel of Fig. 2.2. Notice the significant variation between measurements at different points of the surface. Much like the trajectory statistics of a scale-free Lévy flight is dominated by single, large-scale displacements, the fractal surface statistics are governed by single, but wide, optically inactive gaps. Disorder realisations hence fluctuate heavily. This contrasts the small variation for "normal diffusion" in an optical medium which is disordered, yet characterised by a finite average scattering length, see the right panel of Fig. 2.2.

Another natural field of application for Lévy flights/walks are turbulent systems. The authors of Ref. [79] demonstrate how to generate, within an annular fluid-tank, a circular chain of flow vortices. This is achieved by a combination of a rapid rotation of the tank with a radial, non-isotropic pumping of fluid. A passive tracer particle tends to stick to flow layers near the vortices, but can suddenly be transported quickly over wide distances when caught by a jet stream connecting the vortices, see Fig. 2.1.4. The chaotic trajectories indeed display distinct Lévy walk behaviour, as indicated by an angular superdiffusion and scale-free displacement distributions. Interestingly, even the sticking time distributions are heavy-tailed, so the motion is also close to the continuous time random walk models discussed in the next section. Further examples for scale-free models based on Hamiltonian chaos can be found in [99]. On the theory side, the first analytic concepts for anomalous diffusion in turbulent media originate back to the early works of Richardson [100]. The relation to Lévy walk processes was recognised by

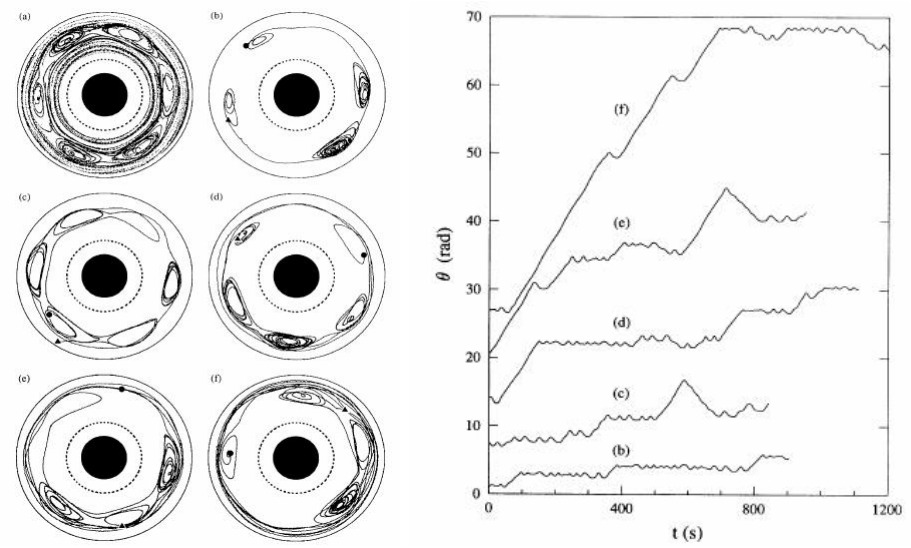


Figure 2.3: Motion of a passive tracer particle in a turbulent rotational flow [79]. *Left panel*: The cross section of the quasi-two-dimensional setup is an annular disc, bounded by inner and outer walls. The fluid is set in motion by a rotation of the walls, and pumps superimpose a radial flow at a non-isotropic rate. The result is the formation of a chain of turbulent vortices, connected by jet streams. One can see the streams of 40 particles in part (a), revealing six almost periodically distributed vortices. (b)-(f) are trajectories of single tracer particles. Periods of long sojourn times near specific vortices are interrupted by long flights, with distances easily ranging to a full rotation. *Right panel*: Angular displacement of the tracer particles (b)-(f) as a function of time.

Shlesinger et al. [97]. Many references for further reading can be found from their paper.

Topological constraints are another common origin for anomalous diffusion mechanisms. The linear particle diffusion along a fast-folding polymer chain [101] is an example for this phenomenon. For an enzyme passively searching for a target site on the polymer, it can be a reasonable strategy to switch between a bounded and an unbounded state. By this, periods of one-dimensional search alternate with periods of free motion in the bulk. Given that the polymer is closely packed, this strategy is advantageous, as it allows the enzyme to occasionally cover large distances with respect to the one-dimensional polymer coordinate. Under certain circumstances (see Ref. [101]), the motion along the chain can bear Lévy flight characteristics. Notice that in this case, the infinite variance of displacements does not contradict fundamental physical principles, since it is simply an effect of the special geometry of the search space.

Finally, there is a deep connection between Lévy flights and continuous time random walks, which will be the focus of the following section. A large portion of the theory on

symmetric stable random variables worked out above can be easily extended and adopted to fit the needs of continuous time random walk theory. Many of the effects encountered here are then to be re-interpreted in the new context. Given these strong parallels, it is not exaggerated to say that the theory of Lévy flights, and of stable processes in general, lies at the heart of any application of continuous time random walk models.

2.2 Continuous time random walks

Continuous time random walks (CTRWs) are – the name suggests it – defined in terms of a microscopic random walk scheme. The position $X(n)$ is a sum of random, IID step lengths $\delta X_n = X(n) - X(n-1)$, $n \geq 1$. The common step length PDF is $\lambda(\delta x)$. Lévy flights resist the attractiveness of Gaussian distributions due to an infinite step length variance σ^2. Continuous time random walks approach anomalous diffusion from a different angle. A key concept in the study of scaling limits is the introduction of a microscopic time scale τ, characterising the time it takes to accumulate a noise impact of the order of δX_n. But in many systems, noise shocks do not occur at a perfectly constant, deterministic rate (e.g. the collisions with a molecule from the solvent, or a propagating wave meeting one of the randomly distributed scattering sites).

A CTRW is specifically designed to account for such a randomness of the microscopic time scale. At each step of the random walk, the generation of a random step distance δX_j is accompanied by the drawing of a random *waiting time* δT_j. The latter are assumed to be IID with a common waiting time PDF $\psi(\delta t)$. For ordinary random walks or Lévy flights the time coordinate is related to the number of steps via the deterministic identity $t = n\tau$. In contrast, for a continuous time random walk, the *real* or *laboratory time* as measured by the observer is a sum of random waiting times $T(n) = \sum_{j=1}^{n} \delta T_j$. This has to be distinguished strictly from the number of steps $n \in \mathbb{N}_0$, also referred to as the *internal time* of the stochastic processes $\{X(n)\}$ and $\{T(n)\}$.

While this additional account for the randomness of noise time scales, for many applications, makes the random walk approach more realistic, it for sure also makes it more complicated to analyse. So is it worthwhile? Intuitively, one might expect that on time scales large as compared to the average waiting time $\langle \delta T_j \rangle$, the fluctuations around the mean become irrelevant. In other words, the scaling limit of the CTRW is the same as for an ordinary random walk with fixed scale $\tau = \langle \delta T_j \rangle$. Indeed, this conjecture can be made on more solid grounds of limit theorems (see the following section), which a posteriori legitimates the concept of *the* microscopic time scale τ, implicated by ordinary random walk approaches. For example, Smoluchowski carefully refers to ν as the "average collision rate", but in his random walk model, section 1.3.2, he uses a non-random $\tau = \nu^{-1}$.

But here again, an observant statistician, acquainted with the concepts of scale-free Lévy flights, may bring up an objection. What if $\langle \delta T_j \rangle$ is infinite? Given the complexity of typical noisy systems, it is not obvious that there exist upper bounds for relaxation time scales, or that one can properly define an average scale, or, even if so, that this scale is small as compared to the observable time scales. The prototype waiting time

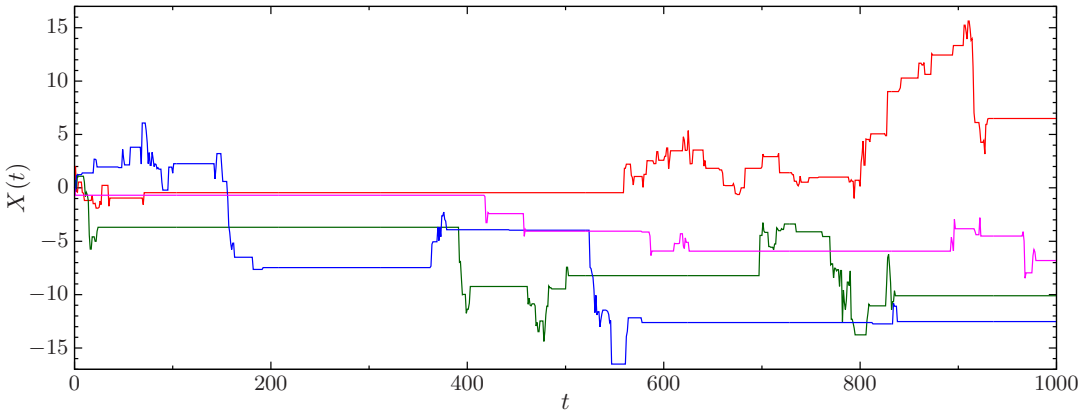


Figure 2.4: Sample paths $X(t)$ of *continuous time random walks (CTRWs)*. The spatially continuous motion is paused for scale-free waiting periods. This considerably slows down the exploration of space as compared to ordinary Brownian motion. Waiting times are heavy-tailed (see text) and thus appear on all time scales, but they are also mutually independent.

PDFs for the study of and infinite mean waiting time are the heavy-tailed functions, with asymptotics

$$\psi(\delta t) \sim B_\alpha \tau^\alpha (\delta t)^{-1-\alpha}, \qquad 0 < \alpha < 1 \tag{2.21}$$

for large waiting times $\delta t \gg \tau$. The prefactor is $B_\alpha = 1/|\Gamma(-\alpha)|$, and τ is from here on coined a *typical* microscopic time scale. All moments $\langle (\delta T_j)^q \rangle$ with $q \geq \alpha$ diverge, which generally includes the mean value. The following section is devoted to the study of this class of CTRW processes and the involved scaling limits.

2.2.1 Flying in time

Continuous time random walks will be subject to an in-depth study in chapter 4.2. Therefore, the analytic details are largely omitted here, and the discussion focuses on the phenomenology. In order to understand the mechanism by which heavy-tailed time scale statistics induce a departure from the familiar random walk behaviour, it is instructive to first look at the non-heavy-tailed case.

As a simple example, let $\psi(\delta t)$ be the PDF for the IID waiting times δT_j, and let both the average $\tau = \langle \delta T_j \rangle$ and variance $\omega^2 = \langle (\delta T_j)^2 \rangle - \tau^2$ be finite. For example, waiting times may be discrete and non-random, or they might be Poissonian. Since the laboratory time $T(n)$ is a sum of the IID random variables δT_j, the central limit theorem applies, see section 1.5.3. For large n, the distribution of $T(n)$ can thus be approximated by a Gaussian. Notice that the mean drift increases linearly with n, $\langle T(n) \rangle = \tau n$; at the same time, the fluctuations around the mean, as measured by the standard deviation, only increase like the square root of time, $\sqrt{\langle T^2(n) \rangle - \tau^2 n^2} = \omega\sqrt{n}$. Indeed, in this

case, fluctuations cease to be relevant for large n, and one might as well approximate $T(n) = \tau n$, or respectively, $\delta T_j = \tau$.

Now turn to heavy-tailed waiting time statistics of the form (2.21). They fall outside the scope of the standard limit theorem, since their average and mean values diverge, and τ is merely a "typical" time scale. The generalised central limit theorem presented in section 2.1.1 for Lévy flights does not apply either, since it is concerned with symmetric random variables. The waiting times δT_j are of course to be restricted to positive values. Such random variables, and also the respective distributions, are said to be *one-sided* or *totally skewed*. Fortunately, Gnedenko and Kolmogorov also accounted [87] for the skewed, heavy-tailed random variables. The respective limit theorem is analogous to the symmetric case. A quasi-continuous parameter $s = n\tau$ is introduced to measure the internal time for the scaling limit process $T(s)$. Remarkably, it turns out that the emerging scaling limit is a *one-sided Lévy α-stable motion*, $T(s) = L_\alpha^+(s)$. Or, to use a more neat, descriptive phrase: the laboratory time of a CTRW exhibits a Lévy flight in the positive direction.

The specific one-sided flights encountered here have many properties in common with their symmetric relatives discussed in the previous sections. Figure 2.4 provides a good impression of what this distinct model ingredient can do to the resulting random walk motion. While the Lévy flights of Fig. 2.1 are conspicuous by large scale displacements, the CTRW trajectories are distinct in terms of extremely long waiting time periods. The most relevant properties of one-sided Lévy flights are listed in the following. While mathematical derivations are omitted, appropriate interpretations in the context of CTRWs are added.

Distribution and scaling. The limiting PDF $g_\alpha(t;s)$ for the laboratory time $T(s) = L_\alpha^+(s)$, a one-sided Lévy flight, is given by

$$g_\alpha(t;s) = (\kappa_\alpha s)^{-1/\alpha}\, \ell_\alpha^+ \left[t(\kappa_\alpha s)^{-1/\alpha} \right] \tag{2.22}$$

in terms of terms of a standard *one-sided α-stable law* ℓ_α^+ and the parameter $\kappa_\alpha = \tau^{\alpha-1}$. Hence, the scaling of the laboratory time is $T(s) \sim s^{1/\alpha}$, which is faster than the linear scaling in case of finite mean waiting times. Notice that the support for the PDFs of all time-related random variables are of course the positive real numbers. It is thus natural to represent the stable PDF ℓ_α^+ in terms of it *Laplace transform*, in place of the characteristic function:

$$\left\langle \exp\left\{ -\theta L_\alpha^+(1) \right\} \right\rangle = \int_0^\infty e^{-\theta t} \ell_\alpha^+(t) \mathrm{d}t = \exp\left(-\theta^\alpha\right). \tag{2.23}$$

To no big surprise, the distribution is heavy-tailed, $\ell_\alpha^+(t) \simeq t^{-1-\alpha}$, so that the expectation value (and higher order moments) of the laboratory time diverges, $\langle T(s) \rangle = \int_0^\infty t g_\alpha(t;s) \mathrm{d}t = \infty$. Note that in the limit $\alpha \to 1$ the PDF in Eq. (2.22) becomes a Dirac δ-distribution, $g_1(t;s) = \delta(t-s)$. Thus, this limiting case restores the equivalence of internal and laboratory time, such that the particle motion is no longer paused for random waiting time periods.

Scale-free waiting times, first passage times and leap-over statistics. In close analogy to the effects of scale-free jump lengths for symmetric Lévy flights, the infinite first moment of the waiting times amounts to immobilisation periods on all time scales, see Fig. 2.4. This calls for a careful study of first-passage and leap-over behaviour.

The first-passage time of $T(s)$ is for every $t > 0$ defined through

$$S(t) = \inf\{s > 0 : T(s) > t\}, \tag{2.24}$$

and plays a very important role in CTRW theory. Recall that $T(s)$ is constructed in terms of a sum of positive random variables δT_n. Therefore, $T(s)$ increases strictly with s. When, at a given laboratory time $t > 0$, the observer measures the state variable of the CTRW, the process $T(s)$ has "passed the boundary t". In this sense, the process $S(t)$ is the *inverse process* of the laboratory time process $T(s)$: it measures the internal time S passed (implicitly, the number of steps executed) at a given laboratory time t. The PDF $h_\alpha(s;t)$ of the internal time $S(t)$ at a given instant t is (see section 4.1, or Refs. [96, 102, 103])

$$h_\alpha(s;t) = \frac{1}{\alpha} t(\kappa_\alpha s)^{-1/\alpha-1}\, \ell_\alpha^+ \left[t(\kappa_\alpha s)^{-1/\alpha}\right]. \tag{2.25}$$

One finds the scaling $S(t) \sim t^\alpha$, consistent with $T(s) \sim s^{1/\alpha}$. In contrast to the symmetric case, the mean first-passage time is finite.[6] But arguably the most important observation is that $S(t)$ has a nontrivial distribution. Recall that for waiting times with finite mean, there is a deterministic relation $t = \tau n = s$ in the scaling limit of large n. One can recover such a behaviour from Eq. 2.25 by letting α tend to 1: $h_1(s;t) = \delta(s-t)$. Hence, randomness of internal time is intimately connected to scale-free waiting times.

The leap-over time, defined by

$$T_1(t) = T(S(t)) - t, \tag{2.26}$$

is also called *forward recurrence time* in the context of CTRW or ageing renewal theory, discussed in chapter 4. Its interpretation is as follows. The CTRW was constructed here such that all dynamic activity commences at time $t = 0$. Consequently, the PDF for the occurrence time of the very first random walk step is given by $\psi(\delta t)$. If an observer starts monitoring the motion not at time 0, but at a later time t_a, then he probably will also have to wait some time period $T_1(t_a)$, until the first *observed* step occurs. However, since the respective previous step might already lie some time in the past, the PDF of $T_1(t_a)$ is *not* $\psi(t_1)$. Instead, it is given by [96, 104, 105, 106, 107]

$$h(t_a, t_1) = \frac{\sin(\pi\alpha)}{\pi} \frac{t_a^\alpha}{t_1^\alpha (t_a + t_1)}. \tag{2.27}$$

[6] The apparent difference to the symmetric Lévy flight or Brownian motion is that the latter is not bounded. The random walker may hence diffuse away arbitrarily far from the first-passage bound. The strictly increasing process $T(s)$ obviously cannot do this, and must hit the "bound" t eventually. Indeed, for Lévy flight or Brownian motion bounded by a reflective wall, the mean first-passage time is also finite.

Thus, the PDF of the forward recurrence time is heavy-tailed, with the same index α that characterises all other waiting times. But the characteristic parameter κ_α (and hence the typical time scale τ) does not appear in the equation. Instead, there is a scaling dependence $T_1(t_a) \sim t_a$. Remarkably, the later an observer starts to look at the motion, the longer he will have to wait to actually witness the onset of activity. This property has severe consequences for the CTRW behaviour, as shown below.

2.2.2 Subordination

The previous discussion focused on the evolution of the laboratory time process $T(n)$, or rather its scaling limit $T(s)$, which was identified as a one-sided Lévy α-stable motion. How can one relate these findings to the actual random walk process? The difficulty of this task depends on the specific model under consideration. The critical question is whether or not the microscopic waiting times δT_n and step lengths δX_n can be treated as mutually independent random sequences. An interesting model class where this is not the case, are the Lévy walks. There, waiting times are actually interpreted as travel times, and are thus coupled to displacement lengths via a more or less complicated velocity relation. Lévy walks are not within the scope of this work. The interested reader can find related references in section 2.1.4.

The present discussion is meant to be a case study on the effect of scale-free waiting times. Consider hence the most simplistic model where waiting times and step distances are mutually independent. Then also the discrete time random walks $X(n)$ and $T(n)$ are independent processes, and so are the respective scaling limits, $X(s)$ and $T(s)$. With the internal time process $S(t)$ as defined above, the pair $X(s)$, $S(t)$ is independent, too.

Consider now the combined CTRW motion[7]

$$Y(t) = X(S(t)), \tag{2.28}$$

i.e. the random, unsteady progression of internal time is modelled by $S(t)$, while independently the spatial displacements during times of dynamic activity are governed by the process $X(s)$. Such a way of combining two random motion is called *subordination* [103, 108]. Ensemble statistics combine the distributional properties of both independent random processes $X(s)$ and $S(t)$. For instance, let $h_\alpha(s;t)$ denote the PDF for internal time S at laboratory time t. Similarly, $p(x;s)$ is the PDF of the position X when the internal time s has passed. Then the PDF $p_\alpha(y;t)$ for the CTRW position Y at time t can be computed as

$$p_\alpha(y;t) = \int_0^\infty p(y;s)\, h_\alpha(s;t)\, ds. \tag{2.29}$$

[7] Equation (2.28) directly subordinates the scaling limit processes $X(s)$ and $S(t)$. But one might also first combine the two time-discrete random walk based processes $X(n)$ and $N(t)$ (the latter being the inverse of $T(n)$), and only then carry out the scaling limit. It is not a priori clear that the result is the same. The distinction is particularly important when it comes to sample path properties. An example where the equivalence can actually be established mathematically can be found in Ref. [103].

Not knowing how much internal time S has passed up to time t, the integral averages over all possible values. The independence of the processes $X(s)$ and $S(t)$ is reflected explicitly in the factorisation into probabilities p and h_α, appearing in the integral. For calculating multipoint PDFs like $p_\alpha(y_1, y_2; t_1, t_2)$ one needs to use the multipoint versions of p and h_α [102, 109].

2.2.3 Fingerprint characteristics of scale-free relaxation times

With CTRW processes being the major topic of section 4.2, the discussion here is limited to a phenomenological list of facts, including intuitive explanations.

Subdiffusion. An immediate effect of scale-free waiting times is the slowing down of diffusion. To see this, consider an ordinary Brownian motion process $X(s) = W(s)$ and subordinate to it an independent internal time process $S(t)$, based on heavy-tailed waiting time statistics with tail index $0 < \alpha < 1$. One then has $X(s) \sim s^{1/2}$ and $S(t) \sim t^\alpha$. It follows directly from Eq. (2.29), that in this case $Y(t) = X(S(t)) \sim t^{\alpha/2}$. Hence, the waiting time process has slowed down the motion from normal diffusion to subdiffusion behaviour.

Ageing. A CTRW $Y(t)$ is an *ageing* process in the following sense. Assume the process starts at time 0 at position $Y(0)$, and at the same instant, an observer starts recording the motion up to time t. His time window of observation is thus $[0; t]$. He does this several times, analyses the data and calculates the ensemble statistics. A second observer looks at the same process, but starts collecting information only after some time delay $t_a > 0$. For him, the origin of motion is $Y(t_a)$, and the observation period is $[t_a; t_a + t]$. If $Y(t)$ is an ordinary Brownian motion (Wiener process), then both observers will obtain the same statistics. The increments of Brownian motion are stationary in distribution: the statistics are invariant to the time shift t_a. The same holds if $Y(t)$ is a Lévy flight. For a CTRW with scale-free waiting times, the situation is different. The two observers, monitoring different time windows, will get different results. This is referred to as non-stationary behaviour. The cause for it lies in the special character of the forward recurrence time $T_1(t_a)$ of the internal time process $S(t)$, as alluded to in section 2.2.1. The late observer starting at $t_a > 0$ has to initially wait a time $T_1(t_a)$ before the random walk is set in motion. $T_1(t_a)$ is statistically very different from all other waiting times, see Eq. (2.27). Since it tends to increase with t_a, a late observer is inclined to diagnose a reduced dynamic activity, an effect referred to as *ageing*. The precise ageing effect for CTRW, and the broader class of ageing renewal processes, is the topic of chapter 4.

Weak ergodicity breaking. Consider now a stationary process $Y(t)$. This does not mean that a single realisation of the process is stuck $Y(t) = \text{const}$; but the statistics of the random variable $Y(t)$ do not depend on time. For example, the motion of a Brownian particle, being confined by an external binding potential $U(y)$, relaxes towards the Boltzmann equilibrium: an ensemble of independent particles is eventually distributed

with PDF $p(y;t) \equiv p(y) = Z^{-1} \exp[-U(y)/(k_B \mathcal{T})]$, where Z is a normalisation constant, the *partition function*. Assume an experimenter is interested in measuring this PDF. To do this, he basically has two options:

The first is to consider the *ensemble statistics*. For this, a large ensemble of trajectories needs to be generated – either by repeatedly executing the measurement in identically prepared systems, or by immersing many particles at once, making sure that mutual interaction is negligible. From this, one can calculate the *ensemble PDF*

$$p^{\mathrm{ens}}(y)\,dy = \frac{\text{number of trajectories terminating between } y \text{ and } y+\,dy}{\text{total number of recorded trajectories}}. \tag{2.30}$$

An average of some observable $f(y)$, derived from the ensemble distribution through $\langle f(Y(t))\rangle = \int f(y)\, p^{\mathrm{ens}}(y)\, dy$, is called an *ensemble average*. The quality of ensemble statistics improves with the size of the ensemble.

An alternative notion are the *occupation time statistics*. From a single trajectory, recorded over a long measurement time, one may calculate the *occupation time PDF*

$$p^{\mathrm{time}}(y)\,dy = \frac{\text{amount of time spent between } y \text{ and } y+\,dy}{\text{total measurement time}}. \tag{2.31}$$

Associated is the *time averaged* observable $\overline{f(Y(t))} = \int f(y)\, p^{\mathrm{time}}(y)\, dy$. The accuracy of occupation time statistics is determined by the total time of measurement.

The process is said to be *ergodic*, if both types of statistics are equivalent, provided their quality is sufficiently good. The underlying hypothesis is that a single realisation, if given enough time, samples the full space of microstates many times, effectively generating by itself the ensemble statistics. Brownian motion is an example of an ergodic process. However, if the mean sojourn time in microstates (e.g. the waiting times of a CTRW) diverges, long time and ensemble averages are not equivalent. While a single process realisation may in principle have access to the complete state space, it takes an infinitely long time to fully explore it. This phenomenon is sometimes called *weak ergodicity breaking* [110].

More concretely, if the waiting times of a CTRW are scale-free, then the statistics of a single trajectory $Y(t)$, observed during $t \in [0;T]$, are being dominated by single, exceptionally large waiting times of the order of T. No matter how large the measurement time T, the resulting occupation time statistics are determined by only a few microstates where the trajectory stayed for an extraordinary amount of time. The occupation times vary largely from one trajectory to the next. Consequently, time averages are inherently random, a seemingly counter-intuitive effect which cannot be reduced by increasing the measurement time T.

2.2.4 Motivation

The following two examples are intended to give an intuitive understanding of how the ubiquitous power-laws inherent to continuous time random walk models may emerge from basic microscopic principles. The first describes a particles diffusing along some

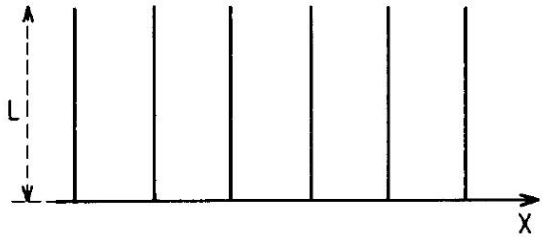


Figure 2.5: Diffusion on a comb-like structure [82]. The particle diffusion is normal, i.e. Brownian, along all sections of the comb. The motion in the x-direction is paused whenever the particle enters one of the spikes. If the spikes extend indefinitely, $L = \infty$, the mean time spent in a spike is infinite, so that the motion along the x-axis is a CTRW.

channel, with the possibility of getting lost in outgoing long, one-way side-arms. The model is likely too simple to describe any realistic system, but gives an impression on the effect of topological traps. The second example deals with the diffusion in a random environment. While being motivated specifically by the study of glassy materials, it generally provides information on how "much" random the characteristic relaxation scales need to be in order to indeed produce scale-free dynamics.

Diffusion on comb-like structures. The following problem can be found in Ref. [82], which is anyway a recommended lecture, as it provides an extensive and understandable overview on contemporary anomalous diffusion approaches and physical applications. The diffusion along a comb-like structure is pictorially represented in Fig. 2.5. Assume the dynamics of interest is the transport along the horizontal x-direction. The basic linear diffusion is taken to be normal Brownian in the whole structure. However, with respect to the motion along the x-axis, diffusion is paused whenever the particle happens to enter one of the perpendicular branches, the "teeth" of the comb. It stays there for a random waiting time δT. If the teeth are taken to be infinitely long, $L = \infty$, then δT is the first return time at the origin (the fork at the entrance of the tooth) of a one-dimensional Brownian motion. The associated PDF is similar to that of the first passage time, Eq. (2.17). In particular, its large time asymptotics are of the same form $\simeq t^{-3/2}$ [82, 60]. Hence, the diffusion transport as projected onto the x-axis is indeed anomalous, namely a simple CTRW with divergent mean waiting time. The heavy-tail parameter in this case is $\alpha = 1/2$. But if the geometry of the branching structures is more complicated than depicted here, other types of power-laws may be encountered. The important lesson here is that anomalous diffusion in general, and CTRW-like trapping behaviour in particular, can emerge from projecting a topologically complicated motion onto a single diffusion coordinate.

Random energy landscapes. Introduced originally in Ref. [111] in the context of relaxation dynamics of glass systems, the random energy landscape or quenched trap models relate scale-free temporal dynamics with disorder. Consider a two- or higher-dimensional random walk, for simplicity, on a lattice structure. Hopping transitions occur between nearest neighbour sites only, with equal probabilities. With each lattice site i, associate an energy $E_i \geq 0$. The latter quantifies the depth of a local potential minimum. The binding potential might represent the interaction with a non-homogeneous lattice structure or the involvement into local chemical reactions. The effect is that the particle, once arriving at a new lattice site, is bound in place until it escapes the potential trap by means of thermal excitation. The escape process can be assumed to be Poissonian. The average trapping time is given by the Arrhenius law $\langle \delta T(E_i) \rangle = \Gamma_0 \exp[E_i/(k_B \mathcal{T})]$, where k_B is the Boltzmann constant, $\mathcal{T}$ denotes temperature, and the rate factor $\Gamma_0 > 0$ is approximately independent of temperature.

Notice that with trapping times being energy- and thus site-dependent, the quenched trap model contrasts the assumption of IID waiting times made for the CTRW models discussed in the previous. With that being said, unbiased, simple random walks in two or higher dimensions are *transient* (Pólya's theorem [3]): the probability of returning to any specific previously visited cite tends to zero at large times. This suggests that, on large time scales, the depth of the traps encountered by the particle are "almost independent". In other words, one can approximate the quenched trap model by a coarse-grained, annealed disorder model, where the trap depths E_i are mutually independent, and site-independent random variables, drawn from a common PDF $\rho(E)$. A natural form for this distribution in glassy materials is a Poissonian law [111]: $\rho(E) = (k_B \mathcal{T}_0)^{-1} \exp[-E/(k_B \mathcal{T}_0)]$, where the critical temperature $\mathcal{T}_0$ is specific for the glass material. In the annealed model approximation, the random walker resides on any given site for a random waiting time δT, which is determined by first drawing an energy E from $\rho(E)$, and then drawing a waiting time $\delta T(E)$ from a Poissonian law $\psi(\delta t; E)$ with the average given by the Arrhenius law above. The waiting time PDF, as an average over disorder, is therefore given by

$$\begin{aligned}
\psi(\delta t) &= \int_0^\infty \rho(E)\, \psi(\delta t; E)\, dE \\
&= \int_0^\infty \frac{e^{-E/(k_B \mathcal{T}_0)}}{k_B \mathcal{T}_0} \frac{e^{-E/(k_B \mathcal{T})}}{\Gamma_0} \exp\left[-\frac{e^{-E/(k_B \mathcal{T})}}{\Gamma_0} \delta t \right] dE \\
&= \Gamma_0 \frac{\mathcal{T}/\mathcal{T}_0}{(\Gamma_0 \delta t)^{1+\mathcal{T}/\mathcal{T}_0}} \int_0^{\Gamma_0 \delta t} y^{\mathcal{T}/\mathcal{T}_0} e^{-y}\, dy \\
&\simeq \Gamma_0^{-\mathcal{T}/\mathcal{T}_0}\, \delta t^{-1-\mathcal{T}/\mathcal{T}_0},
\end{aligned} \tag{2.32}$$

where the substitution used is $y = \delta t\, \Gamma_0 \exp[-E/(k_0 \mathcal{T})]$, and the asymptotics are for $\delta t \gg \Gamma_0^{-1}$. Hence, by comparison with Eq. (2.21), the waiting time distribution is heavy-tailed for temperatures below the critical temperature, $\mathcal{T}/\mathcal{T}_0 < 1$. The resulting random walk dynamics are therefore qualitatively described by a CTRW. This is remarkable,

because both the average depth of a trap $\langle E \rangle$ and the average waiting time in a given trap $\langle \delta T(E) \rangle$ are finite. But as time progresses, the particle probes its environment, averaging over the disorder found in the energy landscape, which *effectively* produces infinite mean trapping times.

Interpreting a CTRW as a random walk in a random energy landscape helps a lot in acquiring an intuitive understanding of its anomalous characteristics: subdiffusion, ageing, weak ergodicity breaking and other phenomena emerge since the particle continuously finds deeper and deeper traps in its environment. One should notice however, that disorder is not a generally sufficient condition for scale-free behaviour. For instance, for temperatures above $\mathcal{T}_0$, the average waiting time is finite. The distribution of energies studied here is exponential. For a Gaussian distribution, the resulting dynamics are ultimately normal at large time scales, but with an intermittent regime of *iterrupted aging*, see Ref. [111]. Finally, a random walk is not transient if the dimension of the embedding space is less than two, or in the presence of a sufficiently strong bias [3]. In this case, the annealed disorder approximation is not applicable. While diffusion in the quenched trap model can still be highly anomalous in these cases, it cannot be represented by an effective CTRW [112].

2.2.5 Applications

CTRWs were originally proposed as a model for charge carrier transport in amorphous semiconductors [76]. In such materials, individual charge carrier reside on specific, randomly and sparsely distributed acceptor sites random time spans, before hopping to a neighbour site. The system therefore falls into the category of random energy landscapes with quenched disorder. The authors put considerable effort into justifying the use of an effective CTRW description. Their theory is able to explain the anomalous properties of the photocurrents, as measured experimentally for semi-conducting materials used in contemporary photocopying machines [77].

Moreover, networks of entangled polymer filaments as found in mammalian cells can cause caging effects for micron-sized objects, as demonstrated in in vitro experiments [113]. The diffusion of a colloidal bead immersed in a semi-dilute solution of polymeric filaments is constrained by the surrounding network structure. The motion characteristics strongly depend on the ratio of the bead size and the average mesh size of the network. If both are of comparable order, the particle "jumps" between between cage-like micro-environments. In this regime, the residence time in cages is indeed heavy-tailed. The reason for this may be conjectured to be the power-law correlations in the network fluctuations (cf. section 2.3.4).

Indeed, CTRW was identified as diffusion process also in living cells [114, 115]. Similarly, the model applies to multiscale trapping times of particles on sticky surfaces [116] and it has been successfully applied to many other physical and geological problems [83, 86].

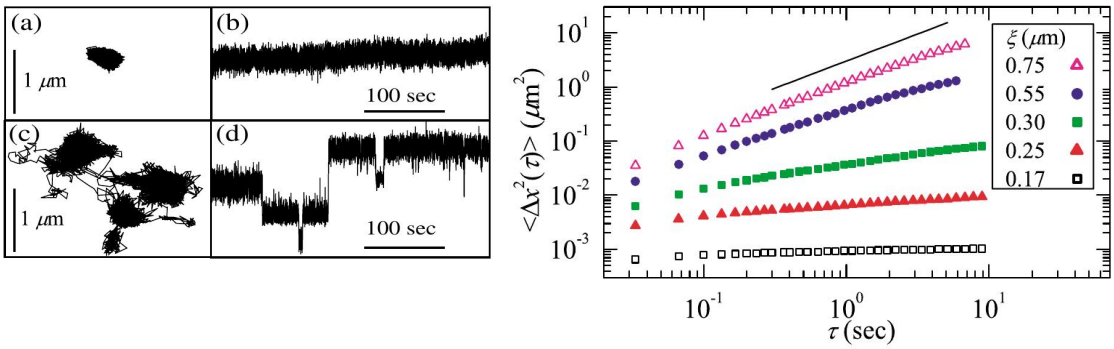


Figure 2.6: Diffusion of a bead in a polymeric network [113]. The network is formed from entangled, semi-flexible actin filements, and has an average mesh size of ξ. *Left*: Images of two individual bead trajectories. (a) and (c) depict the motion as projected onto the x-y-plane, while (b) and (d) plot x versus time. The bead in (a),(b) is trapped in network cage for the full duration of observation. (c),(d) show a bead which jumps between local micro-environments. *Right*: Mean square displacement versus time for a bead of radius 0.25 μm. Depending on the average mesh size ξ (see key), the diffusion ranges from fully confined over subdiffusion to normal.

2.3 Fractional Brownian motion

For standard CTRW and Lévy flight models, the anomaly with respect to ordinary Brownian motion comes through the special scale-free statistics defining jump lengths or waiting times or both. The *fractional Brownian motion* (FBM) describes a totally different sort of anomaly.

Individual noise impacts may be "small", in the sense that fluctuations can be quantified by a small, finite variance σ^2; they may also happen "fast", in the sense that they are separated by a well-defined, microscopic, average time scale τ. Both preliminaries do not rule out in principle the presence of notable, long-lasting correlations within the noise dynamics. After all, he term "noise source" is typically a placeholder for a large collection of individual actors (molecules, charges, spins, humans,...) who produce the noisy signal. While the dynamics of individuals might be fast, and the single effect be small, correlations can induce *collective* dynamics, which vary slowly and have a macroscopically observable effect.

To see the immediate consequences of such a memory component in a diffusion model, study a simple random walk $X(n) = \sum_{j=1}^{n} \delta X_j$, where the random step lengths δX_j are unbiased and identical in distribution with finite variance σ^2. Consider now three cases. (i) Normal, memory-less diffusive behaviour is generated if step lengths are, in addition, statistically independent: in this case, by virtue of the central limit theorem, the distribution of $X(n)$ approaches a Gaussian for large n, and the mean squared

displacement evolves according to $\langle X^2(n)\rangle = \sigma^2 n$. (ii) An extremely persistent random walk is obtained by requiring that all step lengths take exactly the same value. Thus, only the first step length δX_1 is chosen at random, any subsequent step has the same length and goes in the same direction, $\delta X_{j+1} = \delta X_j$ for all $1 \leq j \leq n$. In this case, $X(n) = n\,\delta X_1$, so the distribution is a rescaled copy of the (possibly non-Gaussian) step length distribution. The "diffusive" behaviour here is ballistic, $\langle X^2(n)\rangle = \sigma^2 n^2$. (iii) In a way the opposite case is a random walk with forced oscillations: all step lengths are identical in absolute value, but the signs alternate, $\delta X_{j+1} = -\delta X_j$. Now, there is no net transport, since the random walk constantly oscillates between the values 0 and δX_1.

In general, the behaviour of a random walk with somehow correlated step lengths can be expected to be intermediate between the three extreme cases (i)-(iii). For *mixing* systems in particular (see section 1.5.3 and references therein), the random walk's scaling limit is still Gaussian, and the diffusion characteristics are anomalous with respect to the dependence structures and the scaling with time. The FBM is constructed as an ideal candidate for the diffusion limit of such anomalous motion. Its formal definition is provided in section 2.3.3. First, it is useful to acquire further understanding of correlation effects and to introduce some basic terminology to describe them.

2.3.1 Assessing memory

In the following, two concepts to characterise dependency are introduced: persistence and dependence range. This is done most conveniently on the level of discrete random walk processes, $X(n) = \sum_{j=1}^{n} \delta X_j$. Although the principal ideas are very general, the precise definitions given here are formulated with respect to finite variance statistics. Thus, for preparation, let $\sigma_j^2 = \langle (\delta X_j)^2\rangle < \infty$ and introduce the *correlation* or *covariance function* $c(j,k) = \langle \delta X_{j+k}\delta X_j\rangle$ of step lengths (a footnote on page 19 contrasts the inconsistent usage of vocabulary in physics and mathematics literature). In addition, for simplicity, let the random walk be unbiased, $\mu_j = \langle \delta X_j\rangle = 0$.

Persistence and anti-persistence. Loosely speaking, whenever certain motion patterns observed in the past tend to be repeated in the future, one speaks of a persistent process. Conversely, an anti-persistent memory seeks to undo previous actions. A common tool designed to detect this kind of dependence is the *Pearson correlation coefficient*. For two random variables, say, two steps of a random walk δX_j and δX_k, it is defined as

$$r(j,k) = \mathrm{Corr}[\delta X_j \delta X_k] = \frac{\langle (\delta X_j - \mu_j)(\delta X_k - \mu_k)\rangle}{\sigma_j \sigma_k} = \frac{\langle \delta X_j \delta X_k\rangle - \mu_j \mu_k}{\sigma_j \sigma_k}. \tag{2.33}$$

For a particular sample of δX_j and δX_k, the product $(\delta X_j - \mu_j)(\delta X_k - \mu_k)$ is positive if both variables fall on the same side of the respective value. For instance, the case of interest here is $\mu_j = \mu_k = 0$, so the product is positive if and only if both random steps go in the same direction. Moreover, the correlation coefficient is normalised such that $-1 \leq r(j,k) \leq 1$. If $r(j,k) = \pm 1$, one says the variables are perfectly correlated;

for $r(j,k) = +1$, the correlation is persistent or *positive*; respectively anti-persistent or negative if $r(j,k) = -1$.

The correlation coefficient has the following advantages and limits. First, it can proof the existence of dependence. If $r(j,k) \neq 0$, then the two variables are statistically dependent. However, the converse is does not hold. Namely, if $r(j,k) = 0$ there can still be dependence. This is because the correlation coefficient only measures *linear* dependence. For example, if $\delta X_j = a\delta X_k + b$, for some constants a, b, then $r(j,k) = \text{sign}(a)$. But if δX_j is symmetric and $\delta X_k = \delta X_j^2$, then $\langle \delta X_j \delta X_k \rangle = \left\langle \delta X_j^3 \right\rangle = 0$, for symmetry reasons. Another counterexample is, ironically, given by the X- and Y-coordinates of Pearson's walk in the plane, see section 1.1. Since $X = \cos(\theta)$, $Y = \sin(\theta)$, the two variables are clearly dependent. But with θ uniformly distributed on $[-\pi; \pi)$, $\langle XY \rangle = (2\pi)^{-1} \int_{-\pi}^{\pi} \cos(\theta)\sin(\theta)\, d\theta = 0$, again, due to symmetry.

Fortunately, for jointly Gaussian random variables – fractional Brownian motion is defined as a Gaussian process – the correlation coefficient is an excellent measure of dependence. One can show that two jointly Gaussian random variables are independent if and only if the $r(j,k) = 0$. In the following discussion, mainly the sign of $r(j,k)$ will matter (indicating negative or positive or no dependence). Therefore, the non-normalised version $c(j,k) = \sigma_j \sigma_k\, r(j,k)$ will be used, while still being referred to as correlation coefficient or function.

Short- and long-range dependence. While the sign of the correlation function $c(j,k)$ indicates the basic nature of dependency (persistence or anti-persistence), it does not provide information on the relevance of the effect. For instance, while a Brownian particle must in principle exhibit persistent motion due to inertia at short time scales, the dynamics which have actually been observable experimentally for decades are perfectly well approximated by the memory-less Wiener process. Thus, a quantitative measure for the temporal range of correlations is desirable.

The following definition is in wide use in the context of diffusion models, see Ref. [82] for examples from physics. The property of *long-range dependence*, also *long memory*, is usually defined with respect to random motions with stationary increments, i.e. the sequences $\{\delta X_j\}_{j\in\mathbb{N}}$ and $\{\delta X_{j+k}\}_{j\in\mathbb{N}}$ obey the same probability laws for any $k \in \mathbb{N}$; non-stationary increments are addressed below. The "relevance" of memory effects is then assessed with respect to anomalous diffusion behaviour. More precisely, the decisive measure is the mean squared displacement asymptotics. Since stationary increments

imply $\sigma_j^2 \equiv \sigma^2$ and $c(j,k) \equiv c(|k|)$, one can write

$$\begin{aligned}\left\langle X^2(n)\right\rangle &= \left\langle \left[\sum_{j=1}^{n} \delta X_j\right]^2 \right\rangle = \sum_{j=1}^{n} \left\langle (\delta X_j)^2 \right\rangle + \sum_{j,k=1;\, j+k\le n}^{n} \langle \delta X_j \delta X_{j+k} \rangle \\ &= \sigma^2 n + 2\sum_{j=1}^{n}\sum_{k=1}^{n-j} c(k) = \sigma^2 n + 2\sum_{k=1}^{n}(n-k)\, c(k) \\ &= \sigma^2 n + 2n\sum_{k=1}^{n} c(k) - 2\sum_{k=1}^{n} k\, c(k). \end{aligned} \tag{2.34}$$

This form allows to establish a connection between the long-term diffusive behaviour $\left\langle X^2(n \gg 1)\right\rangle$ and large-lag correlations $c(|k| \gg 1)$.

For example, consider correlations which are asymptotically persistent, $c(|k| \gg 1) > 0$, but decay fast enough as to be *summable*, i.e.[8]

$$\lim_{n\to\infty} \sum_{k=1}^{n} |c(k)| < \infty \tag{2.35}$$

Then the dominant contributions to the mean squared displacement, Eq. (2.34), for a large number of steps n, are the first two terms, both growing linearly with n. Hence, diffusion ultimately becomes normal, $\left\langle X^2(n)\right\rangle \sim \bar{\sigma}^2 n$; correlations solely enter the associated coefficient $\bar{\sigma}^2 = \sigma^2 + 2\sum_{k=1}^{\infty} c(k)$. For this reason, in situations where Eq. (2.35) holds, one speaks of a *short memory* or *short-range dependence*. It occurs for instance when correlations decay exponentially fast, $c(k) \simeq e^{-\lambda k}$ with $\lambda > 0$. But also power-laws $c(k) \simeq k^{-\gamma}$ with $\gamma > 1$, come to mind. Conversely, if the correlations are not summable, i.e. the sum in Eq. (2.35) grows with n indefinitely, then the second term in Eq. (2.34) increases faster than linear with n. It eventually dominates the mean squared displacement behaviour, so from this point of view, the correlations induce anomalous diffusion. Consequently, non-summability of correlations can be used as an indicator of long-range dependence. A simplistic example is the extremely persistent random walk introduced above: since all step lengths are equal, the correlation function is simply $c(k) \equiv \left\langle (\delta X_1)^2 \right\rangle = \sigma^2$, independently of k, which is clearly not summable. But remarkably, anomalous diffusion can arise even when correlations do decrease for large lags k. If the decay is slow enough, e.g. a weak power-law $c(k) \simeq k^{-\gamma}$, $\gamma \le 1$, then the correlations are not summable. This type of persistence drives a superdiffusive motion. The persistent FBM, introduced in section 2.3.3, has exactly this feature.

Thus, the decay behaviour of the correlation function $c(k)$ can be defined as the indicator for short- or long-range dependence. This has become a common habit, mainly because the computation of $c(k)$ is comparatively easy both in terms of theoretical and

[8] The limit in Eq. (2.35) is here assumed to be either finite or $+\infty$ or $-\infty$. Cases where the sequence of partial sums oscillate are not covered by the present discussion. Notice that the latter can happen only if the correlation function $c(k)$ itself oscillates indefinitely.

data analysis. The present work will not break this habit. Still, the definition comes with several caveats, which should not go unnoticed. To start with, the situation is more subtle if random walk steps are asymptotically anti-persistent, $c(k \gg 1) < 0$. In this case, the summability condition (2.35) is automatically fulfilled; otherwise, the mean squared displacement would turn negative eventually. Anti-persistent correlations are hence, using above criterion, to be called short-ranged. However, this does *not* necessarily imply that diffusion is normal, nor that the effect of memory is irrelevant. Notice that the leading order coefficient $\bar{\sigma}^2$ can turn out to be zero, so that for the mean squared displacement, Eq. (2.34), the third term returns the leading order behaviour. For a simple example with exponential decay, consider $c(0) = \sigma^2$ and $c(|k| \geq 1) = -\sigma^2 e^{-\lambda|k|}$ with $\lambda = \ln 3$. A straight-forward calculation, using geometric series formulae, in this case yields $\bar{\sigma}^2 = 0$ and $\left\langle X^2(n) \right\rangle \to 3\sigma^2/2$ as $n \to \infty$. Despite their "fast" exponential decay and "short range", such correlations do confine the diffusion dynamics to stationary limit. A less drastic example is the anti-persistent FBM, introduced in section 2.3.3. It is characterised by negative correlations which decay like a power-law, resulting in subdiffusive dynamics. In general, the only thing one can say for sure about asymptotically anti-persistent memory is that it limits the mean squared displacement $\left\langle X^2(n) \right\rangle$ to increase *at most* linearly for large n.

This kind of complications is in fact typical when it comes to assessing memory effects. It turns out that "memory" is a complex aspect and has multiple facets, which are not easily characterised by single parameters, like the tail exponent of a correlation function. Samorodnitzky provides a thorough treatment [117] of the issue of long-range dependence, including a historical motivation and an extensive overview of several alternative notions of the concept. The section should be closed, however, with an appeasing remark. For FBM, which is the central process of the following discussion, the approach introduced above is applicable and useful. This is, first, because FBM is a Gaussian process, so any information on dependence structures is indeed encoded in the correlation function. Second, it is a self-similar process, which excludes degenerate cases like a stationary or oscillating diffusion limit.

2.3.2 Gaussian random walks

A random walk $X(n) = \sum_{j=1}^{n} \delta X_j$ is called a *Gaussian random walk*, if any d-dimensional random state vector $\boldsymbol{X}(n_1, ..., n_d) = (X(n_1), ..., X(n_d))$ is jointly Gaussian. The simplest case is where the step distances δX_j are IID. The random walk is then said to be uncorrelated or memory-less.[9] An uncorrelated, unbiased Gaussian random walk can be interpreted as the time-discrete version of a Brownian motion.

[9] To be precise, the IID step lengths δX_j are uncorrelated: $\langle \delta X_j \delta X_{j+k} \rangle = \langle \delta X_j \rangle \langle \delta X_{j+k} \rangle$ for all $k, j \in \mathbb{N}_0$. The process $X(n)$ itself is not: clearly, knowing the position of the walker at a time n hints to where to find it at time $n + m$. Still, the common habit is to refer to a random walk $X(n)$ as an (un-)correlated process, whenever its step lengths δX_j have this property. Analogously, a random walk can be said to be (anti-)persistent or to exhibit long- or short-range dependence. The same slight abuse of language can be found in the context of time-continuous processes $X(t)$, where the correlation attribute refers to increments $[X(t + s) - X(t)]$ from non-overlapping time intervals.

Conversely, a random walk $\widetilde{X}(n)$ may carry a certain degree of memory, implying that step distances are not drawn independently. The motion can be characterised by a potential drift $\langle \widetilde{X}(n) \rangle$ and the correlation function $\langle \widetilde{X}(n)\widetilde{X}(n+m) \rangle$. Such a specification is concise and, in the Gaussian case, also complete. However, an alternative definition shall be presented here which helps in understanding the fundamental structure of Gaussian process. The idea is to construct a correlated process from a memory-less one by means of a linear transformation.

Let δX_j be IID Gaussian random variables as above, for simplicity again unbiased. Now introduce a nonrandom function M_j, the *correlation kernel*, and define new step lengths via $\delta\widetilde{X}_k = \sum_{j=1}^{k} M_{k-j+1}\delta X_j$. Recall that Gaussian random variables are stable (see section 2.1.1). Hence, the Gaussianity of the independent variables δX_j carries over to the step lengths $\delta\widetilde{X}_k$. They also remain centred at zero. However, as a first consequence of the linear transformation, the correlated sequence in general becomes non-stationary: the variances $\langle (\delta\widetilde{X}_k)^2 \rangle = \sigma^2 \sum_{j=1}^{k} M_j^2 = \tilde{\sigma}_k^2$ depend on k. In order to generate a step length sequence which is at least stationary in the long time limit, require from here on that the correlation kernel M_j decays with j fast enough as to be square summable, that is

$$\tilde{\sigma}^2 = \lim_{k\to\infty} \tilde{\sigma}_k^2 = \sigma^2 \sum_{j=1}^{\infty} M_j^2 < \infty. \tag{2.36}$$

In addition, it is more convenient to slightly modify the definition of step lengths, namely

$$\delta\widetilde{X}_k = \sum_{j=-\infty}^{k} M_{k-j+1}\delta X_j \tag{2.37}$$

Notice that the sum now starts at $j = -\infty$. The benefit is to obtain a sequence $\delta\widetilde{X}_k$ which is stationary in distribution from the start, $\left\langle (\delta\widetilde{X}_k)^2 \right\rangle = \tilde{\sigma}^2$ for all $k \geq 1$. The definition is meaningful due to the fast decay of the correlation kernel. The drawback is that technically an infinite number of IID variables δX_j is needed to generate even a single step length $\delta\widetilde{X}_k$.

At this point, both sequences δX_j and $\delta\widetilde{X}_k$ are stationary and centred Gaussian. The crucial difference between the two lies in their memory. For the independent variables, one can write $\langle \delta X_j \delta X_{j+k} \rangle = \sigma^2\, \delta_{0k}$, with δ_{jk} denoting the Kronecker symbol. But for

the step lengths $\delta\widetilde{X}_k$, the correlation kernel M_j introduces correlations of the form

$$\begin{aligned}\langle\delta\widetilde{X}_k\delta\widetilde{X}_{k+m}\rangle &= \sum_{j=-\infty}^{k}\sum_{i=-\infty}^{k+m} M_{k-j+1}M_{k+m-i+1}\langle\delta X_j\delta X_i\rangle \\ &= \sigma^2\sum_{j=-\infty}^{k} M_{k-j+1}M_{k+m-j+1} \\ &= \sigma^2\sum_{j=1}^{\infty} M_jM_{j+m} \equiv c(m)\end{aligned} \tag{2.38}$$

for all $k \geq 1, m \geq 0$. Depending on the exact behaviour of the correlation kernel M_j, this correlation function can have either a negative or positive sign, signalling persistence or anti-persistence, respectively. Also short- or long-range dependent processes can be generated by a suitable choice of M_j. To see this, write

$$\begin{aligned}\sum_{m=1}^{\infty}|c(m)| &= \sigma^2\sum_{m=1}^{\infty}\sum_{j=1}^{\infty}|M_jM_{j+m}| \\ &= \frac{\sigma^2}{2}\left(\sum_{k=1}^{\infty}\sum_{j=1}^{\infty}|M_jM_k| - \sum_{j=1}^{\infty}|M_j|^2\right) \\ &= \frac{\sigma^2}{2}\left(\sum_{j=1}^{\infty}|M_j|\right)^2 - \frac{\tilde{\sigma}^2}{2}\end{aligned} \tag{2.39}$$

Hence, random walk steps are long-range dependent in the sense of section 2.3.1, if and only if the correlation kernel is summable. The decay rate of the correlation function is determined implicitly by the asymptotics of the correlation kernel. Oscillations in the kernel M_j naturally induce oscillating correlations $c(m)$. A few instructive examples are given in table 2.1 at the end of this section.

Finally, the random walk process $\widetilde{X}(n) = \sum_{k=1}^{n}\delta\widetilde{X}_k$ associated with such correlated jump lengths has the following series representation:

$$\begin{aligned}\widetilde{X}(n) &= \sum_{k=1}^{n}\delta\widetilde{X}_k = \sum_{k=1}^{n}\sum_{j=-\infty}^{k} M_{k-j+1}\delta X_j \\ &= \sum_{j=-\infty}^{0}\left(\widetilde{M}_{n-j+1} - \widetilde{M}_{-j+1}\right)\delta X_j + \sum_{j=1}^{n}\widetilde{M}_{n-j+1}\delta X_j\end{aligned} \tag{2.40}$$

where

$$\widetilde{M}_k = \sum_{j=1}^{k} M_j. \tag{2.41}$$

Thus, the correlations indeed preserve the Gaussian nature of the process: $\widetilde{X}(n)$ can itself be written as a linear transformation of the IID Gaussian variables δX_j, $j \in \mathbb{Z}$. The transformation coefficients are collected by the random walk kernel $\widetilde{M}_k$, which is simply the sum of the step length correlation kernel M_j.

An important observation is that correlations in a Gaussian random walk generally modify its scaling behaviour. They are even capable of inducing anomalous diffusion. Equations (2.34), (2.38) and (2.41) tightly tie together the asymptotics of the mean squared displacement $\left\langle \widetilde{X}^2(n) \right\rangle$, the step length correlation function $c(m)$ and correlation kernel M_j and the random walk kernel $\widetilde{M}_k$. For instance, if the random walk is directly defined in terms of the random walk kernel $\widetilde{M}_k$, then the mean squared displacement, according to Eq. (2.40), can be derived through

$$\begin{aligned} \left\langle \widetilde{X}^2(n) \right\rangle &= \sigma^2 \left[\sum_{j=-\infty}^{0} \left(\widetilde{M}_{n-j+1} - \widetilde{M}_{-j+1} \right)^2 + \sum_{j=1}^{n} \widetilde{M}_{n-j+1}^2 \right] \\ &= \sigma^2 \left[\sum_{j=1}^{\infty} \left(\widetilde{M}_{j+n} - \widetilde{M}_j \right)^2 + \sum_{j=1}^{n} \widetilde{M}_j^2 \right] \end{aligned} \tag{2.42}$$

Table 2.1 lists a few instructive examples.

2.3.3 Definition of the process

There are essentially two equivalent ways of defining the FBM – as a stochastic integral or in terms of the correlation function – which are both presented in the following.

Stochastic integral representation. Benoit Mandelbrot, who introduced the FBM to the community of applied probability theory with his seminal 1968 paper [118], started from a representation in terms of a stochastic integral. Let $B(t)$ be a standard Brownian motion – in other words, a standard Wiener process, see section 1.4.1. From this, derive a standard FBM $B_H(t)$ via a stochastic integral with respect to $dB(t)$:

$$B_H(t) = \frac{1}{C_H} \left\{ \int_{-\infty}^{0} \left[(t-t')^{H-1/2} - (-t')^{H-1/2} \right] dB(t') + \int_{0}^{t} (t-t')^{H-1/2} \, dB(t') \right\}, \tag{2.43}$$

where the constant $C_H > 0$ is defined below. In addition, let $B_H(0) = 0$ almost surely. The parameter H is called *scaling* or *Hurst exponent*, for reasons which will become apparent below. A complete analysis of this process of course requires some expertise in stochastic calculus. But a few key properties can already be anticipated by recognising the strong formal analogy with the Gaussian random walks discussed in the previous section. In particular, compare the stochastic integral definition (2.43) with the correlated walk $\widetilde{X}(n)$ defined through the sum (2.40). The parallels are obvious: the integral with respect to $dB(t')$ is the time-continuous counterpart to the sum with the IID Gaussian

random walk type	correlation kernel M_j	correlation function $c(m)$	random walk kernel $\widetilde{M}_k$	scaling of $\langle \widetilde{X}^2(n) \rangle$
uncorrelated	$= \delta_{1j}$	$= \sigma^2 \, \delta_{0m}$	$= 1$	$= \sigma^2 n$
exponential correlations (persistent)	$= e^{-\lambda j}$, $\lambda > 0$	$= \tilde{\sigma}^2 e^{-\lambda \lvert m \rvert}$	$\sim \tilde{\sigma}/\sigma$	$\sim \tilde{\sigma}^2 n$
exponential correlations (anti-persistent)	$A\,\delta_{1j} + B(1-\delta_{1j})e^{-\lambda j}$, $\lambda, A, B > 0$ appropriately	$= \sigma^2 \, [\delta_{0m} - (1-\delta_{0m})] \, e^{-\lambda \lvert m \rvert}$	$\simeq -1$	$\simeq 1$
power-law correlations (persistent, short-ranged)	$= j^{-\gamma}$, $\gamma > 1$	$\simeq \lvert m \rvert^{-\gamma}$	$\simeq 1$	$\sim \tilde{\sigma}^2 n$
power-law correlations (persistent, long-ranged)	$= j^{-\gamma}$, $1/2 < \gamma < 1$	$\simeq \lvert m \rvert^{1-2\gamma}$	$\simeq k^{1-\gamma}$	$\simeq n^{3-2\gamma}$

Table 2.1: Comparison of various Gaussian random walks. The processes are specified in terms of the step length correlation kernel M_j, which implicitly defines the step length correlation function $c(m)$, the random walk kernel $\widetilde{M}_k$ and the random walk scaling behaviour $\langle \widetilde{X}^2(n) \rangle$ via Eqs. (2.34), (2.38) and (2.41). The asymptotics for some entries are for large arguments. In some cases, numeric prefactors are omitted for clarity, but in a way that the overall sign is preserved correctly.

random variables δX_i. Both are linear transformations, and hence it is natural to expect that FBM is a Gaussian process. Moreover, the sum (2.40) was constructed such that the random step lengths $\delta \widetilde{X}_k$ are stationary in distribution, but mutually correlated. Likewise, the increments $B_H(t_2) - B_H(t_1)$ of FBM are stationary in distribution, but not independent even for non-overlapping time intervals.

Finally, the diffusive behaviour of FBM can be obtained from a formal conclusion by analogy. The "random walk" kernel for the FBM is a power-law $\widetilde{M}(t) = t^{H-1/2}$. Now replace the sum in Eq. (2.42), to get

$$\begin{aligned} \left\langle B_H^2(t) \right\rangle &= \frac{1}{C_H} \left\{ \int_0^\infty \left[\widetilde{M}(t+t') - \widetilde{M}(t') \right]^2 dt' + \int_0^t \widetilde{M}^2(t') \, dt' \right\} \\ &= \frac{1}{C_H} \left\{ \int_0^\infty \left[(t+t')^{H-1/2} - (t')^{2H-1} \right]^2 dt' + \int_0^t (t')^{H-1/2} \, dt' \right\} \\ &= \frac{t^{2H}}{C_H} \left\{ \int_0^\infty \left[(1+y)^{H-1/2} - y^{H-1/2} \right]^2 dy + \frac{1}{2H} \right\}. \end{aligned} \tag{2.44}$$

Hence, H is indeed a scaling exponent, and the mean squared displacement indicates subdiffusion for $H < 1/2$ and superdiffusion for $H > 1/2$. Ordinary Brownian motion is recovered by setting $H = 1/2$, which can also be seen directly at Eq. (2.43). The constant C_H may at this point conveniently be chosen to equal the expression in the curly brackets of Eq. (2.44), such that FBM is standardised to $\langle B_H^2(1)\rangle = 1$. Notice however that for C_H to be well-defined one has to require $0 < H < 1$. Otherwise, the integral diverges either at $y \to 0$ or ∞.[10]

Correlation function representation. Being a Gaussian process, FBM is of course perfectly well defined in terms of its position correlation function. Thus, let $B_H(t)$ be a Gaussian process with

$$\begin{aligned} \langle B_H(t)\rangle &= 0 \\ \langle B_H(t_1)B_H(t_2)\rangle &= \frac{1}{2}\left[|t_1|^{2H} + |t_2|^{2H} - |t_2 - t_1|^{2H}\right]. \end{aligned} \tag{2.45}$$

It is possible to prove the equivalence with the stochastic integral definition (2.43) [118, 88]. The latter is more constructive and gives a better understanding on path properties. The correlation function in turn is for sure useful for distributional analysis.

For example, the mean squared displacement can be obtained directly by setting $t_1 = t_2 = t$, recovering the familiar result $\langle B_H^2(t)\rangle = t^{2H}$. Furthermore one can study the properties of the motion memory in terms of the unit increments $\delta B_H(t) = B_H(t+1) - B_H(t)$. The increment correlation function can be derived from Eq. (2.45). On the one hand,

$$\begin{aligned} \left\langle [\delta B_H(t)]^2 \right\rangle &= \left\langle [B_H(t+1) - B_H(t)]^2 \right\rangle \\ &= \left\langle B_H^2(t+1)\right\rangle + \left\langle B_H^2(t)\right\rangle - 2\left\langle B_H(t)B_H(t+1)\right\rangle = 1. \end{aligned} \tag{2.46}$$

Thus, indeed increments are stationary in distribution. On the other hand,

$$\begin{aligned} c(t) \equiv \langle \delta B_H(s+t)\delta B_H(s)\rangle &= \langle \delta B_H(t)\delta B_H(0)\rangle \\ &= \langle [B_H(t+1) - B_H(t)]B_H(1)\rangle \\ &= \langle B_H(t+1)B_H(1)\rangle - \langle B_H(t)B_H(1)\rangle \\ &= \frac{1}{2}\left[|t+1|^{2H} - 2|t|^{2H} + |t-1|^{2H}\right], \\ &\sim H(2H-1)|t|^{2H-2} \qquad , |t| \to \infty. \end{aligned} \tag{2.47}$$

Interpret now, for the moment, the unit increments $\delta B_H(t)$ as the steps of a random walk (essentially a FBM on the integers $t \in \mathbb{N}$). In this context, one can apply the vocabulary worked out in section 2.3.1. It turns out that the single parameter H in

[10] The constant C_H and the stochastic integral in Eq. (2.43) is also well-defined for the special value $H = 1$. However, this case is degenerate: one can show [118, 88] that it produces a straight line process, i.e. $B_H(t) = t\, B_H(1)$ almost surely.

fact determines the complete anomalous diffusion characteristics of the FBM. One has to distinguish three regimes. For $1/2 < H < 1$, the correlation function $c(t)$ is positive and not summable, $\sum_{t=1}^{\infty} |c(t)|$. This implies persistence, long-range memory and superdiffusion. For $H = 1/2$, the motion is ordinary Brownian, hence memory-less and normal. For $0 < H < 1/2$, memory effects anti-persistence, since $c(t) < 0$. Although the dependence is short-ranged in the language of section 2.3.1, this regime is anomalous, too, because $\bar{\sigma}^2 = 1 + 2\sum_{t=1}^{\infty} c(t) = 0$. Consequently, FBM in this case is subdiffusive.

2.3.4 Generalisations and applications

For physically motivated problems, it often not the pure FBM that comes into use. Instead, the Langevin equation approach (see section 1.4.2) can be generalised to include the memory. The white noise $\xi(t)$ is to be replaced by a standard *fractional Gaussian noise* $\xi_H(t)$, that is, a stationary Gaussian process with [118, 119]

$$\begin{aligned} \langle \xi_H(t) \rangle &= 0 \\ \langle \xi_H(t_1) \xi_H(t_2) \rangle &= H(2H-1)|t_2 - t_1|^{2H-2} \\ &\quad + 2H|t_2 - t_1|^{2H-1} \delta(t_2 - t_1). \end{aligned} \tag{2.48}$$

For $H = 1/2$, this collapses to a white noise process, see Eq. (1.13). For $H < 1/2$ or $H > 1/2$, there are negative or positive correlations, respectively. Fractional Gaussian noise is the derivative of FBM in the sense that averaging over the integrated noise

$$B_H(t) = \int_0^t \xi_H(t')\, dt' \tag{2.49}$$

recovers the correlation function (2.45) which defines a standard FBM $B_H(t)$.

However, simply exchanging the white noise in the Langevin equation (1.12) produces processes which have unacceptable properties for most physical applications. Instead, the *fractional Langevin equation* reads, in the one-dimensional case,

$$m\ddot{x}(t) = F(x,t) - \overline{\gamma} \int_0^t (t - t')^{2H-2}\, \dot{x}(t')\, dt' + \eta\, \xi_H(t), \tag{2.50a}$$

where

$$\eta = \sqrt{\frac{k_B \mathcal{T} \gamma}{H\, \Gamma(2H)}}, \qquad \overline{\gamma} = \gamma \Gamma(2H-1). \tag{2.50b}$$

The dots indicate derivatives with respect to time, $\overline{\gamma} > 0$ is a generalised friction constant (units are $\mathrm{kg\,sec^{-2H}}$), $k_B\mathcal{T} > 0$ is the thermal energy of the environment, and $F(x,t)$ is an external force. Most importantly, concurrently with introducing a non-white noise, also the linear friction force has to be replaced by a non-linear expression. The involved integral samples over the full history of velocities $\dot{x}(t')$ starting from time $t' = 0$, where the process began. The exponent of the friction memory kernel is linked to the noise

Hurst exponent H. While the noise process itself can be defined, in principle, for any $0 < H < 1$, the friction kernel in Eq. (2.50) diverges for values $H \leq 1/2$. Hence, one has to restrict the parameter to $1/2 < H < 1$, implying that noise correlations are of the long-range, persistent type.

There are two approaches to constitute a Langevin equation of this type. First, notice that some kind of relation between friction and noise terms is in fact natural, since both originate in the interaction with the heat bath. This statement can be made much more precise in terms of the *fluctuation dissipation theorem*, which was particularly promoted and extended by the work of Kubo [120]. He established a general relation between the friction kernel of the noise, and the correlation function of the noise. Only in the special white noise case does the friction term reduce to a linear relation. The key argument in the derivation of the fluctuation dissipation theorem is that Eq. (2.50) should admit a stationary solution. A second pathway to the fluctuation dissipation theorem goes along the Mori-Zwanzig formalism, see section 1.5.2. Under quite general conditions, one can derive a generalised Langevin equation – with more general noise and friction forces – for a large class of Hamiltonian systems [65].

The FBM as encapsulated in the fractional Langevin equation, has found wide application in physics and other natural sciences. This anomalous diffusion aspect stands for the interaction with a complex environment, where relaxation of correlations is slow and cannot be characterised in terms of a single, exponential or otherwise slow relaxation rate.

For example, a Brownian particle experiences collisions with individual molecules of an ambient fluid at a rapid rate, and the impulse transferred by each collision is minute. But if the fluid is dense, the movement of the Brownian particle might generate local convection flow or perturbation in the fluid. The effect falls back onto the Brownian particle dynamics, a phenomenon referred to as *hydrodynamic memory* [65].

Similar correlation phenomena come about when accounting for obstructed motion (e.g. single-file diffusion [121] or other many-body systems [122, 123, 124]) or an interaction with a viscoelastic network [125, 126]. Biological cells feature highly complex, crowded environments, crossed by filament networks. FLE dynamics have thus found wide application in biological physics, describing diffusive motion processes within the cell [126, 127, 128, 129, 130], but also conformational dynamics of individual protein complexes [131].

Historically, one of the first empirical phenomena that called for a convincing theory of long-range dependence came from hydrology. Harold Hurst, who was interested in damn design (the Hurst exponent of FBM is named after him), investigated data on the flow of Nile river [132, 117].

Further inspiration came from economic data sets, and finance is still a prominent field of application for FBM related processes. Several examples can be found in Mandelbrot's book, Ref. [133]. The early Brownian motion theory of Bachelier was a pioneering concept at the time; but Mandelbrot comments:

> "Alas, the theory is elegant but flawed, as anyone who lived through the

> booms and busts of the 1990s can now see. The old financial orthodoxy was founded on two critical assumptions in Bachelier's key model: Price changes are statistically independent, and they are normally distributed. The facts, as I vehemently argued in the 1960s and many economists now acknowledge, show otherwise.
>
> Why this should be is not certain; but one can speculate. What a company does today – a merger, a spin-off, a critical product launch – shapes what the company will look like a decade hence; in the same way, its stock-price movements today may take a long time to absorb and fully price information. When confronted by bad news, some quick-triggered investors react immediately while others, with different financial goals and longer time-horizons, may not react for another month or year. Whatever the explanation, we can confirm the phenomenon exists—and it contradicts the random-walk model. Second, contrary to orthodoxy, price changes are very far from following the bell curve. [...] In fact, the bell curve fits reality very poorly. From 1961 to 2003, the daily index movements of the Dow Jones Industrial Average do not spread out on graph paper like a simple bell curve. The far edges flare too high: too many big changes. [...] Truly, a calamitous era that insists on flaunting all predictions. Or, perhaps, our assumptions are wrong."

Hence, long-range memory as in FBM is only one of many crucial ingredients for the modelling of financial data. This raises the general question whether a separate, independent study of anomalous diffusion paradigms is sufficient for the treatment and understanding of complex empiric phenomena.

3 Correlated continuous time random walks

Anomalous diffusion arises in a wide range of systems across disciplines and is usually characterised in terms of the mean squared displacement $\left\langle [X(t) - X(0)]^2 \right\rangle \simeq t^{2H}$ of a random variable $X(t)$, where the anomalous diffusion or Hurst exponent H distinguishes subdiffusion ($0 < H < \frac{1}{2}$) from superdiffusion ($H > \frac{1}{2}$) [82, 83]. In general, anomalous diffusion processes are not universal and thus their definition through the deviation from the normal diffusive scaling $H = \frac{1}{2}$ is not unique. Instead, this form may be caused by multiple physical mechanisms, some of which are very distinct conceptually. Several pathways to anomalous diffusion have been discussed in section 2: (iii) paths with scale-free displacements are generated by Lévy flights and walks (section 2.1; (ii) trapping mechanisms leading to long sojourn times are modelled in terms of continuous time random walks (CTRW, section 2.2); (iii) long-ranged memory, induced by interaction with a complex surrounding is captured by fractional Brownian motion (FBM) and fractional Langevin equations (section 2.3). These approached are are paradigmatic in the sense that they are designed to tackle one specific aspect of anomalous diffusion.

However, in complex, disordered environments one should expect that more than one of the patterns (i) to (iii) emerge, compete and collude to generate anomalous diffusion patterns, and it remains an open challenge to identify and differentiate them. Thus, the global properties of dispersion in amorphous media can be related to the microscale flow dynamics by adding a memory component to the standard CTRW description [134]. Modern single particle tracking techniques in experiment and simulations indeed corroborate the co-existence of different diffusion mechanisms [78, 84, 115, 128, 135]. For instance, for the motion of individual granules in the intracellular fluid of living cells characteristics of CTRW-style trapping and FBM-like anti-persistence were observed [78].

This section is devoted to the study of a unified stochastic model, namely the correlated CTRW (CCTRW), that merges and extends the classical paradigmatic models of CTRW, FBM, and Lévy flights. Two quantities which are typically accessible experimentally are discussed in detail: the scaling of the particle position X with time t, and the shape of the probability density function (PDF) $p(x;t)$ of the particle displacement x at some instant of time t. The result is a very flexible stochastic model, that will be of use for the data analysis of stochastic processes in complex systems. One immediate lesson is the interplay of the underlying stochastic modes, the blend of which may lead to indistinguishable forms for the PDF $p(x;t)$ for different sets of model parameters.

The content of this section has been published previously to a large extent, so further details on the topic can be found in Ref. [136].

3.1 Model Definition

In standard CTRW and Lévy flight models, individual jump lengths and waiting times, respectively, are independent of each other. The correlated CTRW aims at extending this theory to systems where highly complex environments induce long-range correlations. The basic theoretical approach is to derive a correlated process from an uncorrelated one in terms of linear transformations. By this method, correlations are introduced without altering the scale-free distributional properties of the processes itself.

As described in section Section 2.3.2, a set of independent, identically distributed (IID) Gaussian random variables can be transformed into a set of correlated Gaussian step lengths by means of a linear correlation kernel. This method of correlating a Gaussian random walk can be readily transferred to Lévy flights and CTRWs as defined in the previous sections. The reason is that the involved Gaussian, symmetric stable and one-sided stable random variables are all belong to the important class of general stable random variables. Recall that "stability" refers to the very property of being closed under linear transformations.

The CCTRW is defined here not as a random walk, but directly on the level of scaling limits.[1] The definition is completely analogous Eq. (2.43) for FBM. While the latter is given in terms of a stochastic integral with respect to a Brownian noise $dB(t)$, scale-free dynamics can be generated by integrating with respect to a *Lévy stable noise* $dL_\mu(s)$. For instance, for a symmetric motion, let $L_\mu(s)$ be a symmetric Lévy flight with stable index $0 < \mu < 2$. Now define a *linear fractional stable motion* via the stable stochastic integral

$$Y(s) := (\mu K)^{1/\mu} \int_0^s (s - s')^{K-1/\mu} \, dL_\mu(s'). \tag{3.1}$$

Here $0 < \mu < 2$, and $K > 0$ will be referred to as the Hurst exponent of the stable motion $Y(s)$. In analogy to the fractional Brownian motion, the choice for the correlation kernel is a power-law, so that correlations are potentially of a long-ranged kind.[2] The detailed properties and mathematical foundations of stable stochastic integrals can be found in Ref. [88]. Most importantly, the process $Y(s)$ stays in the domain of μ-stable processes, just like FBM remains a Gaussian process. In particular, at given time s, the probability density function (PDF) $p_{\mu,K}(y; s)$ of the position $Y(s)$ is of the stable form (2.9), albeit with an altered time scaling $Y(s) \sim s^K$,

$$p_{\mu,K}(y; s) = s^{-K} \ell_\mu(y s^{-K}). \tag{3.2}$$

[1] How and when a time-discrete correlated random walk *converges* to a time-continuous correlated motion is taken up in references [137, 138].

[2] There is an apparent discrepancy between the stable motion defined here through Eq. (3.1) and FBM as given by Eq. (2.43). Most notably, the Gaussian integral extends over negative values down to $-\infty$. The motivation for this is that FBM is desired to have stationary increments. In contrast, the more simple definition (3.1), where the integral starts at time $s = 0$, implies that increments are not stationary. For reasons laid out below, the model definition presented here is deliberately extended to include processes with potentially non-stationary increments. A complete discussion on this issue is given in section 3.2.

The scaling prefactor $(\mu K)^{1/\mu}$ in Eq. (3.1) makes sure that the scaling function ℓ_μ is again exactly represented by the characteristic function in Eq. (2.10).

What really sets the process $Y(s)$ apart from the ordinary Lévy motion $L_\mu(s)$ is the stochastic dependence of increments. One may also say the noise related to $Y(s)$ is strongly correlated,[3] or coloured. However, to assess the nature of interdependence here, one cannot use the covariance or correlation function like in Eq. (2.45). While the latter is a meaningful and precise measure of dependence for Gaussian processes, $\mu = 2$, it is ill-defined for stable processes $\mu < 2$. In reference [88], several alternative concepts to deal with the stable cases are introduced and discussed, such as covariation or codifference functions. In short, applying these analytic tools to the correlated process $Y(s)$, yields positive, long-range dependence when $K > 1/\mu$, and negative, short-range dependence when $K < 1/\mu$. (Compare this to the analogous discussion on linear fractional stable motion in [88]. An extensive discussion of the notion of long-range dependence can be found in [117]).

These considerations can be supplemented by spectral analysis arguments, compare also Ref. [138]. Consider a sample path of a Lévy flight $L_\mu(s)$ and denote its Fourier transform by $\widehat{L}_\mu(\omega)$. Now since the stable stochastic integral (3.1) is of a convolution form, there is a simple relation in Fourier space between the correlated noise $dY(s)$ and the Lévy stable noise $dL_\mu(s)$, namely $d\widehat{Y}(\omega) \propto d\widehat{L}_\mu(\omega)/(-i\omega)^{K-1/\mu}$. When comparing the two noise types in the case $K > 1/\mu$, one thus finds that the correlation kernel in the stable integral (3.1) emphasises the low frequency components of the correlated noise. In a sample path of $Y(s)$, this might be conceived as a comparatively steady motion, even in the form of long-term periodic cycles. Conversely, when $K < 1/\mu$, high frequencies are amplified. A sample path $Y(s)$ is then fluctuating violently as compared to an ordinary Lévy flight.

Hence, both the analysis in terms of covariation/codifference functions and the spectral analysis support the idea of an either persistent or anti-persistent motion $Y(s)$. For $K > 1/\mu$, persistence effects long cycles of seemingly steady, biased motion. If $K < 1/\mu$, anti-persistent motion is observed as being wildly fluctuating, since strong, short range, negative memory leads to a quick succession of directional turns. The special case $K = 1/\mu$ recovers ordinary Lévy flights with mutually independent jump lengths.

The last step in the definition of the CCTRW model is the introduction of correlations of waiting times. Analogously to the above, define

$$T(s) := (\alpha G)^{1/\alpha} \int_0^s (s - s')^{G-1/\alpha} \, dL_\alpha^+(s'), \tag{3.3}$$

in terms of a stable integral with respect to one-sided Lévy α-stable noise $dL_\alpha^+(s)$. Here, $0 < \alpha < 1$ and $G \geq 1/\alpha$. The corresponding PDF $g_{\alpha,G}(t; s)$ at given internal time s in

[3] Calling a stable process "correlated" is a bit misleading, since the correlation function is ill-defined if $\mu \neq 2$. Here, being "(un-)correlated" should be interpreted as equivalent to being "statistically (in-)dependent", wich is (some might say unfortunately) a common habit in the physics literature and some other fields of applied probability theory.

this case reads

$$g_{\alpha,G}(t;s) = s^{-G}\ell_\alpha^+(ts^{-G}), \tag{3.4}$$

where the basic shape is still provided by a one-sided α-stable law ℓ_α^+ as defined in equation (2.23). The scaling with internal time s in this case reads $T(s) \sim s^G$. While $T(s)$ is still an α-stable motion, waiting times are no longer independent. Note that for $T(s)$ to be an increasing process, one has to require that $G \geq 1/\alpha$. Thus, correlations in waiting times are necessarily of the persistent type, and have a tendency to increase with s. The only exception to this rule is $G = 1/\alpha$, a parameter setting which leads back to heavy-tailed but uncorrelated waiting times.

In complete analogy to the uncorrelated case, introduce now the inverse process $S(t)$ according to equation (2.24) and combine it with a correlated stable motion, $X(t) = Y(S(t))$, via subordination, see section 2.2.1. The PDF for the particle position X at real time t is then given by

$$p_{\mu,\alpha,K,G}(x;t) = \int_0^\infty p_{\mu,K}(x;s)h_{\alpha,G}(s;t)\,ds, \tag{3.5}$$

where $h_{\alpha,G}(s;t)$ denotes the PDF of internal time S at real time t. The latter is discussed extensively in section 3.3.

Several sample trajectories of a CCTRW can be studied in Fig. 3.1. In addition, the figure includes the trajectories of the uncorrelated Brownian motion, Lévy flight and CTRW processes, also shown in the respective sections, in order to facilitate direct comparison. For the correlated process, the spatially continuous motion is paused for large-scale waiting periods, which appear on all time scales. Waiting times are *not* independent but persistent here: long rests are directly followed by periods of reduced dynamic activity, but then slowly turn into vivid almost Brownian-like motion. Note however that also spatial displacements are persistent.[4]

This completes the definition of the CCTRW model presented here. A discontinuous progression of spatial displacements and laboratory time is modelled in terms of the stable noises $dL_\mu(s)$ and $dL_\alpha^+(s)$. Correlations are separately introduced by power-law correlation kernels to both the spatial dynamics $Y(s)$ and the time evolution $T(s)$. The full model is defined in terms of four parameters: $0 < \mu < 2$ and $0 < \alpha < 1$ determine the respective distributional properties of individual jump lengths δX and waiting times δT. In particular, they define the heavy tails $\lambda(\delta x) \simeq |\delta x|^{-1-\mu}$ and $\psi(\delta t) \simeq \delta t^{-1-\alpha}$. The special cases of continuous spatial and/or temporal evolution are included in the correlated CTRW model on a distribution level as the limits $\mu \to 2$ and $\alpha \to 1$. The parameters $K > 0$ and $G \geq 1/\alpha$ directly measure the scaling exponents with respect to internal time, $Y(s) \sim s^K$ and $T(s) \sim s^G$. Finally, the nature of the correlations can be

[4] All trajectories in Figs. 1.5-3.1 were generated by simulating long random walk trajectories and rescaling temporal and spatial coordinates appropriately. This method approximates the time-continuous motion defined in the respective sections. Heavy-tailed jump lengths or waiting times, respectively, are realised by drawing stable random variables [139]. Correlations as in Eqs. (3.1) and (3.3), are introduced as described in Ref. [88], section 7.11.

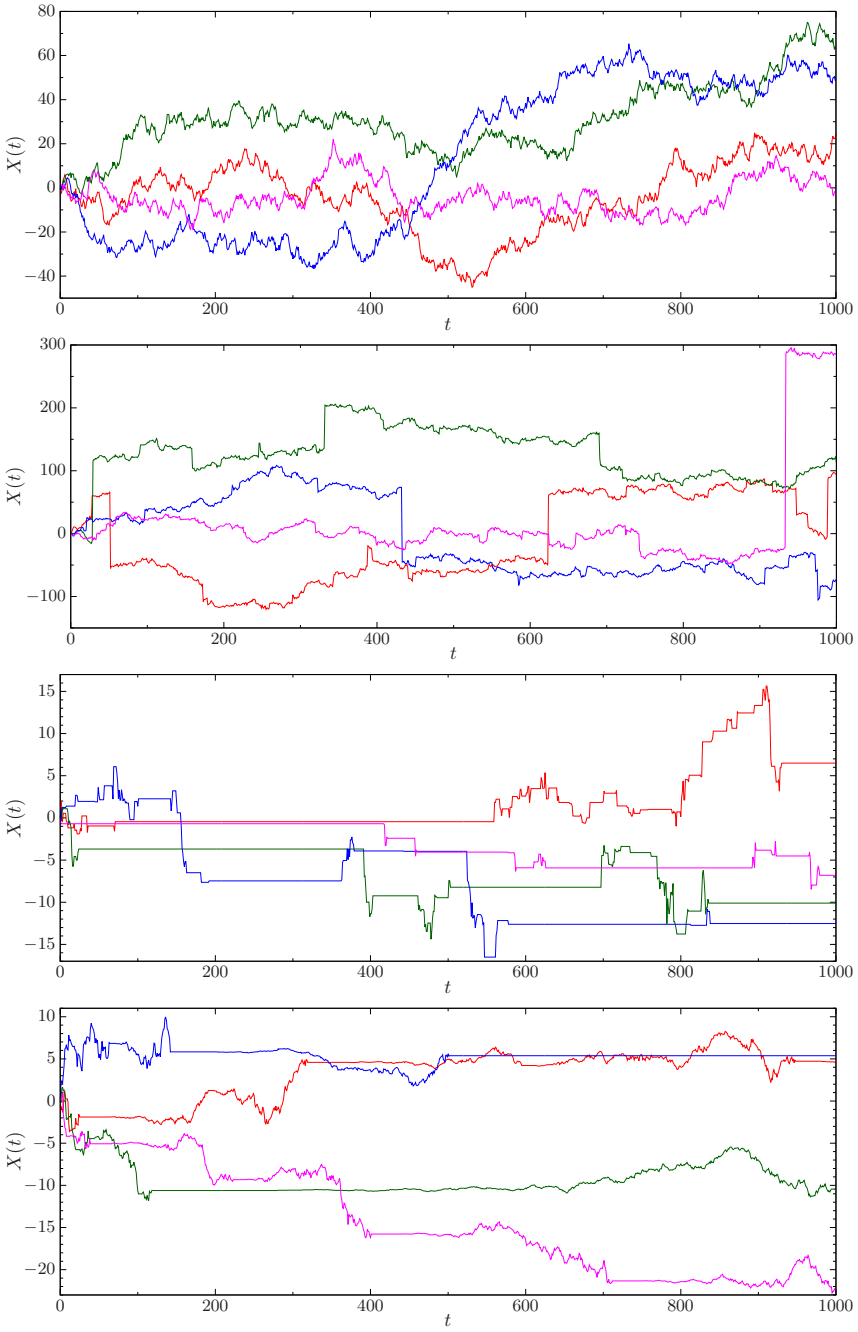


Figure 3.1: Sample paths $X(t)$ of *correlated CTRWs*. Parameters, from top to bottom:

Brownian motion	$\mu = 2$,	$K = 1/2$,	$\alpha = 1$,	$G = 1$
Lévy flight,	$\mu = 3/2$,	$K = 2/3$,	$\alpha = 1$,	$G = 1$
CTRW,	$\mu = 2$,	$H = 1/2$,	$\alpha = 1/2$,	$G = 2$
CCTRW,	$\mu = 2$,	$K = 1.1/2$,	$\alpha = 1/2$,	$G = 2.2$.

assessed by comparing respective parameter pairs: jump distances (waiting times) are persistent if $K > 1/\mu$ ($G > 1/\alpha$), uncorrelated if $K = 1/\mu$ ($G = 1/\alpha$), or anti-persistent if $K < 1/\mu$ (impossible for waiting times).

3.2 Stationarity

The stable processes $Y(s)$ and $T(s)$ are defined directly in terms of their distributional, scaling and correlation properties, as characterised through the parameters μ, α, K, and G, respectively. This quite large class of processes can further categorised through their stationarity properties. For this, apply the preliminary definition of the nth order increments of a stochastic process $Y(s)$,

$$\begin{aligned}
\Delta^{(1)}Y(s;\tau) &= Y(s+\tau) - Y(s) \\
\Delta^{(2)}Y(s;\tau_1,\tau_2) &= \Delta^{(1)}Y(s+\tau_2;\tau_1) - \Delta^{(1)}Y(s;\tau_1) \\
&\vdots \\
\Delta^{(n)}Y(s;\tau_1,\ldots,\tau_n) &= \Delta^{(n-1)}Y(s+\tau_n;\tau_1,\ldots,\tau_{n-1}) \\
&\quad - \Delta^{(n-1)}Y(s;\tau_1,\ldots,\tau_{n-1}).
\end{aligned} \tag{3.6}$$

Thus, $\Delta^{(1)}Y$ is the usual process increment while $\Delta^{(2)}Y$ is an increment of increments, *etc.* If $Y(s)$ is meant to stand for the position of a particle at time s, then the ratio $\Delta^{(1)}Y(s;\tau)/\tau$ can be viewed as the average velocity (bearing in mind that the one-time velocity, i.e., the limit $\tau \to 0$, in general does not exist for the processes discussed here). Likewise, $\Delta^{(2)}Y(s;\tau_1,\tau_2)/(\tau_1\tau_2)$ corresponds to the intuitive notion of an acceleration, and higher order increments represent higher levels of temporal evolution.

The nth order increments of $Y(s)$ in now said to be asymptotically stationary in distribution (ASD), if the random variable $\Delta^{(n)}Y(s;\tau_1,\ldots,\tau_n)$ has a nontrivial limiting distribution for large times, $s \to \infty$. The degrees of stationarity for the stable processes $Y(s)$ and $T(s)$ as defined in the previous section shall be determined in the following. Note that this classification is not a purely academic one. For the application and interpretation of a stochastic process as a real world model system, stationarity properties are highly relevant. Let, for instance, $Y(s)$ model an animal foraging process. Then stationarity of first order increments is an indication for a time-independent search strategy: the distance $\Delta^{(1)}Y$ travelled during, say, $\tau = 1\,\mathrm{day}$ is statistically indistinguishable from one day to the next. Conversely, non-stationary statistics of travel distances can be a signature for an adaptive search strategy, an ageing animal, or changes in the environment. In this case, one would further ask whether or not such internal or external variations are stationary. This relates to second order increments. On smaller scales, $Y(s)$ could be a model for particle diffusion in a heat bath. There, non-stationarity of first order increments is the fingerprint either of an inhomogeneous environment (i.e., the particle displacement statistics changes as the particle explores various spatial regions) or a non-equilibrated environment (i.e., the noise imposed by interaction with the surrounding heat bath is itself non-stationary). Then, analysis of second and higher order

increments yields information on the precise nature of the spatial or temporal variations in the surroundings.

The displacement process $Y(s)$ as defined through the stable integral (3.1) is a non-stationary process, as indicated by the time scaling $Y(s) \sim s^K$. As the particle explores its surrounding space, the probability to find it in any region of fixed size around the origin of motion is decaying with time. Now, the integral representation of first order increments reads

$$\Delta^{(1)}Y(s;\tau) = (\mu K)^{1/\mu} \left\{ \int_0^s \left[(s+\tau-s')^{K-1/\mu} - (s-s')^{K-1/\mu} \right] dL_\mu(s') + \int_s^{s+\tau} (s+\tau-s')^{K-1/\mu}\, dL_\mu(s') \right\}, \tag{3.7}$$

so that its distribution is given in terms of the characteristic function

$$\begin{aligned} &\langle \exp(ik\Delta^{(1)}Y(s;\tau)) \rangle = \\ &\quad = \exp\left[-\mu K|k|^\mu \int_0^s \left| (s+\tau-s')^{K-1/\mu} - (s-s')^{K-1/\mu} \right|^\mu ds' \right. \\ &\qquad \left. -\mu K|k|^\mu \int_s^{s+\tau} \left| (s+\tau-s')^{K-1/\mu} \right|^\mu ds' \right] \\ &\quad = \exp\left[-|k|^\mu I^{(1)}_{\mu,K}(s) - |k\tau^K|^\mu \right], \end{aligned} \tag{3.8}$$

using the abbreviation

$$I^{(1)}_{\mu,K}(s) = \mu K \int_0^s \left| (s'+\tau)^{K-1/\mu} - (s')^{K-1/\mu} \right|^\mu ds'. \tag{3.9}$$

Non-stationarity is indicated by the explicit s-dependence of the integral $I^{(1)}_{\mu,K}$. The latter vanishes identically if $K = 1/\mu$. This is natural, since these cases are the symmetric Lévy stable motions, $Y(s) = L_\mu(s)$, which have stationary increments by definition. Conversely, for any $K \neq 1/\mu$, the integral differs from zero, so in general the first order increments of the stable motion $Y(s)$ are non-stationary. However, they can still be asymptotically stationary, depending on the parameters. The expression in the integral $I^{(1)}_{\mu,K}$ behaves, for large s', like $(s')^{\mu K-\mu-1}$. The asymptotics at large times $s \gg \tau$ are therefore given through

$$I^{(1)}_{\mu,K}(s) \simeq \begin{cases} \text{const}, & \text{for } 0 < K < 1, \\ \log(s), & \text{for } K = 1, \\ \tau^\mu s^{\mu(K-1)}, & \text{for } K > 1. \end{cases} \tag{3.10}$$

First order increments are hence ASD whenever $0 < K < 1$, while spreading indefinitely when $K \geq 1$. One can readily extend the procedure to the study of increments of arbitrary order, see Appendix of Ref. [136]. In general, two classes of parameter settings can be distinguished.

If there is a non-negative integer m such that $K = 1/\mu + m$, then all increments of order $n > m$ are stationary in distribution, and lower order increments, $n \leq m$ on average broaden. This includes the Lévy stable motions, $m = 0$, $K = 1/\mu$, with stationary increments of all orders. To understand this, recall that correlated and Lévy stable noises are related in Fourier space through $d\widehat{Y}(\omega) = d\widehat{L}_\mu(\omega)/(-i\omega)^{K-1/\mu}$. Now for $K = 1/\mu + m$, this suggests that $Y(s)$ can be interpreted as an m-fold repeated integration of a Lévy stable noise. In other words, for $m = 0$, $Y(s)$ is a Lévy flight, so increments are stationary; for $m = 1$, the noise generating $Y(s)$ is already a Lévy flight, therefore only second and higher order increments of $Y(s)$ are stationary; for $m = 2$, the noise generating the noise of $Y(s)$ is a Lévy flight, so one has stationary third order increments; *etc.*

The opposite case is $K \neq 1/\mu + m$ for all nonnegative integers m. Interestingly, here the result is μ-independent: all increments of order $n > K$ are ASD, while lower order increments, $n \leq K$, are spreading indefinitely. An extensive and mathematically rigorous treatment of stochastic processes with stationary nth order increments can be found in [140].

The one-sided α-stable process (3.3), which describes the evolution of laboratory time with respect to the internal time has completely analogous properties. Increments of any order are stationary if $G = 1/\alpha$, since then $T(s) = L_\alpha^+(s)$ is a one-sided Lévy stable motion. If there is a non-negative integer m such that $G = 1/\alpha + m$, then only increments of order $n > m$ are stationary in distribution. If there is no such m, increments of orders $n > G$ are ASD. Lower order increments are non-stationary at all times. Note, however, that for the waiting time process one is forced to require $G \geq 1/\alpha$ for the following reason. Writing out the integral representation for the first order increments,

$$\begin{aligned} \Delta^{(1)}T(s;\tau) &= (\alpha G)^{1/\alpha} \left\{ \int_0^s \left[(s+\tau-s')^{G-1/\alpha} - (s-s')^{G-1/\alpha} \right] dL_\alpha^+(s') \right. \\ &\quad \left. + \int_s^{s+\tau} (s+\tau-s')^{G-1/\alpha}\, dL_\alpha^+(s') \right\}, \end{aligned} \tag{3.11}$$

it becomes obvious that the first integral could potentially give a negative contribution when $G < 1/\alpha$. This is clearly unacceptable in terms of causality: negative increments in laboratory time $T(s)$ would correspond to waiting times finishing earlier than they began. Therefore, one can consider only $G \geq 1/\alpha$, which has two implications. On the one hand, as mentioned above, correlated motions are necessarily persistent. On the other hand, first order increments—reflecting waiting time statistics—are non-stationary. More precisely, they are, in a statistical sense, increasing beyond all bonds, as their (one-sided!) distribution persistently broadens with internal time s.

This section is concluded with a general remark on stationarity in CTRW models. The inverse process $S(t)$ measuring the internal time at fixed laboratory time t, Eq. (2.24), is a highly non-stationary process, as are all of its increments. This holds even when $G = 1/\alpha$, i.e. when waiting times are not correlated. This phenomenon has been discussed extensively in the CTRW literature, where it is commonly referred to as ageing [141,

142, 143, 144] and is closely related to other peculiar effects such as weak ergodicity breaking [144, 145]. The deeper reason behind this non-stationarity are the scale-free characteristics of waiting times. Thus, CTRW models are by definition highly non-stationary stochastic processes, and it is indeed natural to extend the common model candidates ($K = 1/\mu$ for uncorrelated, stationary jump distances and $G = 1/\alpha$ for uncorrelated, stationary waiting times) to the larger class of stable, but correlated and potentially non-stationary motions considered here.

3.3 Time scaling analysis and probability density function

For ordinary Lévy flights or CTRWs, the tail parameters μ and α determine both the distributional and the scaling properties of the process. The present correlated model is slightly more complex in this respect. While the shape of the PDF depends on all four parameters, only the Hurst parameters K and G determine the time scaling. To see this, recall that for the Lévy stable motions the characteristic scalings are $L_\mu(s) \sim s^{1/\mu}$ and $L_\alpha^+(s) \sim s^{1/\alpha}$. From Eqs. (3.1) and (3.3) it follows that $Y(s) \sim s^K$ and $T(s) \sim s^G$. Consequently, the internal time scales as $S(t) \sim t^{1/G}$, and for the correlated motion one gets

$$X(t) = Y(S(t)) \sim t^H, \quad \text{where } H = K/G. \tag{3.12}$$

The parameter H is therefore the scaling or Hurst exponent of the correlated motion $X(t)$. Interestingly, from the point of view of time scaling, persistence in waiting times competes with persistence in jump distances, and the process can turn out to be either sub- ($H < 1/2$), or superdiffusive ($H > 1/2$), or exhibit a normal diffusive scaling ($H = 1/2$). Conversely, measuring the Hurst exponent H alone does not reveal specific information on the time scaling of correlated waiting times (G) and correlated jumps (K), but only on their ratio.

This ambiguity actually goes beyond a simple time scaling analysis and extends to the analysis of the PDF. Let $h_{\alpha,G}(s;t)$ denote the probability density for the internal time S at given laboratory time t. Recall that $T(s) \sim s^G$ is a monotonically increasing process. This implies [103] $S(t) \stackrel{d}{=} (t/T(1))^{1/G}$ for any fixed laboratory time t. Therefore,

$$h_{\alpha,G}(s;t) = Gts^{-G-1}\ell_\alpha^+(ts^{-G}). \tag{3.13}$$

One can now combine Eqs. (3.2), (3.5) and (3.13) to write the PDF $p_{\mu,\alpha,K,G}(x;t)$ for the

correlated CTRW $X(t) = Y(S(t))$ at time t in terms of stable densities,[5]

$$\begin{aligned} p_{\mu,\alpha,K,G}(x;t) &= \int_0^\infty p_{\mu,K}(x;s) h_{\alpha,G}(s;t) ds \\ &= \int_0^\infty \frac{1}{s^K} \ell_\mu \left(\frac{x}{s^K} \right) \frac{Gt}{s^{G+1}} \ell_\alpha^+ \left(\frac{t}{s^G} \right) ds \\ &= \frac{1}{t^H} \int_0^\infty \ell_\mu \left(\frac{x}{(st)^H} \right) \frac{1}{s^{H+2}} \ell_\alpha^+ \left(\frac{1}{s} \right) ds \\ &=: \frac{1}{t^H} q_{\mu,\alpha,H} \left(\frac{x}{t^H} \right) \end{aligned} \tag{3.14}$$

This representation demonstrates that the qualitative shape of the PDF can be classified in terms of only three parameters: the tail parameters μ and α and the scaling exponent $H = K/G$. This means that two processes may seemingly be the same when only studying their PDF and time scaling behaviour, although they are inherently different with respect to their correlations. Apparently, persistence in jump distances can balance persistence in waiting times, similar to the previously observed twin paradox [146]. Consider, for instance, the stochastic process $X(t)$ defined by $\mu = 2$, $K = 1/2$, $\alpha = 1/2$ and $G = 2$. This special case has been studied extensively in the literature, as it represents the simplest type of a CTRW process and is bare of correlations both in jump distances and waiting times. For comparison, now define $X'(t)$ by choosing $\mu' = 2$, $K' = 1.1/2$, $\alpha' = 1/2$ and $G' = 2.2$. Obviously, $X'(t)$ is different from the ordinary CTRW $X(t)$, since both its jump distances and its waiting times are persistent. This is clearly visible when investigating a few sample trajectories, as provided in Fig. 3.1[6] However, on the level of time scaling analysis, Eq. (3.12), and PDF, Eq. (3.14), the random motions are indistinguishable, since $H = H'$.

To study the PDF $p(x;t)$ analytically (subscript parameters are dropped from here on), it is natural to first study equation (3.14) in Fourier-Laplace domain. Making direct use of equations (2.10) and (2.23), one finds

$$\begin{aligned} p(k;u) &\equiv \int_{-\infty}^\infty \int_0^\infty e^{ikx-ut} p(x;t) dt dx \\ &= \int_0^\infty \exp\left(-|k|^\mu s^{H\mu/\alpha} \right) u^{\alpha-1} \exp\left(-u^\alpha s \right) ds. \end{aligned} \tag{3.15}$$

The integral can be interpreted as a Laplace transform with respect to internal time s, while expressing the exponential in terms of a Fox H-function [148],

$$\exp(-z) = H_{0,1}^{1,0} \left[z \,\middle|\, \begin{matrix} - \\ (0,1) \end{matrix} \right]. \tag{3.16}$$

[5] The subordination integral (3.14) is evaluated numerically to generate the PDF plots in Fig. 3.2. For the associated stable densities ℓ_μ [Eq. (2.9)] and ℓ_α^+ [Eq. (2.22)], numerical evaluation tools are available for computer programs such as Mathematica or MATLAB.

[6] Methods to estimate such parameters from empirical CCTRW trajectory data are discussed in [147].

After some straightforward manipulations of the H-function [148], this yields the following representation in Fourier-Laplace space,

$$p(k;u) = \frac{\alpha}{u\mu H} H_{1,2}^{1,1}\left[\frac{u^\alpha}{|k|^{\alpha/H}}\,\middle|\,\begin{matrix}(1,\alpha/(\mu H))\\(1,1)\end{matrix}\right]. \tag{3.17}$$

Inverting to laboratory time t and real space x, one gets [148]

$$p(x;t) = \frac{t^{-H}}{2\mu\sqrt{\pi}} H_{2,3}^{2,1}\left[\frac{|x|}{2t^H}\,\middle|\,\begin{matrix}(1-1/\mu,1/\mu);(1-H,H)\\(0,1/2),(1-H/\alpha,H/\alpha);(1/2,1/2)\end{matrix}\right]. \tag{3.18}$$

Since for H-functions, series representations for small and large arguments are known, one can now analyse in detail the behaviour around the origin and in the tails. Series expansions can in principle be evaluated up to any order, see Refs. [148, 149]. Here, the focus is on the leading order contributions to the PDF, or equivalently, to the scaling function $q(z) = p(z;1)$.

In the vicinity of the starting position, $z \approx 0$, the qualitative shape depends highly on the ratio α/H, if waiting times are heavy tailed, $\alpha < 1$:

$$q(z\approx 0) \sim \begin{cases} \text{const}\cdot|z|^{-1+\alpha/H}, & \alpha/H < 1,\\ \text{const}\cdot\log|z/2|, & \alpha/H = 1,\\ q(0) - \text{const}\cdot|z|^{-1+\alpha/H}, & 1 < \alpha/H < 3,\\ q(0) - \text{const}\cdot z^2\log|z/2|, & \alpha/H = 3,\\ q(0) - \text{const}\cdot z^2, & \alpha/H > 3. \end{cases} \tag{3.19}$$

The constants depend on the parameters μ, α, H, but not on the scaling variable z. Thus, the behaviour around the origin can be divergent ($\alpha/H \leq 1$), continuous with divergent derivative ($1 < \alpha/H < 2$), continuous with discontinuous first derivative ($2 \leq \alpha/H \leq 3$), and continuous with vanishing first derivative ($\alpha/H > 3$). While the cusp-like shape for low values of α/H is reminiscent of CTRW propagators, the increasingly smoother shape for higher values of α/H is imitating Gaussian distributions. Also note that in the absence of heavy-tailed waiting times, corresponding to $\alpha \to 1$, the scaling function returns to the class of stable laws, which are completely smooth (i.e., infinitely differentiable) everywhere. Example plots are given in Fig 3.2.

In contrast, if $\mu < 2$, the heavy tails are directly inherited from the underlying jump length distribution,

$$q(z\to\infty) \simeq |z|^{-1-\mu}, \quad \text{for } \mu < 2. \tag{3.20}$$

This holds regardless of which type of correlations or waiting time distributions characterise the motion, see also Fig. 3.2. In the special case of Gaussian jump lengths, the tails of the PDF are of exponential type, $\log[q(z\to\infty)] \simeq -|z|^{1/2+H(1-\alpha)/\alpha}$.

Finally, notice an interesting, but maybe not intuitively expected property of the scaling function $q(z)$. From Eq. (3.18) one can derive [148]

$$q(z)|_{\alpha\to 1} = q(z)|_{H\to 0}. \tag{3.21}$$

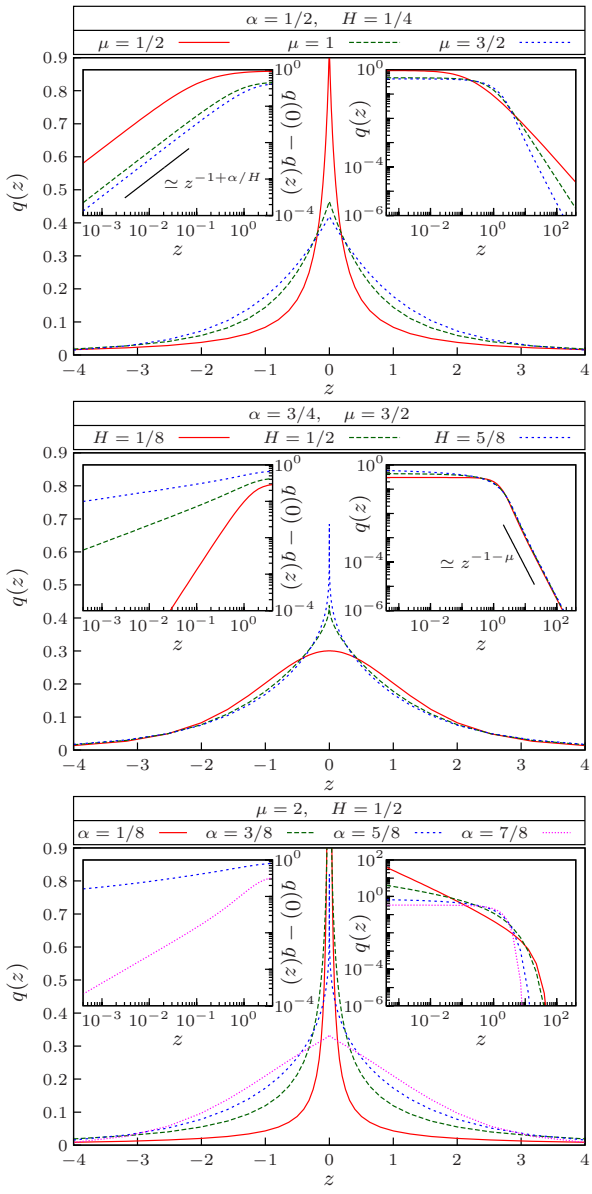


Figure 3.2: Scaling function $q(z)$ for the propagator $p(x;t) = t^{-H}q(xt^{-H})$, numerically evaluated through equation (3.14). The insets detail the behaviour around the origin (*left*) and in the tails (*right*). *Top:* The exponent of power-law tails varies with μ, but the behaviour at the origin is universally $q(0) - q(z) \simeq |z|^{-1-\alpha/H}$. *Centre:* Conversely, a fixed exponent $\mu < 2$ defines the tail properties, $q(z) \simeq |z|^{-1-\mu}$. By varying the ratio α/H, the shape of the maximum turns from a distinct cusp to a smooth Gaussian-like bell. *Bottom:* With $\mu = 2$, tails are stretched exponentials. When $\alpha < H$, $q(z)$ diverges at the origin. With $H = 1/2$, $X^2(t) \sim t$ universally indicates normal diffusion.

The limit $\alpha \to 1$ leads back to a steady time progression, ultimately rendering internal time and laboratory time equivalent. Interestingly, in terms of the shape of the PDFs, this is effectively the same as choosing H very small. Thus, if either waiting times are sufficiently persistent, or jump distances are sufficiently anti-persistent, then the shape of the propagator indicates dynamics devoid of any stalling or trapping mechanisms.

3.4 Comparison with other models of correlated motions

In this section, the correlated continuous time random walk model discussed in the previous sections are compared, contrasted and connected to other existing models of correlated motion.

First, CCTRWs are distinct from the correlated (persistent) random walk models as discussed in [16, 17, 18, 19]. The latter are two-dimensional random walk models, aiming at describing animal foraging and movements patterns. Angular correlations are introduced by means of non-uniform angular distributions, governing the directional evolution of the random walk at each step (akin to the persistent Smoluchowski random walk, see section 1.3.2). Angular and step length distribution define characteristic correlation scales, beyond which the dynamics are essentially Brownian.

The present CCTRW model is a direct continuation of the CTRW with correlated waiting times presented in [150]. The authors discuss a laboratory time process (see Eqs. (23) and (24) in [150], the notation is adopted slightly)

$$T(s) = \int_0^s m(s - s')\, dL_\alpha^+(s'), \qquad \text{with} \qquad m(s) = \int_0^s M(s')\, ds'.$$

While the the correlation kernel $m(s)$ defines the integral representation of laboratory time $T(s)$, the function $M(s) = dm/\,ds$ can be interpreted as a correlation kernel for the noise or waiting time process "$dT/\,ds$" (see Eq. (20) of [150]). Two different types of correlation kernels are taken into consideration. First, power law correlated waiting times, $M(s) \propto s^{-\beta}$, $\beta < 1$, lead to a power law correlated laboratory time process, $m(s) \propto s^{1-\beta}$. By identifying $G = 1 - \beta + 1/\alpha$, $G > 1/\alpha$, one exactly recovers the process definition used here, Eq. (3.3). Since jump lengths in the model of [150] are Gaussian and independent (which, using the language of this chapter, means $\mu = 2$, $K = 1/2$) one can expect a scaling relation $X(t) \sim t^H = t^{K/G} = t^{\alpha/[2\alpha(1-\beta)+2]}$. This is fully consistent with the mean squared displacement analysis in Eq. (43) of [150]. A second interesting choice for the kernel behaviours is an exponentially decaying one, i.e. $M(s) \propto \exp(-\Delta s)$, $\Delta > 0$, corresponding to $m(s) \propto 1 - \exp(-\Delta s)$. While the full scaling behaviour is difficult to calculate explicitly, one can look at the limiting cases $t \ll \Delta$ and $t \gg \Delta$. By virtue of the monotonic increase of the process $T(s)$, this is equivalent to studying approximations with respect to internal time s. For small s, one has $m(s) \simeq s = s^{1-0}$, while for large s, $m(s) \simeq 1 = s^{1-1}$. Hence, one can expect a turnover from the scaling $X(t) \sim t^{\alpha/[2\alpha+2]}$ at $t \ll \Delta$ to $X(t) \sim t^{\alpha/2}$ at $t \gg \Delta$. This is in perfect agreement with the mean squared displacement results Eqs. (36) and (38) in [150].

The authors of [147] discuss a subordinated process close in spirit to the one presented here. In their case, the model ingredients are scale-free, correlated jump statistics as in Eq. (3.1). Waiting times are not correlated, yet have an interesting distributional property: tail statistics are intermediately power law distributed, but an exponential cut-off introduces an intrinsic time scale and ensures finiteness of all moments. Consequently, the diffusion process $X(t)$, defined via subordination, behaves very differently during different temporal regimes, as separated by the average waiting time. Of particular interest is the important discussion on the estimation of parameters from sample trajectory data.

While correlations in CTRW waiting times are not discussed explicitly in [151], the author establishes an intimate principal connection between correlated waiting times, fluctuating waiting time distributions, and rate fluctuations in the underlying higher-dimensional Markovian dynamics. The methods and concepts from this work may thus provide a useful approach to build a microscopic foundation of correlated CTRWs on the one hand, and to find reasonable model extensions on the other.

Finally, there is a close connection to the correlated CTRW introduced in Ref. [152]. On the discrete random walk level, the basic idea is to define a non-stationary and correlated sequence of jump lengths or waiting times in terms of separate random walk processes. For instance, correlated jump lengths δY_n are derived from a Lévy flight in jump length space. In other words,

$$\begin{aligned} \delta Y_n &= \sum_{j=1}^{n} \xi_j, \\ Y_n &= \sum_{j=1}^{n} \delta Y_j = \sum_{j=1}^{n} \sum_{k=1}^{j} \xi_k, \end{aligned} \tag{3.22}$$

where the ξ_j are independent, symmetric μ-stable random variables. The intuitive way of guessing a long time limit approximation of this process can be found by replacing sums with integrals:

$$Y(s) = \int_0^s \int_0^{s'} dL_\mu(s'')\, ds' = \int_0^s L_\mu(s')\, ds'. \tag{3.23}$$

Indeed the convergence in distribution of the discrete random walk Y_n to the continuous process $Y(s)$ was proved in [153]. The latter can be thought of, according to above equation, as an integrated symmetric Lévy flight. Now according to the spectral analysis discussion brought up in section 3.2, such a process should actually be included in the class of correlated motions discussed here. Indeed, one could also rewrite the double sum in Eq. (3.22) as

$$Y_n = \sum_{j=1}^{n} (n-j)\, \xi_j. \tag{3.24}$$

The analogous step for continuous time is formal integration by parts of Eq. (3.23):

$$Y(s) = \int_0^s (s - s')\ dL_\mu(s'). \tag{3.25}$$

Up to a constant prefactor, this exactly corresponds to definition (3.1) with $K = 1/\mu+1$. One thus finds, in complete accordance with [152, 153], that the integrated Lévy flight is a μ-stable process with superdiffusive scaling $Y(s) \sim s^K = s^{1/\mu+1}$.

Correlated waiting times however, are defined in [152, 153] in a slightly different manner. There, consecutive waiting times δT_n are taken from a symmetric Lévy flight, subject to a reflecting boundary condition at $\delta T_n = 0$. In short,

$$\begin{aligned} \delta T_n &= \left| \sum_{j=1}^{n} \zeta_j \right|, \\ T_n &= \sum_{j=1}^{n} \delta T_j = \sum_{j=1}^{n} \left| \sum_{k=1}^{j} \zeta_k \right|, \end{aligned} \tag{3.26}$$

where the ζ_j are independent, symmetric α-stable random variables with $0 < \alpha \leq 2$. The continuous version is

$$T(s) = \int_0^s \left| \int_0^{s'} dL_\alpha(s'') \right| ds' = \int_0^s \left| L_\alpha(s') \right| ds', \tag{3.27}$$

which is *not* a stable process [153], and hence cannot be represented by any of the correlated laboratory time processes (3.3). Still, there is a formal analogy in scaling behaviours. It is easy to show that the integrated Lévy flight on the positive half-line, Eq. (3.27), is self-similar with $T(s) \sim s^{1/\alpha+1}$. The present model yields the same scaling for $G = 1/\alpha + 1$; interestingly, this corresponds to a single integration of a one-sided α-stable motion. In the case of independent Gaussian jump lengths, $\mu = 2$ and $K = 1/2$, such scaling produces subdiffusive dynamics $X(t) \sim t^{K/G} = t^{\alpha/[2(1+\alpha)]}$, as previously found in [152].

4 Ageing renewal theory

The continuous time random walk (CTRW) was introduced in section 2.2 as a model for anomalous diffusion. This class of processes sets itself apart from standard models by the lack of a limiting time scale for microscopic dynamics. Instead of discussing the full CTRW diffusion process, one can extract and focus on this core anomaly. The single steps of a random walk are interpreted as the events of a *renewal process*. This approach has two advantages. On the one hand, renewal theory is very versatile, owing to its abstract formulation. Renewals can alternatively be interpreted as switching events in two-state models, domain crossings of a random motion, the arrival of a shot-noise signal, etc. In complex, disordered media, processes with scale-free waiting times play a particularly prominent role. It is therefore desirable to find a unified analytic foundation for such anomalous dynamics. First steps in this direction are made in the course of section 4.1, in terms of a detailed discussion of the probability density function of the ageing (i.e. scale-free) renewal process, its ageing properties and consequences for ensemble measurements.

On the other hand, anomalous diffusion as observed in experiments is usually a multi-layered phenomenon, where different types of anomaly are encountered; see the examples given in the introduction to section 3. Knowing precisely the effects of scale-free relaxation dynamics – ageing and weak ergodicity breaking among others – helps to single them out from other anomalous mechanisms like long-range memory. The discussion of section 4.2 lays out in detail how the characteristic behaviour of the ageing renewal process translates to the anomalous diffusion properties of CTRWs. In particular, section 4.2.5 provides a specific example for a combined anomalous diffusion model, which captures both the memory effects of a fractional Langevin equation, and the scale-free relaxation of a CTRW.

This chapter summarises results preciously published in Refs. [142, 143], where the reader can find additional information.

4.1 The Ageing renewal process

4.1.1 Motivation

A stochastic process $n(t)$ counting the number of some sort of events occurring during a time interval $[0, t]$ is called a renewal process, if the time spans between consecutive events are independent, identically distributed random variables [154]. Renewal theory does not specify the exact meaning or effect of a single event. It could be interpreted as the appearance of a head in a coin tossing game, the arrival of a bus or of a new customer in a queue. In a mathematical formulation, events remain abstract objects characterised by the time of their occurrence. Thus, not surprisingly, renewal processes are at the core of

many stochastic problems found throughout all fields of science. Maybe the most obvious physical application is the counting of decays from a radioactive substance. This is an example of a Poissonian renewal process: the random time passing between consecutive decay events, the waiting time, has an exponential probability density function $\psi(t) = \tau^{-1} \exp(-t/\tau)$. In other words, events here are observed at a constant rate τ^{-1} (that is, if the sample is sufficiently large and the half-life of individual atoms sufficiently long).

A physical problem of more contemporary interest is subrecoil laser cooling [155, 156]. Two counter-propagating electromagnetic waves can cool down individual atoms to an extent where they randomly switch between a trapped (i.e. almost zero momentum) state and a photon emitting state. Successive life times of individual states are found to be independent and stationary in distribution. Hence, the transitions from the trapped to the light emitting state form the events of a renewal process. Similar in spirit, colloidal quantum dots [157, 158] switch between bright states and dark states under continuous excitation. In contrast to the Poissonian decay process, the latter two examples feature a power law distribution of occupation times t, whose long time asymptotics are of the heavy-tailed form (2.21), that is

$$\psi(t) \sim \frac{1}{|\Gamma(-\alpha)|} \frac{\tau^\alpha}{t^{1+\alpha}}, \qquad 0 < \alpha < 1. \tag{4.1}$$

This type of distributions is not uncommon for physically relevant renewal processes. To see this, consider a simple unbounded, one-dimensional Brownian motion, and let $n(t)$ count the number of times the particle crosses the origin. Then the waiting time between two crossings is of the form (4.1) with $\alpha = 1/2$. Indeed, a random walk of electron-hole pairs either in physical space, or in energy space, was proposed as a mechanism leading to the power law statistics of quantum dot blinking [159]. In general, whenever events are triggered by domain crossings of a more complex, unbounded process, power law distributed waiting times are to be expected. In addition, the latter can be interpreted as a superposition of exponential transition times with an (infinitely) wide range of rate constants τ. The power-law form (4.1) for the inter-event statistics is a typical ansatz to explain such renewal dynamics in highly disordered or heterogeneous media such as spin glasses [160], amorphous semiconductors [77] or biological cells [114, 115].

Distributions as in (4.1) imply a divergent average waiting time, $\langle t \rangle = \int_0^\infty t\, \psi(t)\, dt = \infty$. Renewal processes of this type are said to be scale-free, since, roughly speaking, statistically dominant waiting times are always of the order of the observational time; see sections 2.2.1 and 2.2.4 for a more complete idea. Hence, their outstanding characteristics play out most severely on long time scales: while for $t \gg \tau$, Poissonian renewal processes behave quasi-deterministically, $n(t) \approx t/\tau$, heavy-tailed distributions lead to nontrivial random properties at all times. Stochastic processes of this type are known to exhibit weak ergodicity breaking [110], i.e. time averages and associated ensemble averages of a physical observable are not equivalent. Moreover, despite the renewal property, the process $n(t)$ is non-stationary [105]: events $n_a(t_a, t)$ counted after an unattended ageing period $t_a > 0$, i.e. within some time window $[t_a, t_a + t]$, are found to be statistically very distinct from countings during the initial period $[0, t]$. Fewer events are counted

during late measurements, in a statistical sense, and thus one also says the process exhibits *ageing*. Deeper analysis reveals that this slowing down of dynamics is due to an increasingly large probability to count no events at all during observation. Intuitively, in the limit of long times $t, t_a \gg \tau$, one would expect to see at least some renewal activity $n_a > 0$. Instead, the probability to have exactly $n_a = 0$ increases steadily, and as $t_a \to \infty$, the system becomes completely trapped.

4.1.2 Model definition

Suppose one is interested ind a series of events that occur starting from time $t = 0$. Later, these events may be specifically identified with the arrival of a bus, the steps of a random walk or the blinking of a quantum dot—for now, they remain abstract. Let $n(t)$ count the number of events that occurred up to time t; refer to it as a *counting process*. The time spans between two consecutive events are called waiting times. They are not necessarily fixed. Instead, they are taken to be independent, identically distributed random variables. In such a case, it is justified to refer to events as *renewals*: The process $n(t)$ is not necessarily Markovian, but any memory on the past is erased with the occurrence of an event—the process is renewed. The probability density function (PDF) of individual waiting times is denoted by $\psi(t)$. Obviously, the nature of this quantity heavily influences the statistics of the overall renewal process. Figure 4.1 shows realisations for deterministic periodic renewals, $\psi(t) = \delta(t - \tau)$, for Poissonian waiting times, $\psi(t) = \tau^{-1} \exp(-t/\tau)$, and for heavy-tailed waiting times, i.e. $\psi(t)$ has long-t asymptotics (4.1). In all cases, the scaling parameter $\tau > 0$ serves as a microscopic scale parameter for individual waiting times.

First, study the inset of Fig. 4.1, which focuses on the initial evolution of these processes at short time scales. The complete regularity of the deterministic renewals is distinct, but the two random counting processes are not clearly discernible by study of such single, short-period observations. Now compare this to the main figure, which depicts realisations of the processes on much longer time scales. Here, the realisations of the deterministic and Poissonian renewal processes look almost identical. Recall that for independent exponential waiting times, the average time elapsing until the nth step is made increases linearly with n, while the fluctuations around this average grow like $n^{1/2}$. Thus, the relative deviation from the average decays to zero on longer scales. Roughly speaking, on time scales that are long as compared to the average waiting time $\langle t \rangle = \tau$ one observes a quasi-deterministic relation $n(t) = t/\tau$.

For heavy-tailed waiting times as in Eq. (4.1), the picture is inherently different. Above scaling arguments fail, since the typical time scale to compare to, the average $\langle t \rangle$ of a single waiting time, is infinite. This is why this type of dynamics are sometimes referred to as scale-free and they are studied in light of generalised central limit theorems [87]. Some of the involved analytic aspects are sketched in the following section. Most importantly, it turns out that in the absence of a typical time scale, waiting time periods persist and are statistically relevant on arbitrarily long time scales; see also the discussions in sections 2.1.2 and 2.2.1. The effect is clearly visible in Fig. 4.1: The renewal

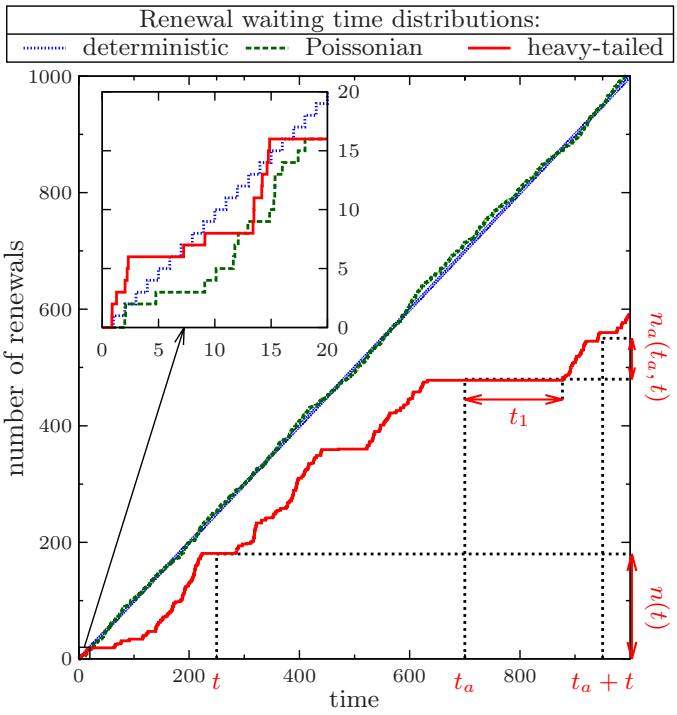


Figure 4.1: Sample realisations for three different types of renewal processes. Events are separated by waiting times, which are independent, identically distributed according to a probability density function $\psi(t)$. Depicted here are deterministic periodic renewals, $\psi(t) = \delta(t-1)$, Poissonian waiting times, $\psi(t) = e^{-t}$, and of heavy-tailed waiting times, $\psi(t) = 4.5 \cdot (5t+1)^{-1.5}$, see key on top. *Inset:* During an observation on short time scales, the randomness of Poissonian and heavy-tailed waiting times contrasts the regular progression of the deterministic process. However, it is difficult to distinguish the two random processes on this level of analysis. *Main figure:* Observation on long time scales reveals the profound statistical difference between the two random processes: The Poissonian renewal process behaves almost deterministically on time scales long as compared to $\langle t \rangle = 1$. At the same time, the random nature of heavy-tailed waiting times is visible on all time scales, as $\langle t \rangle = \infty$. On the one hand, this means that counting the number of renewals n up to time t yields different results from one process realisation to the next. On the other hand, ageing should be taken into account: The statistics of the number of renewals $n(t)$ during observation time $[0, t]$ can turn out to be different from the statistics of $n_a(t_a, t)$, the number of renewals in $[t_a, t_a + t]$. The conceptual difference of the two counting processes lies in the forward recurrence time t_1, which measures the time span between the start of observation at time t_a and the counting of the very first event, see graph.

process $n(t)$ remains a nontrivial random process, even when $t \gg \tau$.

At this point, also introduce the concept of an *aged* measurement: while the renewal process starts at time 0, an observer might only be willing to or capable of counting events starting from a later time $t_a > 0$. In place of the total number of renewals $n(t)$ he then studies the counted fraction $n_a(t_a, t) = n(t_a+t)-n(t_a)$. The fundamental statistical difference between the renewal processes n and n_a stems from the statistics of the time period t_1 which passes between start of the measurement at t_a and the observation of the very first event. It is referred to her as a *forward recurrence time* (see Ref. [105] and section 2.2.1) and denote its PDF by $h(t_a, t_1)$. If the observer counts starting at time $t_a = 0$, the forward recurrence time is simply distributed like any other waiting time, $h(0, t_1) = \psi(t_1)$. But for later, aged measurements, $t_a > 0$, the distribution is different, as indicated in Fig. 4.1.

The dependence of the statistical properties of the counted renewals n_a on the starting time of the measurement t_a is called an *ageing effect*. Its impact crucially depends on the waiting time distribution in use. For instance, a Poissonian renewal process is a Markov process, meaning here that events at all times occur at a constant rate. In this case, $h(t_a, t) \equiv \psi(t)$, so there the process does not age. For any other distribution, $h(t_a, t) \neq \psi(t)$; yet, if the average waiting time is finite, then on time scales long in comparison to the average waiting time $\langle t \rangle$, the renewal process behaves quasi-deterministically (details below). Thus ageing is in effect, but becomes negligible at long times. But scale-free waiting times as in (4.1) result in nontrivial renewal dynamics, with distinct random properties, and ageing effects should be taken into careful consideration. Thus, in the following section, the statistics of the renewal processes $n(t)$ and $n_a(t_a, t)$ are studied and compared in detail in terms of their probability distributions and the consequences for calculating aged ensemble averages.

4.1.3 Long time scaling limit

Several authors have studied the ageing renewal process as defined above and its long time approximation, see [3, 102, 103, 105] and references therein. The basic concept of a scaling limit is demonstrated in the following, focusing on the calculation of the rescaled PDF.

The probability $p(n; t)$ of the random number of events n taking place up to time t takes a simple product form in Laplace space[1]

$$\begin{aligned} p(n; s) = \mathscr{L}_{t\to s}\{p(n; t)\} &= \int_0^\infty e^{-st} p(n; t)\, dt \\ &= \psi(s)^n \, \frac{1 - \psi(s)}{s}, \end{aligned} \tag{4.2}$$

[1]The Laplace transform $f(s) = \mathscr{L}_{t\to s}\{f(t)\} = \int_0^\infty f(t) \exp(-st)\, dt$ of a function $f(t)$ is expressed, throughout the complete chapter 4, by explicit dependence on the Laplace variable s. Likewise, $f(s_a)$ and $f(t_a)$ are Laplace pairs, and Laplace inversion is occasionally indicated explicitly as $f(t) = \mathscr{L}^{-1}_{s\to t}\{f(s)\}$.

which is a direct consequence of the renewal property of the process. It can be read as the probability to count exactly n steps at some arbitrary intermediate points in time and not seeing an event ever since.

In addition, the measurement may be made after some time period t_a during which the process evolved unattendedly. Therefore, consider $n_a(t_a, t) = n(t_a + t) - n(t_a)$, the number of events that happen during the time interval $[t_a, t_a + t]$. The corresponding probability $p_a(n_a; t_a, t)$ reads in double Laplace space, $(t_a, t) \to (s_a, s)$, [105]

$$p_a(n_a; s_a, s) = \begin{cases} (s_a s)^{-1} - h(s_a, s)s^{-1}, & n_a = 0 \\ h(s_a, s)\psi^{n_a - 1}(s)(1 - \psi(s))s^{-1}, & n_a \geq 1 \end{cases} \tag{4.3a}$$

introducing

$$h(s_a, s) = \frac{\psi(s_a) - \psi(s)}{s - s_a} \frac{1}{1 - \psi(s_a)}, \tag{4.3b}$$

the PDF of the forward recurrence time t_1 as described phenomenologically in the preceding section, or defined explicitly through Eq. (2.24). The interpretation of Eq. (4.3) is straightforward: The probability to see any events at all during the period of observation equals the probability that $t_1 \leq t$. Furthermore, the observer counts exactly n_a events if the first event at time $t_a + t_1$ is followed by $(n_a - 1)$ events at intermediate times $t_a + t_i$ and an uneventful time period until the measurement ends at $t_a + t$.

For a check, notice that for Poissonian waiting times, $\psi(t) = \tau^{-1} \exp(-t/\tau)$, one has $\psi(s) = 1/(s\tau + 1) = s_a h(s_a, s)$, and hence $h(t_a, t) \equiv \psi(t)$. The Poissonian (i.e. Markovian) renewal process is unique in this respect.

Now assume that waiting times are heavy-tailed, i.e. their PDF is of the form (4.1). Waiting times of this type have a diverging mean value, which has severe consequences for the resulting renewal process, even in the scaling limit of long times. To see this, introduce a scaling constant $c > 0$ and rescale time as $t \mapsto t/c$. Then write the following approximation in Laplace space, where, by virtue of Tauberian theorems [3], small Laplace variables correspond to long times,

$$\mathscr{L}_{t \to s}\{c\psi(ct)\} = \psi(s/c) \approx 1 - (s\tau/c)^\alpha. \tag{4.4}$$

By rescale the counting process accordingly, meaning here $n \mapsto n/c^\alpha$, one can take the limit $c \to \infty$ to arrive at a long time limiting version of the renewal process. For instance, the probability distribution in Eq. (4.2) takes the following form:

$$\begin{aligned} p(n; s) &\mapsto \mathscr{L}_{t \to s}\{c^\alpha p(nc^\alpha; ct)\} = c^{\alpha - 1} p(nc^\alpha; s/c) \\ &\approx \tau^\alpha s^{\alpha - 1} \left[1 - \frac{n(s\tau)^\alpha}{nc^\alpha}\right]^{nc^\alpha} \\ &\to \tau^\alpha s^{\alpha - 1} \exp[-n(s\tau)^\alpha] \quad (c \to \infty). \end{aligned} \tag{4.5}$$

Note that in these equations, n was turned from an integer to a continuous variable, characterised by a PDF. (Still, for simplicity, one may continue to refer to this variable

as "number of events"). For the sake of notational simplicity, set $\tau = 1$ in what follows, bearing in mind that the rescaled time variables t and t_a are measured in units which are by definition large as compared to the microscopic time scale of individual waiting times.

The evolution of the probability density with respect to real time t is now given through

$$p(n;t) = \frac{1}{\alpha} t n^{-1-1/\alpha} \ell_\alpha^+(nt^{-\alpha}), \tag{4.6a}$$

where

$$\ell_\alpha^+(z) = \mathscr{L}_{s\to z}^{-1}\{\exp(-s^\alpha)\}\,. \tag{4.6b}$$

This perfectly coincides with Eq. (2.25) (set $\kappa_\alpha = \tau^{\alpha-1} = 1$), obtained by a different argument. Remarkably, $n(t)$ thus remains a nontrivial random quantity even after the rescaling procedure. The special limit $\alpha \to 1$ is representative for finite average waiting times. In this case, the Laplace transform in Eq. (4.4) is a moment generating function, and thus returns $\tau = \langle t \rangle$. In the scaling limit such a process collapses to a deterministic counting process: Eq. (4.6) then implies $p(n;t) = \delta(n-t)$.

In contrast, for any $0 < \alpha < 1$, $n(t)$ obeys a scaling relation $n(t) \sim t^\alpha$ and follows a Mittag-Leffler law [102], directly related to the one-sided stable density[2] $\ell_\alpha^+(z)$ [87]. The latter is a fully continuous function on the positive half-line $z \geq 0$. This implies in particular that for $t > 0$, the probability to have exactly $n(t) = 0$ is infinitely small. Apparently, for the long time scaling limit of the counting process $n(t)$, the length of the very first single waiting time is negligible and the observer starts counting events immediately after initiation of the process.

The procedure for finding the PDF p_a of counted events n_a in an aged measurement, $t_a \geq 0$, is analogous. In the long time scaling limit defined above, Eqs. (4.3) and (4.3b) turn into

$$p_a(n_a;s_a,s) = \delta(n_a)\left[\frac{1}{s_a s} - m_\alpha(s_a,s)\right] + h(s_a,s)p(n_a;s), \tag{4.7a}$$

$$m_\alpha(s_a,s) = h(s_a,s)/s, \tag{4.7b}$$

$$h(s_a,s) = \frac{s_a^\alpha - s^\alpha}{s_a^\alpha(s_a - s)}. \tag{4.7c}$$

Eq. (4.7) demonstrates the ageing time's distinct influence on the shape of the PDF of the number of events. Most remarkably, as $t, t_a > 0$, the occurrence of a term proportional to $\delta(n_a)$ indicates a nonzero probability for counting exactly $n_a(t_a,t) = 0$. This means that possibly no events at all are observed in the time interval $[t_a, t_a + t]$. This is a quite distinct ageing effect, contrasting the immediate increase of the non-aged counting $n(t)$. Only the limit $\alpha \to 1$ leads back to a trivial deterministic, non-ageing counting process, and consequently $p_a(n_a;t_a,t) \equiv p(n_a;t) = \delta(n_a - t)$.

[2] Numerical tools for computing stable densities are available for common programs such as Mathematica or Matlab.

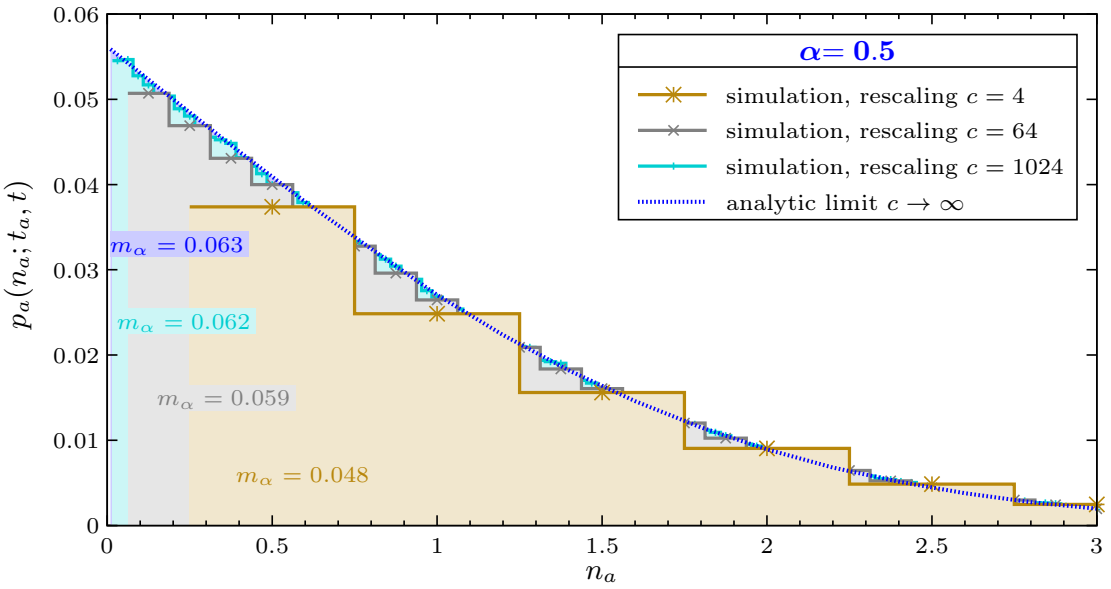


Figure 4.2: Scaling convergence of the ageing renewal process. A renewal process with a waiting time distribution $\psi(t) = (4\pi)^{-1/2}(\pi^{-1} + t)^{-3/2}$ is generated by numeric simulation. The distribution is of the heavy-tailed form (4.1) with $\alpha = 0.5$, $\tau = 1$ (a.u.). Time is rescaled as $t \mapsto t/c$ and renewals as $n_a \mapsto n_a/c^\alpha$. The plot shows the PDF $p_a(n_a; t_a, t)$ for the number of renewals within the time interval $[t_a, t_a + t]$, with $t_a = 100$ and $t = 1$, in terms of the rescaled quantities. The area below each step in the graph represents the probability to count a certain number of (rescaled) renewals; the respective values are indicated by symbols. As the scaling constant c increases (see key), p_a converges to the analytic, smooth scaling limit, i. e. the PDF given through Eqs. (4.8) to (4.8c) (same as in Figs. 4.4-4.6, right centre panel). Since here, only renewal probabilities for $n_a > 0$ are shown, the total area below each graph equals the probability m_α to count any events at all. Ensemble statistics are based on data from 10^7 independent renewal process realisations.

For the aged PDF p_a, Laplace inversion of Eqs. (4.5) to (4.7c) to real time t_a, t yields [106, 107]

$$p_a(n_a; t_a, t) = \delta(n_a)\,[1 - m_\alpha(t_a, t)] + h(t_a, t) *_t p(n_a; t) \tag{4.8a}$$

with [24, 104, 105]

$$\begin{aligned} m_\alpha(t_a, t) &= \int_0^t h(t_a, t')\,dt' \\ &= \frac{B\big([1 + t_a/t]^{-1};\, 1-\alpha, \alpha\big)}{\Gamma(1-\alpha)\Gamma(\alpha)} \equiv m_\alpha(t_a/t), \end{aligned} \tag{4.8b}$$

and where

$$h(t_a, t) = \frac{\sin(\pi\alpha)}{\pi} \frac{t_a^\alpha}{t^\alpha (t_a + t)}. \tag{4.8c}$$

Here the asterisk $*_t$ indicates a Laplace convolution with respect to time t. Figure 4.2 gives a first example of how such an aged PDF behaves. It depicts the case $\alpha = 0.5$ at high ages, $t_a/t = 100$. In addition the plot demonstrates scaling convergence: if a renewal process with simple power-law waiting time distribution is monitored on increasingly long scales for time and event numbers, then its statistics approach the continuous limit described by Eqs. (4.8) to (4.8c).

At this point, a brief remark on the issue of broadly distributed waiting times is in place. As such are designated waiting times characterised by the algebraic decay (4.1), albeit with an exponent $1 < \alpha < 2$. They result in renewal dynamics which are, in a sense, intermediate between simple Poissonian and scale-free behaviour. On the one hand, the average waiting time $\langle t \rangle$ is finite. Hence, by virtue of above scaling arguments, the renewal process $n(t)$ behaves quasi-deterministically on time scales which are arbitrarily large as compared to $\langle t \rangle$. On the other hand, the fluctuations around the average waiting time are considerable, since $\langle t^2 \rangle = \infty$. This has a remarkable consequence for the forward recurrence time. The PDF of the latter in this case becomes, at infinite age [105], $h(t_a = \infty, t) \simeq \langle t \rangle^{-1} (t/\tau)^{-\alpha}$, which is in fact heavy-tailed. In other words, although the average time passing between events is finite, the average time passing before the very first event is infinite. The inferred ageing effect might thus still be notable on time scales t several orders larger than $\langle t \rangle$. Therefore, the parameter regime $1 < \alpha < 2$ certainly deserves a thorough investigation; see also the related discussion in [161, 162, 163]. However, this is not the scope of the present work, and in what follows, the tail parameter is restricted to $0 < \alpha < 1$.

In this case, Eqs. (4.8) to (4.8c) relate the aged PDF p_a to the non-aged PDF p via the PDF h of the forward recurrence time. m_α is the probability to count any events at all during observation. Its representation in terms of an incomplete Beta function $B(z;\, a, b)$ (see Appendix of Ref. [143]) is found by a simple substitution $u = t'/(t' + t_a)$. It can be written as a function of the ratio t_a/t alone and the latter may be used as a more precise and quantitative notion of the *age* of the measurement. In particular, the

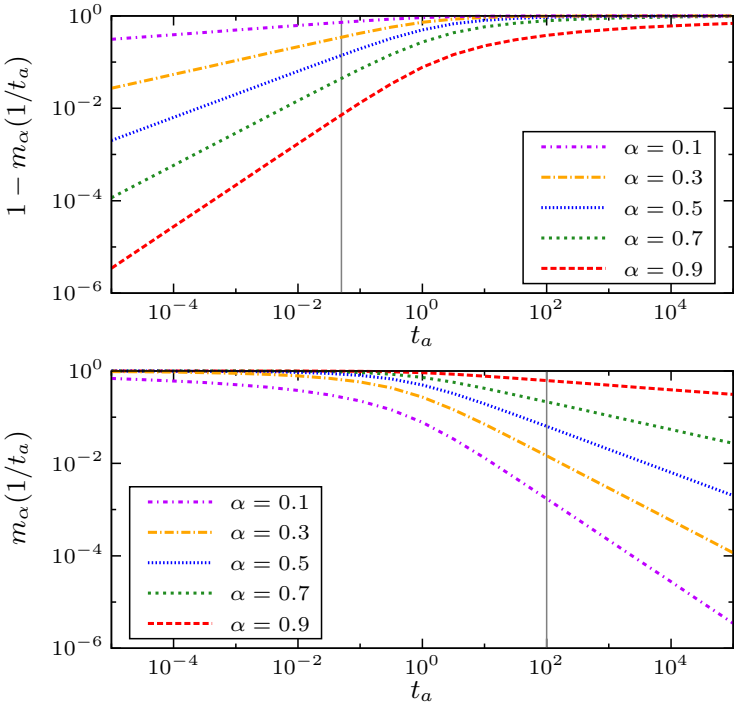


Figure 4.3: Double logarithmic plots of the probability m_α to observe any events during the measurement period $[t_a, t_a + t]$ (*bottom*) and the complementary probability $1 - m_\alpha$ (*top*), both as a function of age t_a at $t = 1$. The full analytic behaviour is given by Eq. (4.8b). Notice the initial power law increase $(1 - m_\alpha) \simeq (t_a/1)^\alpha$ and the final power law decay $m_\alpha \simeq (1/t_a)^{1-\alpha}$. Vertical lines indicate the values of t_a used for the plots in Figs. 4.4-4.6.

process or measurement or observation is called *slightly aged*, if $t_a \ll t$. Conversely, it is *strongly* or *highly aged*, if $t_a \gg t$. These two limiting regimes are now looked into more deeply.

4.1.4 Ageing probability distribution

Slightly aged PDF. First, notice that in Eqs. (4.8) to (4.8c), the PDF h of the forward recurrence time appears inside integrals, and should be interpreted in a distributional sense. For instance, in the limit $t_a \to 0$ one should recover the PDF of a non-aged system, $p_a \to p$. To confirm this, study the limit $s_a \to \infty$ in Laplace space. There, $h(s_a, s) \sim s_a^{-1}$. Thus one should write $h(0, t) = \delta(t)$ in terms of a Dirac δ-distribution, consistently implying $m_\alpha(0) = 1$ and $p_a(n_a; 0, t) \equiv p(n_a; t)$. Again, the result is that only an observer counting from the initiation of the (rescaled) renewal process, $t_a = 0$, witnesses the onset of activity instantly.

One can go one step beyond this limit approximation and study the properties of a

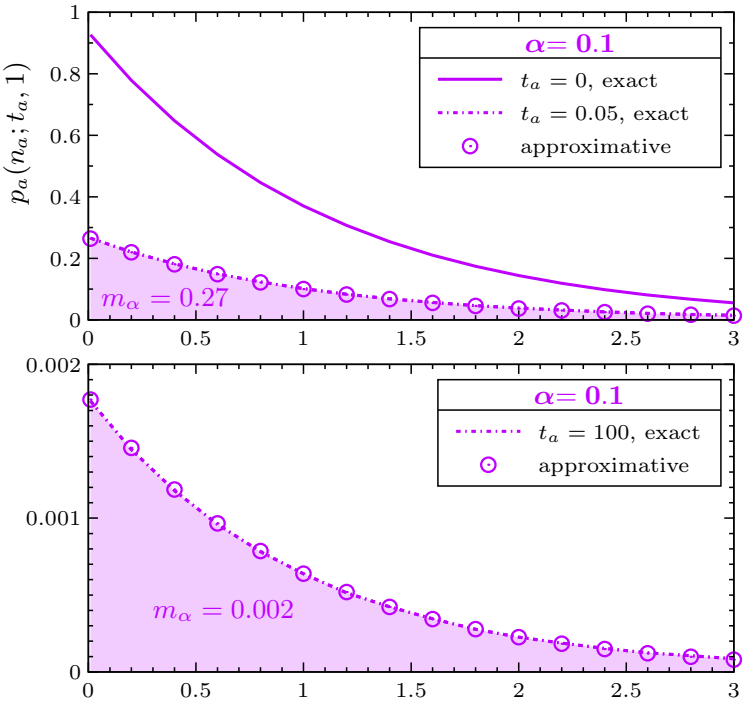


Figure 4.4: Continuous part of the PDF $p(n_a; t_a, t)$ of the number of events n_a counted during the measurement period $[t_a, t_a + t]$. Exact results are provided in terms of a numerical evaluation of the convolution $h(t_a, t) *_t p(n_a; t)$ as defined through Eqs. (4.6) and (4.8) to (4.8c). Sample graphs for $t = 1$ and $\alpha = 0.1$. The top graphs show the non-aged case ($t_a = 0$), the slightly aged case ($t_a = 0.05$) and the leading order approximation for the slightly aged PDF [$t_a = 0.05$, see Eq. (4.9d)]. The bottom graph depicts the highly aged case ($t_a = 100$) and the approximation thereto [$t_a = 100$, see Eq. (4.10d)]. Notice the significantly different vertical scales in the panels, owing to the age-sensitive probability m_α to count any events at all during observation (cf. Fig. 4.3).

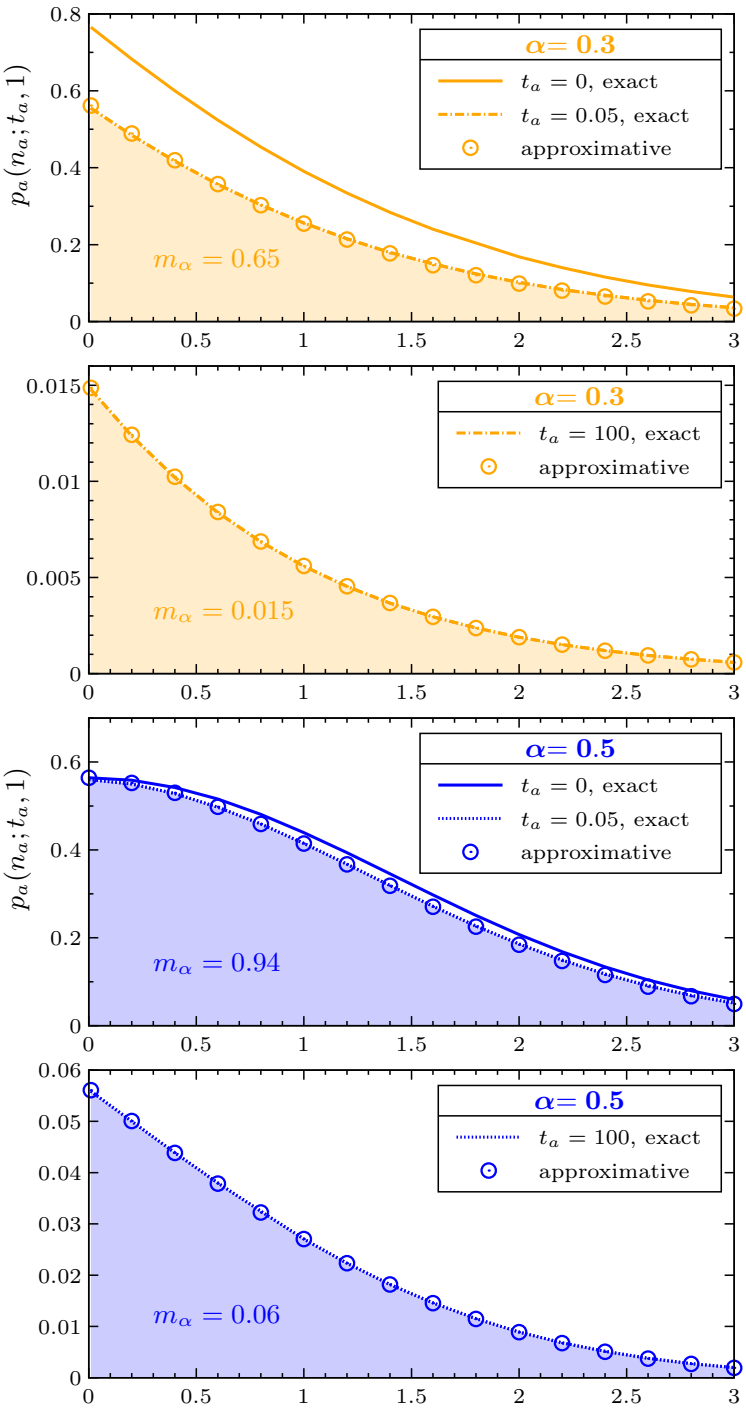


Figure 4.5: Continuation of Fig.4.4 for $\alpha = 0.3, 0.5$.

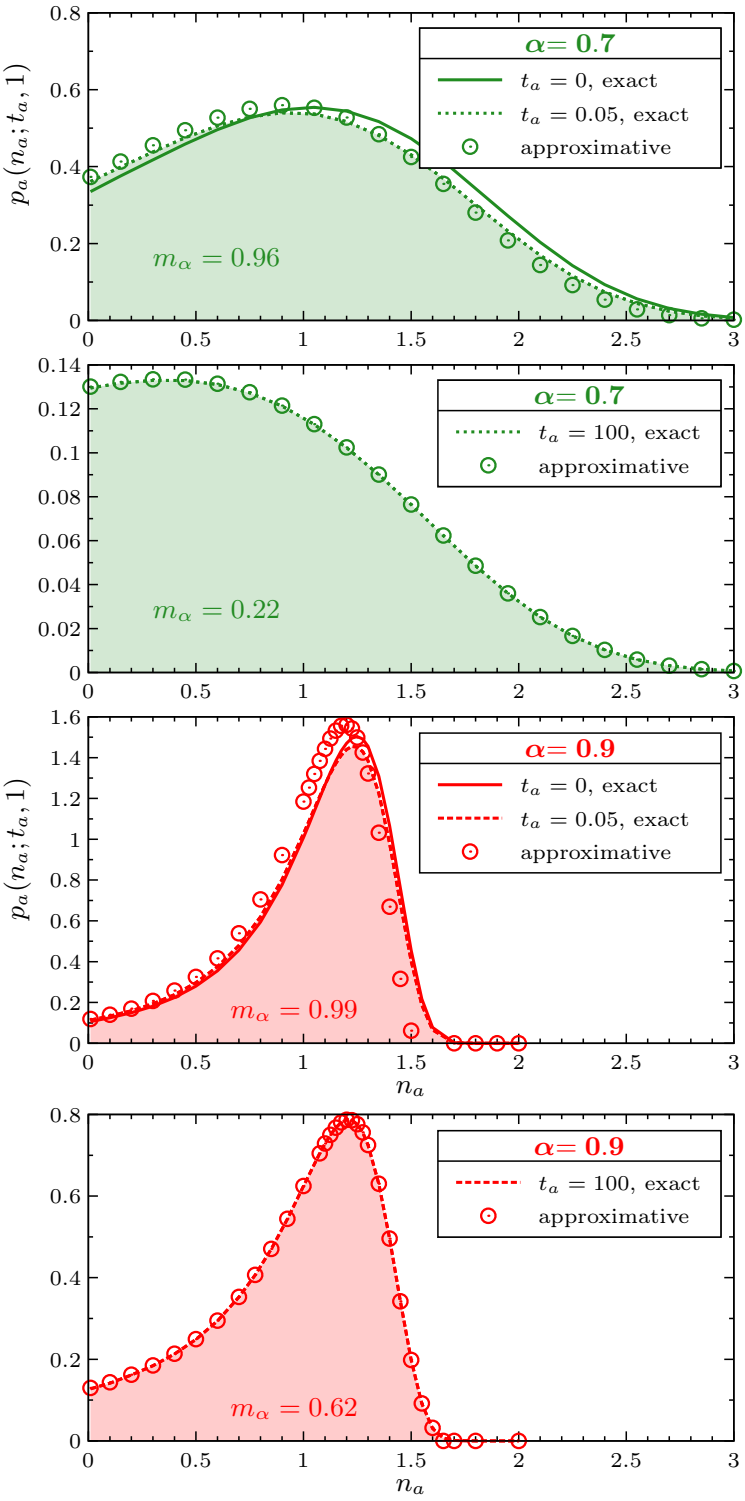


Figure 4.6: Continuation of Fig.4.5 for $\alpha = 0.7, 0.9$.

slightly aged system ($t_a \ll t$ and $s_a \gg s$). Write $h(s_a, s) \sim s_a^{-1} - s_a^{-1-\alpha} s^\alpha$ to find, by use of Tauberian theorems, that

$$1 - m_\alpha(t/t_a) \sim \frac{(t_a/t)^\alpha}{\Gamma(1+\alpha)\Gamma(1-\alpha)}, \tag{4.9a}$$

and

$$\begin{aligned} h(t_a, t) *_t p(n_a; t) &\sim \mathscr{L}_{s\to t}^{-1}\{s^{\alpha-1} e^{-n_a s^\alpha}\} - \frac{(t_a/t)^\alpha}{\Gamma(1+\alpha)} \mathscr{L}_{s\to t}^{-1}\{s^{2\alpha-1} e^{-n_a s^\alpha}\} \\ &\sim p(n_a; t) - \frac{t_a^\alpha}{\Gamma(1+\alpha)} \mathscr{L}_{s\to t}^{-1}\{s^{2\alpha-1} e^{-n_a s^\alpha}\}. \end{aligned} \tag{4.9b}$$

Here, first order corrections are provided in terms of a Laplace inversion. For the analytic discussion, one can alternatively express them as

$$h(t_a, t) *_t p(n_a; t) \sim p(n_a; t) + \frac{t_a^\alpha}{\Gamma(1+\alpha)} \frac{\partial p(n_a; t)}{\partial n_a}, \tag{4.9c}$$

relating them to the non-aged PDF, and thus to the familiar stable density, see Eq. (4.6). Finally, one can also interpret these contributions as Fox-H functions, for which series expansions for small arguments and asymptotics for large arguments (see Appendix of Ref. [143] and [148]; the connection between Fox-H functions, stable densities and fractional calculus is established in [164]) are known:

$$\begin{aligned} h(t_a, t) *_t p(n_a; t) \quad \sim \quad & H_{11}^{10}\left[\frac{n}{t^\alpha} \,\middle|\, \begin{matrix} (1-\alpha, \alpha) \\ (0, 1) \end{matrix} \right] \\ & - \frac{(t_a/t)^\alpha}{\Gamma(1+\alpha)} \frac{1}{t^\alpha} H_{11}^{10}\left[\frac{n}{t^\alpha} \,\middle|\, \begin{matrix} (1-2\alpha, \alpha) \\ (0, 1) \end{matrix} \right]. \end{aligned} \tag{4.9d}$$

From any of these representations, one can get the leading order corrections to the non-aged PDF, which are of the form $(t_a/t)^\alpha$. An intuitive reasoning for this can be given as follows. A slightly late observer generally has to wait the forward recurrence time before seeing the first event. The corrections thus have to account for waiting times which start earlier than the beginning of the observation t_a, but reach far into the observation time window $[t_a, t_a + t]$. Note that the number of (still few) waiting times drawn before time t_a is measured by $n(t_a) \sim t_a^\alpha$, while the probability for any of them to be statistically relevant during an observation of length t is proportional to $\int_t^\infty \psi(t)\, dt \simeq t^{-\alpha}$. The expected number of statistically relevant waiting times starting before but reaching into the time window $[t_a, t_a + t]$ therefore scales like $t_a^\alpha \times t^{-\alpha}$, and so do leading order corrections.

The precise nature of the modifications to the non-aged PDF due to ageing can be separated into two aspects. On the one hand, the continuous part of the aged PDF, $h *_t p$, looses weight to the discrete $\delta(n_a)$-part with growing age t_a of the counting

process. This reflects an increasing probability to have waiting times that not only reach into, but actually span the full observation time window, so that no events at all are observed. A graphical study of the early age-dependence of $1 - m_\alpha$, i.e. the weight of the discrete contribution, is provided in the left panel of Fig. 4.3. The double logarithmic plot clearly demonstrates the initial power law growth $\simeq (t_a/t)^\alpha$. Moreover, for any fixed age t_a/t, the value of $1 - m_\alpha$ is higher for lower values of α. This was to be expected, since lower values of α are related to broader waiting time distributions, and thus stronger ageing effects.

Interestingly, on the other hand, the modification to the continuous part of the aged PDF, $h *_t p$, goes beyond this weight transfer: It is not proportional to the non-aged PDF itself, but, according to Eq. (4.9c), to its slope. With increasing age t_a/t, regions with negative (positive) slope increase (decrease), so that local maxima have a tendency to shift towards $n_a = 0$. Furthermore, one can deduce from the Fox-H function representations, Eq. (4.9d), the behaviour around the origin, $0 < n_a t^{-\alpha} \ll 1$, and the tail asymptotics, $n_a t^{-\alpha} \gg 1$. See Appendix of Ref. [143] for details. For $\alpha > 1/2$, the initial slope of $h *_t p$ is negative, and hence the early ageing effect is an increase of $h *_t p$ between the origin and the local maximum. This is notable since on the long run (i.e. for sufficiently long t_a), the probability to have any $n_a > 0$ tends to zero. For $\alpha < 1/2$, the converse holds: the initial slope is negative, and the PDF in the vicinity of the origin starts dropping from early ages. The PDF tails are, for any value of α, of a compressed exponential form, meaning here $\log(h *_t p) \simeq n^{1/(1-\alpha)}$.

The validity of this early-age approximation can be assessed by studying the left panels of Figs. 4.4-4.6. Depicted are the non-aged PDF, Eq. (4.6) (*full lines*), a numerical evaluation of the continuous part of the aged PDF, $h *_t p$, as given through the full convolution integral in Eq. (4.8) (*dotted/dashed lines*), and the early-age approximation by expanding the Fox-H functions in Eq. (4.9d) as power series (*symbols*). All qualitative statements from the previous paragraph are confirmed by the sample plots. Still, the leading order approximation, Eqs. (4.9b) to (4.9d), is apparently not equally suitable for all values of α. Deviations occur when α gets close to 1. Indeed, one can show that the leading order terms in Eqs. (4.9b) to (4.9d), which account for corrections of $\mathcal{O}([t_a/t]^\alpha)$, are followed by higher order terms of $\mathcal{O}(t_a/t)$. Hence, in general, these "almost leading order" corrections need to be taken into account when α is close to 1. Yet again, in this case, the total corrections with respect to the non-aged PDF, as measured in terms of the weight loss $1 - m_\alpha$, are relatively small anyway.

Highly aged PDF. Conversely, one can approximate for $s_a \ll s$: $h(s_a, s) \sim s_a^{-\alpha} s^{\alpha-1}$. This yields the leading order behaviour for highly aged measurements, $t_a \gg t$:

$$1 - m_\alpha(t/t_a) \sim 1 - \frac{(t/t_a)^{1-\alpha}}{\Gamma(\alpha)\Gamma(2-\alpha)} \tag{4.10a}$$

and

$$h(t_a, t) *_t p(n_a; t) \sim \frac{1}{\Gamma(\alpha)} \frac{1}{t_a^{1-\alpha}} \mathscr{L}_{s\to t}^{-1}\left\{ s^{2\alpha-2} e^{-n_a s^\alpha} \right\}. \tag{4.10b}$$

Above Laplace inversion can be related to the non-aged PDF (4.6) through

$$h(t_a,t) *_t p(n_a;t) \sim -\frac{1}{\Gamma(\alpha)}\frac{1}{t_a^{1-\alpha}}\int_0^t \frac{\partial p(n_a;t')}{\partial n_a}\, dt', \tag{4.10c}$$

or expressed as Fox-H function by

$$h(t_a,t) *_t p(n_a;t) \sim \frac{(t/t_a)^{1-\alpha}}{\Gamma(\alpha)}\frac{1}{t^\alpha} H_{11}^{10}\left[\frac{n}{t^\alpha}\,\middle|\,\begin{matrix}(2-2\alpha,\alpha)\\(0,1)\end{matrix}\right]. \tag{4.10d}$$

Again, modifications account for waiting times that start before but reach into the time window of observation. But in this late time regime, the initiation of the renewal process already lies far in the past, so not all waiting times before t_a have a realistic chance to do that. Instead, assume again that statistically relevant pre-measurement waiting times need to be of the order of t (implying a probability $\simeq t^{-\alpha}$). To estimate their number, note that they must follow events which occur roughly within a time period $[t_a - t, t_a]$. However, at this (late) stage of the renewal process, the average event rate has dropped to $(dn/dt)_{t\approx t_a} \sim t_a^{\alpha-1}$. Thus the expected number of waiting times in the (comparatively short) time period preceding the measurement scales like $\sim t_a^{\alpha-1}t$. This heuristic line of argument explains why for highly aged measurements, leading order contributions are of the order $t_a^{\alpha-1}t \times t^{-\alpha} = (t/t_a)^{1-\alpha}$.

Graphical examples for this regime are included in Figs. 4.3-4.6. Here, the continuous part $h *_t p$ of the aged PDF is not proportional to the slope of the non-aged PDF, but, according to Eq. (4.10c), rather its time integral over the duration of the measurement. Moreover, its Fox-H function representation (4.10d) reveals that the initial slope of the late age PDF is positive for $\alpha > 2/3$ and negative if $\alpha < 2/3$. The far tail behaviour however persists at late ageing stages, as still $\log(h *_t p) \simeq n^{1/(1-\alpha)}$. Notice that in the case of high ages, in Figs. 4.4-4.6 realised as $t_a/t = 100$, the leading order terms (4.10) to (4.10d) are satisfactorily approximating the exact convolution (4.8) for all values of α.

4.1.5 Ageing ensemble averages

From the non-aged PDF in Eq. (4.5) one can derive the expected time behaviour of any function f of the number of events n counted since the initiation of the renewal process:

$$\begin{aligned}\langle f(n(t))\rangle &= \int_0^\infty f(n)\, p(n;t)\, dn \\ &= \mathscr{L}^{-1}_{s\to t}\left\{s^{\alpha-1}\mathscr{L}_{n\to s^\alpha}\{f(n)\}\right\}.\end{aligned} \tag{4.11}$$

Concrete examples are given below. How such an ensemble average is altered if evaluated for the aged counting process? To answer this, substitute n by n_a and p by p_a and insert

Eq. (4.8):

$$\begin{aligned}\langle f(n_a(t_a,t))\rangle &= \int_0^\infty f(n_a)\, p_a(n_a;t_a,t)\, dn_a \\ &= f(0)[1-m_\alpha(t_a/t)] + h(t_a,t) *_t \langle f(n(t))\rangle,\end{aligned} \tag{4.12}$$

with the limiting behaviour provided as in Eq. (4.9c)

$$h(t_a,t) *_t \langle f(n(t))\rangle \sim \begin{cases} \langle f(n(t))\rangle + t_a^\alpha \langle f'(n(t))\rangle/\Gamma(1-\alpha), & t_a \ll t \\ t_a^{\alpha-1}\int_0^t \langle f'(n(t'))\rangle\, dt'/\Gamma(\alpha), & t_a \gg t. \end{cases} \tag{4.13}$$

These equations relate an ensemble average taken for the observation window $[t_a, t_a+t]$ to the respective quantity measured in $[0,t]$. Interestingly, the modifications due to ageing are rather related to the ensemble average of the derivative of the observable, $f'(n) = (\partial f/\partial n)(n)$.

As an example, consider qth order moments of the number of renewals, $f(n) = n^q$, $q > 0$. They are given by

$$\langle n^q(t)\rangle = \mathscr{L}^{-1}_{s\to t}\left\{ s^{\alpha-1}\frac{\Gamma(q+1)}{s^{\alpha q+\alpha}}\right\} = A_0 t^{\alpha q} \tag{4.14a}$$

and

$$\begin{aligned}\langle n_a^q(t_a,t)\rangle &= \frac{\Gamma(q+1)}{\Gamma(\alpha)\Gamma(1+\alpha q-\alpha)}\,(t+t_a)^{\alpha q}\, B\big([1+t_a/t]^{-1};\, 1+\alpha q-\alpha, \alpha\big) \\ &\sim \begin{cases} \langle n^q(t)\rangle + A_1 t_a^\alpha t^{\alpha q-\alpha}, & t_a \ll t \\ B_1 t_a^{\alpha-1} t^{1-\alpha+\alpha q}, & t_a \gg t. \end{cases}\end{aligned} \tag{4.14b}$$

The incomplete Beta function comes into play again by substituting $u = t'/(t'+t_a)$ in the convolution integral in expression (4.12). The time-independent coefficients are given by

$$A_0 = \frac{\Gamma(q+1)}{\Gamma(\alpha q+1)}, \quad A_1 = \frac{\Gamma(q+1)}{\Gamma(1-\alpha)\Gamma(\alpha q+1-\alpha)}, \quad B_1 = \frac{\Gamma(q+1)}{\Gamma(\alpha)\Gamma(\alpha q+2-\alpha)}. \tag{4.14c}$$

For the non-aged system, moments evolve like $t^{\alpha q}$, reflecting the characteristic scaling property of the renewal process, $n(t) \sim t^\alpha$. But for aged systems, $t_a > 0$, the scaling is broken, and moments behave more complex with respect to time. Only at very high ages, $t_a \gg t$, one can approximate again by a single power law. When comparing the growth of the counting processes for the two limiting regimes, Eq. (4.14b) is somewhat ambivalent. At high ages, the probability for observing no events at all tends to one. Consequently, a prefactor $t_a^{\alpha-1}$ lets all moments decay to zero as t_a goes to ∞. However note that for a fixed, large but finite value of t_a, the t-dependence is $\simeq t^{1-\alpha+\alpha q}$ (in accordance with [165]), so the power law exponent is actually larger than for the non-aged moments.

In summary, at higher ages t_a, the absolute number n_a of counted events drops, but it increases faster with measurement time t. In particular, consider the average number of events during observation, $q = 1$: Counting from the start of the process, there is a sublinear behaviour $\langle n \rangle \simeq t^\alpha$; but at fixed, high age of the process, the average rate of events is approximately constant, $\langle n_a \rangle \simeq t$, like in a non-ageing, Poisson type of renewal process.

Another useful expression is the Laplace transform of Eq. (4.6) with respect to the number of renewals, $n \to \lambda$, and respectively Eq. (4.8) with $n_a \to \lambda$, which is obtained through the same techniques. Thus,

$$\langle \exp[-\lambda n(t)] \rangle = \mathscr{L}^{-1}_{s \to t} \left\{ \frac{s^{\alpha-1}}{s^\alpha + \lambda} \right\} = E_\alpha(-\lambda t^\alpha)\,, \tag{4.15a}$$

and

$$\langle \exp[-\lambda\, n_a(t_a, t)] \rangle \sim \begin{cases} \langle \exp[-\lambda\, n(t)] \rangle \, [1 - \lambda t_a^\alpha / \Gamma(1-\alpha)]\,, & t_a \ll t \\ 1 - (t/t_a)^{1-\alpha} E_{\alpha,2-\alpha}(-\lambda t^\alpha)\,/\Gamma(\alpha), & t_a \gg t \end{cases}, \tag{4.15b}$$

where E_α and $E_{\alpha,2-\alpha}$ are (generalized) Mittag-Leffler functions (see Ref.[187] or Appendix of Ref. [143]). Interestingly, here the early first order corrections due to ageing do not significantly alter the t-dependence. At low age, the Mittag-Leffler function interpolates between $1 - \text{const} \cdot t^\alpha$ for $t \ll \lambda^{-1/\alpha}$ and $\text{const} \cdot t^{-\alpha}$ at $t \gg \lambda^{-1/\alpha}$. At high age, the transition is from $1 - \text{const} \cdot t^{1-\alpha}$ to $1 - \text{const} \cdot t^{1-2\alpha}$.

4.1.6 Conditional ensemble averages

The section concludes with a discussion of the following question: how do counting statistics change when selectively evaluating only realisations of the process where $n_a > 0$? In other words, what if all the data where no events happen during the complete time of observation $[t_a, t_a + t]$ is discarded? This is, on the one hand, a relevant question when it comes to the application of renewal theory: An observer who is unaware of the underlying counting mechanism, might misinterpret realisations with $n_a = 0$ as a separate, dynamically different process, since the statistics are so distinct from the remaining continuum $n_a > 0$. A process realisation during which no events occur at all might even not be visible to the observer in the first place. An example follows in the next section. On the other hand, this study also helps us to distinguish two aspects of the ageing PDF: One can deliberately neglect the effect of having single waiting times that cover the full observation window, leading to a weight transfer from the continuous to the discrete part of the PDF. By instead specifically only accounting for waiting times that start before but finish during the period of observation, one can study the modifications of the continuous part beyond its loss of weight. To this intent, consider

the conditional ensemble average

$$\begin{aligned}\langle f(n_a(t_a,t))\rangle_m &\equiv \int_0^\infty f(n_a)p_a(n_a|n_a>0;t_a,t)\;dn_a \\ &= \frac{h(t_a,t)*_t\langle f(n(t))\rangle}{m_\alpha(t/t_a)} \\ &\sim \begin{cases}\langle f(n(t))\rangle, & t_a=0 \\ t^{\alpha-1}\int_0^t\langle f'(n(v))\rangle\;dv/\Gamma(2-\alpha), & t_a/t\to\infty.\end{cases}\end{aligned} \tag{4.16}$$

For $t_a = 0$, as mentioned above, counting of events starts instantly, so the restriction to $n_a > 0$ is redundant. But in the limit of late ages, a possibly nonzero forward recurrence time affects the measurement. Ensemble averages conditioned to $n_a > 0$ have a well defined, nontrivial limiting time dependence, even when t_a/t tends to infinity. This is in contrast to the full ensemble, where at infinite ages, $\langle f(n_a(t_a,t))\rangle \to f(0)$. As an example consider again the time evolution of qth order moments, restricted to $n_a > 0$. The behaviour is

$$\langle n_a^q(t_a,t))\rangle \to 0, \quad \text{as} \quad t_a/t\to\infty, \tag{4.17a}$$

but

$$\langle n_a^q(t_a,t))\rangle_m \sim \begin{cases}A_0\,t^{\alpha q}, & t_a=0, \\ C_0\,t^{\alpha q}, & t_a/t\to\infty,\end{cases} \tag{4.17b}$$

where A_0 is given by Eq. (4.14c) and

$$C_0 = \frac{\Gamma(q+1)\Gamma(2-\alpha)}{\Gamma(\alpha q+2-\alpha)}. \tag{4.17c}$$

Thus, as opposed to the unrestricted ensemble measurement, conditional moments scale like $\sim t^{\alpha q}$ in *both* non-aged and extremely aged systems. Note however, that prefactors are different, and a behaviour deviating from a simple power law is still observable at intermediate ages.

4.2 Ageing continuous time random walks

4.2.1 From renewal theory to continuous time random walks

The theory of CTRWs (see section 2.2) directly builds on the concept of the renewal theory. Many of the intriguing features of this random motion can be viewed in the light of the abstract analytic renewal theory ideas developed above. Consider a random walk process (one dimensional, for the sake of simplicity) where steps do not occur at a fixed deterministic rate, but are instead separated by random, real valued waiting times. The idea is to model the random motion in a complex environment where sticking, trapping or binding reactions are to be taken into account. In the simplest scenario, sticking or

trapping mechanisms are decoupled from diffusion dynamics. On the level of theoretical modelling, this means one can take jump distances of the base random walk to be independent from waiting times. The renewal theory of the previous section is readily extended to describe such idealised systems. Let $x(n)$ be the random walk process as a function of the number of steps n. Then by means of subordination (see section 2.2.1 or Refs.[103, 108, 109, 166]), one can construct a CTRW as $x(t) = x(n(t))$, where $n(t)$ is a renewal counting process. Each step of the random walk is hence interpreted as an event of the renewal process. The properties of $x(t)$ follow from the combined statistics of $x(n)$ and $n(t)$. For example, if $W_{\mathrm{RW}}(x; n)$ denotes the PDF for the position coordinate x after n steps, starting with $x(0) = 0$, then, according to Eq. (2.29),

$$W_{\mathrm{CTRW}}(x; t) = \int_0^\infty W_{\mathrm{RW}}(x; n)\, p(n; t)\, dn \tag{4.18}$$

is the PDF of the associate CTRW process at time t. Note that n is seen as a continuous variable here, so the discussion is on the level of long time scaling limits. In the simplest case, $x(n)$ would be an ordinary Brownian motion so that W_{RW} would be a Gaussian.

Now imagine that a particle is injected into a complex environment, beginning a CTRW like motion at time 0. In general, the experimenter might start its observation at a later time $t_a > 0$. The reason for this could be experimental limitations, or maybe the goal is to study a maximally relaxed system, which makes it necessary to wait for relaxation. In either case, the particle motion is initially unattended. At time t_a the particle is tracked down, and its position at this instant serves as the origin of motion for the following observations. Instead of the full CTRW $x(t)$, the experimenter then monitors the aged CTRW $x_a(t_a, t) = x(t_a + t) - x(t_a)$. Thus, as worked out in the previous section, if the dynamics are characterised by heavy-tailed trapping or sticking times, the issue of ageing has to be taken into careful consideration.

4.2.2 Population splitting

Arguably the most striking ageing effect in CTRW theory is the emergence of an apparent population splitting [107, 142]. The aged renewal process $n_a(t_a, t)$ controls the dynamic activity of the aged CTRW $x_a(t_a, t)$. Consequently, the forward recurrence time t_1 marks the onset of dynamic motion in the monitored window of time. The analysis of ageing renewal theory revealed that for increasingly late measurements, one should expect t_1 to assume considerable values. Physically, this reflects the possibility for the particle to find ever deeper traps or to undergo more complex binding procedures, when given more time to probe its environment before the beginning of observation. In particular, the forward recurrence time is more and more likely to span the full observation time window, $t_1 > t$, so that the particle does not visibly exhibit dynamic activity.

From an external, experimental point of view, an ensemble measurement in such system appears to indicate a splitting of populations. Let, for instance, the base random walk $x(n)$ be Markovian, and let the PDF $W_{\mathrm{RW}}(x; n)$ be a continuous function of x (e.g. Gaussian or stable). In this case one can supplement the ensemble PDF W_{CTRW} from

Eq. (4.18) by its aged counterpart W_{ACTRW}, using renewal theory Eq. (4.12):

$$\begin{aligned} W_{\mathrm{ACTRW}}(x_a; t_a, t) &= \int_0^\infty W_{\mathrm{RW}}(x_a; n_a)\, p_a(n_a; t_a, t)\, dn_a \\ &= \delta(x_a)[1 - m_\alpha(t_a/t)] + h(t_a, t) *_t W_{\mathrm{CTRW}}(x_a; t). \end{aligned} \tag{4.19}$$

This propagator was discussed previously in [107]. The ensemble statistics have a sharply peaked δ-contribution, caused by a fraction $1 - m_\alpha$ of particles which remain immobile at the origin of the observed motion. They contrast the mobile fraction m_α of particles, since the PDF of their arrival coordinates, $h *_t W_{\mathrm{CTRW}}$, varies continuously along the accessible regions of space $0 < x_a < \infty$ (as derived, by virtue of Eq. (4.18), from the continuous nature of the PDFs h, W_{RW} and p). For fixed evaluation time t, the size of the immobile subpopulation increases with growing age t_a, as ageing renewal dynamics terminally come to a complete halt. An exhaustive discussion on the shape of the aged propagator W_{ACTRW} can be found in [107] for both unbiased motion and in the presence of a drift.

Indeed, splitting into subpopulations is a phenomenon encountered in complex environments such as biological cells. Such dynamics was observed for the motion of lipids in phospholipid membranes [167], single protein molecules in the cell nucleus [168], H-Ras on plasma membranes [169], and of membrane proteins [170]. The immobile fraction is also often found in fluorescence recovery after photobleaching experiments [167]. The authors of Ref. [113], who conducted the in vitro experiment on the diffusion in a polymeric network described in section 2.2.5 (see in particular Fig. 2.6), indeed report that a considerable fraction of beads, roughly 80%, remain caged during the full observation time.

It is important to understand that for CTRW types of motion, this effect emerges without assuming non-identical particle dynamics. Even during the evolution of the process, stochastic motion of individual particles in an ensemble is independent and identical. In particular, all particles in principle exhibit their dynamic activity for an indefinite amount of time. The 'immobility attribute' can only be assigned when the evolution of the (aged) process is studied within a *finite time* window. Then, a certain fraction of particles—the immobile ones—stand out *statistically* from the rest. The displacement propagator (4.19) with its conspicuous, discontinuous contribution serves as a statistical indicator for the population splitting, if evaluated at finite times $t < \infty$ (more precisely, the effect is most noticeable while t remains short as compared to the age t_a). Similarly, population splitting is particularly relevant when assessing time average data on a per-trajectory basis, as shown in the following section. Of course, in any case, observations of real physical systems are finite by nature. It is hence important to know the characteristics of the ageing population splitting, as to set it apart from separation mechanisms due to physically non-identical particle dynamics.

4.2.3 Analysis of mean squared displacements

An alternative way to assess particle spreading in the solvent is to study the time evolution of the mean squared displacement. This is particularly useful when ensemble data is not extensive enough as to deduce reliable propagator statistics W_{ACTRW}. There are two common ways of defining such mean squared displacements: either in terms of an ensemble average or a single-trajectory time average. Analysis and comparison of these two types of observables reveals several fingerprint phenomena of CTRW motion, such as subdiffusion and weak ergodicity breaking. The following section collects known results and extends the discussion to implications for aged systems.

Ensemble average. As a simple example, take $x(n)$ to be unbounded, unbiased Brownian motion and consider an ensemble measurement of the mean squared displacement. In this case, the PDF W_{RW} is Gaussian, and consequently

$$\langle [x(k+n) - x(k)]^2 \rangle = \langle x^2(n) \rangle = \int_{-\infty}^{\infty} x^2 \, W_{\mathrm{RW}}(x; n) \, dx = 2Dn \tag{4.20}$$

for all $k, n > 0$. Spatial units are arbitrary from here, $2D = 1$. Since $x(n)$ is a process with stationary increments, the calculation of moments is independent of the number of steps k made before start of the measurement. Thus, if the process the experimenter studies is Brownian motion, there are no ageing effects: at all initial times k, one observes a linear scaling with respect to observational time n, $\langle x^2(n) \rangle \simeq n$, a behaviour commonly classified as normal diffusion. However, if the motion is paused irregularly for heavy-tailed waiting time periods, the dynamics are of CTRW type and the picture is quite different. For the process $x(t) = x(n(t))$, one has

$$\begin{aligned}
\langle [x(t_a + t) - x(t_a)]^2 \rangle &= \langle x_a^2(t_a, t) \rangle \\
&= \int_{-\infty}^{\infty} x_a^2 W_{\mathrm{ACTRW}}(x_a; t_a, t) \, dx_a \\
&= \int_{-\infty}^{\infty} \int_{0}^{\infty} x_a^2 W_{\mathrm{RW}}(x_a; n_a) p_a(n_a; t_a, t) \, dn_a \, dx_a \\
&= \int_{0}^{\infty} n_a p_a(n_a; t_a, t) \, dn_a \\
&= \langle n_a(t_a, t) \rangle = \langle n(t_a + t) \rangle - \langle n(t_a) \rangle \\
&= \frac{1}{\Gamma(1+\alpha)} \left[(t_a + t)^\alpha - t_a^\alpha \right] \\
&\sim \begin{cases} t^\alpha / \Gamma(\alpha + 1) + t_a^\alpha / \Gamma(1 - \alpha), & t_a \ll t \\ t_a^{\alpha - 1} t / \Gamma(\alpha), & t_a \gg t \end{cases},
\end{aligned} \tag{4.21}$$

in accordance with Ref. [107]. As expected, the non-stationarity of the ageing renewal process carries over to the CTRW. If monitored at $t_a = 0$, the mean squared displacement grows like $\langle x_a^2 \rangle = \langle x^2 \rangle \simeq t^\alpha$. Since the increase is less than linear in time, the CTRW

is subdiffusive. For $t_a > 0$, the mean squared displacement is no longer described in terms of a single power law. The process looks more like diffusion in a non-equilibrium environment and in fact t_a might be conceived as an internal relaxation time scale. Interestingly, if the measurement takes place in the highly aged regime $t_a \gg t$, the mean squared displacement as a function of the observation time t indicates normal diffusion with an age dependent diffusion coefficient, $\langle x_a^2 \rangle \simeq t_a^{\alpha-1} t$. Only if data on an extensive range of time scales is available, one can identify the complete turnover from one ageing regime to the other as a fingerprint of CTRW dynamics. Still, for as long as the experimenter cannot control the age t_a of the measurement, the ageing effect can be misinterpreted as internal relaxation mechanism. The situation gets even more complicated, if t_a is possibly random.

Time average. Now consider the alternative time average notion of the mean squared displacement. For a single particle trajectory $x(t)$, recorded at times $t' \in [t_a, t_a + T]$, it is defined in terms of the sliding average

$$\overline{\delta^2(\Delta; t_a, T)} = \frac{1}{T - \Delta} \int_{t_a}^{t_a + T - \Delta} \left[x(t' + \Delta) - x(t') \right]^2 dt'. \tag{4.22}$$

Here, Δ is called lag time, and parameters defining the time window of observation are also referred to as age t_a and measurement time T. While the ensemble mean (4.21) is evaluated in terms of squared displacements from a multitude of independent process realisations, the time average (4.22) uses data from within a single trajectory at several points in time. The latter is thus a useful alternative whenever measurements on long (i.e. $T \gg \Delta$) but relatively few trajectories are available. For ergodic, stationary processes, both types of averages are equivalent. For example, for a Brownian motion, the time average $\overline{\delta^2}$ is by definition a random quantity differing from one trajectory to the next; but in the limit of long trajectory measurements, $T \gg \Delta$, the time average converges to the corresponding ensemble value, $\overline{\delta^2} \to 2D\Delta$, and fluctuations become negligible [144].

In contrast, the CTRW $x(t) = x(n(t))$ violates both ergodicity ($\overline{\delta^2}$ remains random for arbitrarily long measurement times) and stationarity ($\overline{\delta^2}$ depends on t_a and T). In the context of ageing, one can now ask two questions. First, does the distribution of the time average change qualitatively when evaluated after the onset of the particle dynamics, $t_a > 0$? This issue is discussed extensively in [142]. In short, the time averaged mean squared displacement is directly related to the number of steps n_a made during the measurement via [144, 145, 171]

$$\frac{\overline{\delta^2(\Delta; t_a, T)}}{\left\langle \overline{\delta^2(\Delta; t_a, T)} \right\rangle} = \frac{n_a(t_a, T)}{\langle n_a(t_a, T) \rangle}. \tag{4.23}$$

Figure 4.7 provides a numerical validation of this relation in terms of explicit CTRW simulations for free and confined motion. Notice that Eq. (4.23) is not a distributional

equality, but meant to be a stronger, per-trajectory equality which is validated here by means of simulations data. Hence, all random properties of the time average in this case are not only related to, but quite literally taken over from the underlying counting process. In particular, the distribution of the time average $\overline{\delta^2}$ is a rescaled version of the renewal theory PDF p_a as discussed in section 4.1.3 and plotted in Figs. 4.3 and 4.4-4.6. This implies that a statistical study of time averages reveals the ageing population splitting: within an ensemble of independent particles, there are some which do not exhibit any dynamic activity during observation, $n_a = 0, \overline{\delta^2} = 0$, and thus apparently stand out as an individual subpopulation from the remaining continuum, $n_a > 0, \overline{\delta^2} > 0$. Likewise, one can study [172] the statistics of microscopic diffusivities $D_\alpha = \delta^2/\Delta^\alpha = [x(t'+\Delta) - x(t')]^2/\Delta^\alpha$ along a single time series $t' \in [0, T]$. These diffusivities are consistently found to have a discrete probability for $D_\alpha = 0$. The latter implies that no steps are made during any Δ-sized lag time interval along the time series, and the probability for this actually grows with measurement time T.

The second ageing-related question addresses the explicit lag time dependence. Albeit being different from one trajectory to the next, $\overline{\delta^2}$ is generally characterised by a universal scaling with respect to lag time Δ. The question now is: Is this scaling age-dependent? To answer this, consider the average (over many individual trajectories) of Eq. (4.22). By help of Eq. (4.21),

$$\left\langle \overline{\delta^2(\Delta; t_a, T)} \right\rangle = \frac{1}{T-\Delta} \int_{t_a}^{t_a+T-\Delta} \langle \big[x(t'+\Delta) - x(t')\big]^2 \rangle \, dt' \tag{4.24}$$

$$\begin{aligned} &= \frac{1}{(T-\Delta)\,\Gamma(2+\alpha)} \big[(t_a+T)^{\alpha+1} \\ &\quad -(t_a+T-\Delta)^{\alpha+1} - (t_a+\Delta)^{\alpha+1} + t_a^{\alpha+1}\big] \\ &\sim \frac{\Lambda_\alpha(t_a/T)}{\Gamma(1+\alpha)} \, \frac{\Delta}{T^{1-\alpha}}, \end{aligned} \tag{4.25}$$

where

$$\Lambda_\alpha(z) = (1+z)^\alpha - z^\alpha. \tag{4.26}$$

The approximation in the fourth line is for the relevant case of long measurement times, $T \gg \Delta$. A few points are remarkable when comparing the time dependence of the ensemble average, Eq. (4.21) to the scaling of the time average, Eq. (4.25). The answer to the question on lag time dependence is a simple one: There is generally a linear scaling $\overline{\delta^2} \sim \Delta$, regardless of age t_a. In this respect, the time average is a less complicated observable than the ensemble average. The latter has a comparatively complicated t-dependence and is characterised by an age-dependent regime separation. Compare this to the time average, where non-stationarity enters in terms of the prefactors Λ_α and $T^{\alpha-1}$. Only the amplitude of time averages depends on the measurement duration and age. More precisely, if either measurement time parameter tends to ∞, the time average itself tends to zero. This is why Λ_α is called an *ageing depression*. It can be expressed

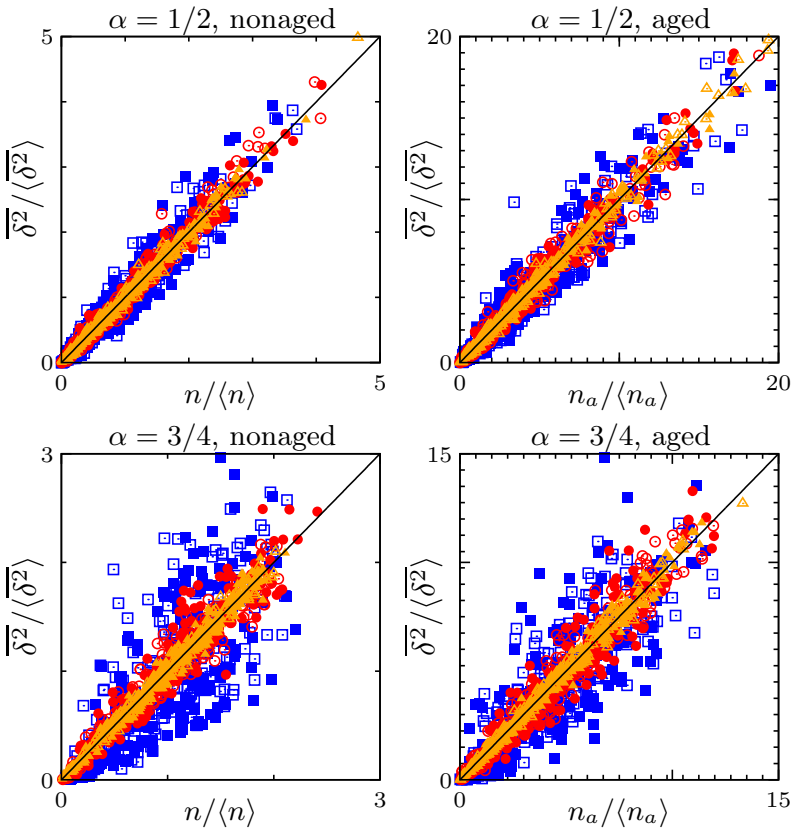


Figure 4.7: Numerical demonstration of the asymptotic identity (4.23) in the limit of long measurement times $T \gg \Delta$. While $\Delta = 100$ remains fixed, the comparison of $T = 2 \times 10^5$ (blue squares ■), $T = 2 \times 10^6$ (red discs ●), and $T = 2 \times 10^7$ (orange triangles ▲) demonstrates convergence. Each point in the graph represents an individual CTRW trajectory. Full symbols represent free CTRW, open symbols are for motion bounded by a box. t_a is either 0 (non-aged), or for specific α and T chosen such that $m_\alpha(t_a/T) = 0.21$ (aged). The simulation data is extensive enough as to ensure that every system is represented by roughly 200 points with $\overline{\delta^2} > 0$.

in terms of the dimensionless ratio t_a/T. In analogy to ensemble measurements one thus speaks of a non-aged ($t_a = 0$), a slightly aged ($t_a \ll T$) and a highly aged ($t_a \gg T$) time average.

Aside from the differences, there is also an interesting parallel between ensemble and time average: the linear scaling of the time average with respect to Δ is reminiscent of the linear t-dependence of the ensemble average at high ages. If, in addition to a very long measurement duration, the preceding ageing period is even longer, $t_a \gg T \gg \Delta$, then the similarity even becomes an equivalence,

$$\left\langle \overline{\delta^2(\Delta; t_a, T)} \right\rangle \sim t_a^{\alpha-1} \frac{\Delta}{\Gamma(\alpha)} \sim \langle x_a^2(t_a, \Delta) \rangle. \tag{4.27}$$

The time scalings of ensemble and time averages are hence identical at high ages. This is quite surprising considering that these quantities are fundamentally distinct in a process that exhibits weak ergodicity breaking. In the following it shall be clarified whether or not these discrepancies and parallels of the two types of averages are specific to the mean squared displacement.

4.2.4 Ageing ensemble and time averages

Consider a random walk $x(n)$ and a stationary observable $F(x_2, x_1)$, meaning that

$$\langle F(x(n+k), x(k)) \rangle = \langle F(x(n), x(0)) \rangle \equiv f(n) \tag{4.28}$$

for any number of steps n or k. In other words, the above ensemble average should not depend on when the observation begins. The example discussed in the previous section falls into this category. There, $x(n)$ was Brownian motion, $F(x_2, x_1) = (x_2 - x_1)^2$ was the squared displacement, and thus $f(n) = 2Dn$. But condition (4.28) would also allow for general order moments or any other function of displacements $F(x_2 - x_1)$. Moreover, $x(n)$ can be replaced by other processes with stationary increments, like fractional Brownian motion (section2.3). If $x(n)$ is even stationary itself (e.g. the stationary limit of confined motion), then any function F is fine (allowing to calculate, e.g. correlation functions, $F(x_2, x_1) = x_2 x_1$; CTRW multi-point correlation functions have been studied extensively in [109, 144, 173, 174]).

Ensemble averages. In any such case, one can ask the question of how the statistical properties of the random motion and the measured observable change when introducing heavy-tailed waiting times between steps. Define $x(t) = x(n(t))$ via subordination, assuming that $x(n)$ and $n(t)$ are stochastically independent processes. In general, the ageing properties of the renewal process $n(t)$ are inherited by the CTRW $x(t)$. The stationarity of the ensemble average is broken. To calculate the magnitude of the effect, conditional averaging can be used by virtue of the independence of the two stochastic processes at work. Denote with $\langle \cdot \rangle_{\mathrm{RW}}$ the average with respect to realisations $x(n)$ and

write

$$\begin{aligned}\langle F(x(t_a+t),x(t_a))\rangle &= \langle F(x(n(t_a+t)),x(n(t_a)))\rangle \\ &= \int_0^\infty\int_0^\infty \langle F(x(n_2),x(n_1))\rangle_{\mathrm{RW}} \\ &\qquad \times \Pr\{n(t_a)=n_1, n_a(t_a,t)=n_2-n_1\}\, dn_1\, dn_2 \\ &= \int_0^\infty\int_0^\infty f(n_2-n_1) \\ &\qquad \times \Pr\{n(t_a)=n_1, n_a(t_a,t)=n_2-n_1\}\, dn_1\, dn_2 \\ &= \int_0^\infty\int_0^\infty f(m)\Pr\{n(t_a)=n_1, n_a(t_a,t)=m\}\, dn_1\, dm \\ &= \int_0^\infty f(m)p_a(m;t_a,t)\, dm \\ &= \langle f(n_a(t_a,t))\rangle. \end{aligned} \tag{4.29}$$

The average on the last line is with respect to $n_a(t_a,t)$, so all knowledge on ensemble averages of the ageing renewal process can come to use. The latter are characterised by a distinct turnover behaviour, see Eqs. (4.11) to (4.13). For the slightly aged CTRW ensemble average, one has

$$\langle f(n(t))\rangle = \mathscr{L}^{-1}_{s\to t}\big\{s^{\alpha-1}\mathscr{L}_{n\to s^\alpha}\{f(n)\}\big\}, \tag{4.30a}$$

$$\langle f(n_a(t_a,t))\rangle = \langle f(n(t))\rangle + \mathcal{O}((t_a/t)^\alpha) \quad (t_a \ll t). \tag{4.30b}$$

For CTRWs, the leading order corrections due to ageing are of the order $(t_a/t)^\alpha$, just as for the underlying renewal process. Conversely, at high ages, slightly rewrite the leading order terms as

$$\begin{aligned}\langle f(n_a(t_a,t))\rangle &\sim f(0) + \frac{t_a^{\alpha-1}}{\Gamma(\alpha)}\int_0^t \left[\langle f'(n(t'))\rangle - f(0)\,\frac{(t')^{-\alpha}}{\Gamma(1-\alpha)}\right] dt' \\ &= f(0) + \frac{t_a^{\alpha-1}}{\Gamma(\alpha)}\,\mathscr{L}^{-1}_{s\to t}\big\{s^{2\alpha-2}\mathscr{L}_{n\to s^\alpha}\{f(n)-f(0)\}\big\} \\ &\equiv C + \frac{t_a^{\alpha-1}}{\Gamma(\alpha)}\,g(t) \qquad\qquad (t_a \gg t), \end{aligned} \tag{4.31}$$

introducing the constant $C = f(0)$ and defining an auxiliary function $g(t)$, either relating it to the analogue stationary average $f(n)$ (third line) or to the corresponding non-aged CTRW average $\langle f'(n(t))\rangle$ (second line). When the measurement of the observable F is taken arbitrarily late, $t_a \to \infty$, one ultimately measures the constant C. For example, if the base random motion $x(n)$ is a random walk with a characteristic scaling $x \sim n^H$, $H > 0$, and one studies moments of displacements, $F(x_2,x_1) = |x_2-x_1|^q$, $q > 0$, then $f(n) \simeq n^{qH}$; in this case, $C = 0$, moments become arbitrarily small at high ages. If $x(n)$ is the stationary limit of a confined motion and one is interested in the correlation

function $F(x_2, x_1) = x_2 x_1$ (as studied in Refs. [144, 142]), then late measurements will be close to the thermal value of the squared position $C = \langle x^2 \rangle$. The decay to this constant value is generally of the order $(t/t_a)^{1-\alpha}$, no matter which observable is being studied.

Time averages. Now turn to the analogue time average, namely

$$\overline{F(\Delta; t_a, T)} = \frac{1}{T-\Delta} \int_{t_a}^{t_a+T-\Delta} F(x(t'+\Delta), x(t'))\, dt', \tag{4.32}$$

and calculate its expectation value

$$\begin{aligned} \left\langle \overline{F(\Delta; t_a, T)} \right\rangle &= \frac{1}{T-\Delta} \int_{t_a}^{t_a+T-\Delta} \langle f(n_a(t, \Delta)) \rangle \, dt \\ &\sim C + \frac{1}{T} \int_{t_a}^{t_a+T} \frac{(t)^{\alpha-1}}{\Gamma(\alpha)} g(\Delta)\, dt \\ &= C + \frac{\Lambda_\alpha(t_a/T)}{\Gamma(1+\alpha)} \frac{g(\Delta)}{T^{1-\alpha}}. \end{aligned} \tag{4.33}$$

The approximation in the second line holds for sufficiently long trajectories, $T \gg \Delta$. Notice that from the full, possibly complicated time dependence of the ensemble average, only the late-age limiting behaviour, Eq. (4.31), entered this approximation. The reason is that with the integrand in the time average of Eq. (4.33) decaying like $(t')^{\alpha-1}$, the integral itself is still an increasing function of T, namely it grows like T^α.

Thus, indeed ageing effects for time averages in CTRW type of random motion are universally described in terms of the constant C and simple prefactors Λ_α and $T^{\alpha-1}$. The full lag time dependence is captured by the function $g(\Delta)$, which is independent of the parameters defining the time window of observation, t_a and T. Conversely, the concrete choice for the model $x(n)$ or the observable F enter only C and g, but not the prefactors.

Moreover, the Δ-dependence of the time average is closely related to the t-dependence of the ensemble average at high ages, $t_a \gg t$. In particular, a universal asymptotic equivalence is found in the time scaling of time and ensemble averages, if both are taken during late measurements:

$$\left\langle \overline{F(\Delta; t_a, T)} \right\rangle \sim \langle F(x(t_a+\Delta), x(t_a)) \rangle \text{ for } t_a \gg T \gg \Delta. \tag{4.34}$$

Thus indeed, averaging at late ages appears to happen under stationary conditions. Keep in mind however, that above identity refers to the expectation value of time averages, and thus to the time scaling behaviour. Since CTRWs exhibit weak ergodicity breaking, the amplitude of the time average of a single trajectory can largely deviate from the expected value, especially at high ages. (See for instance the discussion on the ergodicity breaking parameter for $\overline{\delta^2}$ in Ref. [142]).

4.2.5 Interplay of ageing and internal relaxation

This section on ageing CTRWs closes with the study of a process which extends the previous examples to a more elaborate model. On the one hand, this serves as exemplary application of the formulae and methods described in this paper. In particular, it gives a concrete meaning to the analytic discussion of time and ensemble averages and their relation to internal relaxation time scales. On the other hand, its complexity and versatility makes it more suitable for possibly describing real world physical phenomena (examples below). The definition of the model is as follows.

With the base random walk $x(n)$ depart now from the simple Brownian motion and instead consider the fractional Langevin equation (FLE, see section 2.3.4)

$$0 = -\lambda x(n) - \overline{\gamma} \int_0^n (n-n')^{2H-2}\, \dot{x}(n')\, dn' + \sqrt{\overline{\gamma} k_B \mathcal{T}}\, \xi_H(n). \tag{4.35}$$

Here, the dot signals a first order derivative, $\overline{\gamma} > 0$ is a generalised friction constant, $k_B\mathcal{T} > 0$ gives the thermal energy of the environment, and $\lambda > 0$ quantifies the strength of an external, harmonic potential $V(x) = \lambda x^2$, centred around $x = 0$. Thus, this Assume that dynamics are overdamped, meaning that the particle mass is so small that one can neglect particle inertia. Consequently, there is no term proportional to $\ddot{x}(n)$ in Eq. (4.35). Recall that fractional Gaussian noise is a stationary Gaussian process defined through (see section 2.3.4 and Refs.[118, 119])

$$\langle \xi_H(n) \rangle = 0 \tag{4.36a}$$

$$\langle \xi_H(n_1)\xi_H(n_2) \rangle = |n_2 - n_1|^{2H-2} + \frac{2}{2H-1} |n_2 - n_1|^{2H-1} \delta(n_2 - n_1), \tag{4.36b}$$

where $1/2 < H < 1$.

Eq. (4.35) admits a solution with a stationary velocity profile, so that the net energy exchange of the particle with the heat bath is zero. Such solution is a Gaussian process defined by [175]

$$\langle x(n) \rangle = 0 \tag{4.37a}$$

$$\langle [x(k+n) - x(k)]^2 \rangle = \frac{2k_B\mathcal{T}}{\lambda} \left[1 - E_{2-2H}\left(-\frac{\lambda}{\gamma} |n|^{2-2H} \right) \right], \tag{4.37b}$$

in terms of a generalised Mittag-Leffler function $E_{2-2H}(z)$ (see Ref. [187] or Appendix of Ref. [143]), introducing $\gamma = \overline{\gamma}\Gamma(2H-1)$. Note that the definition of a Gaussian process in terms of its correlation function is equivalent to the definition in terms of squared increments, since $\langle [x_2 - x_1]^2 \rangle = \langle x_1^2 \rangle + \langle x_1^2 \rangle - 2\langle x_1 x_2 \rangle$. The non-equilibrium solutions to the FLE are discussed in Refs. [176, 177], including an interesting discussion on their transient ageing behaviour [177]. In the limit $H \to 1/2$, Eq. (4.37) defines a stationary Ornstein-Uhlenbeck process, implying exponential relaxation. Physically, the latter models overdamped motion in an harmonic potential where friction and noise forces are memoryless.

The stationary Gaussian process (4.37) is the starting point for the discussion on ageing induced by heavy-tailed waiting times. There are, of course, various ways to introduce an ageing, non-ergodic, CTRW-like model component, and the resulting stochastic processes differ largely. For instance, one can combine the Gaussian dynamics with an independent CTRW motion in a purely additive manner, as discussed in [178]. Moreover, for a non-overdamped, inertial motion, the FLE velocity process $v(n) = \dot{x}(n)$ can be modified by adding periods of constant velocity with heavy-tailed statistics [179]. Here, the standard subordination approach is followed, as described in the preceding sections: stalling dynamics are introduced by defining $x(t) = x(n(t))$, where the ageing renewal process $n(t)$ is assumed to be statistically independent from $x(n)$. Physically, this scenario implies that the particle under observation is governed by the FLE (4.35). Eventually, it becomes trapped for a random waiting time governed by the probability density $\psi(t)$. After release from the trap the particle motion quickly thermalises and the particle again follows the stationary Gaussian dynamics (4.37) until the next trapping event.

The independence of $x(n)$ and $n(t)$, physically implies that stalling dynamics are neither affected by the external binding potential, nor intertwined with heat bath relaxation mechanisms. The theory presented here may thus be considered as an effective or approximated approach or merely a simplified case study; the implications of waiting time parameters τ or α responding to external forces are discussed for two-state models in Refs. [180, 181]; the interplay of ageing renewals and adapted non-stationary external forces are highlighted in Ref. [182, 183]. Finally, the origin of the time coordinate is set at a time where the particle enters a trap (i.e. $n(t)$ starts at $t = 0$), and by the time t_a the observation starts, the FLE dynamics $x(n)$ have already relaxed to the stationary equilibrium state. Note that for $t_a = 0$, this implies that one either has means to measure or control the onset of trapping dynamics.

There are basically three motivations to study this random motion. First, CTRWs provide one approach to model diffusive motion in biological cells, where waiting times represent multi-scale binding or caging dynamics. The sheer complexity of this kind of environment however brings the necessity to further introduce aspects of anomalous diffusion not contained in the renewal process $n(t)$ [165, 135, 124, 78, 84]. While the overdamped FLE dynamics (4.37) adds both friction and external binding forces, it also introduces the concept of long-time correlations within an equilibrated environment.

From the point of view of a theoretical discussion of ageing CTRWs, the second motive to discuss this particular model is the stationarity of increments of $x(n)$. It allows to exercise the methods introduced in the previous section by calculating ensemble and time averages—mean squared displacements in particular—of the ageing process $x(t)$.

Third, aside from its didactic purpose, the definition of the process $x(n)$ also extends the ordinary Brownian motion by introduction of an intrinsic time scale $n^* = (\gamma/\lambda)^{1/(2-2H)}$, characterising the competition between binding and friction forces. Ac-

cording to Eq. (4.37), stationary Langevin dynamics exhibit a turnover

$$\langle [x(k+n)-x(k)]^2\rangle \sim \frac{2k_B\mathcal{T}}{\lambda}\begin{cases}[\Gamma(3-2H)]^{-1}(n/n^*)^{2-2H}, & n\ll n^*\\ 1, & n\gg n^*,\end{cases} \tag{4.38}$$

from subdiffusion on short time scales to the stationary Boltzmann-limit $2k_B\mathcal{T}/\lambda$ on long scales. Now for the ageing process $x(t)$, the age of a measurement t_a itself can be conceived as a time scale separating the time dependence into early and late age behaviour. It would be interesting to know how complex the process $x(t)$ behaves with respect to both time scales. Are all conceivable time regimes clearly distinct? Are time and ensemble averages equally sensitive to transitions from one regime to the other?

A partial answer is given by Eqs. (4.29) to (4.31), which focus on the calculation of ensemble averages. The mean squared displacement is obtained by setting $F(x_1,x_2)=(x_1-x_2)^2$ and $f(n)$ as defined through Eq. (4.37). Inserting the latter into Eq. (4.30), yields an approximation for low ages, $t_a \ll t$:

$$\begin{aligned}\langle x_a^2(t_a,t)\rangle &= \langle [x(t_a+t)-x(t_a)]^2\rangle\\ &= \langle [x(t)-x(0)]^2\rangle + \mathcal{O}((t_a/t)^\alpha)\\ &\sim \frac{2k_B\mathcal{T}}{\lambda}\times\\ &\quad \mathscr{L}^{-1}_{s\to t/\tau}\left\{s^{\alpha-1}\mathscr{L}_{n\to s^\alpha}\left\{\left[1-E_{2-2H}\left(-\frac{\lambda}{\gamma}|n|^{2-2H}\right)\right]\right\}\right\}\\ &= \frac{2k_B\mathcal{T}}{\lambda}\,\mathscr{L}^{-1}_{s\to t/\tau}\left\{\frac{1}{s}-\frac{s^{(2-2H)\alpha-1}}{s^{(2-2H)\alpha}-\lambda/\gamma)}\right\}\\ &= \frac{2k_B\mathcal{T}}{\lambda}\,\left[1-E_{(2-2H)\alpha,2-\alpha}\left(-\frac{\lambda}{\gamma_\alpha}t^{(2-2H)\alpha}\right)\right].\end{aligned} \tag{4.39}$$

In order to obtain reasonable physical units, the parameter τ from Eq. (4.1) was reintroduced, bearing the dimension of seconds s. Moreover, the new constant $\gamma_\alpha=\gamma\tau^{(2-2H)\alpha}$ has physical units kg $\mathrm{s}^{-(2-2H)\alpha-2}$.

For increasingly aged ensemble measurements, ageing corrections of the order $(t_a/t)^\alpha$ come into play. The detailed time behaviour of the ensemble mean squared displacement is found by combining Eqs. (4.37), (4.29) and (4.12). Here $f(0)=0$, and thus one can write

$$\langle x_a^2(t_a,t)\rangle = \frac{2k_B\mathcal{T}}{\lambda}h(t_a,t)\ast_t\left[1-E_{(2-2H)\alpha,2-\alpha}\left(-\frac{\lambda}{\gamma_\alpha}t^{(2-2H)\alpha}\right)\right]. \tag{4.40}$$

The full time dependence of the ensemble average comes as a convolution of the forward recurrence time PDF (4.8c) with a generalized Mittag-Leffler function. Graphical examples are provided below.

The behaviour at high ages, $t_a \gg t$ can be calculated analytically by use of Eqs. (4.37) and (4.30):

$$\langle x_a^2(t_a,t)\rangle \sim \frac{t_a^{\alpha-1}}{\Gamma(\alpha)}\, g(t),$$

where in this case $C = f(0) = 0$ and

$$\begin{aligned} g(t) &= \frac{2k_B\mathcal{T}}{\lambda}\mathscr{L}^{-1}_{s\to t/\tau}\left\{s^{2\alpha-2}\mathscr{L}_{n\to s^\alpha}\left\{1 - E_{2-2H}\left(-\frac{\lambda}{\gamma}|n|^{2-2H}\right)\right\}\right\} \\ &= \frac{2k_B\mathcal{T}}{\lambda}t^{1-\alpha}\left[\frac{1}{\Gamma(2-\alpha)} - E_{(2-2H)\alpha,2-\alpha}\left(-\frac{\lambda}{\gamma_\alpha}t^{(2-2H)\alpha}\right)\right]. \end{aligned} \tag{4.41}$$

The time dependence of the mean squared displacement, Eq. (4.40), is relatively complicated. In both the limits of slightly and highly aged measurements there is a Mittag-Leffler behaviour, but with different parameters. From Eqs. (4.39) and (4.41), four time regimes can be extracted, where the diffusive behaviour with respect to time t is described in terms of single power laws; regimes are separated by a time scale t_a induced by ageing and an intrinsic relaxation time scale $\tau_\alpha^* = (\gamma_\alpha/\lambda)^{1/[(2-2H)\alpha]}$:

$$\langle x_a^2(t_a,t)\rangle \sim \frac{2k_B\mathcal{T}}{\lambda}\begin{cases} A_\alpha^*\, t_a^{\alpha-1}t^{1-(2H-1)\alpha}, & t \ll t_a, \tau_\alpha^*, \\ B_\alpha^*\, t^{(2-2H)\alpha}, & t_a \ll t \ll \tau_\alpha^*, \\ C_\alpha^*\, t_a^{\alpha-1}t^{1-\alpha}, & \tau_\alpha^* \ll t \ll t_a, \\ 1, & t_a, \tau_\alpha^* \ll t, \end{cases} \tag{4.42}$$

with coefficients

$$\begin{aligned} A_\alpha^* &= \left[\tau_\alpha^{*(2-2H)\alpha}\Gamma(\alpha)\Gamma(2-(2H-1)\alpha)\right]^{-1}, \\ B_\alpha^* &= \left[\tau_\alpha^{*(2-2H)\alpha}\Gamma(1-(2-2H)\alpha)\right]^{-1}, \\ C_\alpha^* &= \left[\Gamma(\alpha)\Gamma(2-\alpha)\right]^{-1}. \end{aligned}$$

Figure 4.8 gives several examples for the detailed turnover behaviour of the mean squared displacement for various values of α and H. At infinite times $t \to \infty$, the ensemble mean squared displacement converges to the Boltzmann limit. At finite times, there is generally a subdiffusive behaviour. The dynamics are slowest when t is short as compared both intrinsic relaxation and ageing time scales. Notice that the behaviour at times $t \gg \tau_\alpha^*$ is independent of the parameter H defining the memory of friction and noise forces. This regime is fully dominated by the ageing transition.

Importantly, the full double turnover behaviour of the ensemble average might not be visible to the observer of a real physical system due to limitations of the experimental setup. In addition, precise knowledge on the ageing time t_a, which preceded the actual ensemble measurement, might not be available. Maybe t_a is even random, varying from

one trajectory to the next. In any such case, the observer can not know which of the power law regimes of (4.42) he is probing by means of mean squared displacement measurements.

The analogue time average measurement is much less complex and thus easier to interpret. According to Eq. (4.33, for $T \gg \Delta$,

$$\begin{aligned}\left\langle \overline{\delta^2(\Delta; t_a, T)} \right\rangle &\sim \frac{\Lambda_\alpha(t_a/T)}{\Gamma(1+\alpha)} \frac{g(\Delta)}{T^{1-\alpha}} \\ &= \frac{\Lambda_\alpha(t_a/T)}{\Gamma(1+\alpha)\, T^{1-\alpha}} \frac{2k_B\mathcal{T}}{\lambda} \\ &\quad \times \Delta^{1-\alpha} \left[\frac{1}{\Gamma(2-\alpha)} - E_{(2-2H)\alpha, 2-\alpha}\left(-\frac{\lambda}{\gamma_\alpha} \Delta^{(2-2H)\alpha} \right) \right]. \end{aligned} \tag{4.43}$$

The dependence on measurement time parameters t_a and T factorises from the lag time dependence. The latter is captured by the function g. Recall that for weakly non-ergodic CTRW dynamics, the amplitude of a single-trajectory time average $\overline{\delta^2}$ is random by nature, while its scaling with lag time Δ is not. Thus, the Δ-scaling of the time averaged mean squared displacement is age-independent. For combined CTRW-FLE dynamics, it is universally given by g in terms of a Mittag-Leffler type single turnover, with (lag) time regimes being separated by the intrinsic time scale $\tau_\alpha^* = (\gamma_\alpha/\lambda)^{1/[(2-2H)\alpha]}$. In particular, a single, long trajectory measurement is in principle sufficient to read off the scaling exponents α and H, and the time scale τ_α^*.

Ageing affects the amplitude of the time average only; as t_a increases, the typical values of $\overline{\delta^2}$ become smaller. For exemplary plots, see Fig. 4.9. Notice that the late lag time behaviour is generally $\Delta^{1-\alpha}$, independently of H, as reported previously [184, 144] for confined, memory-less CTRWs (i.e. for $H = 1/2$). Moreover, in the limit $\alpha \to 1$, ageing becomes negligible, $\Lambda_1 \equiv 1$, recovering the ergodic FLE result $\langle \overline{\delta^2(\Delta)} \rangle = \langle [x(n = \Delta/\tau) - x(0)]^2 \rangle$, with $x(n)$ as in Eq. (4.37).

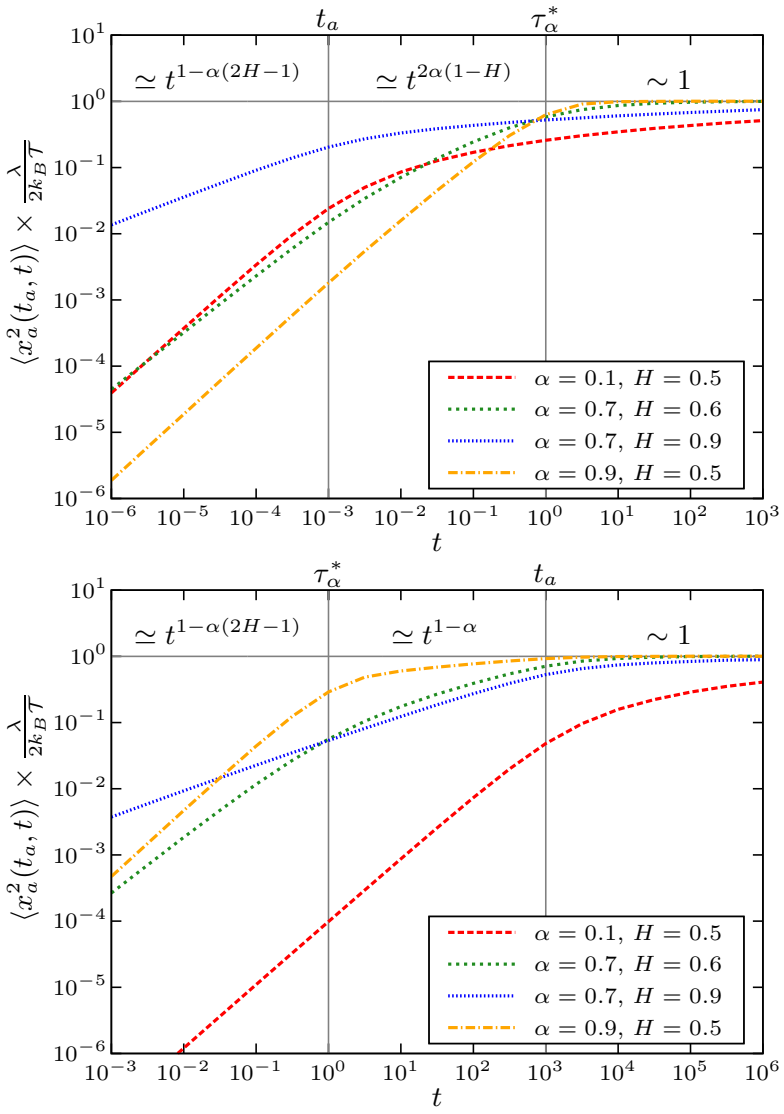


Figure 4.8: Time evolution of the ensemble averaged mean squared displacement for combined CTRW and overdamped, confined FLE dynamics. Plots are numerical evaluations of the convolution integral (4.40). Several parameter configurations for H and α are considered. The behaviour is characterised by a double turnover between several power law regimes as labelled above the graphs, see also Eq. (4.42). Respective turnover time scales τ^*_α and t_a are indicated as vertical lines. A horizontal line gives the infinite time stationary limit $\langle x_a^2\rangle \to 2k_b\mathcal{T}/\lambda$. *Top:* The ensemble measurement starts at a time which is small compared to the FLE relaxation time scale, i.e. $t_a \ll \tau^*_\alpha$. *Bottom:* Opposite case, $t_a \gg \tau^*_\alpha$.

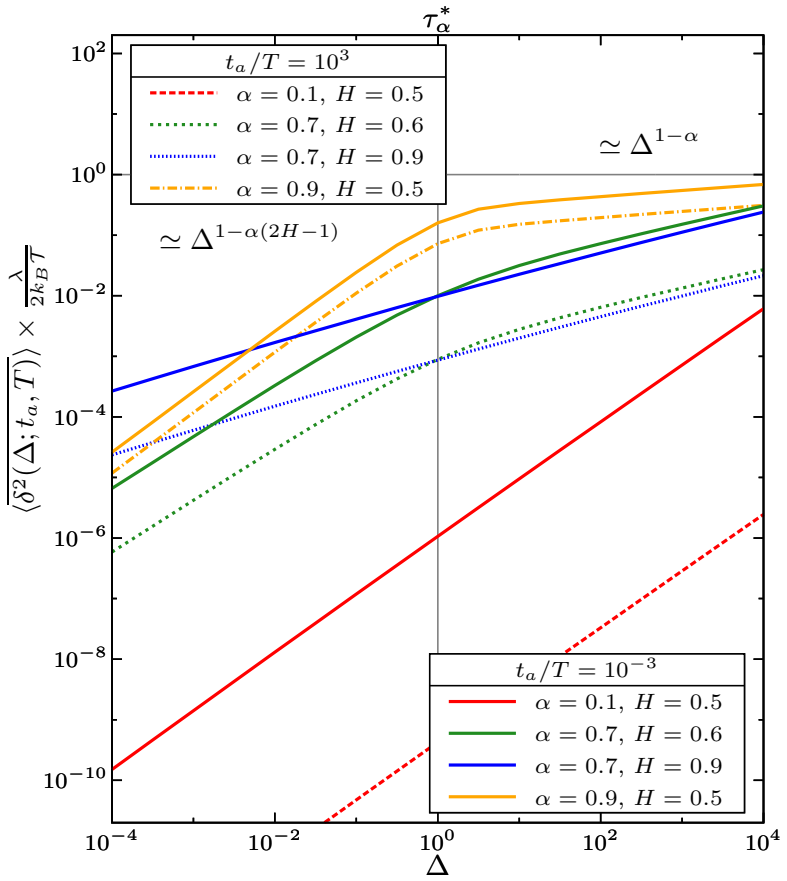


Figure 4.9: Expectation value of the time averaged squared displacement $\overline{\delta^2}$ as a function of lag time Δ. Plots graphically represent Eq. (4.43), for the same system parameters α and H as in Fig. 4.8. In all cases the duration of the measurement is $T = 10^6$. In contrast to the analogue ensemble average, the behaviour is given by a single turnover at the internal relaxation time scale τ_α^*, no matter at which time t_a the recording of the trajectory was initiated. Results obtained at late age ($t_a/T = 10^3$, dashed/dotted lines) differ from those at early age ($t_a/T = 10^{-3}$, continuous lines of same colour) merely by a prefactor (displaying as a constant shift in the double logarithmic plot).

5 Discussion and outlook

Anomalous diffusion is a term that really covers a plethora of phenomena. The paradigmatic models presented in chapter 2 and the following, are prominent examples, with various, distinctive traits – but they are only the tip of the iceberg. Several refined, modified or independent models have been and still are discussed in the literature – random walks on fractals, tempered heavy-tailed statistics, models of obstructed diffusion, crowding models, just to name a few. Even more, the class of anomalous diffusion processes grows tremendously when starting to combine these models, in order to build multi-layered processes capable of describing realistic complex environments.

At this point, it becomes imperative to find the deeper structure behind these models. In which aspect are they distinct/where do they overlap? Is there maybe a certain redundancy, some re-invention of the wheel? Do some even contradict each other? Such questions need to be answered in order to find sensible categories, simplified approaches, and finally, to develop an understanding for the mechanisms behind anomalous diffusion processes. Basically two pathways can lead to such profound understanding.

One option is to follow a phenomenological approach. After "blindly" mixing several model ingredients, one can study the effects for certain experimentally observable quantities. By means of this rather positivistic method, one finds categories for anomalous diffusion *effects*. For example, the correlated CTRW introduced in chapter 3 combines the effects of scale-free displacements, scale-free sojourn times and long-range correlations for both. It is thus applicable to a wide range of complex, heterogeneous systems. The probability density function is very distinct from an ordinary Gaussian distribution. A detailed study of its shape reveals information contained in the tail properties and the detailed behaviour around the origin. However, care must be taken when assessing the effects of correlations: processes with contrasting jump length and waiting time correlations can be indiscernible on the level of scaling and propagator analysis. In other words, the overall effect of competing correlations is not visible in terms of this physical observable.

Moreover, correlated CTRWs can be classified in the context of processes with stationary increments of higher order. Such considerations indicate an intimate connection between strong correlations and higher-, possibly fractional-order integrals of stochastic noise processes.

To see the exact nature of correlations, further studies of this process need to include an in-depth discussion of second order quantities of the correlated model. This question is particularly intricate, on the one hand, for scale-free displacements ($\mu < 2$), since the ordinary correlation function $\langle X(t_1)X(t_2)\rangle$ is ill-defined. But even for CTRW dynamics with $\mu = 2$, on the other hand, the issue of inter-dependencies is a subtle one. By means of simple subordination arguments, one can show that the increments – and

by this any notion of "velocities"– of an unbounded, possibly correlated CTRW are uncorrelated. Yet, they are not stochastically independent (cf. e.g. [103]). It thus remains an open question how (or if) correlation functions can be used to determine and assess dependencies, especially within correlated CTRW data.

It would also be interesting to study a process where correlations within displacements and waiting times are coupled, such as Lévy walks. Finally, the physical properties of these processes, such as ageing or (non-)ergodicity, will be of interest.

The second strategy to handle the overwhelmingly complex class of anomalous diffusion processes is to pick one such model, deconstruct it and search for its core anomaly. This helps finding the analytic *structure* of stochastic models and the deep relation between them. Thus, the continuous time random walk can be reduced to a renewal process, which is probably the most fundamental representative of a scale-free relaxation process. It has several nontrivial properties, one of which is ageing, as discussed in chapter 4: the statistics of renewals counted within a finite observation time window strongly depends on the specific instant at which one starts to count. The most remarkable ageing effect is a growing, discrete probability to not count a single event during observation. Concurrently, also the continuous distribution of nonzero number of renewals is transformed. In section 4.1, several analytic expressions for this distribution can be found, including and analysis in terms of series expansions, tail behaviour and monotonicity. Moreover, expressions for related ensemble averages are given, for instance, arbitrary moments of the number of renewals in terms of incomplete Beta functions. Fewer renewals are counted at increasingly late ages. However, this is mainly due to the high probability of counting zero renewals. Restricting the counting statistics to eventful measurements, see Eq. (4.17), yields a finite distribution even at infinite age, characterised by the same time scaling as the non-aged process.

Ageing renewal theory controls the CTRW motion: renewal events are identified with steps of the random walk. The ageing effects translate from the renewal to the random walk process. Thus the increasing probability for the complete absence of dynamic activity is conceived as a population splitting effect. In an ensemble of identical particles, a certain fraction remains fully immobile during a finite time observation. The total size of the mobile population decreases with age, but also their detailed statistics change.

The implications for ensemble and time averaged observables in CTRW theory were discussed in section 4.2. Ageing affects the distributions of time averages, and population splitting has to be considered in particular. Remarkably, ensemble averages behave very differently with respect to ageing effects than related time averages. The age of a measurement modifies the complete time dependence of an ensemble average, mimicking an internal relaxation mechanism. In contrast, ageing enters the associate time average only as a distributional modification, while its scaling with respect to (lag) time remains indifferent. The precise ageing modifications for the ensemble averaged time average do not depend on details of the definition of the process or the observable. Instead they are captured by universal constants, and the age enters in particular through the ageing depression Λ_α. Despite this fundamental conceptual difference between time and ensemble averages, their time scalings are identical in highly aged measurements. This

is a surprising result since CTRWs are notorious for weak ergodicity breaking, i.e. the general *in*equivalence of the two types of averages.

Clearly, an advantage of the de-constructive strategy towards anomalous process, is that, once sufficient understanding has grown, the constituents can be rearranged to conceive other theories. Thus, while CTRW is a very natural application of ageing renewal theory, it is far from unique. All ageing effects such as population splitting and altered ergodic behaviour have their analogue phenomenon in any physical system where the mean sojourn time in micro-states is infinite. Thus one can expect a certain fraction of blinking quantum dots, or cool atoms, to be constantly stuck in one state during a delayed observation period. At the same time the statistics of the switching population are ageing. Power spectral densities obtained from signals from such systems should display related ageing properties, as discussed briefly in [185]. Aside from stochastic process studies, also weakly chaotic systems were shown to exhibit analogous ageing behaviours [186]. When it comes to diffusion dynamics, a study of alternative CTRW-like models might turn out worthwhile. Examples are the noisy CTRW [178], ageing renewals on a velocity level [179], Lévy walks [97, 161] and correlated CTRW [152]. Moreover, further studies of the ageing renewal process, e. g. with respect to higher dimensional probability distributions, will reveal additional insight into ageing mechanisms of such physical systems.

Acknowledgments

I had a terrific time during the years of my PhD studies, and I fully owe this to the people whom I worked with and who accompanied me.

My supervisor Ralf Metzler supported me on every level of my work. I am grateful for his kind guidance and positive attitude, which gave me the confidence I needed to progress and develop.

It was always great fun to work with the people in my group. I like to remember the early days with Jae-Hyung Jeon, Leila Sereshki, and Rouhollah Abdolvahab. The later years were a fun time thanks to Maximilian Bauer, Vladimir Palyulin, Jochen Kursawe and Mahsa Vahabi. I am especially thankful to Vladimir, who was a perfect host and an excellent sommelier during my many visits to the Potsdam Group. I thank Maximilian, Mahsa and Jochen for being such pleasant room-mates, and Mahsa and Jochen for working with me on nice scientific projects.

I thank Aleksei Chechkin for an excellent collaboration and inspiring conversations during work and evening sessions.

My second, very fruitful collaboration was with Eli Barkai. He taught me a great deal about doing science and about how to publish. Even more, I owe to him six very exciting weeks I spent in his country.

I am grateful to Friedrich Simmel and Martin Zacharias for taking the time and patience to study my thesis and to conduct my defence.

I also appreciate the financial support from the CompInt graduate school at TUM in the course of the Elitenetzwerk Bayern program.

Finally, a big thanks goes to all the people who supported me outside the university: my friends for sometimes helping me forget about work, and my mother, for sometimes helping me handle every-day life. And then there is Lisa, who, by her love, her unconditional support and endless patience, helped me finish my work and made the last year one of the happiest in my life.

List of Figures

Bibliography

[1] K. Pearson, Nature **72**, 294 (1905)

[2] K. Pearson, *Mathematical contributions to the theory of evolution. XV. A mathematical theory of random migration [with the assistance of J. Blakeman], Drapers' Company Research Memoirs, Biometric Series III*, London, Dulau and Co, 1906.

[3] B. D. Hughes, Random Walks and Random Environments, Vol. I: Random Walks (Oxford University Press, Oxford, UK, 1995).

[4] J. W. Strutt, Lord Rayleigh, Philosophical Magazine **10**, 73 (1880); reprinted in Scientific Papers **1**, 491 (1964), New York: Dover.

[5] J. W. Strutt, Lord Rayleigh, Nature **72**, 318 (1905)

[6] M. Slack, Journal of the Institution of Electrical Engineers, Part III **93**, 76 (1946).

[7] R. Barakat, Optica Acta **21**, 903 (1974).

[8] R. Barakat, Journal of the Optical Society of America **71**, 86 (1981).

[9] J. W. Goodman, *Laser Speckle and Related Phenomena*, Heidelberg: Springer-Verlag, 1975.

[10] J. W. Goodman, Journal of the Optical Society of America **66**, 1145 (1976).

[11] F. Horner, Philosophical Magazine (VII) **37**, 145 (1946).

[12] E. Jakeman, and R. J. A. Tough, Advances in Physics **37**, 471 (1988).

[13] J. W. Strutt, Lord Rayleigh, Philosophical Magazine (6) **37**, 321 (1919); reprinted in Scientific Papers **6**, 604 (1964), New York: Dover.

[14] R. J. Nossal, J. Stat. Phys. **30**, 391 (1983).

[15] R. M. Macnab, and D. E. Koshland, Jr., Proc. Nat. Acad. Sci. USA **69**, 2509 (1972).

[16] P. Turchin, Quantitative analysis of movement: measuring and modeling population redistribution in plants and animals (Sinauer, Sunderland, MA, 1998);

[17] P. M. Kareiva, N. Shigesada, Oecologia **56**, 234 (1983).

[18] P. Bovet, P. A. Benhamou, J. Theor. Biology **131**, 419 (1988).

[19] F. Bartumeus, and S.A. Levin, Proc. Natl. Acad. Sci. USA **105**, 19702 (2008).

[20] W. Kuhn, Kolloidzeitschrift **68**, 2 (1934).

[21] E. Guth, and H. Mark, Sitzungsberichte d. mathem.-naturw. Kl. **143**, 445 (1934).

[22] P. J. Flory, *Principles of Polymer Chemistry*, Ithaca: Cornell Uni-University Press (1953).

[23] A. A. Markov, *Wahrscheinlichkeitsrechnung*, Leipzig: B. G. Teubner, 1912, translation from the Russian.

[24] W. Feller, An Introduction to Probability Theory and its Application Wiley, New York, 1970, Vol. 2, Chaps. 13 and 14.

[25] K. Pearson, Nature **72**, 342 (1905)

[26] G. Ehrhardt, J. Stat. Edu. **21**, 2 (2013).
[27] J. C. Maxwell, Phil. Mag. **19**, 19 (1860); Phil. Mag. **20**, 21 (1860)
[28] S. G. Brush, *History of the Kinetic Theory of Gases*, available online:`http://punsterproductions.com/~sciencehistory/pdf/ITALENC.pdf`; Original text of article published in Italian in *Storia della Scienza* **7**, LOttocento, Chapter 44 (with omissions and additional material), Istituto della Enciclopedia Italiana (published 2004)
[29] R. Brown, Phil. Mag. **4**, 161 (1828); Ann. Phys. Chem. **14**, 294 (1828).
[30] C. Wiener, Poggendorffs Annalen **118**, 79 (1863).
[31] E. Kappler, Ann. d. Phys. (Leipzig) **11** 233, (1931).
[32] P. Pearle, B. Collett, K. Bart, D. Bilderback, D. Newman, and S. Samuels, Am. J. Phys. **78**, 1278 (2010).
[33] K. von Nägeli, Sitz.ber. Kgl. Bayerische Akad. Wiss. München, Math.-phys. Kl. **9**, 389 (1879);
[34] J. Renn, Ann. Phys. (Leipzig) **14**, Supplement, 23 (2005) .
[35] J. Perrin, Annales de chimie et de physique VIII **18**, 5 (1909).
[36] A. Einstein, in: R. Fürth (Ed.), Investigations on the theory of the Brownian movement, Dover, New York, 1956.
[37] M. Smoluchowski, Annalen der Physik **21** (14), 756 (1906).
[38] Henk Tijms, *Understanding Probability: Chance Rules in Everyday Life*, Cambridge: Cambridge University Press, 2004.
[39] J. Perrin, *Les Atomes*, 1ère éd. Paris (1913), 2de éd., Paris, Alcan (1936).
[40] M. Nye, *Molecular Reality: A Perspective on the Scientific Work of Jean Perrin*, London: MacDonald, 1972.
[41] N. Wiener, J. Math. and Phys. (MIT) 2 **131** (1923); Bull. Soc. math. (France) 52 (1924) 569; Acta Math. 55 (1930) 117; Amer. J. Math. 60 (1938) 897
[42] B. Oksendal, *Stochastic Differential Equations: An Introduction with Applications*, Springer, Berlin, 2003.
[43] K. Itō, Proc. Imperial Acad. Tokyo 20, 519 (1944); Memoirs, American Mathematical Society **4**, 1 (1951).
[44] P. Protter, Notices of the American Mathematical Society **54** (6), 744 (2007).
[45] P. Langevin, C. R. Acad. Sci. Paris **146**, 530 (1908); english translation: Am. J. Phys. **65** (11), November 1997.
[46] G. E. Uhlenbeck, and L. S. Ornstein, Phys. Rev. **36**, 823 (1930); reprinted in *Selected Papers on Noise and Stochastic Processes*, N. Wax, Editor, Dover, New York, 1954.
[47] M. C. Wang, and G. E. Uhlenbeck, Rev. Mod. Phys., 17, 323 (1945); reprinted in *Selected Papers on Noise and Stochastic Processes*, N. Wax, Editor, Dover, New York, 1954.
[48] J. L. Doob, Ann. Math. **43**, 351 (1942); reprinted in *Selected Papers on Noise and Stochastic Processes*, N. Wax, Editor, Dover, New York, 1954.
[49] W. T. Coffey, Yu. P. Kalmykov, and J. T. Waldron, *The Langevin Equation: With Applications to Stochastic Problems in Physics, Chemistry and Electrical Engi-*

neering, Singapore: World Scientific, 2004.
[50] H. Risken, *The Fokker-Planck equation*, Springer, Berlin (1989)
[51] C.W. Gardiner, *Handbook of Stochastic Methods: for Physics, Chemistry and the Natural Sciences*, Berlin: Springer, 1983.
[52] K. S. Miller, *Complex Stochastic Processes*, Addison Wesley, Reading, Mass. 1974
[53] A. Kolmogorov, Mathematische Annalen **104**, 415 (1931)
[54] A. D. Fokker, Ann. Phys. **348**, 810 (1914)
[55] M. Planck, Sitzber. Preu. Akad. Wiss., p. 324 (1917)
[56] M. v. Smoluchowski, Ann. Phys. **353**, 1103 (1915)
[57] J. W. Strutt, Lord Rayleigh, Phil. Mag. **32**, 424 (1891).
[58] O. Klein, Arkiv Mat. Astr. Fys. **16**, 5 (1922)
[59] H. A. Kramers, Physica **7**, 284 (1940)
[60] S. Redner, *A Guide to First-Passage Processes*, Cambridge: Cambridge University Press, 2001
[61] P. Atkins, *Physical Chemistry* (6th ed.), New York: Freeman, (1998).
[62] J. Johnson, Phys. Rev. **32**, 97 (1928).
[63] H. Nyquist, Phys. Rev. **32**, 110 (1928).
[64] J. R. Barry, E. A. Lee, and D. G. Messerschmitt, *Digital Communications*, Sprinter, p. 69 (2004).
[65] R. Mazo, *Brownian Motion: Fluctuations, Dynamics and Applications*, Oxford University Press, Oxford, 2002
[66] B. Lindnera, and E.M. Nicola, Eur. Phys. J. Special Topics **157**, 43 (2008)
[67] F. Black, M. Scholes. J. Polit. Econ. **81** (3), 637 (1973);
Z. Bodie, A. Kane, and A. J. Marcus, *Investments* (7th ed.), New York: McGraw-Hill/Irwin (2008).
[68] R. W. Zwanzig, J. Chem. Phys **33**, 1388 (1960).
[69] H. Mori, Prog. Theo. Phys. **33**, 423 (1965).
[70] J. M. Deutch, and R. Silbey, Phys. Rev. A **3**, 2049 (1971).
[71] Billingsley, Patrick (1995), Probability and Measure (Third ed.), John Wiley & sons, p. 362
[72] J. W. Lindeberg, Mathematische Zeitschrift **15** 211225 (1922).
[73] R. Bradley, Probability Surveys **2**. 107 (2005).
[74] A. Hald, *A History of Mathematical Statistics from 1750 to 1939*, John Wiley & sons;
H. Fischer, *A History of the Central Limit Theorem: From Classical to Modern Probability Theory, Sources and Studies in the History of Mathematics and Physical Sciences*, New York: Springer (2011);
L. Le Cam, Statistical Science **1** (1): 7891 (1986).
[75] F. Galton, *Natural Inheritance*, p.66 (1889), available online `http://galton.org/cgi-bin/searchImages/galton/search/books/natural-inheritance/pages/natural-inheritance_0073.htm` .
[76] H. Scher, and M. Lax, Phys. Rev. B **7**, 4491 (1973); H. Scher, and M. Lax, Phys. Rev. B **7**, 4502 (1973).

[77] H. Scher and E. W. Montroll, Phys. Rev. B **12**, 2455 (1975).
[78] S. M. A. Tabei, S. Burov, H. Y. Kim, A. Kuznetsov, T. Huynh, J. Jureller, L. H. Philipson, A. R. Dinner, and N. F. Scherer, Proc. Natl. Acad. Sci. USA **110**, 4911 (2013).
J.-H. Jeon, V. Tejedor, S. Burov, E. Barkai, C. Selhuber-Unkel, K. Berg-Sørensen, L. Oddershede, and R. Metzler, Phys. Rev. Lett. **106**, 048103 (2011).
[79] T. H. Solomon, E. R. Weeks, and H. L. Swinney, Phys. Rev. Lett. **71**, 3975 (1993)
[80] H. Scher, G. Margolin, R. Metzler, J. Klafter, B. Berkowitz, Geophys. Res. Lett. **29**, 1061
[81] J.-P. Bouchaud and M. Potters, Theory of financial risks (Cambridge University Press, Cambridge, UK, 2000).
[82] J.-P. Bouchad, A. Georges, Phys. Rep. **195**, 127 (1990).
[83] R. Metzler and J. Klafter, Phys. Rep. **339**, 1 (2000).
R. Metzler and J. Klafter, J. Phys. A **37**, R161 (2004).
[84] F. Höfling and T. Franosch, Rep. Prog. Phys. **76**, 046602 (2013).
[85] J. Klafter, S.-C. Lim, and R. Metzler, Editors 2011 *Fractional Dynamics in Physics* (World Scientific, Singapore).
[86] R. Metzler, E. Barkai and J. Klafter, Euro. Phys. Lett. **46**, 431 (1999)
[87] B. V. Gnedenko and A. N. Kolmogorov, Limit Distributions for Sums of Independent Random Variables (Addison-Wesley, Cambridge, MA, 1954).
[88] G. Samorodnitsky and M. S. Taqqu, Stable Non-Gaussian Random Processes: Stochastic Models with Infinite Variance (Chapman & Hall, 1994)
[89] S. Eule, R. Friedrich, F. Jenko, and D. Kleinhans, J. Phys. Chem. B **111**, 11474 (2007)
[90] A. V. Chechkin, R. Metzler, V. Yu. Gonchar, J. Klafter, and L. V. Tanatarov, J. Phys. A **36**, L537 (2003).
[91] N. Krepysheva, and L. Di Pietro, Phys. Rev. E **73**, 021104 (2006)
[92] M. Vahabi, J. H. P. Schulz, B. Shokri, and R. Metzler, Phys. Rev. E **87**, 042136 (2013)
[93] K. Falconer, Fractal Geometry: Mathematical Foundations and Applications (Wiley, Chichester, UK, 1990).
[94] E. Sparre-Andersen, Math. Scand. **1**, 263 (1953)
[95] E. Sparre-Andersen, Math. Scand. **2**, 195 (1954)
[96] T. Koren, M. A. Lomholt, A. V. Chechkin, J. Klafter, and R. Metzler, Phys. Rev. Lett. **99**, 160602 (2007).
[97] M. F. Shlesinger, J. Klafter, and Y. M. Wong, J. Stat. Phys. **27**, 499 (1982).
[98] P. Barthelemy, J. Bertolotti, D. S. Wiersma, Nature **453**, 495 (2008).
[99] G. M. Zaslavsky, Hamiltonian Chaos and Fractional Dynamics (Oxford University Press, Oxford, UK, 2008).
[100] L.F. Richardson, Proc. Roy. Soc. **110**, 709 (1926).
[101] M. A. Lomholt, T. Ambjörnsson, and R. Metzler, Phys. Rev. Lett. **95**, 260603 (2005).
D. Brockmann and T. Geisel, Phys. Rev. Lett. **91**, 048303 (2003).

I. M. Sokolov, J. Mai, and A. Blumen, Phys. Rev. Lett. **79**, 857 (1997).
[102] N. H. Bingham, Z. Wahrscheinlichkeitsth. **17**, 1-22 (1971).
[103] M. Meerschaert, and H.-P. Seffler, J. Appl. Prob. **41**, 623-638 (2004).
[104] E. B. Dynkin, Selected Translations in Mathematical Statistics and Prob- ability American Mathematical Society, Providence, **1**, 249 (1961).
[105] C. Godrèche and J.M. Luck, J. Stat. Phys. **104**, 489 (2001).
[106] E. Barkai, Phys. Rev. Lett. **90**, 104101 (2003).
[107] E. Barkai and Y.-C. Cheng, J. Chem. Phys. **118**, 167 (2003).
[108] H. C. Fogedby, Phys. Rev. E **50**, 1657 (1994).
[109] A. Baule and R. Friedrich, Phys. Rev. E **71**, 026101 (2005).
[110] A. Rebenshtok and E. Barkai, J. Stat. Phys. **133**, 565 (2008).
[111] C. Monthus, and J.-P. Bouchaud, J. Phys. A **29**, 3847 (1996)
[112] S. Burov, and E. Barkai, Phys. Rev. Lett. **98**, 250601 (2007)
[113] Y. Wong, M. L. Gardel, D. R. Reichman, E. R. Weeks, M. T. Valentine, A. R. Bausch, and D. A. Weitz, Phys. Rev. Lett. **92**, 178101 (2004).
[114] J.-H. Jeon, V. Tejedor, S. Burov, E. Barkai, C. Selhuber-Unkel, K. Berg-Sørensen, L. Oddershede, and R. Metzler, Phys. Rev. Lett. **106**, 048103 (2011).
[115] A. V. Weigel, B. Simon, M. M. Tamkun, and D. Krapf, Proc. Natl. Acad. Sci. USA **108**, 6438 (2011).
[116] Q. Xu, L. Feng, R. Sha, N. C. Seeman, and P. M. Chaikin, Phys. Rev. Lett. **106**, 228102 (2011).
[117] G. Samorodnitsky, Foundations and Trends in Stochastic Systems **1**, 163257 (2006)
[118] B. B. Mandelbrot and J. W. van Ness, SIAM **10**, 4 (1968).
[119] G. Gripenberg and I. Norros, J. Appl. Prob. **33**, 400 (1996).
[120] R. Kubo, Rep. Prog. Phys. **29**, 255 (1966).
[121] L. P. Sanders and T. Ambjörnsson, J. Chem. Phys. **136**, 175103 (2012).
[122] L. Lizana, T. Ambjörnsson, A. Taloni, E. Barkai, and M. A. Lomholt, Phys. Rev. E **81**, 051118 (2010).
[123] R. Kupferman, J. Stat. Phys. **114**, 291 (2004).
[124] M. Weiss, Phys. Rev. E **88**, 010101(R) (2013).
[125] I. Goychuk, Phys. Rev. E **80**, 046125 (2009); Adv. Chem. Phys. **150**, 187 (2012).
[126] S. C. Weber, A. J. Spakowitz, and J. A. Theriot, Phys. Rev. Lett. **104**, 238102 (2010).
[127] K. Burnecki, E. Kepten, J. Janczura, I. Bronshtein, Y. Garini, and A. Weron, Biophys. J. **103**, 1839 (2012).
[128] J. Szymanski and M. Weiss, Phys. Rev. Lett. **103**, 038102 (2009).
[129] J.-H. Jeon, H. Martinez-Seara Monne, M. Javanainen, and R. Metzler, Phys. Rev. Lett. **109**, 188103 (2012); G. R. Kneller, K. Baczynski, and M. Pasenkiewicz-Gierula, J. Chem. Phys. **135**, 141105 (2011).
[130] J.-H. Jeon, N. Leijnse, L. B. Oddershede, and R. Metzler, New J. Phys. **15**, 045011 (2013)
[131] W. Min, G. Luo, B. J. Cherayil, S. C. Kou, and X. S. Xie, Phys. Rev. Lett. **94**, 198302 (2005).

[132] H. Hurst, Trans. Am. Soc. Civ. Eng., **116**, 770 (1951);
H. Hurst, Proc. Inst. Civ. Eng., Part I, 519 (1955).
[133] B. Mandelbrot, and R. Hudson, *The (mis)behavior of markets: a fractal view of risk, ruin and reward*, Basic Books (2004).
[134] P. de Anna, T. Le Borgne, M. Dentz, A. M. Tartakovsky, D. Bolster, and P. Davy, Phys. Rev. Lett. **110**, 184502 (2013)
T. Le Borgne, D. Bolster, M. Dentz, P. de Anna, A. Tartakovsky, Water Resources Research, **47**, W12538 (2011)
[135] T. Akimoto, E. Yamamoto, K. Yasuoka, Y. Hirano, and M. Yasui, Phys. Rev. Lett. **107**, 178103 (2011).
[136] J. H. P. Schulz, A. Chechkin, and R. Metzler, 2013 J. Phys. A: Math. Theor. 46 475001
[137] M. M. Meerschaert, E. Nane, and Y. Xiao, Stat. Probab. Lett. **79**, 1194 (2009)
[138] A. V. Chechkin, and V. Yu. Gonchar, Physica A, **277**, 312 (2000).
[139] J. M. Chambers, C. L. Mallows, B. W. Stuck, Journal of the American Statistical Association **71**, 340 (1976)
[140] A. M. Yaglom, Matematicheskii sbornik 37 (79), No.1, 141-196 (1955) (in Russian).
A. M. Yaglom, M. S. Pinsker, Doklady Akademii nauk SSSR (Proceedings of the USSR Academy of Science) **90**, 731 (1953).
M. S. Pinsker, A. M. Yaglom, Doklady Akademii nauk SSSR (Proceedings of the USSR Academy of Science) **94**, 385 (1954).
[141] E. Barkai and Y.-C. Cheng, J. Chem. Phys. **118**, 6167 (2003).
[142] J. H. P. Schulz, E. Barkai, and R. Metzler, Phys. Rev. Lett. **110**, 020602 (2013).
[143] J. H. P. Schulz, E. Barkai, and R. Metzler, arXiv:1310.1058 [cond-mat.stat-mech]; *submitted to* Phys. Rev. X.
[144] S. Burov, J.-H. Jeon, R. Metzler, and E. Barkai, Phys. Chem. Chem. Phys. **13**, 1800 (2011);
S. Burov, R. Metzler, and E. Barkai, Proc. Natl. Acad. Sci. USA **107**, 13228 (2010).
[145] Y. He, S. Burov, R. Metzler, and E. Barkai, Phys. Rev. Lett. **101**, 058101 (2008).
J.-H. Jeon, V. Tejedor, S. Burov, E. Barkai, C. Selhuber-Unkel, K. Berg-Sørensen, L. Oddershede, and R. Metzler, Phys. Rev. Lett. **106**, 048103 (2011).
E. Barkai, Y. Garini, and R. Metzler, Phys. Today **65**(8), 29 (2012).
[146] Y. Meroz, I. M. Sokolov, and J. Klafter, Phys. Rev. Lett. **107**, 260601 (2011).
I. M. Sokolov and R. Metzler, J. Phys. A **37**, L609 (2004).
[147] M. Teuerle, A. Wyłomańska and G. Sikora, J. Stat. Mech. **2013**, 5016
[148] A. M. Mathai, R. K. Saxena, and H. J. Haubold, The H-Function, Theory and Applications, (Springer 2009).
[149] A. A. Kilba, M. Saigo, J. Appl. Math. Stoch. Anal. **12**, 191 (1999).
[150] A. V. Chechkin, M. Hofmann, and I. M. Sokolov, Phys. Rev. E **80**, 031112 (2009)
[151] A. I. Sushin, Phys Rev E **77**, 031130 (2008).
[152] V. Tejedor and R. Metzler, J. Phys. A **43** (2010) 082002;
[153] M. Magdziarz, R. Metzler, W. Szczotka, and P. Zebrowski, Phys. Rev. E **85**, 051103 (2012);

M. Magdziarz, R. Metzler, W. Szczotka, and P. Zebrowski, *J. Stat. Mech.* **2012**, P04010.

[154] Cox, David, Renewal Theory, London: Methuen & Co. (1970).

[155] F. Bardou, J.-P. Bouchaud, A. Aspect, and C. Cohen-Tannoudji, Lévy Statistics and Laser Cooling (Cambridge University Press, Cambridge, England, 2002);
F. Bardou, J.-P. Bouchaud, O. Emile, A. Aspect, and C. Cohen-Tannoudji, Phys. Rev. Lett. **72**, 203 (1994) .

[156] E. Lutz, Phys. Rev. Lett. **793**, 190602 (2004).

[157] X. Brokmann, J.-P. Hermier, G. Messin, P. Desbiolles, J.-P. Bouchaud, and M. Dahan, Phys. Rev. Lett. **90**, 120601 (2003) .

[158] G. Margolin and E. Barkai, J. Chem. Phys. **121**, 1566 (2004).

[159] K. T. Shimizu et al., Phys. Rev. B **63**, 205316 (2001).

[160] J.-P. Bouchaud, J. Phys. I (Paris) **2**, 1705 (1992).

[161] P. Allegrini, J. Bellazzini, G. Bramanti, M. Ignaccolo, P. Grigolini, and J. Yang, Phys. Rev. E **66**, 015101 (2002).

[162] P. Allegrini, G. Aquino, P. Grigolini, L. Palatella, and A. Rosa. Phys. Rev. E **68**, 056123 (2003)

[163] P. Allegrini, G. Aquino, P. Grigolini, L. Palatella, A. Rosa, and B. J. West, Phys. Rev. E **71**, 066109 (2005)

[164] W.R. Schneider, in: S. Albeverio, G. Casati, D. Merlini (Eds.), Stochastic Processes in Classical and Quantum Systems, Lecture Notes in Physics, Vol. 262, Springer, Berlin, 1986

[165] Y. Meroz, I. M. Sokolov, and J. Klafter, Phys. Rev. E **81**, 010101(R) (2010).

[166] M. Magdziarz, A. Weron, and K. Weron, Phys. Rev. E **75**, 016708 (2007).

[167] G. T. Schütz, H. Schneider, and T. Schmidt, Biophys. J. **73**, 1073 (1997) and Refs. therein.

[168] T. Kues, R. Peters, and U. Kubitschek, Biophy. J. **80**, 2954 (2001).

[169] P. H. M. Lommerse, B. E. Snaar-Jagalska, H. P. Spaink, and T. Schmidt, J. Cell Science **118**, 1799 (2005).

[170] S. Manley *et al.*, Nature Methods 5, **155** (2008).

[171] J.-H. Jeon and R. Metzler, J. Phys. A **43**, 252001 (2010).

[172] T. Albers and G. Radons, Europhys. Lett. **102**, 40006 (2013)

[173] F. Šanda, and S. Mukamel, Phys. Rev. E **72**, 031108 (2005).

[174] E. Barkai and I. M. Sokolov, J. Stat. Mech. **2007**, P08001

[175] N. N. Pottier, Physica A **317**, 371 (2003).

[176] W. Deng and E. Barkai, Phys. Rev. E **79**, 011112 (2009).

[177] J. Kursawe, J. H. P. Schulz, and R. Metzler, Phys. Rev. E **88**, 062124 (2013).

[178] J.-H. Jeon, E. Barkai, and R. Metzler, J. Chem. Phys. **139**, 121916 (2013).

[179] R. Friedrich, F. Jenko, A. Baule, and S. Eule, Phys. Rev. Lett. **96**, 230601 (2006);
S. Eule, R. Friedrich, F. Jenko, and D. Kleinhans, J. Phys. Chem. B **111**, 11474 (2007)

[180] G. Aquino, P. Grigolini, and B. J. West, Euro. Phys. Lett. **80**, 10002 (2007)

[181] P. Allegrini, M. Bologna, P. Grigolini, and B. J. West, Phys. Rev. Lett. **99**, 010603

(2007)
[182] P. Allegrini, M. Bologna, L. Fronzoni, P. Grigolini, and L. Silvestri1, Phys. Rev. Lett. **103**, 030602 (2009)
[183] G. Aquino, M. Bologna, P. Grigolini, and B. J. West, Phys. Rev. Lett. **105**, 040601 (2010)
[184] T. Neusius, I. M. Sokolov, and J. C. Smith, Phys. Rev. E **80**, 011109 (2009).
[185] M. Niemann, H. Kantz, and E. Barkai, Phys. Rev. Lett. **110**, 140603 (2013).
[186] T. Akimoto, and E. Barkai, Phys. Rev. E **87**, 032915 (2013)
[187] M. Abramowitz and I. Stegun, *Handbook of Mathematical Functions*, Dover, New York (1971).